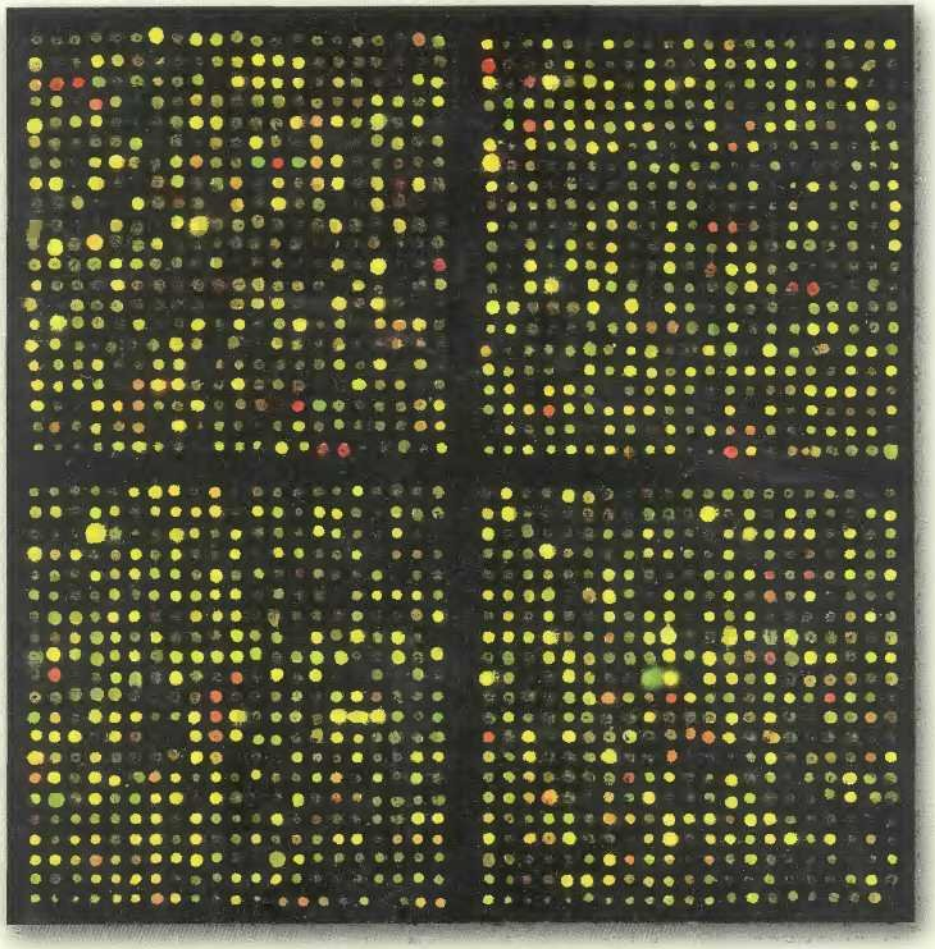
GENETICS

ANALYSIS OF GENES AND GENOMES

Fifth Edition



Daniel L. Hartl
Elizabeth W. Jones

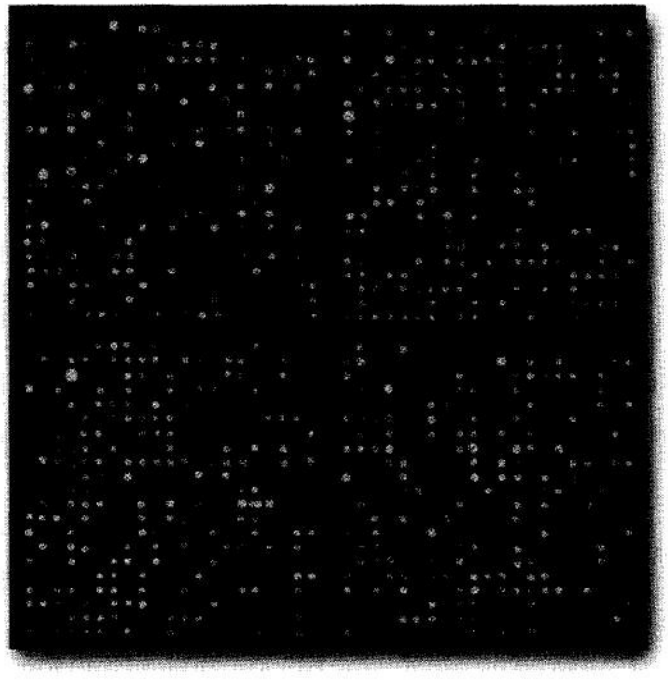
GENETICS

ANALYSIS OF GENES AND GENOMES

Fifth Edition

Daniel L. Hartl
Harvard University

Elizabeth W. Jones
Carnegie Mellon University



JONES AND BARTLETT PUBLISHERS
Sudbury, Massachusetts
BOSTON TORONTO LONDON SINGAPORE

TO THE BEST TEACHERS WE EVER HAD—OUR PARENTS AND OUR STUDENTS

ABOUT THE AUTHORS

Daniel L. Hartl is Higgins Professor of Biology at Harvard University. He received his B.S. degree and Ph.D. from the University of Wisconsin. His research interests include molecular genetics, molecular evolution, and population genetics. Elizabeth W. Jones is a Professor of Biological Sciences at Carnegie Mellon University. She received her B.S. degree and Ph.D. from the University of Washington in Seattle. Her research interests include gene regulation and the genetic control of cellular form. Currently she is studying the function and assembly of organelles in the yeast *Saccharomyces*.

ABOUT THE COVER

DNA chip showing four 21 × 21 arrays of microspots of yeast DNA, each from a different gene, placed with precision by a robotic microarrayer. The chip has been hybridized with samples of fluorescently labeled DNA copied from messenger RNA extracted from a control culture and an experimental culture. The colors indicate the relative levels of gene expression: red, greater expression in the experimental culture; green, less expression in the experimental culture; and yellow, approximately equal expression. Each spot is 150–200 micrometers in diameter. The actual size of the entire array is approximately one square centimeter. [Courtesy of Jeffrey P. Townsend, Duccio Cavalieri, and the Harvard Center for Genomics Research]

ABOUT THE PUBLISHER

World Headquarters
Jones and Bartlett Publishers
40 Tall Pine Drive
Sudbury, MA 01776
978-443-5000
info@jbpub.com
www.jbpub.com

Jones and Bartlett Publishers Canada
2406 Nikanna Road
Mississauga, ON L5C 2W6
CANADA

Jones and Bartlett Publishers International
Barb House, Barb Mews
London W6 7PA
UK

Chief Executive Officer: Clayton Jones
Chief Operating Officer: Don W. Jones, Jr.
Executive Vice President and Publisher: Tom Manning
V.P., Senior Managing Editor: Judith H. Hauck
V.P., Design and Production: Anne Spencer
V.P., National Sales Director: Paul Shepardson
V.P., Manufacturing and Inventory Control: Therese Bräuer
Marketing Manager: Elizabeth Pearson
Special Projects Editor: Mary Hill
Associate Editor: Tim Gleeson
Production Assistant: Tara McCormick

Director of Media Services: W. Scott Smith
Interactive Technology Project Editor: Nicole Healey
Web Product Manager: Adam Alboyadjian
Web Site Design: Kristin Ohlin
Text and Cover Design: Anne Spencer
Production and Editorial Services: Thompson Steele, Inc.
Artwork: J/B Woolsey Associates
Photo Research: Billie Porter, Victoria Jones
PrePress: Eastern Rainbow, Inc.
Book Manufacture: Courier Kendallville
Cover Manufacture: John Pow Company

Library of Congress Cataloging-in-Publication Data

Hartl, Daniel L.
Genetics : analysis of genes and genomes / Daniel L. Hartl, Elizabeth W. Jones. — 5th ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-7637-0913-1
1. Genetics. I. Jones, Elizabeth W. II. Title.

QH430 .H3733 2000
576.5—dc21 00-038440

COPYRIGHT ©2001 BY JONES AND BARTLETT PUBLISHERS, INC.

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the copyright owner.

Credits for photos repeated in the Table of Contents: p. iv James Holmes/Photo Researchers, Inc.; p. v Richard G. Rawlins/Custom Medical Stock Photo; p. vi E. Winzeler, D. Richards, R. Davis; p. vii Agricultural Research Service USDA; p. ix C. Schubert, J. Priess; p. x Roslin Institute/PA Photos. Full credits for these and all interior photos appear with the images and are an extension of this copyright page.

Printed in the United States
04 03 02 01 9 8 7 6 5 4 3 2

BRIEF CONTENTS

Chapter	1	Introduction to Molecular Genetics and Genomics	1
Chapter	2	DNA Structure and DNA Manipulation	36
Chapter	3	Transmission Genetics: The Principle of Segregation	86
Chapter	4	Genes and Chromosomes	132
Chapter	5	Genetic Linkage and Chromosome Mapping	174
Chapter	6	Molecular Biology of DNA Replication and Recombination	222
Chapter	7	Molecular Mechanisms of Mutation and DNA Repair	264
Chapter	8	Molecular Organization of Chromosomes	310
Chapter	9	Human Karyotypes and Chromosome Behavior	344
Chapter	10	Genetics of Bacteria and Their Viruses	392
Chapter	11	Molecular Biology of Gene Expression	444
Chapter	12	Molecular Mechanisms of Gene Regulation	488
Chapter	13	Genetic Engineering and Genomics	540
Chapter	14	Genetic Control of Development	590
Chapter	15	Molecular Genetics of the Cell Cycle and Cancer	634
Chapter	16	Extranuclear Inheritance	676
Chapter	17	Population Genetics and Evolution	706
Chapter	18	The Genetic Basis of Complex Inheritance	748

CONTENTS

Preface xiii

Chapter 1 Introduction to Molecular Genetics and Genomics 1

1.1 **DNA: The Genetic Material** 2
Experimental Proof of the Genetic Function of DNA 3
Genetic Role of DNA in Bacteriophage 6
Connection: Shear Madness 8
1.2 **DNA Structure: The Double Helix** 9
1.3 **An Overview of DNA Replication** 10
1.4 **Genes and Proteins** 11
Inborn Errors of Metabolism as a Cause of Hereditary Disease 12
Connection: The Black Urine Disease 12
Mutant Genes and Defective Proteins 15
1.5 **Gene Expression: The Central Dogma** 16
Transcription 18
Translation 18
The Genetic Code 19
1.6 **Mutation** 21
Protein Folding and Stability 23
1.7 **Genes and Environment** 25
1.8 **Evolution: From Genes to Genomes, From Proteins to Proteomes** 27
The Molecular Unity of Life 27
Natural Selection and Diversity 28
Chapter End Material 29
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 2 DNA Structure and DNA Manipulation 36

2.1 **Genomes and Genetic Differences Among Individuals** 39
DNA Markers as Landmarks in Chromosomes 39
2.2 **The Molecular Structure of DNA** 41
Polynucleotide Chains 41
Base Pairing and Base Stacking 44
Antiparellel Strands 47
Connection: The Double Helix 48
DNA Structure as Related to Function 48
2.3 **The Separation and Identification of Genomic DNA Fragments** 49
Restriction Enzymes and Site-Specific DNA Cleavage 50
Gel Electrophoresis 52
Nucleic Acid Hybridization 54
The Southern Blot 57
2.4 **Selective Replication of Particular DNA Fragments** 58
Constraints on DNA Replication: Primers and 5′-to-3′ Strand Elongation 59
The Polymerase Chain Reaction 60
2.5 **The Terminology of Genetic Analysis** 64
2.6 **Types of DNA Markers Present in Genomic DNA** 65
Single Nucleotide Polymorphisms (SNPs) 66
Restriction Fragment Length Polymorphisms (RFLPs) 66
Connection: Origin of the Human Genetic Linkage Map 68
Random Amplified Polymorphic DNA (RAPD) 68
Amplified Fragment Length Polymorphisms (AFLPs) 71
Simple Tandem Repeat Polymorphisms (STRPs) 72
2.7 **Applications of DNA Markers** 73
Genetic Markers, Genetic Mapping, and "Disease Genes" 73
Other Uses for DNA Markers 74
Chapter End Material 76
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web



Chapter 3 Transmission Genetics: The Principle of Segregation 86

3.1 **Morphological and Molecular Phenotypes** 88

3.2 **Segregation of a Single Gene** 90
Phenotypic Ratios in the F_2 Generation **93**
The Principle of Segregation **94**
Verification of Segregation **95**
Connection: What Did Gregor Mendel Think He Discovered? 96
The Testcross and the Backcross **98**

3.3 **Segregation of Two or More Genes** 98
The Principle of Independent Assortment **99**
The Testcross with Unlinked Genes **101**
The Big Experiment **102**

3.4 **Human Pedigree Analysis** 104
Characteristics of Dominant and Recessive Inheritance **104**
Molecular Markers in Human Pedigrees **106**

3.5 **Pedigrees and Probability** 107
Connection: Troubadour 108
Mutually Exclusive Possibilities **109**
Independent Possibilities **110**

3.6 **Incomplete Dominance and Epistasis** 111
Multiple Alleles **111**
Human ABO Blood Groups **112**
Epistasis **115**

3.7 **Genetic Analysis: Mutant Screens and the Complementation Test** 117
The Complementation Test in Gene Identification **118**
Why Does the Complementation Test Work? **122**

Chapter End Material 123
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 4 Genes and Chromosomes 132

4.1 **The Stability of Chromosome Complements** 134

4.2 **Mitosis** 135

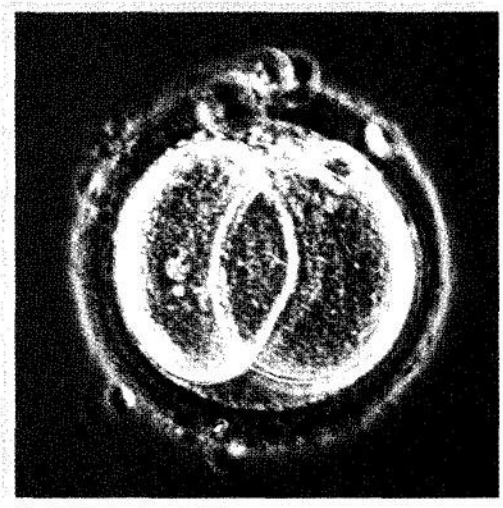
4.3 **Meiosis** 139
The First Meiotic Division: Reduction **141**
Connection: Grasshopper, Grasshopper 145
The Second Meiotic Division: Equation **148**

4.4 **Chromosomes and Heredity** 148
Chromosomal Determination of Sex **148**
X-linked Inheritance **150**
Pedigree Characteristics of Human X-linked Inheritance **154**
Connection: The White-Eyed Male 151
Heterogametic Females **155**
Nondisjunction as Proof of the Chromosome Theory of Heredity **155**
Sex Determination in *Drosophila* **157**

4.5 **Probability in the Prediction of Progeny Distributions** 158
Using the Binomial Distribution in Genetics **158**

4.6 **Testing Goodness of Fit to a Genetic Hypothesis** 161
The Chi-square Method **161**

4.7 **Are Mendel's Data Too Good to Be True?** 164
Connection: The Case Against Mendel's Gardener 166



Chapter End Material 166
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 5 Genetic Linkage and Chromosome Mapping 174

5.1 **Linkage and Recombination of Genes in a Chromosome 176**
Coupling versus Repulsion of Syntenic Alleles **178**
The Chi-Square Test for Linkage **179**
Each Pair of Linked Genes Has a Characteristic Frequency of Recombination **180**
Recombination in Females versus Males **180**

5.2 **Genetic Mapping 181**
Map Distance and Frequency of Recombination **182**
Crossing-Over **186**
Crossing-Over Takes Place at the Four-Strand Stage of Meiosis **187**
Multiple Crossovers **189**
Connection: Genes All in a Row 190

5.3 **Genetic Mapping in a Three-Point Testcross 192**
Chromosome Interference in Double Crossovers **194**
Genetic Mapping Functions **195**
Genetic Map Distance and Physical Distance **196**

5.4 **Genetic Mapping in Human Pedigrees 197**
Connection: Human Gene Map 198
Maximum Likelihood and Lod Scores **198**

5.5 **Mapping by Tetrad Analysis 201**
Analysis of Unordered Tetrads **202**
Analysis of Ordered Tetrads **206**

5.6 **Recombination Within Genes 209**

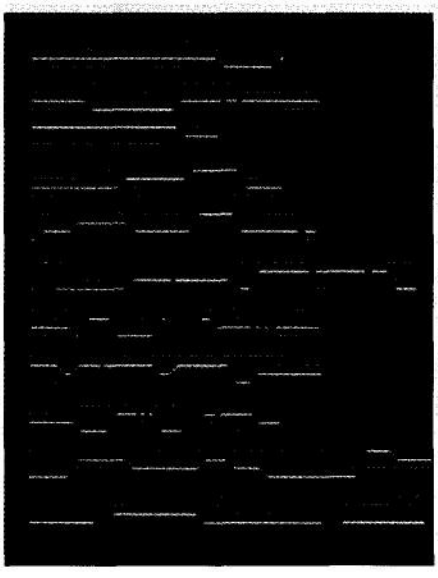
5.7 **Mitotic Recombination 209**

Chapter End Material 211
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 6 Molecular Biology of DNA Replication and Recombination 222

6.1 **Problems of Initiation, Elongation, and Incorporation Error 224**

6.2 **Semiconservative Replication of Double-Stranded DNA 225**
The Meselson-Stahl Experiment **225**
Connection: Replication by Halves 228
Semiconservative Replication of DNA in Chromosomes **228**
Theta Replication of Circular DNA Molecules **230**
Rolling-Circle Replication **231**
Multiple Origins and Bidirectional Replication in Eukaryotes **232**



6.3 **Unwinding, Stabilization, and Stress Relief 234**

6.4 **Initiation by a Primosome Complex 235**

6.5 **Chain Elongation and Proofreading 236**

6.6 **Discontinuous Replication of the Lagging Strand 238**
Fragments in the Replication Fork **239**
The Joining of Precursor Fragments **240**

6.7 **Terminator Sequencing of DNA 242**
The Sequencing Procedure **243**
Shotgun Sequencing **244**
Clinical Use of Dideoxynucleoside Analogs **247**

6.8 **Molecular Mechanisms of Recombination 248**
Gene Conversion and Mismatch Repair **248**
The Holliday Model of Recombination **250**
Single-Strand Break and Repair Model **253**
Connection: Happy Holliday 254
Double-Strand Break and Repair Model **254**

Chapter End Material 256
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 7 Molecular Mechanisms of Mutation and DNA Repair 264

7.1 **Types of Mutations 266**
Germ-Line and Somatic Mutations **266**
Conditional Mutations **266**
Classification by Function **267**

7.2 **The Molecular Basis of Mutation 268**
Nucleotide Substitutions **268**
Missense Mutations: The Example of Sickle-Cell Anemia **269**
Insertions, Deletions, and Frameshift Mutations **271**
Dynamic Mutation of Trinucleotide Repeats **272**

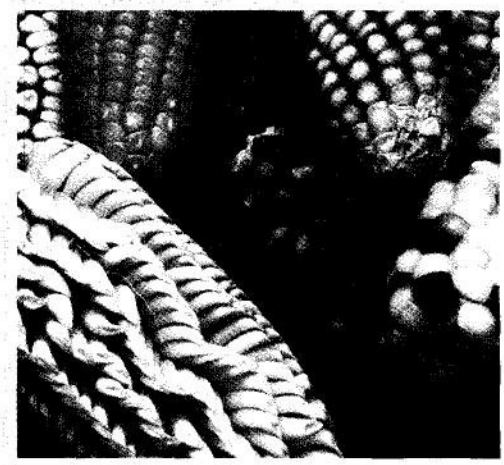
7.3 **Transposable Elements 274**
Connection: Transposable Elements 274
Molecular Mechanisms of Transposition **275**
Transposable Elements as Agents of Mutation **278**
RIP: A Defense Against Transposons **280**

7.4 **Spontaneous Mutation 280**
The Nonadaptive Nature of Mutation **281**
Estimation of Mutation Rates **282**
Hot Spots of Mutation **284**

7.5 **Mutagens 286**
Base-Analog Mutagens **286**
Chemical Agents That Modify DNA **288**
Misalignment Mutagenesis **289**
Ultraviolet Irradiation **289**
Ionizing Radiation **290**
Connection: X-Ray Daze 291
Genetic Effects of the Chernobyl Nuclear Accident **293**

7.6 **Mechanisms of DNA Repair 295**
Mismatch Repair **295**
Connection: Replication Slippage in Unstable Repeats 297
AP Repair **298**
Photoreactivation **299**
Excision Repair **299**
Postreplication Repair **299**
The SOS Repair System **300**

7.7 **Reverse Mutations and Suppressor Mutations 300**
Intragenic Suppression **301**
Intergenic Suppression **302**
The Ames Test for Mutagen/Carcinogen Detection **302**



Chapter End Material 303
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 8 Molecular Organization of Chromosomes 310

8.1 **Genome Size and Evolutionary Complexity: The C-Value Paradox 312**

8.2 **The Supercoiling of DNA 314**
Topoisomerase Enzymes **315**

8.3 **The Structure of Bacterial Chromosomes 316**

8.4 **The Structure of Eukaryotic Chromosomes 318**
The Nucleosome—The Structural Unit of Chromatin **318**
The Nucleosome Core Particle **318**
The Arrangement of Chromatin Fibers in a Chromosome **320**

8.5 **Polytene Chromosomes 323**

8.6 **Repetitive Nucleotide Sequences in Eukaryotic Genomes 324**
Kinetics of DNA Renaturation **324**
Analysis of Genome Size and Repetitive Sequences by Renaturation Kinetics **327**
Connection: Post-Genomics Genetics 328

8.7 **Unique and Repetitive Sequences in Eukaryotic Genomes 329**
Unique Sequences **329**
Highly Repetitive Sequences **329**
Middle-Repetitive Sequences **330**

8.8 **Centromere and Telomere Structure** 330
Molecular Structure of the Centromere **331**
Molecular Structure of the Telomere **332**
Connection: Telomeres: The Beginning of the End 334

Chapter End Material 338
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 9 Human Karyotypes and Chromosome Behavior 344

9.1 **The Human Karyotype** 346
Standard Karyotypes **347**
The Centromere and Chromosome Stability **350**
Dosage Compensation of X-linked Genes **351**
The Calico Cat Is a Result of X-Chromosome Inactivation **353**
Connection: Lyonization of an X Chromosome 354
Pseudoautosomal Inheritance **355**
Gene Content and Evolution of the Y Chromosome **355**

9.2 **Chromosome Abnormalities in Spontaneous Abortion** 357
Down Syndrome and Other Viable Trisomies **358**
Connection: The First Human Chromosomal Disorder Identified 358
Trisomic Segregation **361**
Sex-Chromosome Abnormalities **362**

9.3 **Chromosomal Deletions and Duplications** 362
Deletions **363**
Duplications **364**
Unequal Crossing-Over in Red-Green Color Blindness **365**

9.4 **Genetics of Chromosomal Inversions** 367
Paracentric Inversion (Not Including the Centromere) **368**
Pericentric Inversion (Including the Centromere) **368**

9.5 **Chromosomal Translocations** 370
Reciprocal Translocations **370**
Translocation Complexes in *Oenothera* **372**
Robertsonian Translocations **373**
Translocations and Trisomy 21 **375**

9.6 **Genomic Position Effects on Gene Expression** 375

9.7 **Polyploidy in Plant Evolution** 377
Autopolyploids and Allopolyploids **379**
Monoploid Oganisms **381**

9.8 **Genome Evolution in the Grass Family (Gramineae)** 382

Chapter End Material 384
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 10 Genetics of Bacteria and Their Viruses 392

10.1 **Mobile DNA** 394
Plasmids **394**
The F plasmid: A Conjugative Plasmid **395**
Insertion Sequences and Transposons **396**
Mobilization of Nonconjugative Plasmids **397**
Integrons and Antibiotic-Resistance Cassettes **398**
Multiple-Antibiotic-Resistant Bacteria **401**

10.2 **Bacterial Genetics** 402
Mutant Phenotypes **402**
Mechanisms of Genetic Exchange **403**

10.3 **DNA-Mediated Transformation** 403

10.4 **Conjugation** 405
Cointegrate Formation and Hfr Cells **405**
Time-of-Entry Mapping **408**
F′ Plasmids **411**

10.5 **Transduction** 412
Connection: The Sex Life of Bacteria 413
The Phage Lytic Cycle **413**
Generalized Transduction **413**
Connection: Origin of Phage Genetics 417

10.6 **Bacteriophage Genetics** 418
Plaque Formation and Phage Mutants **418**
Genetic Recombination in the Lytic Cycle **418**
Fine Structure of the *rII* Gene in Bacteriophage T4 **424**

10.7 **Lysogeny and Specialized Transduction** 427
Connection: Artoo 428

Site-Specific Recombination and Lysogeny **429**
Specialized Transduction **433**

Chapter End Material 435
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 11 Molecular Biology of Gene Expression 444

11.1 Amino Acids, Polypeptides, and Proteins 446
Protein Folding **446**
What Are the Minimal Genetic Functions Needed for Life? **449**

11.2 Colinearity Between Coding Sequences and Polypeptides 449
Connection: One Gene, One Enzyme 450

11.3 Transcription 452
Overview of RNA Synthesis **452**
Types of RNA Polymerase **452**
Mechanism of Transcription **454**
Genetic Evidence for Promoters and Terminators **456**

11.4 Messenger RNA 456
Connection: Messenger "Light" 457

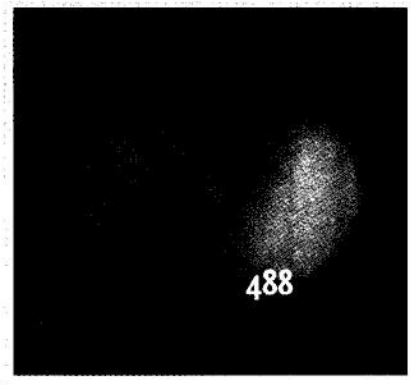
11.5 RNA Processing in Eukaryotes 458
5′ Capping and 3′ Polyadenylation **458**
Splicing of Intervening Sequences **458**
Mutations in Introns **462**
Exon Shuffle in the Origin of New Genes **462**

11.6 Translation 463
Initiation by mRNA Scanning **464**
Elongation **466**
Release **468**

11.7 Complex Translation Units 468
Polysomes **468**
Polycistronic mRNA **470**

11.8 The Standard Genetic Code 470
Connection: Uncles and Aunts 472
Genetic Evidence for a Triplet Code **472**
How the Code Was Cracked **475**
Features of the Standard Code **476**
Transfer RNA and Aminoacyl-tRNA Synthetase Enzymes **477**
Redundancy and Wobble **478**
Nonsense Suppression **479**

Chapter End Material 482
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web



Chapter 12 Molecular Mechanisms of Gene Regulation

12.1 Transcriptional Regulation in Prokaryotes 490
Inducible and Repressible Systems of Negative Regulation **490**
Positive Regulation **491**

12.2 Lactose Metabolism and the Operon System of Gene Regulation 492
Lac⁻ Mutants **492**
Inducible and Constitutive Synthesis and Repression **493**
The Repressor **494**
The Operator Region **495**
The Promoter Region **495**
The Operon System of Transcriptional Regulation **495**
Positive Regulation of the Lactose Operon **497**

12.3 Regulation of the Tryptophan Operon 499
Connection: Operator? Operator? 500
Attenuation **502**

12.4 Regulation in Bacteriophage Lambda 505

12.5 Transcriptional Regulation in Eukaryotes 507
Galactose Metabolism in Yeast **508**
Transcriptional Activator Proteins **510**
Transcriptional Enhancers and Transcriptional Silencers **510**
The Eukaryotic Transcription Complex **511**
Chromatin-Remodeling Complexes **514**
Alternative Promoters **516**

12.6 Epigenetic Mechanisms of Transcriptional Regulation 518

Cytosine Methylation **518**
Methylation and Transcriptional Inactivation **519**
Cosuppression Through Transcriptional Silencing **520**
Genomic Imprinting in the Female and Male Germ Lines **520**

12.7 **Regulation Through RNA Processing and Decay 522**
Alternative Splicing **522**
Messenger RNA Stability **522**
Cosuppression Through RNA Turnover **523**
RNA Interference (RNAi) **524**

12.8 **Translational Control 524**

12.9 **Programmed DNA Rearrangements 524**
Gene Amplification **525**
Antibody and T-Cell Receptor Variability **525**
Mating Type Interconversion **528**
Transcriptional Control of Mating Type **529**
Connection: Sex-Change Operations 530

Chapter End Material 532
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 13 Genetic Engineering and Genomics 540

13.1 **Restriction Enzymes and Vectors 542**
Production of Defined DNA Fragments **542**
Recombinant DNA Molecules **544**
Plasmid, Lambda, and Cosmid Vectors **545**

13.2 **Cloning Strategies 546**
Joining DNA Fragments **547**
Insertion of a Particular DNA Molecule into a Vector **549**
The Use of Reverse Transcriptase: cDNA and RT-PCR **550**



13.3 **Detection of Recombinant Molecules 551**
Gene Inactivation in the Vector Molecule **551**
Screening for Particular Recombinants **554**
Positional Cloning **554**

13.4 **Transgenic Organisms 556**
Germ-Line Transformation in Animals **556**
Connection: Hello, Dolly 558
Genetic Engineering in Plants **560**
Transformation Rescue **561**

13.5 **Site-Directed Mutagenesis and Knockout Mutations 562**

13.6 **Some Applications of Genetic Engineering 564**
Giant Salmon with Engineered Growth Hormone **564**
Nutritionally Engineered Rice **565**
Production of Useful Proteins **566**
Genetic Engineering with Animal Viruses **566**

13.7 **Analysis of Complex Genomes 567**
Electrophoretic Separation of Large DNA Fragments **567**
Cloning of Large DNA Fragments **569**
Connection: Catch-22 572
Physical Mapping **572**
High Throughput Genomic Sequencing and the Human Genome Project **574**
What Can We Learn From Genomic Sequencing? **576**

13.8 **Functional Genomics 578**
DNA Chips **578**

Chapter End Material 581
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 14 Genetic Control of Development 590

14.1 **Genetic Determinants of Development 592**

14.2 **Early Embryonic Development in Animals 594**
Autonomous Development and Intercellular Signaling **595**
Early Development and Activation of the Zygote Genome **597**
Composition and Organization of Oocytes **598**

14.3 **Genetic Analysis of Development in the Nematode 598**

Analysis of Cell Lineages **599**
Mutations Affecting Cell Lineages **600**
Types of Lineage Mutations **600**
Connection: Distinguished Lineages 602
Loss-of-Function and Gain-of-Function Alleles **604**

14.4 **Genetic Control of Development in *Drosophila* 607**
Maternal-Effect Genes and Zygotic Genes **609**
Genetic Basis of Pattern Formation in Early Development **610**
Coordinate Genes **612**
Connection: Embryo Genesis 613
Gap Genes **614**
Pair-Rule Genes **614**
Segment-Polarity Genes **616**
Metamorphosis of the Adult Fly **616**
Homeotic Genes **618**
HOX Genes in Evolution **620**

14.5 **Genetic Control of Development in Higher Plants 622**
Flower Development in *Arabidopsis* **623**
Combinatorial Determination of the Floral Organs **624**

Chapter End Material 626
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 15 Molecular Genetics of the Cell Cycle and Cancer 634

15.1 **The Cell Cycle 636**
Key Events in the Cell Cycle **636**
Transcriptional Program of the Cell Cycle **638**

15.2 **Genetic Analysis of the Cell Cycle 640**
Mutations Affecting Progression Through the Cell Cycle **640**

15.3 **Progression Through the Cell Cycle 642**
Cyclins and Cyclin-Dependent Protein Kinases **642**
Connection: Cycle-Ops 644
Targets of the Cyclin CDK Complexes **644**
Triggers for the G_1/S and G_2/M Transitions **646**
Protein Degradation Helps Regulate the Cell Cycle **647**

15.4 **Checkpoints in the Cell Cycle 648**
The DNA Damage Checkpoint **649**
The Centrosome Duplication **653**
The Spindle Checkpoint **654**

15.5 **Cancer Cells 655**
Oncogenes and Proto-Oncogenes **656**
Tumor-Suppressor Genes **659**

15.6 **Hereditary Cancer Syndromes 660**
Defects in Cell-Cycle Regulation and Checkpoints **660**
Defects in DNA Repair **664**
Connection: Two Hits, Two Errors 652

15.7 **Genetics of the Acute Leukemias 666**

Chapter End Material 667
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 16 Extranuclear Inheritance 676

16.1 **Patterns of Extranuclear Inheritance 678**
Mitochondrial Genetic Diseases **678**
Heteroplasmy **680**
Maternal Inheritance and Maternal Effects **678**

16.2 **Organelle Heredity 682**
RNA Editing **683**
The Genetic Codes of Organelles **683**
Leaf Variegation in Four-O'clock Plants **684**
Connection: *Chlamydomonas* Moments 686
Drug Resistance in *Chlamydomonas* **686**
Respiration-Defective Mitochondrial Mutants **688**
Cytoplasmic Male Sterility in Plants **688**

16.3 **The Evolutionary Origin of Organelles 691**

16.4 **Cytoplasmic Transmission of Symbionts 691**
Connection: A Coming Together 694

16.5 **Maternal Effect in Snail Shell Coiling 695**

16.6 **In Search of Mitochondrial "Eve" 696**

Chapter End Material 699
Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 17 Population Genetics and Evolution 706

17.1 **Allele Frequencies and Genotype Frequencies** 708

17.2 **The Matter of Mating** 711

Random Mating and the Hardy–Weinberg Principle **712**

Implications of the Hardy–Weinberg Principle **713**

A Test for Random Mating **713**

Frequency of Heterozygous Genotypes **714**

Multiple Alleles **715**

Connection: Twin Bill 715

DNA Typing **716**

DNA Exclusions **720**

Differences Among Populations **720**

X-Linked Genes **721**

17.3 **Inbreeding** 722

The Inbreeding Coefficient **722**

Allelic Identity by Descent **724**

Calculation of the Inbreeding Coefficient from Pedigrees **725**

Effects of Inbreeding **726**

17.4 **Genetics and Evolution** 727

17.5 **Mutation and Migration** 728

Irreversible Mutation **728**

Reversible Mutation **729**

17.6 **Natural Selection** 730

Selection in a Laboratory Experiment **731**

Selection in Diploid Organisms **732**

Components of Fitness **732**

Selection–Mutation Balance **733**

Heterozygote Superiority **735**

Connection: Resistance in the Blood 736

17.7 **Random Genetic Drift** 736

Chapter End Material 740

Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Chapter 18 The Genetic Basis of Complex Inheritance 748

18.1 **Complex Traits** 750

Continuous, Meristic, and Threshold Traits **751**

The Normal Distribution **751**

18.2 **Causes of Variation** 754

Connection: A Maize'n Grass 754

Genotypic Variation **755**

Environmental Variation **757**

Genetics and Environment Combined **758**

Genotype–Environment Interaction and Association **759**

18.3 **Genetic Analysis of Complex Traits** 760

The Number of Genes Affecting Complex Traits **761**

Broad-Sense Heritability **762**

Twin Studies **762**

18.4 **Artificial Selection** 763

Narrow-Sense Heritability **764**

Phenotypic Change with Individual Selection: A Prediction Equation **764**

Long-Term Artificial Selection **766**

Inbreeding Depression and Heterosis **766**

18-5 **Correlation Between Relatives** 767

Covariance and Correlation **767**

The Geometrical Meaning of a Correlation **768**

Estimation of Narrow-Sense Heritability **769**

18-6 **Heritabilities of Threshold Traits** 770

18-7 **Identification of Genes Affecting Complex Traits** 772

Linkage Analysis in the Genetic Mapping of Quantitative Trait Loci **772**

Connection: Win, Place, or Show? 774

Candidate Genes for Complex Traits **774**

Chapter End Material 777

Chapter Summary • Key Terms • Review the Basics • Guide to Problem Solving • Analysis and Applications • Challenge Problems • Further Reading • GeNETics on the Web

Answers to Even-Numbered Problems 786

Concise Dictionary of Genetics 804

Index 834

PREFACE

GENETICS: ANALYSIS OF GENES AND GENOMES, Fifth Edition, embodies our belief that the time is ripe to teach transmission genetics, the molecular biology of gene expression and regulation, and evolutionary genetics as integrated subjects. For our own convenience in teaching genetics, and to fit the manner in which the overwhelming majority of college courses in genetics have been organized (including our own), there is a group of chapters that most geneticists would label "classical" genetics, another group that would qualify as "molecular" genetics, and still another group that would undoubtedly be called "population" genetics. We distinguish between these subfields because we agree with most geneticists and teachers of genetics that the distinction is still useful. In our opinion this organizational framework is optimal, not only for convenience in teaching, but also for transmitting to students some sense of the history of advances in genetics. Nevertheless, most modern geneticists recognize that the distinctions between these subfields have become increasingly artificial in the post-genomics era.

Classical or Molecular Genetics: Where to Start?

The distinction between classical genetics and molecular genetics began to appear about a half-century ago when the three-dimensional structure of DNA burst upon the scene. Methods for the direct study of DNA (and proteins) were then quite new and still in the earliest stages of development. The subfield of "classical" genetics became defined as the study of the patterns in which alternative organismic traits (phenotypes) are transmitted in family histories (pedigrees). The critical generalization was the fruit of Mendel's experiments: the physical entities that are transmitted are not the traits themselves but "genes." Classical genetics took upon itself the role of recognizing different forms of genes through their effects on visible traits and on mapping the genes along the chromosomes. Those working in the subfield of "molecular" genetics could not have cared less about genetic transmission. In this arena the questions were: "At the molecular level, precisely how do alternative forms of a gene differ, how do they result in different expressed traits, and how does one alternative of a gene mutate into another alternative?" At this time the subfield of "population" genetics was almost divorced from the other two. Population genetics set for itself the task of understanding the patterns of transmission of complex traits whose expression depends on multiple genes as well as upon environmental factors, and not knowing either the genetics or the molecular biology, an adherence to statistical means, variances, and covariances was the only way to stay afloat.

But we are now (quite by coincidence, and happily) in a different millenium from that in which genetics as a science developed, and the previously

separate subdisciplines have been integrated. For example, many modern geneticists do population studies of DNA markers in individuals in pedigrees in which an inherited risk factor for a disease may be present. The goal is to find a genetic linkage between the occurrence of the disease and a genetic marker in the chromosomes. Then an examination of the complete genomic sequence of the chromosome yields a set of candidate genes in the region, which the geneticist uses to propose hypotheses for the molecular basis of the increased risk; these are then tested by further pedigree, population, and molecular studies. Is this approach "classical," "molecular," or "population" genetics? It is, in fact, all of the above, integrated.

Integration. That has been our goal in this edition. We discovered that an integration of the subfields can be achieved almost effortlessly and without sacrificing any of the traditional emphasis on such classical subjects as Mendel's experiments. One practical advantage of this approach is that instructors no longer need to decide "Should I teach classical genetics first? Or should I teach molecular genetics first?" With an integrated approach, there is no longer a need to make this choice. Closely related aspects of both classical and molecular genetics are taught together, and they illuminate and reinforce each other. The appeal to students is tremendous. In previous years both of us have watched our undergraduates' eyes glaze over when we first mentioned Mendel (as if, judging from the drooping eyelids, they were thinking, "Oh, we know this round and wrinkled stuff; wake us up when you come to something relevant to those of us in the modern world.") This book is it: an integrated view of the modern world of genetics, with classical genetics, molecular genetics, and population genetics each taught with regard to its separate goals, aspirations, and history, but taught as one unified field that, even in our post-genomics era, still goes by the name "genetics."

Instructional Goals

The overall aim of *Genetics: Analysis of Genes and Genomes*, Fifth Edition, is to provide a clear, comprehensive, rigorous, and balanced introduction to genetics and genomics at the college level. It is a guide to learning critically important and sometimes difficult subjects. It is our belief that a good course should maintain the right balance between two important aspects of genetics. The first aspect is that genetics is a body of knowledge pertaining to genetic transmission, function, and mutation; the second is that genetics is an experimental approach, or a kit of "tools," for the study of biological processes such as development or behavior. The rationale of the book is that any student claiming a knowledge of genetics must:

- Understand the basic processes of gene transmission, mutation, expression, and regulation;
- Be familiar with the principal experimental methods that geneticists and molecular biologists use in their studies, and recognize the advantages and limitations of these approaches;
- Be able to think like a geneticist at the elementary level of being able to formulate genetic hypotheses, work out their consequences, and test the results against observed data;
- Be able to solve problems of several types, including problems that ask the student to verbalize genetic principles in his or her own words, single-concept exercises that require application of definitions or the basic principles of genetics, problems in genetic analysis in which several concepts must be applied in logical order, and problems in quantitative analysis that call for some numerical calculation;
- Gain some sense of the social and historical context in which genetics and genomics has developed and is continuing to develop; and
- Have some familiarity with the genetic resources and information that are available through the Internet.

Special Features

This edition incorporates many special features to help students achieve these learning goals. The text is clearly and concisely written in a somewhat relaxed prose style without being chummy or excessively familiar. Each chapter is headed by a list of Principles that are related at numerous points to the larger whole. Each chapter contains two or three "Connections" in which the text material is connected to excerpts of classic papers that report key experiments in genetics or that raise important social, ethical, or legal issues in genetics. Each Connection has a brief introduction of its own, explaining the importance of the experiment and the historical context in which it was carried out. At the end of each chapter is a complete Summary, Key Terms, GeNETtics on the Web exercises that guide students in the use of Internet resources in genetics, and several different types and levels of Problems. These features are discussed individually below.

In recent decades, both the amount of genetic knowledge and its rate of growth have exploded. Many of the new discoveries have personal and social relevance through applications of genetics to human affairs in prenatal diagnosis, testing for carriers, and

identification of genetic risk factors for complex traits, such as breast cancer and heart disease. There are also ethical controversies: Should genetic manipulation be used on patients for the treatment of disease? Should human fetuses be used in research? Should human beings be cloned? There are also social controversies—for example, should there be laws governing genetic privacy, and how far should they go? Who should have access to genetic testing records, and for what purpose?

Inspired in part by the controversies and the publicity, many of today's students come to a course in genetics with great enthusiasm. The challenges for the teacher are:

- To sustain this enthusiasm;
- To help motivate a desire to understand the principles of genetics in a comprehensive and rigorous way;
- To guide students in gaining an understanding that genetics is not only a set of principles but also an experimental approach to solving a wide range of biological problems; and
- To help students learn to think about genetic problems and about the wider social and ethical issues arising from genetics.

While addressing these challenges, we have also tried to show the beauty, logical clarity, and unity of the subject. Endlessly fascinating, genetics is the material basis of the continuity of life.

Chapter Organization

In order to help the student keep track of the main issues and avoid being distracted by details, each chapter begins with a list of the Principles that provide the main focus of the chapter. There is also an Outline, showing step by step the path along the way. An opening paragraph gives an overview of the chapter, illustrates the subject with some specific examples, and shows how the material is connected to genetics as a whole. The text makes liberal use of numbered lists and "bullets" in order to help students organize their learning, as well as summary statements set off in special type in order to emphasize important principles. Each chapter ends with a Summary and list of Key Terms as well as the Problems. There is a Concise Dictionary of Genetics at the end of the book for students to check their understanding of the Key Terms or look up any technical terms they may have forgotten. The Dictionary includes not only the Key Terms but also genetic terms that students are likely to encounter in exploring the Internet or in their further reading.

Contents and Organization

The organization of the chapters is that favored by the majority of instructors who teach genetics. It is the organization we use in our own courses. An important feature is the presence of two introductory chapters providing a broad overview of DNA, genes, and genomes—what they are, how they function, how they change by mutation, and how they evolve through time. Today, most students learn about DNA in grade school or high school; in our teaching, we have found it artificial to pretend that DNA does not exist until the middle of the term. The introductory chapters serve to connect the more advanced concepts that students are about to learn with what they already know. It also serves to provide each student with a solid framework for integrating the material that comes later.

Throughout each chapter, there is a balance between observation and theory, between principle and concrete example, and between challenge and motivation. Molecular, classical, and evolutionary genetics are integrated throughout. Present throughout the book are frequent references to human genetics, including discussions of metabolic diseases, the fragile-X syndrome, imprinting, the genetic basis of cancer, expansion of unstable repeats in diseases such as Huntington disease, the relationship of DNA repair enzymes to hereditary colon cancer, the controversial mitochondrial "Eve," genetic diseases associated with defective mitochondria, DNA typing in individual identification, the Human Genome Project, and many other special topics. Although applications to human genetics are often stressed, the book is also liberally supplied with applications to other animals as well as plants.

A number of points related to organization and coverage should be noted:

- **Chapter 1** is an overview of genetics designed to bring students with disparate backgrounds to a common level of understanding. This chapter enables classical, molecular, and evolutionary genetics to be integrated in the rest of the book. Included in Chapter 1 are the basic concepts of molecular genetics: DNA structure, replication, expression, and mutation. Chapter 1 also includes a discussion of the experimental evidence that DNA is the genetic material, and introduces the concepts of genome and proteome.
- **Chapter 2** emphasizes that the primary tools of the modern geneticist derive from the experimental manipulation of DNA. It includes a more detailed look at DNA structure, and it introduces the principal methods of DNA manipulation including restriction enzymes, electrophoresis, DNA hy-

bridization, Southern blotting, and the polymerase chain reaction. We also discuss how and why these methods are used to study DNA markers such as SNPs, RFLPs, and other key approaches that form the basis of modern genetic analysis.

- **Chapters 3 through 5** are the core of Mendelian genetics, including segregation and independent assortment, the chromosome theory of heredity, mitosis and meiosis, linkage and chromosome mapping, and tetrad analysis in fungi. Also included is the basic probability framework of Mendelian genetics and the testing of genetic models by means of the chi-square test. Unique in Chapter 3 is the integration of molecular genetics with Mendel's experiments. We describe the molecular basis of the wrinkled mutation and show how a modern geneticist would carry out Mendel's study, examining the molecular phenotypes on the one hand and the morphological phenotypes on the other. This pedagogy provides a solid basis for understanding not only Mendel's experiments as he actually performed and interpreted them, but also for understanding how modern molecular approaches are used in genetic analysis. Molecular markers are also integrated in the discussion of human genetic analysis.

  An important principle of genetics, too often ignored or given inadequate treatment, is that of the complementation test and how complementation differs from segregation or other genetic principles. Chapter 3 includes a clear and concise description of complementation, with examples, showing how complementation is used in mutation screens to group mutations into categories corresponding to genes.
- **Chapters 6 through 8** deal with the molecular structure and replication of DNA and the molecular organization of chromosomes, as well as the molecular basis of mutation and the effects of mutagens, including the genetic effects of the Chernobyl nuclear accident. Chapter 6 also covers the molecular mechanisms of recombination and Chapter 7 the rapidly growing field of DNA repair. Chapter 8 includes a discussion of repetitive DNA sequences in eukaryotic genomes and the molecular structures of centromeres and telomeres.
- **Chapter 9** covers the principles of cytogenetics, including the human genome with special reference to human chromosome number and structure and the types of aberrations that are found in human chromosomes. The genetic implications of chromosome abnormalities—duplications, deficiencies, inversions, and translocations—are also discussed. We emphasize the importance of polyploidy in plant evolution, and examine genome evolution in the cereal grains as an example.
- **Chapter 10** deals with the principles of genetics in prokaryotes, beginning with the genetics of mobile DNA, plasmids, and integrons, and their relationships to the evolution of multiple antibiotic resistance. There is a thorough discussion of mechanisms of genetic recombination in microbes, including transformation, conjugation, and transduction, as well as discussion of temperate and virulent bacteriophages.
- **Chapters 11 and 12** deal with molecular genetics in the strict sense. These chapters include the classical principles of gene expression and gene regulation. For the first time we include broader aspects of gene regulation that are topics of much current research: chromatin remodeling complexes, imprinting and other epigenetic modifications n gene expression, transcriptional and post-transcriptional cosuppression, and RNA interference (RNAi).
- **Chapter 13** focuses on recombinant DNA and genomics. Included are the use of restriction enzymes and vectors in recombinant DNA, cloning strategies, site-directed mutagenesis, the production of genetically defined transgenic animals and plants, and applications of genetic engineering. Also discussed are methods used in large-scale genomic sequencing and a summary of what new principles have emerged from genomic sequencing. Functional genomics is introduced by examining how DNA microarrays ("DNA chips") are used to study global patterns of coordinated gene expression.
- **Chapters 14 and 15** deal with the genetic control of development. In Chapter 14 we focus on genetic analysis of development in nematodes (*Caenorhabditis elegans*) and *Drosophila*, and also include is a thorough examination of the genetic basis of floral development in *Arabidopsis thaliana*. Chapter 15 stresses cancer from the standpoint of the genetic control of cell division, with emphasis on the checkpoints that, in normal cells, result either in inhibition of cell division or in programmed cell death (apoptosis). Cancer results from a series of successive mutations, usually in somatic cells, that overcome the normal checkpoints that control cellular proliferation.
- **Chapter 16** covers organelle genetics, including genetic defects in human mitochondrial DNA.
- **Chapters 17 and 18** deal with population and evolutionary genetics. The discussion includes the population genetics of the CCR5 receptor muta-

tion that confers resistance to infection by HIV virus, as well as DNA typing in criminal investigations, paternity testing, the effects of inbreeding, and the evolutionary mechanisms that drive changes in allele frequency. The approach to quantitative genetics includes a discussion of how particular genes influencing quantitative traits (QTLs, or quantitative-trait loci) may be identified and mapped by linkage analysis. There is also a section on the genetic determinants of human behavior with examples of the approach using "candidate" genes that led to the identification of the "natural Prozac" polymorphism in the human serotonin transporter gene.

Connections

A unique special feature of this book is found in boxes called Connections. Each chapter has two or three of these boxes. They are our way of connecting genetics to the world outside the classroom. All of the Connections include short excerpts from the original literature of genetics, usually papers, each introduced with a short explanatory passage. Many of the Connections are excerpts from classic papers, such as Mendel's paper, but by no means all of the "classic" papers are old papers. Many of them are very recent, including the report of the first human chromosome sequenced, chromosome 22.

The pieces are called Connections because each connects the material in the text to something that broadens or enriches its implications. Some of the Connections raise issues of ethics in the application of genetic knowledge, social issues that need to be addressed, or issues related to the proper care of laboratory animals. They illustrate other things as well. Because each Connection names the place where the research was carried out, the student will learn that great science is done in many universities and research institutions throughout the world. Some of the pieces were published originally in French, others in German. These appear in English translation. In papers that use outmoded or unfamiliar terminology, or that use archaic gene symbols, we have substituted the modern equivalent because the use of a consistent terminology in the text and in the Connections makes the material more accessible to the student.

GeNETics on the Web

More than in most fields of biology, genetic resources and genetic information are abundant on the Internet. The most useful sites are not always easy to find. A recent search of Internet sites using the AltaVista search engine and the keyword genetics yielded about 800,000 hits. Most of these are of limited usefulness, but quite a few are invaluable to the student and to the practicing geneticist. The problem is how to find the really useful ones among the 800,000 sites.

To make the genetic information explosion on the Internet available to the student, we developed for the Fourth Edition and have updated for the Fifth Edition the feature called GeNETics on the Web, which make use of Internet resources. Genetics is a dynamic science, and through this venue we can introduce the newest discoveries as soon as they appear and keep the textbook up to date.

The addresses of the relevant genetic sites are not printed in the book. Instead, the sites are accessed through the use of key words that are highlighted in each exercise. The key words are maintained as hot links at the publisher's web site

http://www.jbpub.com/genetics

and are kept constantly up to date, tracking the address of each site if it should change. The use of key words also allows an innovation: one exercise in each chapter makes use of a mutable site that changes frequently.

Problems

Each chapter provides about 50 problems for solution, graded in difficulty, for the students to test their understanding. The problems are of several different types:

Guide to Problem Solving demonstrates problems worked in full. The concepts needed to solve the problem, and the reasoning behind the answer, are explained in detail. This feature serves as a review of the important concepts used in working problems. It also highlights some of the most common mistakes made by beginning students and gives pointers on how the student can avoid falling into these conceptual traps.

Review the Basics problems ask for genetic principles to be restated in the student's own words; some are matters of definition or call for the application of elementary principles.

Analysis and Applications problems are more traditional types of genetic problems in which several concepts must be applied in logical order and often require some numerical calculation. The level of mathematics is that of arithmetic and elementary probability as it pertains to genetics. None of the problems uses mathematics beyond elementary algebra.

Challenge Problems are similar to those in Analysis and Applications, but they are a degree more challenging, often because they require a more extensive analysis of data before the question can be answered.

Solutions

The answers to even-numbered Analysis and Applications problems are included in the answer sec-

tion at the end of the book. The answers are complete. They explain the logical foundation of the solution and lay out the methods. The answers to the rest of the Analysis and Applications and Challenge Problems are available in the *Student Solutions Manual and Supplemental Problems* book.

We find that many of our students, like students everywhere, often sneak a look at the answer before attempting to solve a problem. This is a pity. Working backward from the answer should be a last resort. This is because problems are valuable opportunities to learn. Problems that the student cannot solve are usually more important than the ones that can be solved, because the sticklers usually identify trouble spots, areas of confusion, or gaps in understanding. So, forever in hope but against all experience, we urge our students to try answering each question before looking at the answer.

Further Reading

Each chapter also includes recommendations for Further Reading for the student who either wants more information or who needs an alternative explanation for the material presented in the book. Some additional "classic" papers and historical perspectives are included.

Illustrations

The art program is spectacular, thanks to the creative efforts of J/B Woolsey Associates, with special thanks to John Woolsey and Patrick Lane. Every chapter is richly illustrated with beautiful graphics in which color is used functionally to enhance the value of each illustration as a learning aid. The illustrations are also heavily annotated with "process labels" explaining step-by-step what is happening at each level of the illustration. These labels make the art inviting as well as informative. They also allow the illustrations to stand relatively independently of the text, enabling the student to review material without rereading the whole chapter.

The art program is used not only for its visual appeal but also to increase the pedagogical value of the book:

- Characteristic colors and shapes have been used consistently throughout the book to indicate different types of molecules—DNA, mRNA, tRNA, and so forth. For example, DNA is illustrated in any one of a number of ways, depending on the level of resolution necessary for the illustration, and each time a particular level of resolution is depicted, the DNA is shown in the same way. It avoids a great deal of potential confusion that DNA, RNA, and proteins are represented in the same manner in every chapter.
- There are numerous full-color photographs of molecular models in three dimensions; these give a strong visual reinforcement of the concept of macromolecules as physical entities with defined three-dimensional shapes and charge distributions that serve as the basis of interaction with other macromolecules.
- The page design is clean, crisp, and uncluttered. As a result, the book is pleasant to look at and easy to read.

Flexibility

There is no necessary reason to start at the beginning and proceed straight to the end. Each chapter is a self-contained unit that stands on its own. This feature gives the book the flexibility to be used in a variety of course formats. Throughout the book, we have integrated molecular and classical principles, so you can begin a course with almost any of the chapters. Most teachers will prefer starting with the overview in Chapter 1 because it brings every student to the same basic level of understanding. Chapter 2 introduces the basic experimental manipulations used in modern genetics and serves to integrate molecular and classical genetics in the discussion of Mendel in Chapter 3. Teachers preferring the Mendel-early format should start with Chapter 1, continue with Chapter 3, then backtrack to Chapter 2. Some teachers are partial to a chromosomes-early format, which would suggest the order Chapter 1, 4, 2, 3, and 5. A novel approach would be a genomes-first format, which could be implemented by beginning with Chapters 1, 2, and 13. The writing and illustration program was designed to accommodate a variety of formats, and we encourage teachers to take advantage of this flexibility in order to meet their own special needs.

Supplements

An unprecedented offering of traditional and interactive multimedia supplements is available to assist instructors and aid students in mastering genetics. Additional information and review copies of any of the following items are available through your Jones and Bartlett Sales Representative.

Instructor's Tool Kit CD-ROM includes:

- The Computerized TestBank contains over a thousand questions and complete answers. It uses Diploma software, a format which allows the instructor to easily generate tests and quizzes from a single chapter or collection of chapters. Diploma allows you to create paper, Internet, or network tests. The questions, authored by Elena Lozovsky of Harvard University, are a mix of factual, descriptive,

and quantitative types. A typical chapter contains multiple-choice, fill-in-the-blank, and short answer questions. In addition to being available in the convenient Diploma format, each chapter's TestBank is also provided as Rich Text files.

- The Lecture Success Image Bank is an easy-to-use multimedia tool which provides many of the illustrations from the text specially enhanced for classroom presentation. You can select images you need by chapter, topic, and figure number and easily generate your own slide shows, print the files for transparency creation, or insert the images into the PowerPoint presentations already prepared for you.
- Microsoft Word files of the *Student Solutions Manual and Supplemental Problems* are provided. This manual is available to your students as a printed supplemental text. It provides detailed answers for all of the problems at the end of chapters. It also contains supplemental problems with explanations to help give your students further practice with difficult genetics concepts.
- A PowerPoint presentation containing the detailed outline for each chapter of *Genetics: Analysis of Genes and Genomes*, 5/e is also included. The presentation, designed to mirror the text, is flexibly constructed in order to meet your lecture's organization. The outline is open, allowing you to provide the elements you deem necessary, whether it be new text or the images from the Lecture Success Image Bank.

GenNETics on the Web

Corresponding to the end-of-chapter GeNETics on the Web exercises, this World Wide Web site offers genetics-related links, articles and monthly updates to other genetics sites on the Web. Material for this site is carefully selected and updated by the authors. Jones and Bartlett Publishers ensures that links for the site are regularly maintained. Visit the GeNETics on the Web site at

http://www.jbpub.com/genetics

Student Solutions Manual and Supplemental Problems

This supplemental text, available to all students, is comprised of two parts. Firstly, it contains a complete set of solutions for all of the end-of-chapter problems in the text. Secondly, it contains hundreds of supplemental problems with complete solutions. These problems, also written by Elena Lozovsky of Harvard University, are a great resource for students seeking an extra advantage. The problems are designed to test their knowledge of each chapter's content, but particularly to provide practice thinking and working through some of the more difficult concepts in genetics.

Cogito: *Electronic Companion to Genetics*™, Version 2.0

Cogito Learning Media Inc.

This Mac/IBM CD-ROM, prepared by Philip Anderson and Barry Ganetzky of the University of Wisconsin, Madison, reviews important genetics concepts covered in class using state-of-the-art interactive multimedia. It consists of hundreds of animations, diagrams, and videos that dynamically explain difficult concepts to students. In addition, it contains over 400 interactive multiple-choice, "drag and drop," true/false, and fill-in problems. Brand new to this version are review screens and self-test questions on genomics, a new HTML format, readable through any current Web browser on any platform, and improved navigation and quicker loading and playing of animations. These resources will prove invaluable to students in a self-study environment and to instructors as a lecture-enhancement tool. This CD-ROM is available for packaging exclusively with Jones and Bartlett Publishers texts.

Video Resource Library

A full complement of quality videos is available to qualified adopters. Genetics-related topics include: Origin and Evolution of Life, Human Gene Therapy, Biotechnology, the Human Genome Project, Oncogenes, and Science and Ethics.

The Gist of Genetics: Guide to Learning and Review

Written by Rowland H. Davis and Stephen G. Weller of the University of California, Irvine, this study aid uses illustrations, tables, and text outlines to review all of the fundamental elements of genetics. It includes extensive practice problems and review questions with solutions for self-check. The Gist helps students formulate appropriate questions and generate hypotheses that can be tested with classical principles and modern genetic techniques.

Visual Genetics Plus: Tutorial and Laboratory Simulations:

This Mac/IBM CD-ROM, created by Alan W. Day and Robert L. Dean of the University of Western Ontario, is already in use at over 200 institutions worldwide. *Visual Genetics* 3.0 continues to provide a unique, dynamic presentation tool for viewing key genetic and molecular processes in the classroom. With this new, greatly expanded version of the Virtual Genetics Lab 2.0, instructors can now assign 17 comprehensive lab simulations. You can also bring the lab into the class-

room, as the program allows you to perform on-screen tasks such as the selection of mutant colonies, using a pipette to make a dilution series, inoculating mutants to petri dishes to test for response to growth factors, and then to analyze and interpret the data. Through the testing feature and presentation capabilities, you can offer a complete lab environment. Site Licenses and Instructor Copies are available.

Acknowledgments

We are indebted to many colleagues whose advice and thoughtful reviews, comments, and advice have helped in the preparation of this and previous editions. Their expert recommendations are reflected in the content, organization, and presentation of the material.

Laura Adamkewicz, George Mason University, Fairfax VA

Jeremy C. Ahouse, Brandeis University, Waltham MA

Mary Alleman, Duquesne University, Pittsburgh PA

Peter D. Ayling, University of Hull, Hull UK

John C. Bauer, Stratagene, Inc., La Jolla CA

Anna W. Berkovitz, Purdue University, West Lafayette IN

Mary K. B. Berlyn, Yale University, New Haven CT

Pierre Carol, Université Joseph Fourier, Grenoble, France

John Celenza, Boston University, Boston MA

Alan C. Christensen, University of Nebraska–Lincoln, Lincoln NE

Leslie Dendy, University of New Mexico, Los Alamos NM

Stephen J. DíSurney, University of Mississippi, University MS

John W. Drake, National Institute of Environmental Health Sciences, Research Triangle Park, NC

Kathleen Dunn, Boston College, Boston MA

Chris Easton, State University of New York, Binghamton NY

Wolfgang Epstein, University of Chicago, Chicago IL

Robert G. Fowler, San Jose State University, San Jose CA

David W. Francis, University of Delaware, Newark DE

Dan Garza, Florida State University, Tallahassee FL

Gail Gasparich, Towson University, Towson MD

Elliott S. Goldstein, Arizona State University, Tempe AZ

Jeffrey C. Hall, Brandeis University, Waltham MA

Mark L. Hammond, Campbell University, Buies Creek NC

Steven Henikoff, Fred Hutchinson Cancer Research Center, Seattle WA

Charles Hoffman, Boston College, Boston MA

Ivan Huber, Fairleigh Dickenson University, Madison NJ

Kerry Hull, Bishop's University, QC Canada

Lynn A. Hunter, University of Pittsburgh, Pittsburgh PA

Richard Imberski, University of Maryland, College Park MD

Joyce Katich, Monsanto, Inc., St. Louis MO

Jeane M. Kennedy, Monsanto, Inc., St. Louis MO

Jeffrey King, University of Berne, Switzerland

Yan B. Linhart, University of Colorado, Boulder CO

K. Brooks Low, Yale University, New Haven CT

Sally A. MacKenzie, Purdue University, West Lafayette IN

Gustavo Maroni, University of North Carolina, Chapel Hill NC

Jeffrey Mitton, University of Colorado, Boulder CO

Robert K. Mortimer, University of California, Berkeley CA

Gisela Mosig, Vanderbilt University, Nashville TN

Steve O'Brien, National Cancer Institute, Frederick MD

Kevin O'Hare, Imperial College, London UK

Ronald L. Phillips, University of Minnesota, St. Paul MN

Robert Pruitt, Purdue University, West Lafayette IN

Peggy Redshaw, Austin College, Sherman TX

Pamela Reinagel, California Institute of Technology, Pasadena CA

Kenneth E. Rudd, National Library of Medicine, Bethesda MD

Thomas F. Savage, Oregon State University, Corvallis OR

David Shepard, University of Delaware, Newark DE

Leslie Smith, National Institute of Environmental Health Sciences, Research Triangle Park NC

Charles Staben, University of Kentucky, Lexington KY

Johan H. Stuy, Florida State University, Talahassee FL

David T. Sullivan, Syracuse University, Syracuse NY

Jeanne Sullivan, West Virginia Wesleyan College, Buckhannon WV

Irwin Tessman, Purdue University, West Lafayette IN

James H. Thomas, University of Washington, Seattle WA

David Ussery, The Technical University of Denmark, Lyngby, Denmark

Denise Wallack, Muhlenberg College, Allentown PA

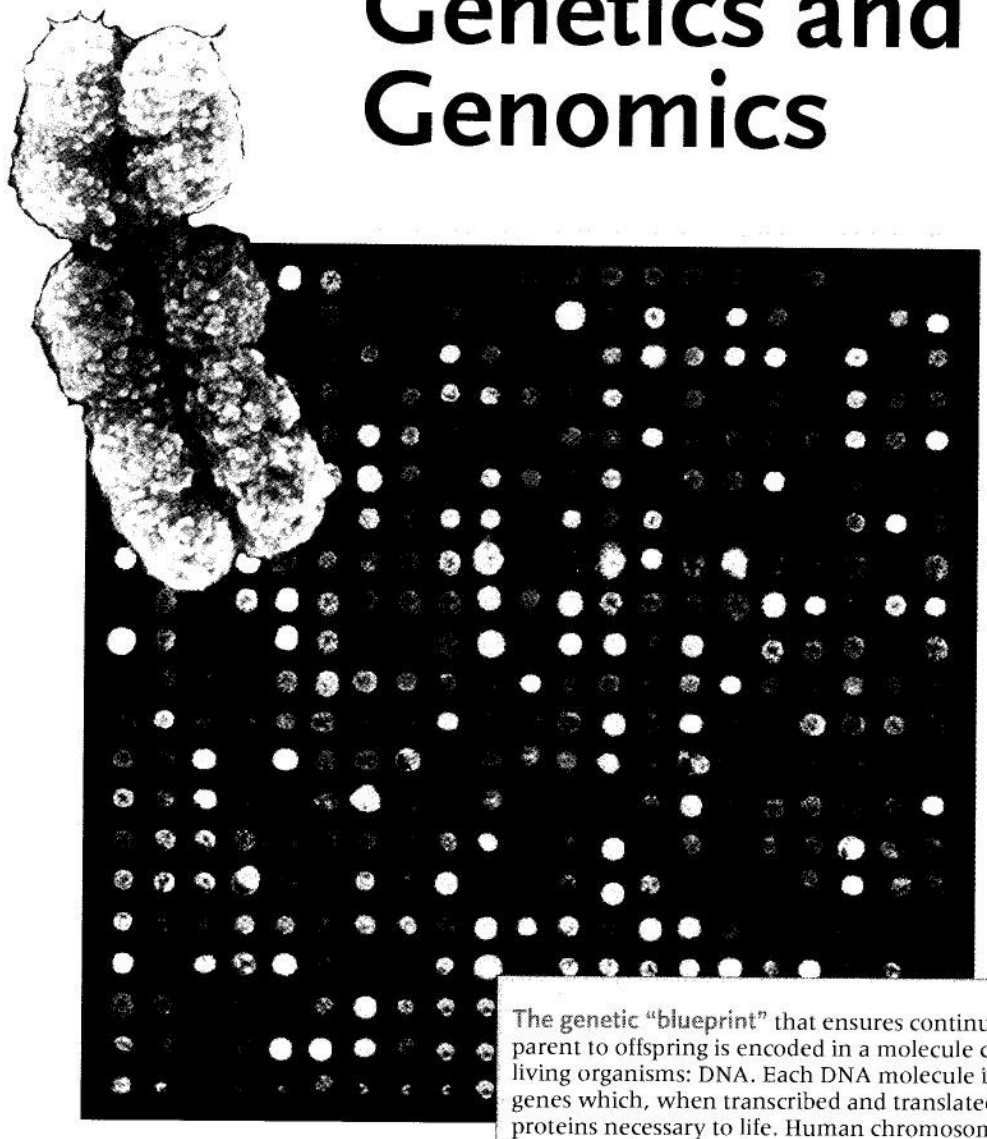
Kenneth E. Weber, University of Southern Maine, Gorham ME

We especially want to acknowledge Elena Lozovsky of Harvard University for her work on the *Testbank, Student Solutions Manual,* and *Supplemental Problems.*

We also wish to acknowledge the superb art, production, and editorial staff who helped make this book possible: Tom Manning, Brian McKean, Judy Hauck, Anne Spencer, Therese Bräuer, Mary Hill, Tim Gleeson, and Elizabeth Pearson of Jones and Bartlett; Andrea Fincke, Bonnie Van Slyke, and Sally Steele of Thompson Steele; John Woolsey and Patrick Lane of J/B Woolsey Associates. Much of the credit for the attractiveness and readability of the book should go to them. Thanks also to Jones and Bartlett, the publishers, for their ongoing commitment to high quality in book production. We are also grateful to the many people, acknowledged in the legends of the illustrations, who contributed photographs, drawings, and micrographs from their own research and publications, especially those who provided color photographs for this edition. Every effort has been made to obtain permission to use copyrighted material and to make full disclosure of its source. We are grateful to the authors, journal editors, and publishers for their cooperation. Any errors or omissions are wholly inadvertant and will be corrected at the first opportunity.

CHAPTER 1

Introduction to Molecular Genetics and Genomics



The genetic "blueprint" that ensures continuity of life from parent to offspring is encoded in a molecule common to all living organisms: DNA. Each DNA molecule is subdivided into genes which, when transcribed and translated, produce the proteins necessary to life. Human chromosome four, shown here, contains approximately 2000 genes. New technologies, like that of the DNA microarray–or "chip"–offer researchers the means to read genetic blueprints. [DNA chip courtesy of Jeffrey P. Townsend, Duccio Cavalieri, and the Harvard Center for Genomics Research.]

CHAPTER OUTLINE

1.1 **DNA: The Genetic Material**
Experimental Proof of the Genetic Function of DNA
Genetic Role of DNA in Bacteriophage

1.2 **DNA Structure: The Double Helix**

1.3 **An Overview of DNA Replication**

1.4 **Genes and Proteins**
Inborn Errors of Metabolism as a Cause of Hereditary Disease
Mutant Genes and Defective Proteins

1.5 **Gene Expression: The Central Dogma**
Transcription
Translation
The Genetic Code

1.6 **Mutation**
Protein Folding and Stability

1.7 **Genes and Environment**

1.8 **Evolution: From Genes to Genomes, from Proteins to Proteomes**
The Molecular Unity of Life
Natural Selection and Diversity

PRINCIPLES

- Inherited traits are affected by genes.
- Genes are composed of the chemical deoxyribonucleic acid (DNA).
- DNA replicates to form copies of itself that are identical (except for rare mutations).
- DNA contains a genetic code specifying what types of enzymes and other proteins are made in cells.
- DNA occasionally mutates, and the mutant forms specify altered proteins that have reduced activity or stability.
- A mutant enzyme is an "inborn error of metabolism" that blocks one step in a biochemical pathway for the metabolism of small molecules.
- Traits are affected by environment as well as by genes.
- Organisms change genetically through generations in the process of biological evolution.

CONNECTIONS

Shear Madness
Alfred D. Hershey and Martha Chase 1952
Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage

The Black Urine Disease
Archibald E. Garrod 1908
Inborn Errors of Metabolism

Each species of living organism has a unique set of inherited characteristics that makes it different from other species. Each species has its own developmental plan—often described as a sort of "blueprint" for building the organism—which is encoded in the DNA molecules present in its cells. This developmental plan determines the characteristics that are inherited. Because organisms in the same species share the same developmental plan, organisms that are members of the same species usually resemble one another, although some notable exceptions usually are differences between males and females. For example, it is easy to distinguish a human being from a chimpanzee or a gorilla. A human being habitually stands upright and has long legs, relatively little body hair, a large brain, and a flat face with a prominent nose, jutting chin, distinct lips, and small teeth. All of these traits are inherited—part of our developmental plan—and help set us apart as *Homo sapiens*.

But human beings are by no means identical. Many traits, or observable characteristics, differ from one person to another. There is a great deal of variation in hair color, eye color, skin color, height, weight, personality traits, and other characteristics. Some human traits are transmitted biologically, others culturally. The color of our eyes results from biological inheritance, but the native language we learned as a child results from cultural inheritance. Many traits are influenced jointly by biological inheritance and environmental factors. For example, weight is determined in part by inheritance but also in part by environment: how much food we eat, its nutritional content, our exercise regimen, and so forth. **Genetics** is the study of biologically inherited traits, including traits that are influenced in part by the environment.

The fundamental concept of genetics is that:

> Inherited traits are determined by the elements of heredity that are transmitted from parents to offspring in reproduction; these elements of heredity are called **genes**.

The existence of genes and the rules governing their transmission from generation to generation were first articulated by Gregor Mendel in 1866 (Chapter 3). Mendel's formulation of inheritance was in terms of the abstract rules by which hereditary elements (he called them "factors") are transmitted from parents to offspring. His objects of study were garden peas, with variable traits like pea color and plant height. At one time genetics could be studied only through the progeny produced from matings. Genetic differences between species were impossible to define, because organisms of different species usually do not mate, or they produce hybrid progeny that die or are sterile. This approach to the study of genetics is often referred to as *classical* genetics, or organismic or morphological genetics. Given the advances of *molecular*, or modern, genetics, it is possible to study differences between species through the comparison and analysis of the DNA itself. There is no fundamental distinction between classical and molecular genetics. They are different and complementary ways of studying the same thing: the function of the genetic material. In this book we include many examples showing how molecular and classical genetics can be used in combination to enhance the power of genetic analysis.

The foundation of genetics as a molecular science dates back to 1869, just three years after Mendel reported his experiments. It was in 1869 that Friedrich Miescher discovered a new type of weak acid, abundant in the nuclei of white blood cells. Miescher's weak acid turned out to be the chemical substance we now call **DNA (deoxyribonucleic acid).** For many years the biological function of DNA was unknown, and no role in heredity was ascribed to it. This first section shows how DNA was eventually isolated and identified as the material that genes are made of.

1.1 DNA: The Genetic Material

That the cell nucleus plays a key role in inheritance was recognized in the 1870s by the observation that the nuclei of male and female reproductive cells undergo fusion in the process of fertilization. Soon thereafter, **chromosomes** were first observed inside the nucleus as thread-like objects that become visible in the light microscope when the cell is stained with certain dyes. Chromosomes were found to exhibit a characteristic "splitting" behavior in which each daughter cell formed by cell division

receives an identical complement of chromosomes (Chapter 4). Further evidence for the importance of chromosomes was provided by the observation that, whereas the number of chromosomes in each cell may differ among biological species, the number of chromosomes is nearly always constant within the cells of any particular species. These features of chromosomes were well understood by about 1900, and they made it seem likely that chromosomes were the carriers of the genes.

By the 1920s, several lines of indirect evidence began to suggest a close relationship between chromosomes and DNA. Microscopic studies with special stains showed that DNA is present in chromosomes. Chromosomes also contain various types of proteins, but the amount and kinds of chromosomal proteins differ greatly from one cell type to another, whereas the amount of DNA per cell is constant. Furthermore, nearly all of the DNA present in cells of higher organisms is present in the chromosomes. These arguments for DNA as the genetic material were unconvincing, however, because crude chemical analyses had suggested (erroneously, as it turned out) that DNA lacks the chemical diversity needed in a genetic substance. The favored candidate for the genetic material was protein, because proteins were known to be an exceedingly diverse collection of molecules. Proteins therefore became widely accepted as the genetic material, and DNA was assumed to function merely as the structural framework of the chromosomes. The experiments described below finally demonstrated that DNA is the genetic material.

Experimental Proof of the Genetic Function of DNA

An important first step was taken by Frederick Griffith in 1928 when he demonstrated that a physical trait can be passed from one cell to another. He was working with two strains of the bacterium *Streptococcus pneumoniae* identified as S and R. When a bacterial cell is grown on solid medium, it undergoes repeated cell divisions to form a visible clump of cells called a **colony.** The S type of *S. pneumoniae* synthesizes a gelatinous capsule composed of complex carbohydrate (polysaccharide). The enveloping capsule makes each colony large and gives it a glistening or smooth (S) appearance. This capsule also enables the bacterium to cause pneumonia by protecting it from the defense mechanisms of an infected animal. The R strains of *S. pneumoniae* are unable to synthesize the capsular polysaccharide; they form small colonies that have a rough (R) surface (**Figure 1.1**). This strain of the bacterium does not cause pneumonia, because without the capsule the bacteria are inactivated by the immune system of the host. Both types of bacteria

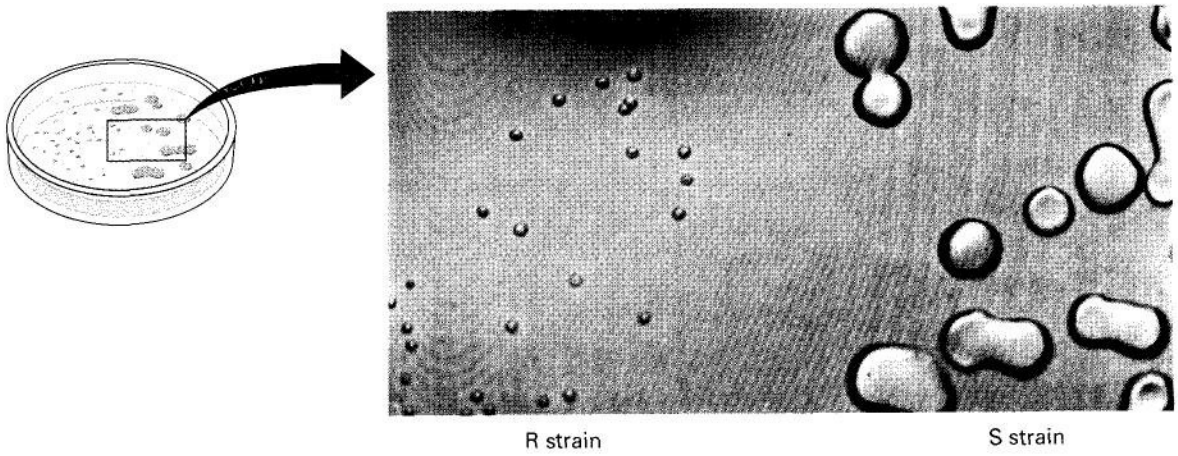


Figure 1.1 Colonies of rough (R, the small colonies) and smooth (S, the large colonies) strains of *Streptococcus pneumoniae.* The S colonies are larger because of the gelatinous capsule on the S cells. [Photograph from O. T. Avery, C. M. MacLeod, and M. McCarty. Reproduced from the *Journal of Experimental Medicine,* 1944, vol. 79, p. 137 by copyright permission of The Rockefeller University Press.]

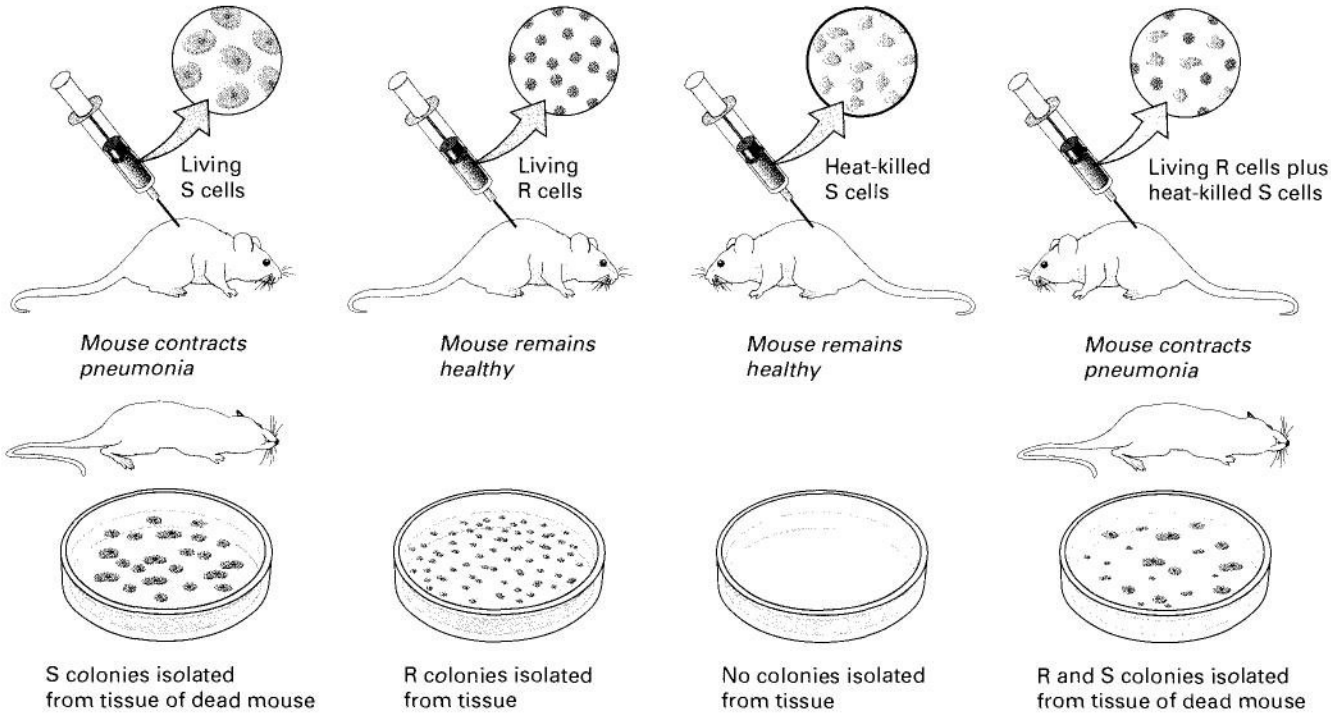


Figure 1.2 The Griffith's experiment demonstrating bacterial transformation. A mouse remains healthy if injected with either the nonvirulent R strain of *S. pneumoniae* or heat-killed cell fragments of the usually virulent S strain. R cells in the presence of heat-killed S cells are transformed into the virulent S strain, causing pneumonia in the mouse.

"breed true" in the sense that the progeny formed by cell division have the capsular type of the parent, either S or R.

Mice injected with living S cells get pneumonia. Mice injected either with living R cells or with heat-killed S cells remain healthy. Here is Griffith's critical finding: mice injected with a *mixture* of living R cells and heat-killed S cells contract the disease—they often die of pneumonia (**Figure 1.2**). Bacteria isolated from blood samples of these dead mice produce S cultures with a capsule typical of the injected S cells, even though the injected S cells had been killed by heat. Evidently, the injected material from the dead S cells includes a substance that can be transferred to living R cells and confer the ability to resist the immunological system of the mouse and cause pneumonia. In other words, the R bacteria can be changed—or undergo **transformation**—into S bacteria. Furthermore, the new characteristics are inherited by descendants of the transformed bacteria.

Transformation in *Streptococcus* was originally discovered in 1928, but it was not until 1944 that the chemical substance responsible for changing the R cells into S cells was identified. In a milestone experiment, Oswald Avery, Colin MacLeod, and Maclyn McCarty showed that the substance causing the transformation of R cells into S cells was DNA. In doing these experiments, they first had to develop chemical procedures for isolating almost pure DNA from cells, which had never been done before. When they added DNA isolated from S cells to growing cultures of R cells, they observed transformation: A few cells of type S cells were produced. Although the DNA preparations contained traces of protein and RNA (ribonucleic acid, an abundant cellular macromolecule chemically related to DNA), the transforming activity was not altered by treatments that destroyed either protein or RNA. However, treatments that destroyed DNA eliminated the transforming activity (**Figure 1.3**). These experiments implied that the substance responsible for genetic transformation was the DNA of the cell—hence that DNA is the genetic material.

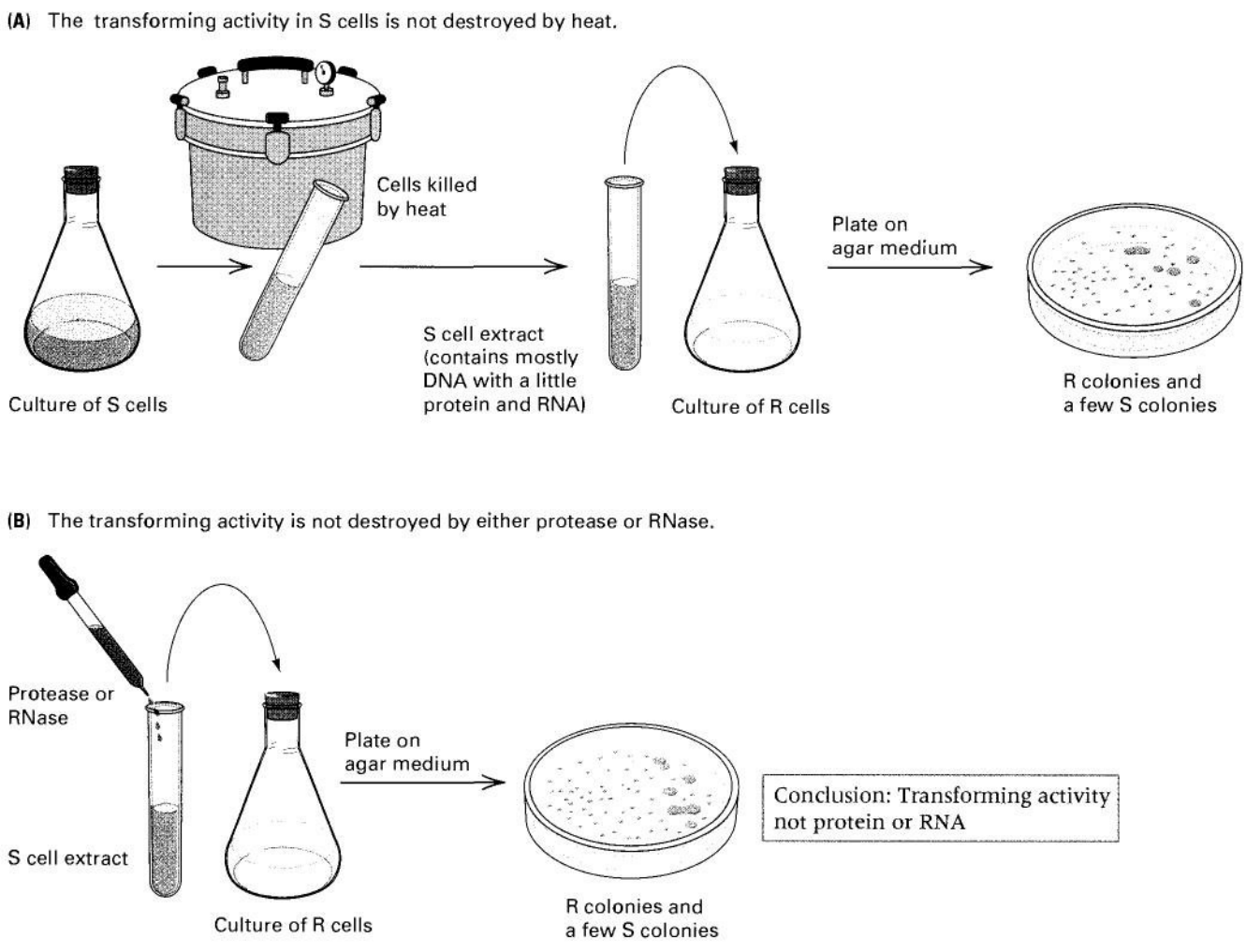


Figure 1.3 A diagram of the Avery–MacLeod–McCarty experiment that demonstrated that DNA is the active material in bacterial transformation. (A) Purified DNA extracted from heat-killed S cells can convert some living R cells into S cells, but the material may still contain undetectable traces of protein and/or RNA. (B) The transforming activity is not destroyed by either protease or RNase. (C) The transforming activity is destroyed by DNase and so probably consists of DNA.

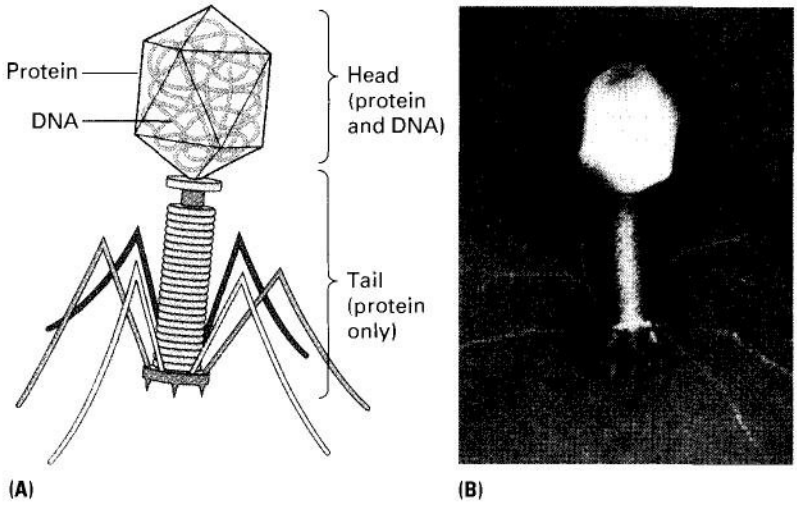


Figure 1.4 (A) Drawing of *E. coli* phage T2, showing various components. The DNA is confined to the interior of the head. (B) An electron micrograph of phage T4, a closely related phage. [Electron micrograph courtesy of Robley Williams.]

Genetic Role of DNA in Bacteriophage

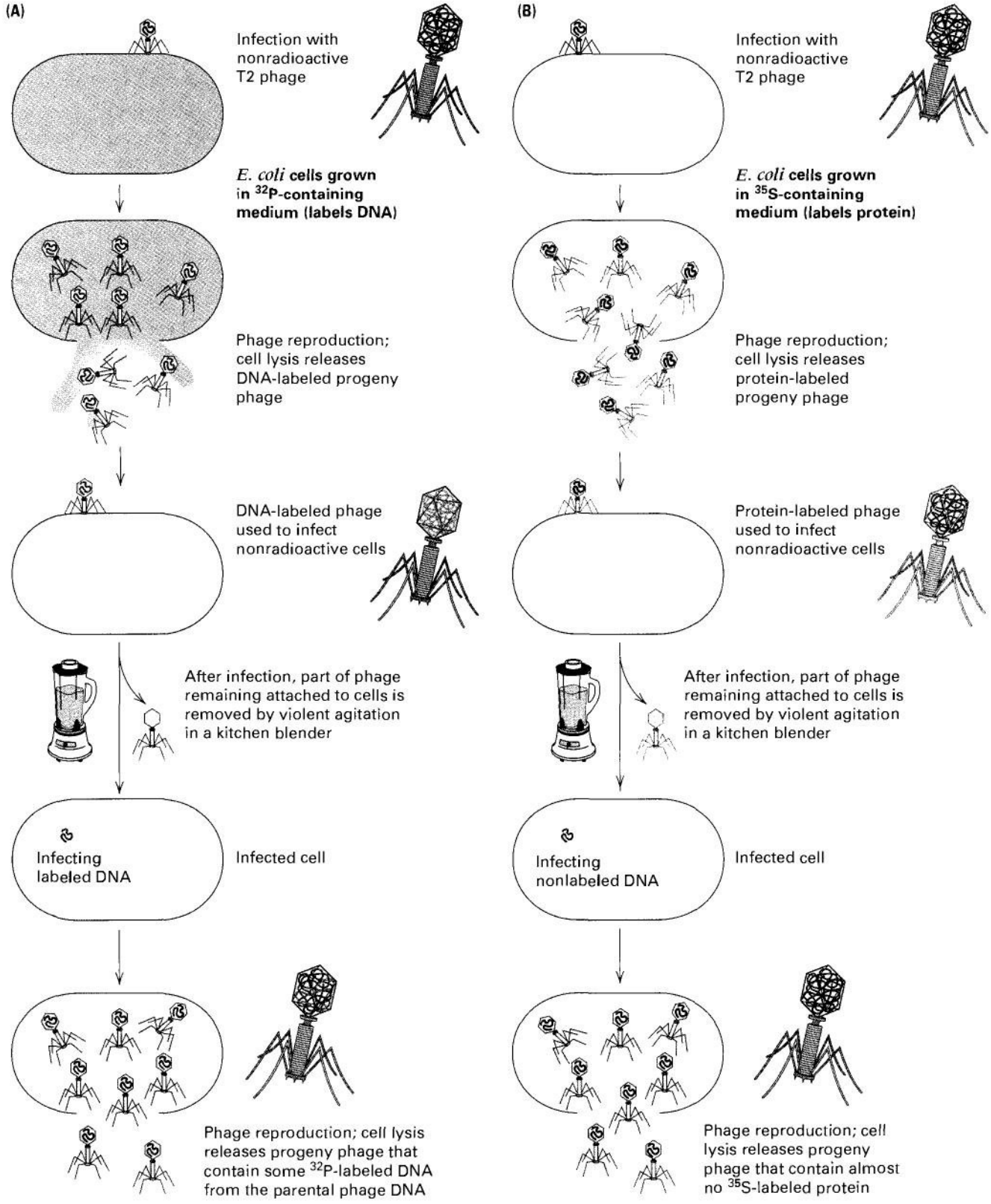
A second pivotal finding was reported by Alfred Hershey and Martha Chase in 1952. They studied cells of the intestinal bacterium *Escherichia coli* after infection by the virus T2. A virus that attacks bacterial cells is called a **bacteriophage,** a term often shortened to **phage.** *Bacteriophage* means "bacteria-eater." The structure of a bacteriophage T2 particle is illustrated in Figure 1.4. It is exceedingly small, yet it has a complex structure composed of head (which contains the phage DNA), collar, tail, and tail fibers. (The head of a human sperm is about 30–50 times larger in both length and width than the head of T2.) Hershey and Chase were already aware that T2 infection proceeds via the attachment of a phage particle by the tip of its tail to the bacterial cell wall, entry of phage material into the cell, multiplication of this material to form a hundred or more progeny phage, and release of the progeny phage by bursting (lysis) of the bacterial host cell. They also knew that T2 particles were composed of DNA and protein in approximately equal amounts.

Because DNA contains phosphorus but no sulfur, whereas most proteins contain sulfur but no phosphorus, it is possible to label DNA and proteins differentially by using radioactive isotopes of the two elements. Hershey and Chase produced particles containing radioactive DNA by infecting *E. coli* cells that had been grown for several generations in a medium that included ^{32}P (a radioactive isotope of phosphorus) and then collecting the phage progeny. Other particles containing labeled proteins were obtained in the same way, by using medium that included ^{35}S (a radioactive isotope of sulfur).

In the experiments summarized in Figure 1.5, nonradioactive *E. coli* cells were infected with phage labeled with *either* ^{32}P (part A) *or* ^{35}S (part B) in order to follow the DNA and proteins individually. Infected cells were separated from unattached phage particles by centrifugation, resuspended in fresh medium, and then swirled violently in a kitchen blender to shear attached phage material from the cell surfaces. This treatment was found to have no effect on the subsequent course of the infection, which implies that the phage genetic material must enter the infected cells very soon after phage attachment. The kitchen blender turned out to be the critical piece of equipment. Other methods had been tried to tear the phage heads from the bacterial cell surface, but nothing had worked reliably. Hershey later explained, "We tried various grinding arrangements, with results that weren't very encouraging. When Margaret McDonald loaned us her kitchen blender, the experiment promptly succeeded."

After the phage heads were removed by the blender treatment, the infected bacteria were examined. Most of the radioactivity from ^{32}P-labeled phage was found to be associated with the bacteria, whereas only a small fraction of the ^{35}S radioactivity was present in the infected cells. The retention of most of the labeled DNA, contrasted with the loss of most of the labeled protein, implied that a T2 phage transfers most of its DNA, but very little of its protein, to the cell it infects. The critical finding (Figure 1.5)

Figure 1.5 (on facing page) The Hershey–Chase ("blender") experiment demonstrating that DNA, not protein, is responsible for directing the reproduction of phage T2 in infected *E. coli* cells. (A) Radioactive DNA is transmitted to progeny phage in substantial amounts. (B) Radioactive protein is transmitted to progeny phage in negligible amounts.



Conclusion: DNA from an infecting parental phage is inherited in the progeny phage

CoNnection

Shear Madness

Alfred D. Hershey and Martha Chase 1952
Cold Spring Harbor Laboratories, Cold Spring Harbor, New York
Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage

Published a full eight years after the paper of Avery, MacLeod, and McCarty, the experiments of Hershey and Chase get equal billing. Why? Some historians of science suggest that the Avery et al. *experiments were "ahead of their time." Others suggest that Hershey had special standing because he was a member of the "in group" of phage molecular geneticists. Max Delbrück was the acknowledged leader of this group, with Salvador Luria close behind. (Delbrück, Luria, and Hershey shared a 1969 Nobel Prize.) Another possible reason is that whereas the experiments of Avery et al. were feats of strength in biochemistry, those of Hershey and Chase were quintessentially genetic. Which macromolecule gets into the hereditary action, and which does not? Buried in the middle of this paper, and retained in the excerpt, is a sentence admitting that an earlier publication by the researchers was a misinterpretation of their preliminary results. This shows that even first-rate scientists, then and now, are sometimes misled by their preliminary data. Hershey later explained, "We tried various grinding arrangements, with results that weren't very encouraging. When Margaret McDonald loaned us her kitchen blender the experiment promptly succeeded."*

The work [of others] has shown that bacteriophages T2, T3, and T4 multiply in the bacterial cell in a non-infective [immature] form. Little else is known about the vegetative [growth] phase of these viruses. The experiments reported in this paper show that one of the first steps in the growth of T2 is the release from its protein coat of the nucleic acid of the virus particle, after which the bulk of the sulfur-containing protein has no further function. . . . Anderson has obtained electron micrographs indicating that phage T2 attaches to bacteria by its tail. . . . It ought to be a simple matter to break the empty phage coats off the infected bacteria, leaving the phage DNA inside the cells. . . . When a suspension of cells with ^{35}S- or ^{32}P-labeled phage was spun in a blender at 10,000 revolutions per minute, . . . 75 to 80 percent of the phage sulfur can be stripped from the infected cells. . . . These facts show that the bulk of the phage sulfur remains at the cell surface during infection. . . . Little or no ^{35}S is contained in the mature phage progeny. . . . Identical experiments starting with phage labeled with ^{32}P show that phosphorus is transferred from parental to progeny phage at yields of about 30 phage per infected bacterium. . . . [Incomplete separation of phage heads] explains a mistaken preliminary report of the transfer of ^{35}S from parental to progeny phage. . . . The following questions remain unanswered. (1) Does any sulfur-free phage material other than DNA enter the cell? (2) If so, is it transferred to the phage progeny? (3) Is the transfer of phosphorus to progeny direct or indirect? . . . Our experiments show clearly that a physical separation of the phage T2 into genetic and nongenetic parts is possible. The chemical identification of the genetic part must wait until some of the questions above have been answered. . . . The sulfur-containing protein of resting phage particles is confined to a protective coat that is responsible for the adsorption to bacteria, and functions as an instrument for the injection of the phage DNA into the cell. This protein probably has no function in the growth of the intracellular phage. The DNA has some function. Further chemical inferences should not be drawn from the experiments presented.

Our experiments show clearly that a physical separation of the phage T2 into genetic and nongenetic parts is possible.

Source: Journal of General Physiology 36: 39–56

was that about 50 percent of the transferred ^{32}P-labeled DNA, but less than 1 percent of the transferred ^{35}S-labeled protein, was inherited by the *progeny* phage particles. Hershey and Chase interpreted this result to mean that the genetic material in T2 phage is DNA.

The experiments of Avery, MacLeod, and McCarty and those of Hershey and Chase are regarded as classics in the demonstration that genes consist of DNA. At the present time, the equivalent of the transformation experiment is carried out daily in many research laboratories throughout the world, usually with bacteria, yeast, or animal or plant cells grown in culture. These experiments indicate that DNA is the genetic material in these organisms as well as

in phage T2. Although there are no known exceptions to the generalization that DNA is the genetic material in all cellular organisms and many viruses, in a few types of viruses the genetic material consists of RNA.

1.2 DNA Structure: The Double Helix

The inference that DNA is the genetic material still left many questions unanswered. How is the DNA in a gene duplicated when a cell divides? How does the DNA in a gene control a hereditary trait? What happens to the DNA when a mutation (a change in the DNA) takes place in a gene? In the early 1950s, a number of researchers began to try to understand the detailed molecular structure of DNA in hopes that the structure alone would suggest answers to these questions. In 1953 James Watson and Francis Crick at Cambridge University proposed the first essentially correct three-dimensional structure of the DNA molecule. The structure was dazzling in its elegance and revolutionary in suggesting how DNA duplicates itself, controls hereditary traits, and undergoes mutation. Even while their tin-and-wire model of the DNA molecule was still incomplete, Crick would visit his favorite pub and exclaim "we have discovered the secret of life."

In the Watson–Crick structure, DNA consists of two long chains of subunits, each twisted around the other to form a double-stranded helix. The double helix is right-handed, which means that as one looks along the barrel, each chain follows a clockwise path as it progresses. You can visualize the right-handed coiling in part A of **Figure 1.6** if you imagine yourself looking up into the structure from the bottom. The dark spheres outline the "backbone" of each individual strand, and they coil in a clockwise direction. The subunits of each strand are **nucleotides,** each of which contains any one of four chemical constituents called **bases** attached to a phosphorylated molecule of the 5-carbon sugar **deoxyribose.** The four bases in DNA are

- **Adenine (A)**
- **Guanine (G)**
- **Thymine (T)**
- **Cytosine (C)**

The chemical structures of the nucleotides and bases need not concern us at this time. They are examined in Chapter 2. A key point for our present purposes is that the bases in the double helix are paired as shown in Figure 1.6B. That is:

> At any position on the paired strands of a DNA molecule, if one strand has an A, then the partner strand has a T; and if one strand has a G, then the partner strand has a C.

The pairing between A and T and between G and C is said to be **complementary;** the complement of A is T, and the complement of G is C. The complementary pairing means that each base along one strand of the DNA is matched with a base in the opposite position on the other strand. Furthermore:

> Nothing restricts the sequence of bases in a single strand, so any sequence could be present along one strand.

This principle explains how only four bases in DNA can code for the huge amount of information needed to make an organism. It

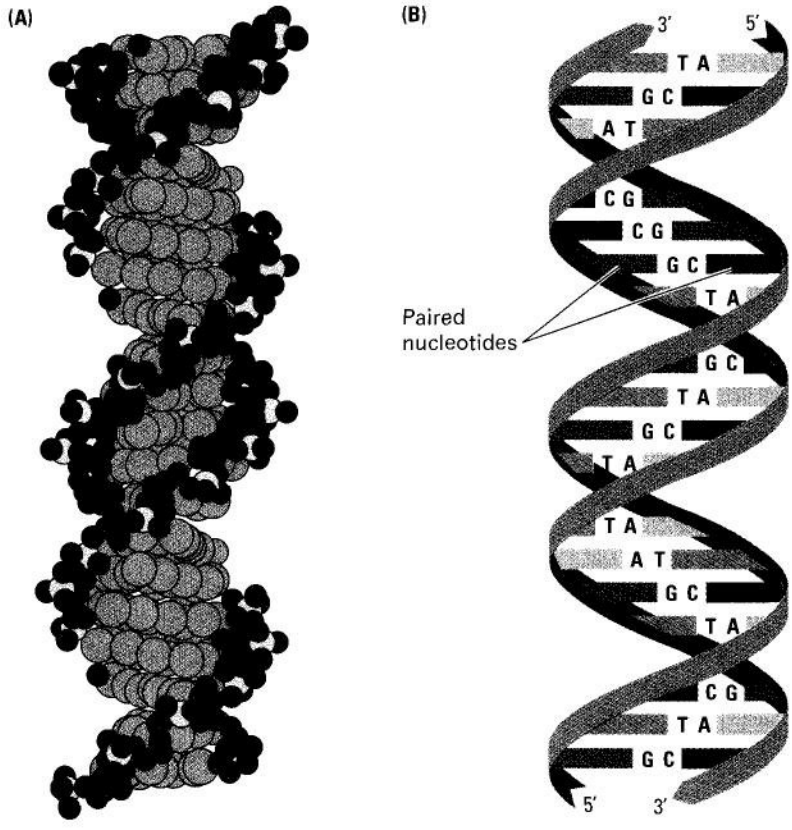


Figure 1.6 Molecular structure of the DNA double helix in the standard "B form." (A) A space-filling model, in which each atom is depicted as a sphere. (B) A diagram highlighting the helical strands around the outside of the molecule and the A–T and G–C base pairs inside.

is the *sequence* of bases along the DNA that encodes the genetic information, and the sequence is completely unrestricted.

The complementary pairing is also called **Watson–Crick pairing.** In the three-dimensional structure in Figure 1.6A, the base pairs are represented by the lighter spheres filling the interior of the double helix. The base pairs lie almost flat, stacked on top of one another perpendicular to the long axis of the double helix, like pennies in a roll. When discussing a DNA molecule, biologists frequently refer to the individual strands as **single-stranded DNA** and to the double helix as **double-stranded DNA** or **duplex DNA.**

Each DNA strand has a **polarity,** or directionality, like a chain of circus elephants linked trunk to tail. In this analogy, each elephant corresponds to one nucleotide along the DNA strand. The polarity is determined by the direction in which the nucleotides are pointing. The "trunk" end of the strand is called the *3' end* of the strand, and the "tail" end is called the *5' end.* In double-stranded DNA, the paired strands are oriented in opposite directions, the 5' end of one strand aligned with the 3' end of the other. The molecular basis of the polarity, and the reason for the opposite orientation of the strands in duplex DNA, are explained in Chapter 2. In illustrating DNA molecules in this book, we use an arrow-like ribbon to represent the backbone, and we use tabs jutting off the ribbon to represent the nucleotides. The polarity of a DNA strand is indicated by the direction of the arrow-like ribbon. The tail of the arrow represents the 5' end of the DNA strand, the head the 3' end.

Beyond the most optimistic hopes, knowledge of the structure of DNA immediately gave clues to its function:

1. The sequence of bases in DNA could be copied by using each of the separate "partner" strands as a pattern for the creation of a new partner strand with a complementary sequence of bases.
2. The DNA could contain genetic information in coded form in the sequence of bases, analogous to letters printed on a strip of paper.
3. Changes in genetic information (mutations) could result from errors in copying in which the base sequence of the DNA became altered.

In the remainder of this chapter, we discuss some of the implications of these clues.

1.3 An Overview of DNA Replication

Watson and Crick noted that the structure of DNA itself suggested a mechanism for its replication. "It has not escaped our notice," they wrote, "that the specific base pairing we have postulated immediately suggests a copying mechanism." The copying process in which a single DNA molecule becomes two identical molecules is called **replication.** The replication mechanism that Watson and Crick had in mind is illustrated in **Figure 1.7.**

As shown in part A of Figure 1.7, the strands of the original (parent) duplex separate, and each individual strand serves as a pattern, or **template,** for the synthesis of a new strand (replica). The replica strands are synthesized by the addition of successive nucleotides in such a way that each base in the replica is complementary (in the Watson–Crick pairing sense) to the base across the way in the template strand (Figure 1.7B). Although the mechanism in Figure 1.7 is simple in principle, it is a complex process that is fraught with geometrical problems and requires a variety of enzymes and other proteins. The details are examined in Chapter 6. The end result of replication is that a single double-stranded molecule becomes replicated into two copies with identical sequences:

```
                5'-ACGCTTGC-3'
                3'-TGCGAACG-5'
               ↙              ↘
5'-ACGCTTGC-3'                  5'-ACGCTTGC-3'
3'-TGCGAACG-5'                  3'-TGCGAACG-5'
```

Here the bases in the newly synthesized strands are shown in red. In the duplex on the left, the top strand is the template from the parental molecule and the bottom strand is newly synthesized; in the duplex on the right, the bottom strand is the template from the parental molecule and the top strand is newly synthesized. Note in Figure 1.7B that in the synthesis of each new strand, new nucleotides are added only to the 3' end of the growing chain:

> The obligatory elongation of a DNA strand only at the 3' end is an essential feature of DNA replication.

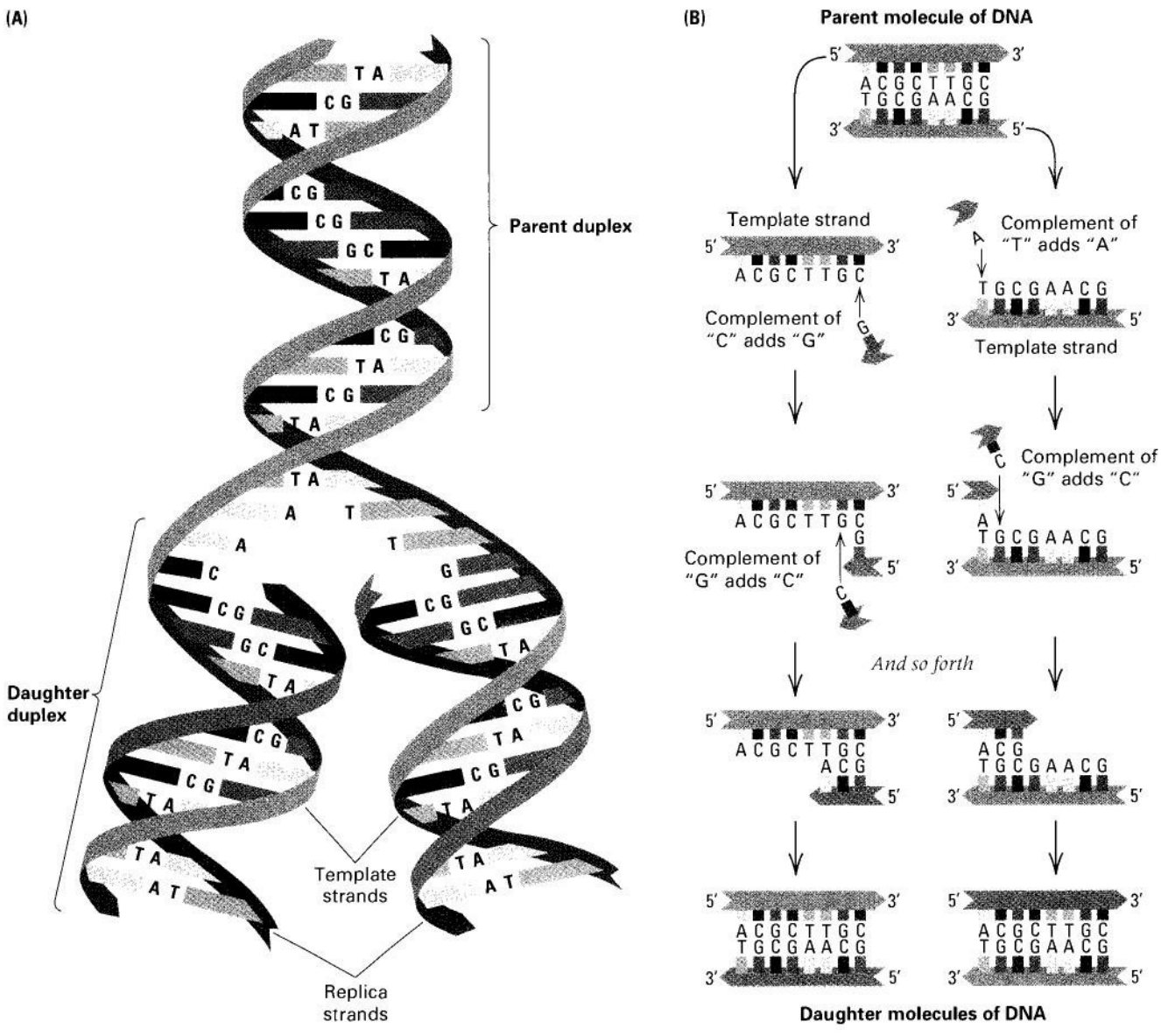


Figure 1.7 Replication of DNA. (A) Replication of a DNA duplex as originally envisioned by Watson and Crick. As the parental strands separate, each parental strand serves as a template for the formation of a new daughter strand by means of A–T and G–C base pairing. (B) Greater detail showing how each of the parental strands serves as a template for the production of a complementary daughter strand, which grows in length by the successive addition of single nucleotides to the 3′ end.

1.4 Genes and Proteins

Now that we have some basic understanding of the structural makeup of the genetic blueprint, how does this developmental plan become a complex living organism? If the code is thought of as a string of letters on a sheet of paper, then the genes are made up of distinct words that form sentences and paragraphs that give meaning to the pattern of letters. What is created from the complex and diverse DNA codes is protein, a class of macromolecules that carries out most of the activities in the cell. Cells are largely made up of proteins: structural proteins that give the cell rigidity and mobility, proteins that form pores in the cell membrane to control the traffic of small molecules into and out of the cell, and receptor proteins that regulate cellular activities in response to molecular signals from the growth medium or from other cells.

Proteins are also responsible for most of the metabolic activities of cells. They are essential for the synthesis and breakdown of organic molecules and for generating the chemical energy needed for cellular activities. In 1878 the term **enzyme** was introduced to refer to the biological catalysts that accelerate biochemical reactions in cells. By 1900, thanks largely to the work of the German biochemist Emil Fischer, enzymes were shown to be proteins. As often happens in science, nature's "mistakes" provide clues as to how things work. Such was the case in establishing a relationship between genes and disease, because a "mistake" in a gene (a mutation) can result in a "mistake" (lack of function) in the corresponding protein. This provided a fruitful avenue of research for the study of genetics.

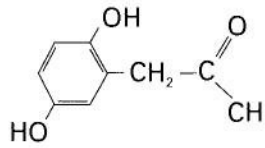
Inborn Errors of Metabolism as a Cause of Hereditary Disease

It was at the turn of the twentieth century that the British physician Archibald Garrod realized that certain heritable diseases followed the rules of transmission that Mendel had described for his garden peas. In 1908 Garrod gave a series of lectures in which he proposed a fundamental hypothesis about the relationship between heredity, enzymes, and disease:

> Any hereditary disease in which cellular metabolism is abnormal results from an inherited defect in an enzyme.

Such diseases became known as **inborn errors of metabolism,** a term still in use today.

Garrod studied a number of inborn errors of metabolism in which the patients excreted abnormal substances in the urine. One of these was **alkaptonuria**. In this case, the abnormal substance excreted is **homogentisic acid**:



An early name for homogentisic acid was *alkapton,* hence the name *alkaptonuria* for the disease. Even though alkaptonuria is

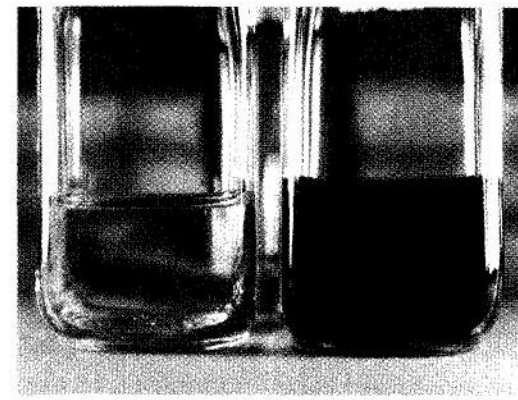


Figure 1.8 Urine from a person with alkaptonuria turns black because of the oxidation of the homogentisic acid that it contains. [Courtesy of Daniel De Aguiar.]

rare, with an incidence of about one in 200,000 people, it was well known even before Garrod studied it. The disease itself is relatively mild, but it has one striking symptom: The urine of the patient turns black because of the oxidation of homogentisic acid (**Figure 1.8**). This is why alkaptonuria is also called *black urine disease.* An early case was described in the year 1649:

> The patient was a boy who passed black urine and who, at the age of fourteen years, was submitted to a drastic course of treatment that had for its aim the subduing of the fiery heat of his viscera, which was supposed to bring about the condition in question by charring and blackening his bile. Among the measures prescribed were bleedings, purgation, baths, a cold and watery diet, and drugs galore. None of these had any obvious effect, and eventually the patient, who tired of the futile and superfluous therapy, resolved to let things take their natural course. None of the predicted evils ensued. He married, begat a large family, and lived a long and healthy life, always passing urine black as ink. (Recounted by Garrod, 1908.)

Garrod was primarily interested in the biochemistry of alkaptonuria, but he took note of family studies that indicated that the disease was inherited as though it were due to a defect in a single gene. As to the biochemistry, he deduced that the problem in alkaptonuria was the patients' inability to break down the phenyl ring of six carbons that is present in homogentisic acid. Where does this ring come from? Most animals

obtain it from foods in their diet. Garrod proposed that homogentisic acid originates as a breakdown product of two amino acids, phenylalanine and tyrosine, which also contain a phenyl ring. An **amino acid** is one of the "building blocks" from which proteins are made. Phenylalanine and tyrosine are constituents of normal proteins. The scheme that illustrates the relationship between the molecules is shown in Figure 1.9. Any such sequence of biochemical reactions is called a **biochemical pathway** or a **metabolic pathway.** Each arrow in the pathway represents a single step depicting the transition from the "input" or **substrate molecule,** shown at the head of the arrow, to the "output" or **product molecule,** shown at the tip. Biochemical pathways are usually oriented either vertically with the arrows pointing down, as in Figure 1.9, or horizontally, with the arrows pointing from left to right. Garrod did not know all of the details of the pathway in Figure 1.9, but he did understand that the key step in the breakdown of homogentisic acid is the breaking open of the phenyl ring and that the phenyl ring in homogentisic acid comes from dietary phenylalanine and tyrosine.

What allows each step in a biochemical pathway to occur? Garrod correctly surmised that each step requires a specific enzyme to catalyze the reaction for the chemical transformation. Persons with an inborn error of metabolism, such as alkaptonuria, have a defect in a single step of a metabolic pathway because they lack a functional enzyme for that step. When an enzyme in a pathway is defective, the pathway is said to have a **block** at that step. One frequent result of a blocked pathway is that the substrate of the defective enzyme accumulates. Observing the accumulation of homogentisic acid in patients with alkaptonuria, Garrod proposed that there must be an enzyme whose function is to open the phenyl ring of homogentisic acid and that this enzyme is missing in these patients. Isolation of the enzyme that opens the phenyl ring of homogentisic acid was not actually achieved until 50 years after Garrod's lectures. In normal people it is found in cells of the liver, and just as Garrod had predicted, the enzyme is defective in patients with alkaptonuria.

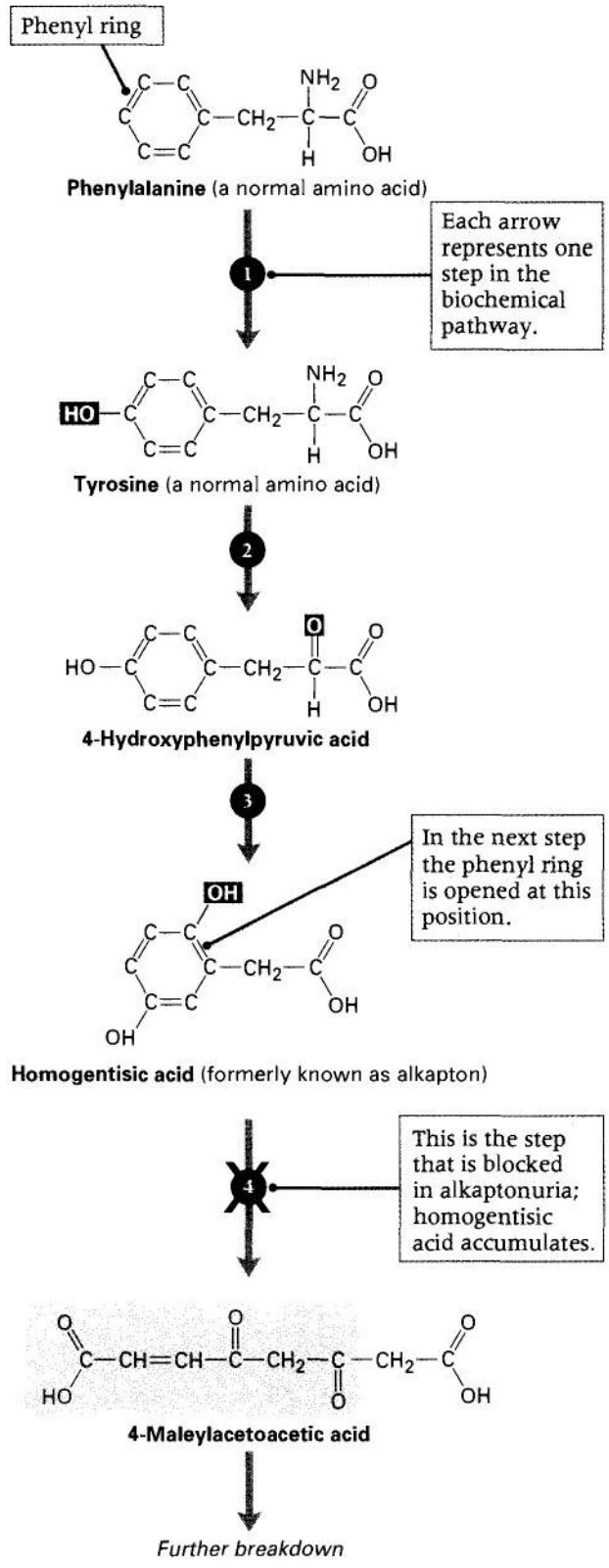


Figure 1.9 Metabolic pathway for the breakdown of phenylalanine and tyrosine. Each step in the pathway, represented by an arrow, requires a specific enzyme to catalyze the reaction. The key step in the breakdown of homogentisic acid is the breaking open of the phenyl ring.

CoNnection

The Black Urine Disease

Archibald E. Garrod 1908
St. Bartholomew's Hospital,
London, England
Inborn Errors of Metabolism

Although he was a distinguished physician, Garrod's lectures on the relationship between heredity and congenital defects in metabolism had no impact when they were delivered. The important concept that one gene corresponds to one enzyme (the "one gene–one enzyme hypothesis") was developed independently in the 1940s by George W. Beadle and Edward L. Tatum, who used the bread mold Neurospora crassa *as their experimental organism. When Beadle finally became aware of* Inborn Errors of Metabolism, *he was generous in praising it. This excerpt shows Garrod at his best, interweaving history, clinical medicine, heredity, and biochemistry in his account of alkaptonuria. The excerpt also illustrates how the severity of a genetic disease depends on its social context. Garrod writes as though alkaptonuria were a harmless curiosity. This is indeed largely true when the life expectancy is short. With today's longer life span, alkaptonuria patients accumulate the dark pigment in their cartilage and joints and eventually develop severe arthritis.*

To students of heredity the inborn errors of metabolism offer a promising field of investigation. . . . It was pointed out [by others] that the mode of incidence of alkaptonuria finds a ready explanation if the anomaly be regarded as a rare recessive character in the Mendelian sense. . . . Of the cases of alkaptonuria a very large proportion have been in the children of first cousin marriages. . . . It is also noteworthy that, if one takes families with five or more children [with both parents normal and at least one child affected with alkaptonuria], the totals work out in strict conformity to Mendel's law, i.e. 57 [normal children] : 19 [affected children] in the proportions 3 : 1. . . . Of inborn errors of metabolism, alkaptonuria is that of which we know most. In itself it is a trifling matter, inconvenient rather than harmful. . . . Indications of the anomaly may be detected in early medical writings, such as that in 1584 of a schoolboy who, although he enjoyed good health, continuously excreted black urine; and that in 1609 of a monk who exhibited a similar peculiarity and stated that he had done so all his life. . . . There are no sufficient grounds [for doubting that the blackening substance in the urine originally called alkapton] is homogentisic acid, the excretion of which is the essential feature of the alkaptonuric. . . . Homogentisic acid is a product of normal metabolism. . . . The most likely sources of the phenyl ring in homogentisic acid are phenylalanine and tyrosine, [because when these amino acids are administered to an alkaptonuric] they cause a very conspicuous increase in the output of homogentisic acid. . . . Where the alkaptonuric differs from the normal individual is in having no power of destroying homogentisic acid when formed—in other words of breaking up the phenyl ring of that compound. . . . We may further conceive that the splitting of the phenyl ring in normal metabolism is the work of a special enzyme and that in congenital alkaptonuria this enzyme is wanting.

We may further conceive that the splitting of the phenyl ring in normal metabolism is the work of a special enzyme and that in congenital alkaptonuria this enzyme is wanting.

Source: Originally published in London, England, by the Oxford University Press. Excerpts from the reprinted edition in Harry Harris. 1963. *Garrod's Inborn Errors of Metabolism.* London, England: Oxford University Press

The pathway for the breakdown of phenylalanine and tyrosine, as it is understood today, is shown in Figure 1.10. In this figure the emphasis is on the enzymes rather than on the structures of the **metabolites,** or small molecules, on which the enzymes act. Each step in the pathway requires the presence of a particular enzyme that catalyzes that step. Although Garrod knew only about alkaptonuria, in which the defective enzyme is homogentisic acid 1,2 dioxygenase, we now know the clinical consequences of defects in the other enzymes. Unlike alkaptonuria, which is a relatively benign inherited disease, the others are very serious. The condition known as **phenylketonuria (PKU)** results from the absence of (or a defect in) the enzyme **phenylalanine hydroxylase (PAH).** When this step in the pathway is blocked, phenylalanine accumulates. The excess phenylalanine is broken down into harmful metabolites that cause defects in myelin formation that damage a child's developing

nervous system and lead to severe mental retardation.

However, if PKU is diagnosed in children soon enough after birth, they can be placed on a specially formulated diet low in phenylalanine. The child is allowed only as much phenylalanine as can be used in the synthesis of proteins, so excess phenylalanine does not accumulate. The special diet is very strict. It excludes meat, poultry, fish, eggs, milk and milk products, legumes, nuts, and bakery goods manufactured with regular flour. These foods are replaced by an expensive synthetic formula. With the special diet, however, the detrimental effects of excess phenylalanine on mental development can largely be avoided, although in adult women with PKU who are pregnant, the fetus is at risk. In many countries, including the United States, all newborn babies have their blood tested for chemical signs of PKU. Routine screening is cost-effective because PKU is relatively common. In the United States, the incidence is about 1 in 8000 among Caucasian births. The disease is less common in other ethnic groups.

In the metabolic pathway in Figure 1.10, defects in the breakdown of tyrosine or of 4-hydroxyphenylpyruvic acid lead to types of tyrosinemia. These are also severe diseases. Type II is associated with skin lesions and mental retardation, Type III with severe liver dysfunction.

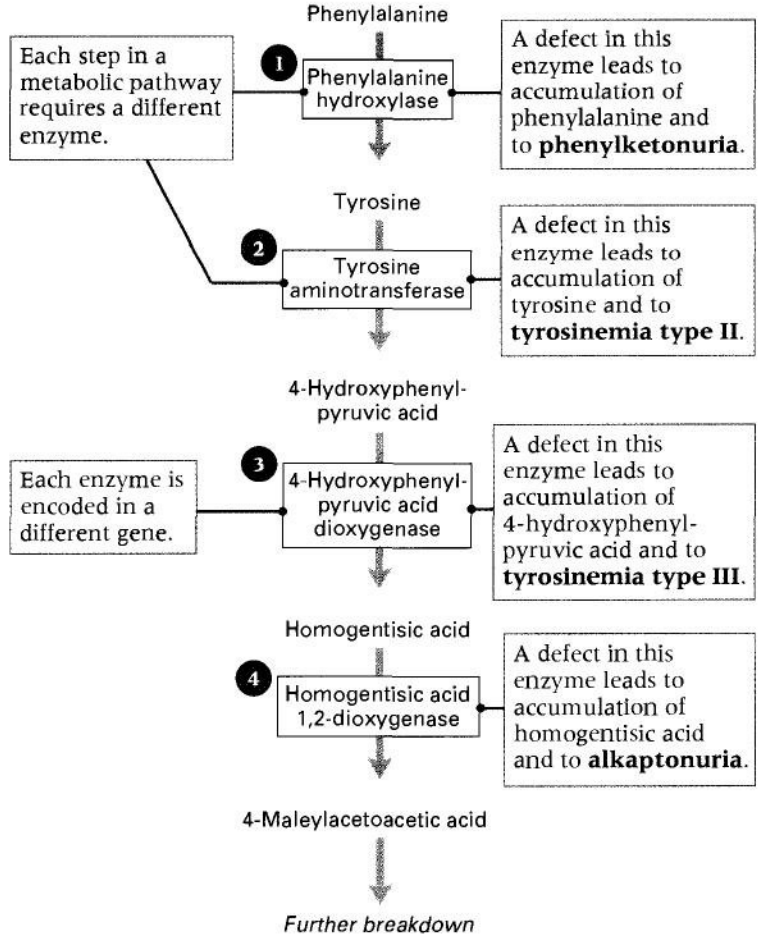


Figure 1.10 Inborn errors of metabolism that affect the breakdown of phenylalanine and tyrosine. An inherited disease results when any of the enzymes is missing or defective. Alkaptonuria results from a mutant homogentisic acid 1,2 dioxygenase; phenylketonuria results from a mutant phenylalanine hydroxylase.

Mutant Genes and Defective Proteins

It follows from Garrod's work that a defective enzyme results from a mutant gene, but how? Garrod did not speculate. For all he knew, genes *were* enzymes. This would have been a logical hypothesis at the time. We now know that the relationship between genes and enzymes is somewhat indirect. With a few exceptions, each enzyme is *encoded* in a particular sequence of nucleotides present in a region of DNA. The DNA region that codes for the enzyme, as well as adjacent regions that regulate when and in which cells the enzyme is produced, make up the "gene" that encodes the enzyme.

The genes for the enzymes in the biochemical pathway in Figure 1.10 have all been identified and the nucleotide sequence of the DNA determined. In the following list, and throughout this book, we use the standard typographical convention that *genes* are written in *italic* type, whereas gene products are not printed in italics. This convention is convenient, because it means that the protein product of a gene can be represented with the same symbol as the gene itself, but whereas the gene symbol is in italics, the protein symbol is not.

- The gene *PAH* on the long arm of chromosome 12 encodes phenylalanine hydroxylase (PAH).
- The gene *TAT* on the long arm of chromosome 16 encodes tyrosine aminotransferase (TAT).
- The gene *HPD* on the long arm of chromosome 12 encodes 4-hydroxyphenylpyruvic acid dioxygenase (HPD).

- The gene *HGD* on the long arm of chromosome 3 encodes homogentisic acid 1,2 dioxygenase (HGD).

Next we turn to the issue of *how* genes code for enzymes and other proteins.

1.5 Gene Expression: The Central Dogma

Watson and Crick were correct in proposing that the genetic information in DNA is contained in the sequence of bases in a manner analogous to letters printed on a strip of paper. In a region of DNA that directs the synthesis of a protein, the genetic code for the protein is contained in only one strand, and it is decoded in a linear order. A typical protein is made up of one or more polypeptide chains; each **polypeptide chain** consists of a linear sequence of amino acids connected end to end. For example, the enzyme PAH consists of four identical polypeptide chains, each 452 amino acids in length. In the decoding of DNA, each successive "code word" in the DNA specifies the next amino acid to be added to the polypeptide chain as it is being made. The amount of DNA required to code for the polypeptide chain of PAH is therefore $452 \times 3 = 1356$ nucleotide pairs. The entire gene is very much longer—about 90,000 nucleotide pairs. Only 1.5 percent of the gene is devoted to coding for the amino acids. The noncoding part includes some sequences that control the activity of the gene, but it is not known how much of the gene is involved in regulation.

There are 20 different amino acids. Only four bases code for these 20 amino acids, with each "word" in the genetic code consisting of three adjacent bases. For example, the base sequence ATG specifies the amino acid methionine (Met), TCC specifies serine (Ser), ACT specifies threonine (Thr), and GCG specifies alanine (Ala). There are 64 possible three-base combinations but only 20 amino acids because some combinations code for the same amino acid. For example, TCT, TCC, TCA, TCG, AGT, and AGC all code for serine (Ser), and CTT, CTC, CTA, CTG, TTA, and TTG all code for leucine (Leu). An example of the relationship between the base sequence in a DNA duplex and the amino acid sequence of the corresponding protein is shown in **Figure 1.11**. This particular DNA duplex is the human sequence that codes for the first seven amino acids in the polypeptide chain of PAH.

The scheme outlined in Figure 1.11 indicates that DNA codes for protein not directly but indirectly through the processes of *transcription* and *translation*. The indirect route of information transfer,

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Protein}$$

is known as the **central dogma** of molecular genetics. The term *dogma* means "set of beliefs"; it dates from the time the idea was put forward first as a theory. Since then the "dogma" has been confirmed experimentally, but the term persists. The central dogma is shown in **Figure 1.12**. The main concept in the central dogma is that DNA

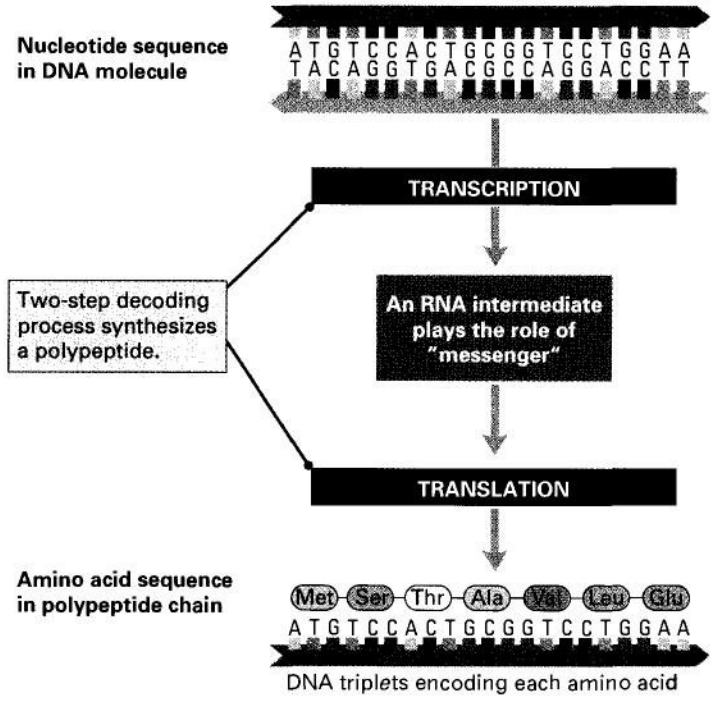


Figure 1.11 DNA sequence coding for the first seven amino acids in a polypeptide chain. The DNA sequence specifies the amino acid sequence through a molecule of RNA that serves as an intermediary "messenger." Although the decoding process is indirect, the net result is that each amino acid in the polypeptide chain is specified by a group of three adjacent bases in the DNA. In this example, the polypeptide chain is that of phenylalanine hydroxylase (PAH).

does not code for protein directly but rather acts through an intermediary molecule of **ribonucleic acid (RNA).** The structure of RNA is similar to, but not identical with, that of DNA. There is a difference in the sugar (RNA contains the sugar **ribose** instead of deoxyribose), RNA is usually single-stranded (not a duplex), and RNA contains the base **uracil (U)** instead of thymine (T), which is present in DNA. Actually, three types of RNA take part in the synthesis of proteins:

- A molecule of **messenger RNA (mRNA),** which carries the genetic information from DNA and is used as a template for polypeptide synthesis. In most mRNA molecules, there is a high proportion of nucleotides that actually code for amino acids. For example, the mRNA for PAH is 2400 nucleotides in length and codes for a polypeptide of 452 amino acids; in this case, more than 50 percent of the length of the mRNA codes for amino acids.
- Several types of **ribosomal RNA (rRNA),** which are major constituents of the cellular particles called **ribosomes** on which polypeptide synthesis takes place.
- A set of **transfer RNA (tRNA)** molecules, each of which carries a particular amino acid as well as a three-base recognition region that base-pairs with a group of three adjacent bases in the mRNA. As each tRNA participates in translation, its amino acid becomes the terminal subunit added to the length of the growing polypeptide chain. The tRNA that carries methionine is denoted $tRNA^{Met}$, that which carries serine is denoted $tRNA^{Ser}$, and so forth.

The central dogma is the fundamental principle of molecular genetics because it summarizes how the genetic information in DNA becomes expressed in the amino acid sequence in a polypeptide chain:

> The sequence of nucleotides in a gene specifies the sequence of nucleotides in a molecule of messenger RNA; in turn, the sequence of nucleotides in the messenger RNA specifies the sequence of amino acids in the polypeptide chain.

Given a process as conceptually simple as DNA coding for protein, what might account for the additional complexity of RNA intermediaries? One possible reason is that an RNA intermediate gives another level for control, for example, by degrading the mRNA for an unneeded protein. Another possible reason may be historical. RNA structure is unique in having both an informational content present in its sequence of bases and a complex, folded three-dimensional structure that endows some RNA molecules with catalytic activities. Many scientists believe that in the earliest forms of life, RNA served both for genetic information and catalysis. As evolution proceeded, the informational role was transferred to DNA and the catalytic role to protein. However, RNA became locked into its central location as a go-between in the processes of information transfer and protein synthesis. This hypothesis implies that the participation of RNA in protein synthesis is a relic of the earliest stages of evolution—a "molecular fossil." The hypothesis is supported by a variety of observations. For example, (1) DNA replication requires an RNA molecule in order to get started (Chapter 6), (2) an RNA molecule is essential in the synthesis of the tips of the chromosomes (Chapter 8), and (3) some RNA molecules act to catalyze key reactions in protein synthesis (Chapter 11).

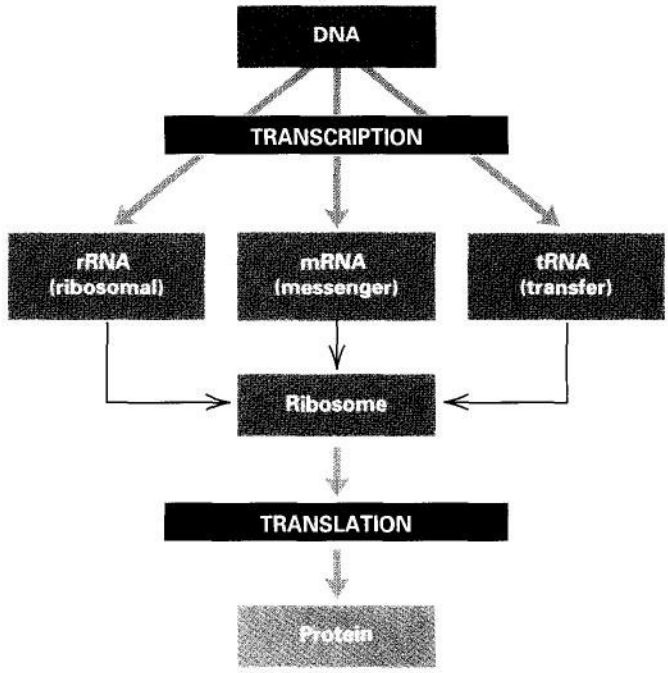


Figure 1.12 The "central dogma" of molecular genetics: DNA codes for RNA, and RNA codes for protein. The DNA → RNA step is transcription, and the RNA → protein step is translation.

Transcription

The manner in which genetic information is transferred from DNA to RNA is shown in **Figure 1.13.** The DNA opens up, and one of the strands is used as a template for the synthesis of a complementary strand of RNA. (How the template strand is chosen is discussed in Chapter 11.) The process of making an RNA strand from a DNA template is **transcription,** and the RNA molecule that is made is the **transcript.** The base sequence in the RNA is complementary (in the Watson–Crick pairing sense) to that in the DNA template, except that U (which pairs with A) is present in the RNA in place of T. The rules of base pairing between DNA and RNA are summarized in **Figure 1.14**. Each RNA strand has a polarity—a 5' end and a 3' end—and, as in the synthesis of DNA, nucleotides are added only to the 3' end of a growing RNA strand. Hence the 5' end of the RNA transcript is synthesized first, and transcription proceeds along the template DNA strand in the 3'-to-5' direction. Each gene includes nucleotide sequences that initiate and terminate transcription. The RNA transcript made from any gene begins at the initiation site in the template strand, which is located "upstream" from the amino acid–coding region, and ends at the termination site, which is located "downstream" from the amino acid–coding region. For any gene, the length of the RNA transcript is very much smaller than the length of the DNA in the chromosome. For example, the transcript of the *PAH* gene for phenylalanine hydroxylase is about 90,000 nucleotides in length, but the DNA in chromosome 12 is about 130,000,000 nucleotide pairs. In this case, the length of the *PAH* transcript is less than 0.1 percent of the length of the DNA in the chromosome. A different gene in chromosome 12 would be transcribed from a different region of the DNA molecule in chromosome 12, and perhaps from the opposite strand, but the transcribed region would again be small in comparison with the total length of the DNA in the chromosome.

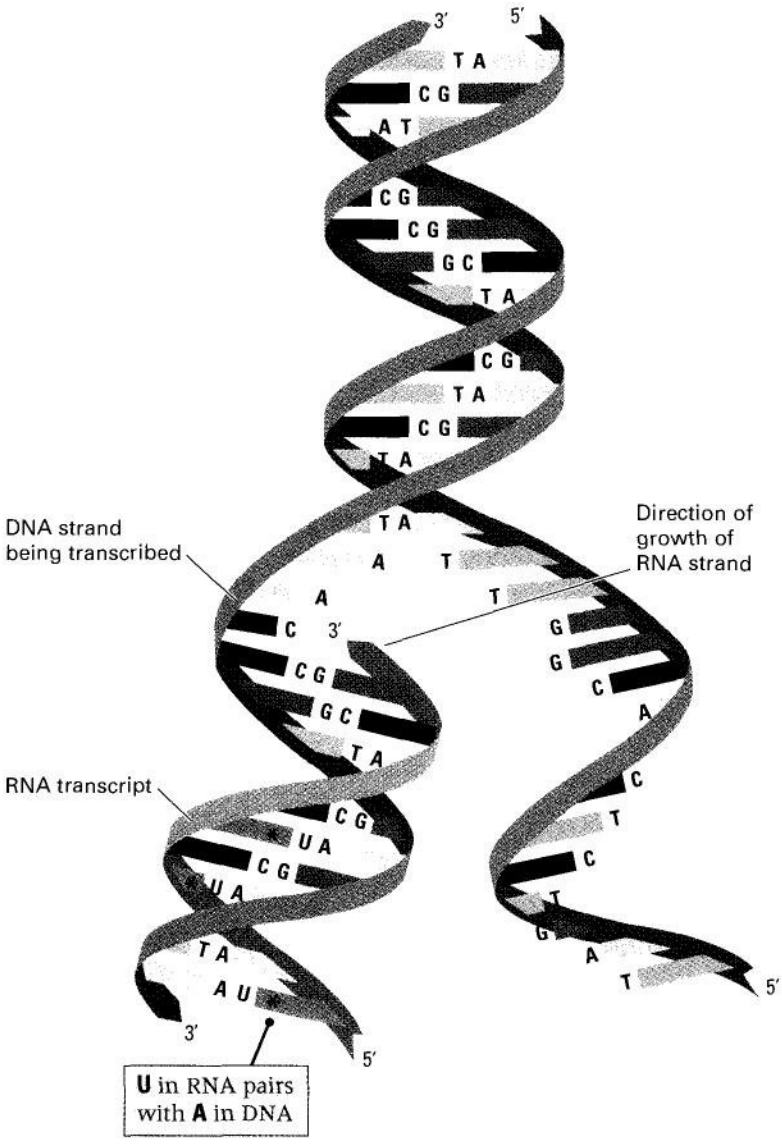


Figure 1.13 Transcription is the production of an RNA strand that is complementary in base sequence to a DNA strand. In this example, the DNA strand at the bottom is being transcribed into a strand of RNA. Note that in an RNA molecule, the base U (uracil) plays the role of T (thymine) in that it pairs with A (adenine). Each A—U pair is marked.

Translation

The synthesis of a polypeptide under the direction of an mRNA molecule is known as **translation.** Although the sequence of bases in the mRNA codes for the sequence of amino acids in a polypeptide, the molecules that actually do the "translating" are the tRNA molecules. The mRNA molecule is translated in nonoverlapping groups of three bases called **codons.** For each codon in the mRNA that specifies an amino acid, there is one tRNA molecule containing a complementary group of three adjacent bases that can pair with those in that codon. The correct amino acid is attached to the other end of the tRNA, and when the tRNA comes into line, the amino acid to which it

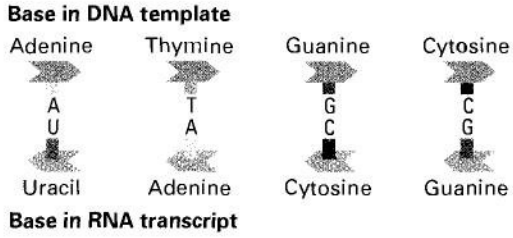


Figure 1.14 Pairing between bases in DNA and in RNA. The DNA bases A, T, G, and C pair with the RNA bases U, A, C, and G, respectively.

is attached becomes the most recent addition to the growing end of the polypeptide chain.

The role of tRNA in translation is illustrated in **Figure 1.15** and can be described as follows:

> The mRNA is read codon by codon. Each codon that specifies an amino acid matches with a complementary group of three adjacent bases in a single tRNA molecule. One end of the tRNA is attached to the correct amino acid, so the correct amino acid is brought into line.

The tRNA molecules used in translation do not line up along the mRNA simultaneously as shown in Figure 1.15. The process of translation takes place on a ribosome, which combines with a single mRNA and moves along it from one end to the other in steps, three nucleotides at a time (codon by codon). As each new codon comes into place, the next tRNA binds with the ribosome. Then the growing end of the polypeptide chain becomes attached to the amino acid on the tRNA. In this way, each tRNA in turn serves temporarily to hold the polypeptide chain as it is being synthesized. As the polypeptide chain is transferred from each tRNA to the next in line, the tRNA that previously held the polypeptide is released from the ribosome. The polypeptide chain elongates one amino acid at each step until any one of three particular codons specifying "stop" is encountered. At this point, synthesis of the chain of amino acids is finished, and the polypeptide chain is released from the ribosome. (This brief description of translation glosses over many of the details that are presented in Chapter 11.)

The Genetic Code

Figure 1.15 indicates that the mRNA codon AUG specifies methionine (Met) in the polypeptide chain, UCC specifies Ser (serine), ACU specifies Thr (threonine), and so on. The complete decoding table is

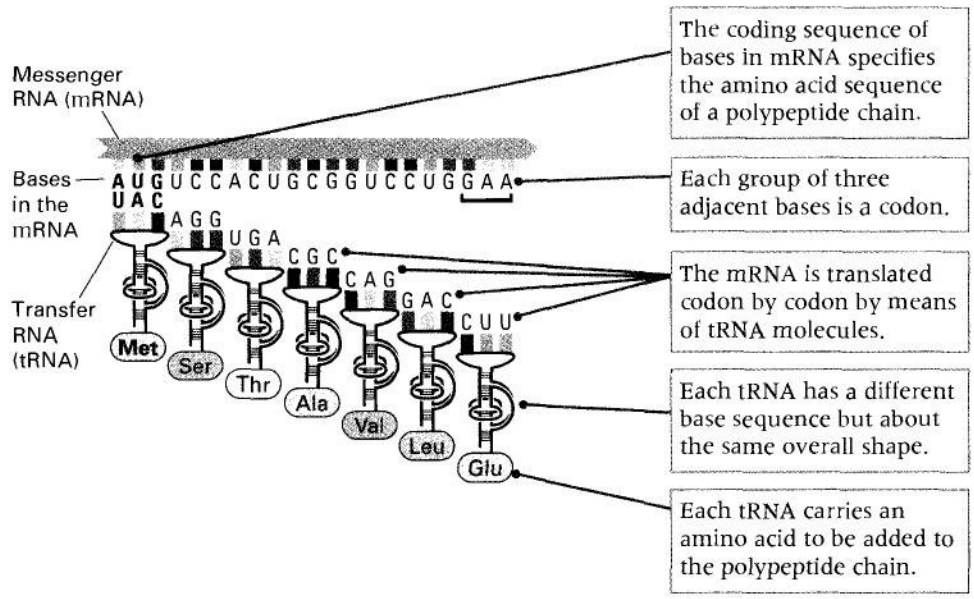


Figure 1.15 The role of messenger RNA in translation is to carry the information contained in a sequence of DNA bases to a ribosome, where it is translated into a polypeptide chain. Translation is mediated by transfer RNA (tRNA) molecules, each of which can base-pair with a group of three adjacent bases in the mRNA. Each tRNA also carries an amino acid. As each tRNA, in turn, is brought to the ribosome, the growing polypeptide chain is elongated.

Table 1.1 The standard genetic code

First nucleotide in codon (5′ end)	Second nucleotide in codon: U				C				A				G				Third nucleotide in codon (3′ end)
U	UUU	Phe	F	*Phenylalanine*	UCU	Ser	S	*Serine*	UAU	Tyr	Y	*Tyrosine*	UGU	Cys	C	*Cysteine*	U
	UUC	Phe	F	*Phenylalanine*	UCC	Ser	S	*Serine*	UAC	Tyr	Y	*Tyrosine*	UGC	Cys	C	*Cysteine*	C
	UUA	Leu	L	*Leucine*	UCA	Ser	S	*Serine*	UAA		Termination		UGA		Termination		A
	UUG	Leu	L	*Leucine*	UCG	Ser	S	*Serine*	UAG		Termination		UGG	Trp	W	*Tryptophan*	G
C	CUU	Leu	L	*Leucine*	CCU	Pro	P	*Proline*	CAU	His	H	*Histidine*	CGU	Arg	R	*Arginine*	U
	CUC	Leu	L	*Leucine*	CCC	Pro	P	*Proline*	CAC	His	H	*Histidine*	CGC	Arg	R	*Arginine*	C
	CUA	Leu	L	*Leucine*	CCA	Pro	P	*Proline*	CAA	Gln	Q	*Glutamine*	CGA	Arg	R	*Arginine*	A
	CUG	Leu	L	*Leucine*	CCG	Pro	P	*Proline*	CAG	Gln	Q	*Glutamine*	CGG	Arg	R	*Arginine*	G
A	AUU	Ile	I	*Isoleucine*	ACU	Thr	T	*Threonine*	AAU	Asn	N	*Asparagine*	AGU	Ser	S	*Serine*	U
	AUC	Ile	I	*Isoleucine*	ACC	Thr	T	*Threonine*	AAC	Asn	N	*Asparagine*	AGC	Ser	S	*Serine*	C
	AUA	Ile	I	*Isoleucine*	ACA	Thr	T	*Threonine*	AAA	Lys	K	*Lysine*	AGA	Arg	R	*Arginine*	A
	AUG	Met	M	*Methionine*	ACG	Thr	T	*Threonine*	AAG	Lys	K	*Lysine*	AGG	Arg	R	*Arginine*	G
G	GUU	Val	V	*Valine*	GCU	Ala	A	*Alanine*	GAU	Asp	D	*Aspartic acid*	GGU	Gly	G	*Glycine*	U
	GUC	Val	V	*Valine*	GCC	Ala	A	*Alanine*	GAC	Asp	D	*Aspartic acid*	GGC	Gly	G	*Glycine*	C
	GUA	Val	V	*Valine*	GCA	Ala	A	*Alanine*	GAA	Glu	E	*Glutamic acid*	GGA	Gly	G	*Glycine*	A
	GUG	Val	V	*Valine*	GCG	Ala	A	*Alanine*	GAG	Glu	E	*Glutamic acid*	GGG	Gly	G	*Glycine*	G

Codon

Three-letter and single-letter abbreviations

called the **genetic code,** and it is shown in **Table 1.1**. For any codon, the column on the left corresponds to the first nucleotide in the codon (reading from the 5′ end), the row across the top corresponds to the second nucleotide, and the column on the right corresponds to the third nucleotide. The complete codon is given in the body of the table, along with the amino acid (or translational "stop") that the codon specifies. Each amino acid is designated by its full name and by a three-letter abbreviation as well as a single-letter abbreviation. Both types of abbreviations are used in molecular genetics. The code in Table 1.1 is the "standard" genetic code used in translation in the cells of nearly all organisms. In Chapter 11 we examine general features of the standard genetic code and the minor differences found in the genetic codes of certain organisms and cellular organelles. At this point, we are interested mainly in understanding how the genetic code is used to translate the codons in mRNA into the amino acids in a polypeptide chain.

In addition to the 61 codons that code only for amino acids, there are four codons that have specialized functions:

- The codon AUG, which specifies Met (methionine), is also the "start" codon for polypeptide synthesis. The positioning of a $tRNA^{Met}$ bound to AUG is one of the first steps in the initiation of polypeptide synthesis, so all polypeptide chains begin with Met. (Many polypeptides have the initial Met cleaved off after translation is complete.) In most organisms, the $tRNA^{Met}$ used for initiation of translation is the same $tRNA^{Met}$ used to specify methionine at internal positions in a polypeptide chain.
- The codons UAA, UAG, and UGA are each a "stop" that specifies the termination of translation and results in release of the completed polypeptide chain from the ribosome. These codons do not have tRNA molecules that recognize them but are instead recognized by protein factors that terminate translation.

How the genetic code table is used to infer the amino acid sequence of a polypep-

tide chain can be illustrated by using PAH again, in particular the DNA sequence coding for amino acids 1 through 7. The DNA sequence is

```
5'-ATGTCCACTGCGGTCCTGGAA-3'
3'-TACAGGTGACGCCAGGACCTT-5'
```

This region is transcribed into RNA in a left-to-right direction, and because RNA grows by the addition of successive nucleotides to the 3' end (Figure 1.13), it is the bottom strand that is transcribed. The nucleotide sequence of the RNA is that of the top strand of the DNA, except that U replaces T, so the mRNA for amino acids 1 through 7 is

```
5'-AUGUCCACUGCGGUCCUGGAA-3'
```

The codons are read from left to right according to the genetic code shown in Table 1.1. Codon AUG codes for Met (methionine), UCC codes for Ser (serine), and so on. Altogether, the amino acid sequence of this region of the polypeptide is

```
5'-AUG|UCC|ACU|GCG|GUC|CUG|GAA-3'
   Met|Ser|Thr|Ala|Val|Leu|Glu
```

or, in terms of the single-letter abbreviations,

```
5'-AUG|UCC|ACU|GCG|GUC|CUG|GAA-3'
   M | S | T | A | V | L | E
```

The full decoding operation for this region of the PAH gene is shown in **Figure 1.16**. In this figure, the initiation codon AUG is highlighted because some patients with PKU have a mutation in this particular codon. As might be expected from the fact that methionine is the initiation codon for polypeptide synthesis, cells in patients with this particular mutation fail to produce any of the PAH polypeptide. Mutation and its consequences are considered next.

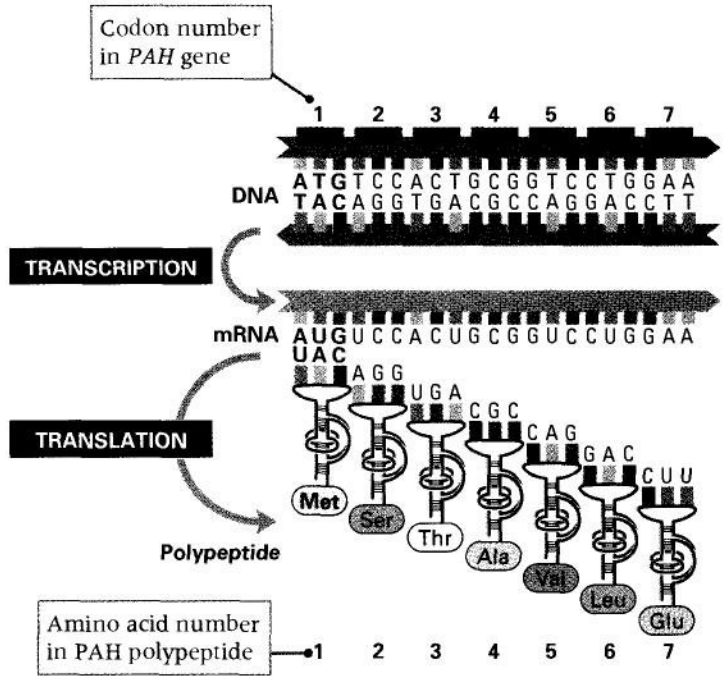


Figure 1.16 The central dogma in action. The DNA that encodes PAH serves as a template for the production of a messenger RNA, and the mRNA serves to specify the sequence of amino acids in the PAH polypeptide chain through interactions with the ribosome and tRNA molecules.

1.6 Mutation

The term **mutation** refers to any heritable change in a gene (or, more generally, in the genetic material) or to the process by which such a change takes place. One type of mutation results in a change in the sequence of bases in DNA. The change may be simple, such as the substitution of one pair of bases in a duplex molecule for a different pair of bases. For example, a C—G pair in a duplex molecule may mutate to T—A, A—T, or G—C. The change in base sequence may also be more complex, such as the deletion or addition of base pairs. These and other types of mutations are considered in Chapter 7. Geneticists also use the term **mutant,** which refers to the result of a mutation. A mutation yields a mutant gene, which in turn produces a mutant mRNA, a mutant protein, and finally a mutant organism that exhibits the effects of the mutation—for example, an inborn error of metabolism.

DNA from patients from all over the world who have phenylketonuria has been studied to determine what types of mutations are responsible for the inborn error. There are a large variety of mutant types. More than 400 different mutations have been described in the gene for PAH. In some cases part of the gene is missing, so the genetic information to make a complete PAH enzyme is absent. In other cases the genetic defect is more subtle, but the result is still either the failure to produce a PAH

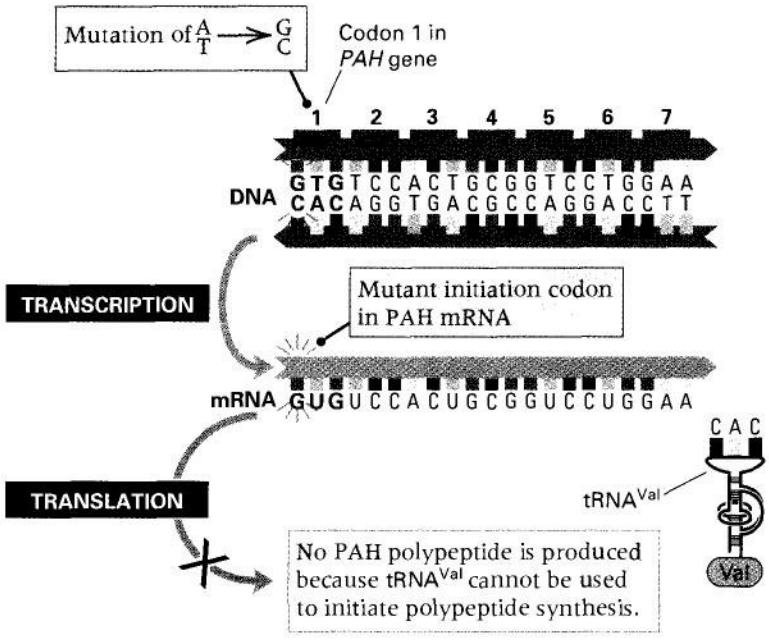


Figure 1.17 The M1V mutant in the *PAH* gene. The methionine codon needed for initiation mutates into a codon for valine. Translation cannot be initiated, and no PAH polypeptide is produced.

protein or the production of a PAH protein that is inactive. In the mutation shown in Figure 1.17, substitution of a G—C base pair for the normal A—T base pair at the very first position in the coding sequence changes the normal codon AUG (Met) used for the initiation of translation into the codon GUG, which normally specifies valine (Val) and cannot be used as a "start" codon. The result is that translation of the PAH mRNA cannot occur, and so no PAH polypeptide is made. This mutant is designated M1V because the codon for M (methionine) at amino acid position 1 in the PAH polypeptide has been changed to a codon for V (valine). Although the M1V mutant is quite rare worldwide, it is common in some localities, such as Québec Province in Canada.

One PAH mutant that is quite common is designated R408W, which means that codon 408 in the PAH polypeptide chain has been changed from one coding for arginine (R) to one coding for tryptophan (W). This mutation is one of the four most common among European Caucasians with PKU. The molecular basis of the mutant is shown in Figure 1.18. In this case, the first base pair in codon 408 is changed from a C—G base pair into a T—A base pair. The result is that the PAH mRNA has a mutant codon at position 408; specifically, it has UGG instead of CGG. Translation does occur in this mutant because everything else about the mRNA is normal, but the result is that the mutant PAH carries a tryptophan (Trp) instead of an arginine (Arg) at posi-



The women in the wedding photograph are sisters. Both have two copies of the same mutant *PAH* gene. The bride is the younger of the two. She was diagnosed just three days after birth and put on the PKU diet soon after. Her older sister, the maid of honor, was diagnosed too late to begin the diet and is mentally retarded. The two-year old pictured in the photo at the right is the daughter of the married couple. They planned the pregnancy: dietary control was strict from conception to delivery to avoid the hazards of excess phenylalanine harming the fetus. Their daughter has passed all developmental milestones with distinction. [Courtesy of Charles R. Scriver.]

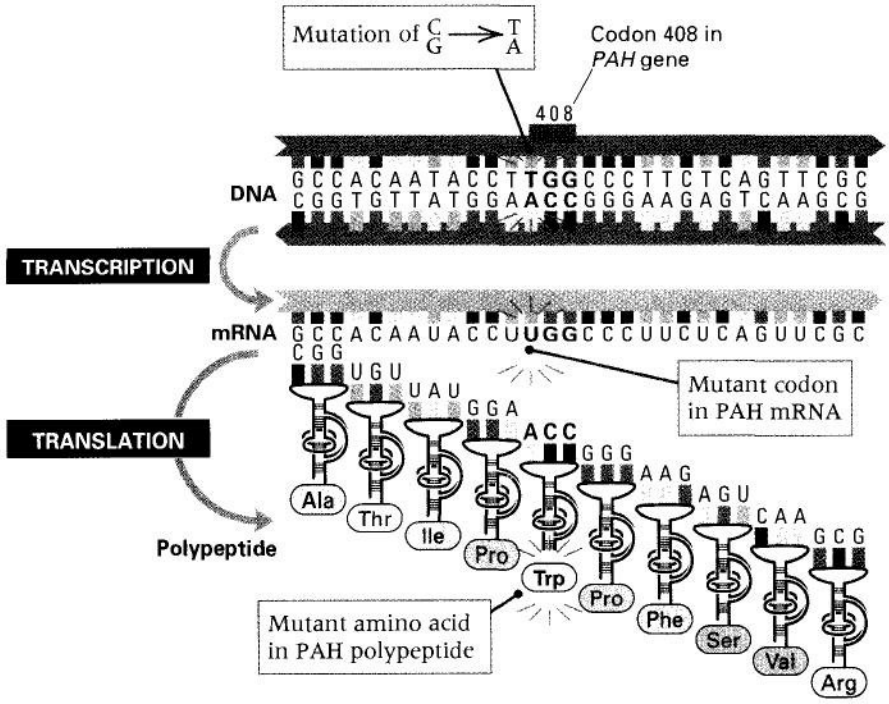


Figure 1.18 The R408W mutant in the *PAH* gene. Codon 408 for arginine (R) is mutated into a codon for tryptophan (W). The result is that position 408 in the mutant PAH polypeptide is occupied by tryptophan rather than by arginine. The mutant protein has no PAH enzyme activity.

tion 408 in the polypeptide chain. The consequence of the seemingly minor change of one amino acid is very drastic. Although the R408W polypeptide is complete, the enzyme has less than 3 percent of the activity of the normal enzyme.

Protein Folding and Stability

More than 400 different mutations in the *PAH* gene have been identified in patients with PKU throughout the world. Many of the mutations affect the level of expression of the gene or processing of the RNA transcript, and some mutations are deletions in which part of the gene is missing. But more than 240 of the mutations are simple amino acid replacements resulting from single nucleotide substitutions in the DNA. Surprisingly, only a minority of amino acid replacements result in a normal amount of PAH protein with a reduced enzyme activity. As a result of most mutations the amount of PAH is reduced, sometimes drastically, and in some other mutations the enzyme activity of the PAH protein that remains is virtually normal. Yet in all these cases the level of expression of the gene, and the amount of mRNA, are within the normal range.

The reason why so many amino acid replacements reduce the amount of protein is that they cause problems in protein folding, or in the coming together of the protein subunits, or in the stability of the folded protein. **Protein folding** is the complex process by which polypeptide chains attain a stable three-dimensional structure through short-range chemical interactions between nearby amino acids and long-range interactions between amino acids in different parts of the molecule. Folding normally occurs as the polypeptide is being synthesized on the ribosome, and the process is facilitated by a class of proteins known as **chaperones.** During the folding process, the polypeptide chain twists and bends until it achieves a minimum energy state that maximizes the stability of the resulting structure, which is referred to as the **native conformation.** For example, one aspect of protein folding is that hydrophobic amino acids, which have low affinity for water molecules, tend to move toward each other and to form a relatively hydrophobic center, or core, in the native conformation. For a polypeptide of realistic length, there are so many short-range and long-range interactions, and so many possible folded conformations, that even the fastest computers cannot calculate and compare all their energy levels. Computer simulation of protein folding has yielded some insights, but the reliable prediction of protein folding is still a major challenge.

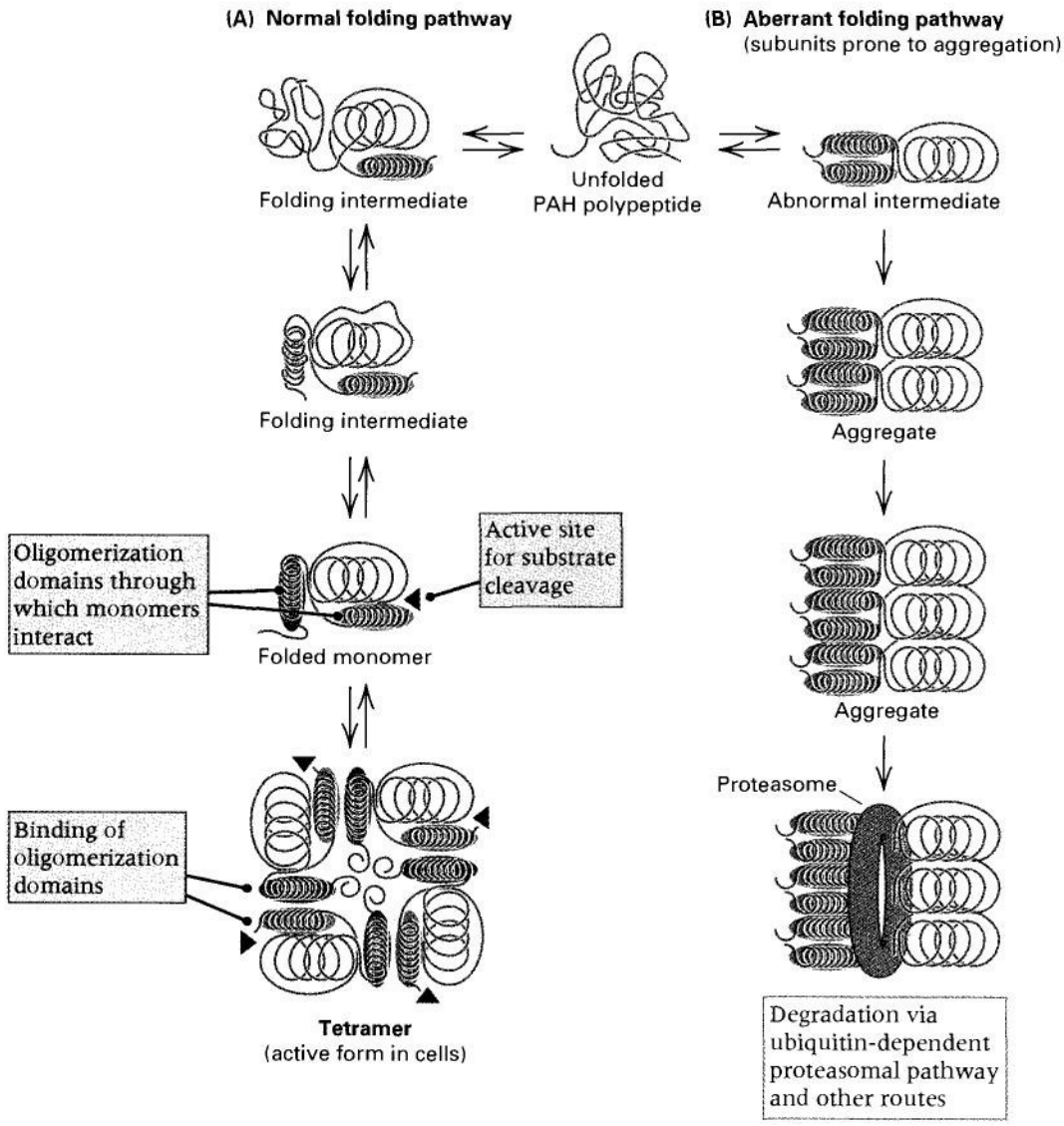
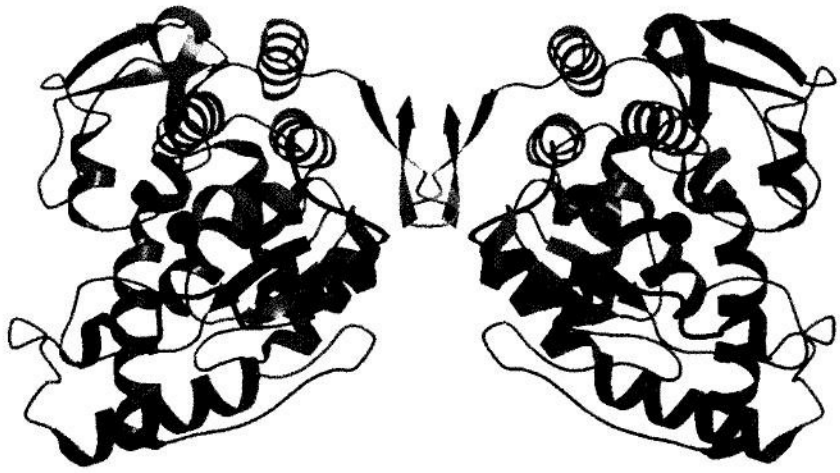


Figure 1.19 Some amino acid replacements perturb the ability of a protein to fold properly. (A) Normal folding in phenylalanine hydroxylase forms the active tetramer. (B) Abnormal folding of a mutant polypeptide chain results in the formation of polypeptide aggregates, which are progressively cleaved into the constituent amino acids through a ubiquitin-dependent proteosomal degradation pathway.

Hypothetical pathways of protein folding in the case of PAH are illustrated in **Figure 1.19**. The normal pathway is shown in part A, including some of the (typically short-lived) intermediates in the folding process. The native conformation of a single PAH polypeptide constitutes the PAH **monomer.** Like many other polypeptides, the PAH monomer contains short regions, called **oligomerization domains** (*oligo* means "a small number"), through which PAH polypeptides undergo stable binding to one another. In the case of PAH, the active form of the PAH enzyme is a **tetramer,** consisting of four identical polypeptide chains held together by interactions between the tetramerization domains. Note that the folding and tetramerization processes are reversible, so any amino acid replacement that decreases the stability of the tetramer or any of the intermediates will cause more of the PAH polypeptides to fold according to pathway B.

Pathway B in the figure is a misfolding pathway in which the folded monomers are prone to undergo *irreversible* aggregation with each other. These aggregates are targeted for enzymatic breakdown into their constituent amino acids, first by becoming covalently bound with a 76–amino acid polypeptide called *ubiquitin,* which is attached through the activity of several proteins, including a *ubiquitin-conjugating enzyme.* The tagged protein is then degraded by the *proteasome,* which is a large multiprotein complex containing proteins with



Three-dimensional structure of two of the four subunits in the active tetramer of phenylalanine hydroxylase. The chains of amino acids are represented as a sequence of curls, loops, and flat arrows, which represent different types of local structure described in Chapter 11. The oligomerization (in this case, tetramerization) domain is shown in green, and the catalytic domain containing the active site is shown in gold. Compare this with the simplified diagram in Figure 1.19. [Courtesy of R. C. Stevens, T. Flatmark, and H. Erlandsen. Reprinted by permission from *Nature Structural Biology* 4:995, H. Erlandsen, F. Fusetti, A Martinez, E. Hough, T. Flatmark, and R. C. Stevens. Copyright 1997 Macmillan Magazines Ltd.]

ubiquitin-binding, protease, and other activities. The ubiquitin/proteasome pathway is required for the degradation of specific proteins during the cell cycle and development, and it is also used to degrade certain proteins that are intrinsically unstable or that become unfolded in response to stress. In the case of PAH, many amino acid replacements decrease the stability of the protein to such an extent that the protein is routed through the misfolding pathway in Figure 1.19B and targeted for degradation. A destabilizing amino acid replacement can be located at virtually any position along the polypeptide chain.

1.7 Genes and Environment

Inborn errors of metabolism illustrate the general principle that genes code for proteins and that mutant genes code for mutant proteins. In cases such as PKU, mutant proteins cause such a drastic change in metabolism that a severe genetic defect results. But biology is not necessarily destiny. Organisms are also affected by the environment. PKU serves as an example of this principle, because patients who adhere to a diet restricted in the amount of phenylalanine develop mental capacities within the normal range. What is true in this example is true in general. Most traits are determined by the interaction of genes and environment.

It is also true that most traits are affected by multiple genes. No one knows how many genes are involved in the development and maturation of the brain and nervous system, but the number must be in the thousands. This number is in addition to the genes that are required in all cells to carry out metabolism and other basic life functions. It is easy to lose sight of the multiplicity of genes when considering extreme examples, such as PKU, in which a single mutation can have such a drastic effect on mental development. The situation is the same as that with any complex machine. An airplane can function if thousands of parts are working together in harmony, but only one defective part, if that part affects a vital system, can bring it down. Likewise, the development and functioning of every trait require a large number of genes working in harmony, but in some cases a single mutant gene can have catastrophic consequences.

In other words, the relationship between a gene and a trait is not necessarily a simple one. The biochemistry of organisms is a complex branching network in which different enzymes may share substrates, yield the same products, or be responsive to the same regulatory elements. The result is

that most visible traits of organisms are the net result of many genes acting together and in combination with environmental factors. PKU affords examples of each of three principles governing these interactions:

1. ***One gene can affect more than one trait.*** Children with extreme forms of PKU often have blond hair and reduced body pigment. This is because the absence of PAH is a metabolic block that prevents conversion of phenylalanine into tyrosine, which is the precursor of the pigment melanin. The relationship between severe mental retardation and decreased pigmentation in PKU makes sense only if one knows the metabolic connections among phenylalanine, tyrosine, and melanin. If these connections were not known, the traits would seem completely unrelated. PKU is not unusual in this regard. Many mutant genes affect multiple traits through their secondary or indirect effects. The various, sometimes seemingly unrelated, effects of a mutant gene are called **pleiotropic effects,** and the phenomenon itself is known as **pleiotropy.** Figure 1.20 shows a cat with white fur and blue eyes, a pattern of pigmentation that is often (about 40 percent of the time) associated with deafness. Hence deafness can be regarded as a pleiotropic effect of white coat and blue eye color. The developmental basis of this pleiotropy is unknown.

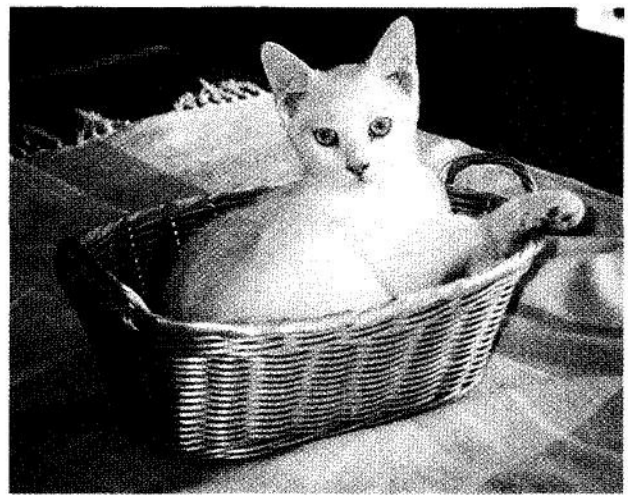


Figure 1.20 Among cats with white fur and blue eyes, about 40 percent are born deaf. We do not know why there is defective hearing nor why it is so often associated with coat and eye color. This form of deafness can be regarded as a pleiotropic effect of white fur and blue eyes. [Courtesy of Janet Klein.]

2. ***Any trait can be affected by more than one gene.*** We discussed this principle earlier in connection with the large number of genes that are required for the normal development and functioning of the brain and nervous system. Among these are genes that affect the function of the blood–brain barrier, which consists of specialized glial cells wrapped around tight capillary walls in the brain, forming an impediment to the passage of most water-soluble molecules from the blood to the brain. The blood–brain barrier therefore affects the extent to which excess free phenylalanine in the blood can enter the brain itself. Because the effectiveness of the blood–brain barrier differs among individuals, PKU patients with very similar levels of blood phenylalanine can have dramatically different levels of cognitive development. This also explains in part why adherence to a controlled-phenylalanine diet is critically important in children but less so in adults; the blood–brain barrier is less well developed in children and is therefore less effective in blocking the excess phenylalanine.

Multiple genes affect even simpler metabolic traits. Phenylalanine breakdown and excretion serve as a convenient example. The metabolic pathway is illustrated in Figure 1.10. Four enzymes in the pathway are indicated, but even more enzymes are involved at the stage labeled "further breakdown." Because differences in the activity of any of these enzymes can affect the rate at which phenylalanine can be broken down and excreted, all of the enzymes in the pathway are important in determining the amount of excess phenylalanine in the blood of patients with PKU.

3. ***Most traits are affected by environmental factors as well as by genes.*** Here we come back to the low-phenylalanine diet. Children with PKU are not doomed to severe mental deficiency. Their capabilities can be brought into the normal range by dietary treatment. PKU serves as an example of what motivates geneticists to try to discover the molecular basis of inherited disease. The hope is that knowing the metabolic basis of the disease will eventually make it possible to develop methods for clinical intervention through diet, medication, or other treatments that will ameliorate the severity of the disease.

1.8 Evolution: From Genes to Genomes, from Proteins to Proteomes

The pathway for the breakdown and excretion of phenylalanine is by no means unique to human beings. One of the remarkable generalizations to have emerged from molecular genetics is that organisms that are very distinct—for example, plants and animals—share many features in their genetics and biochemistry. These similarities indicate a fundamental "unity of life":

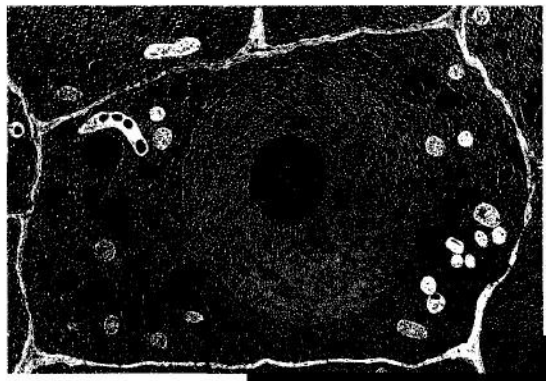
> All creatures on Earth share many features of the genetic apparatus, including genetic information encoded in the sequence of bases in DNA, transcription into RNA, and translation into protein on ribosomes with the use of transfer RNAs. All creatures also share certain characteristics in their biochemistry, including many enzymes and other proteins that are similar in amino acid sequence, three-dimensional structure, and function.

The totality of DNA in a single cell is called the **genome** of the organism. In sexual organisms, the genome is usually regarded as the DNA present in a reproductive cell. The human genome, which is contained in the chromosomes of a sperm or egg, includes approximately 3 billion nucleotide pairs of DNA. The complete set of proteins encoded in the genome is known as the **proteome**. The study of genomes constitutes *genomics;* the study of proteomes constitutes *proteomics*. The fundamental unity of life can be seen in the similarity of proteins in the proteomes of diverse types of organisms. For example, the proteome of the fruit fly *Drosophila melanogaster* includes 13,601 proteins; these can be grouped into 8065 different *families* of proteins that are similar in amino acid sequence. For comparison, the proteome produced by the nematode worm *Caenorhabditis elegans* contains 18,424 proteins that can be grouped into 9453 families. These two proteomes share about 5000 proteins that are sufficiently similar to be regarded as having a common function. Among the protein families that are common to flies and worms, about 3000 are also found in the proteome of the yeast *Saccharomyces cerevisiae*. Based on these data from these three completely sequenced complex genomes, it seems likely that all organisms whose cells have a nucleus and chromosomes will turn out to share several thousand protein families. Furthermore, about a thousand of these protein families are shared with organisms as distantly related as bacteria.

The Molecular Unity of Life

Why do organisms share a common set of similar genes and proteins? Because all creatures share a common origin. The process of **evolution** takes place when a population of organisms descended from a common ancestor gradually changes in genetic composition through time. From an evolutionary perspective, the unity of fundamental molecular processes is derived by inheritance from a distant common ancestor in which the molecular mechanisms were already in place.

Not only the unity of life but also many other features of living organisms become comprehensible from an evolutionary perspective. For example, the interposition of an RNA intermediate in the basic flow of



Dr. Jeremy Burgess/Science Photo Library/Photo Researchers, Inc.



Dr. Gopal Murti/Science Photo Library/Photo Researchers, Inc.

Membrane-bound organelles, similar in form and function, fill the body of both plant (top) and animal (right) cells. Despite the fact that these cells come from two very different organisms, their similarities speak to a fundamental unity of life. Taking up most of each cell is the nucleus, where the genetic material resides.

genetic information from DNA to RNA to protein makes sense if the earliest forms of life used RNA for both genetic information and enzyme catalysis. The importance of the evolutionary perspective in undestanding aspects of biology that seem pointless or needlessly complex is summed up in a famous aphorism of the evolutionary biologist Theodosius Dobzhansky: "Nothing in biology makes sense except in the light of evolution."

One indication of the common ancestry among Earth's creatures is illustrated in Figure 1.21. The tree of relationships was inferred from similarities in nucleotide sequence in an RNA molecule found in the small subunit of the ribosome. Three major kingdoms of organisms are distinguished:

1. ***The kingdom Bacteria*** This group includes most bacteria and cyanobacteria (formerly called blue-green algae). Cells of these organisms lack a membrane-bounded nucleus and mitochondria, are surrounded by a cell wall, and divide by binary fission.
2. ***The kingdom Archaea*** This group was initially discovered among microorganisms that produce methane gas or that live in extreme environments, such as hot springs or high salt concentrations. They are widely distributed in more normal environments as well. Like those of bacteria, the cells of archaeans lack internal membranes. DNA sequence analysis indicates that the machinery for DNA replication and transcription in archaeans resembles that of eukaryans, whereas their metabolism strongly resembles that of bacteria. About half of the genes found in the kingdom Archaea are unique to this group.
3. ***The kingdom Eukarya*** This group includes all organisms whose cells contain an elaborate network of internal membranes, a membrane-bounded nucleus, and mitochondria. Their DNA is organized into true chromosomes, and cell division takes place by means of mitosis (discussed in Chapter 4). The eukaryotes include plants and animals as well as fungi and many single-celled organisms, such as amoebae and ciliated protozoa.

The members of the kingdoms Bacteria and Archaea are often grouped together into a larger assemblage called **prokaryotes,** which literally means "before [the evolution of] the nucleus." This terminology is convenient for designating prokaryotes as a group in contrast with **eukaryotes,** which literally means "good [well-formed] nucleus."

Natural Selection and Diversity

Although Figure 1.21 illustrates the unity of life, it also illustrates life's diversity. Frogs are different from fungi, and beetles are different from bacteria. As a human being, it is sobering to consider that complex, multicellular organisms are relatively recent ar-

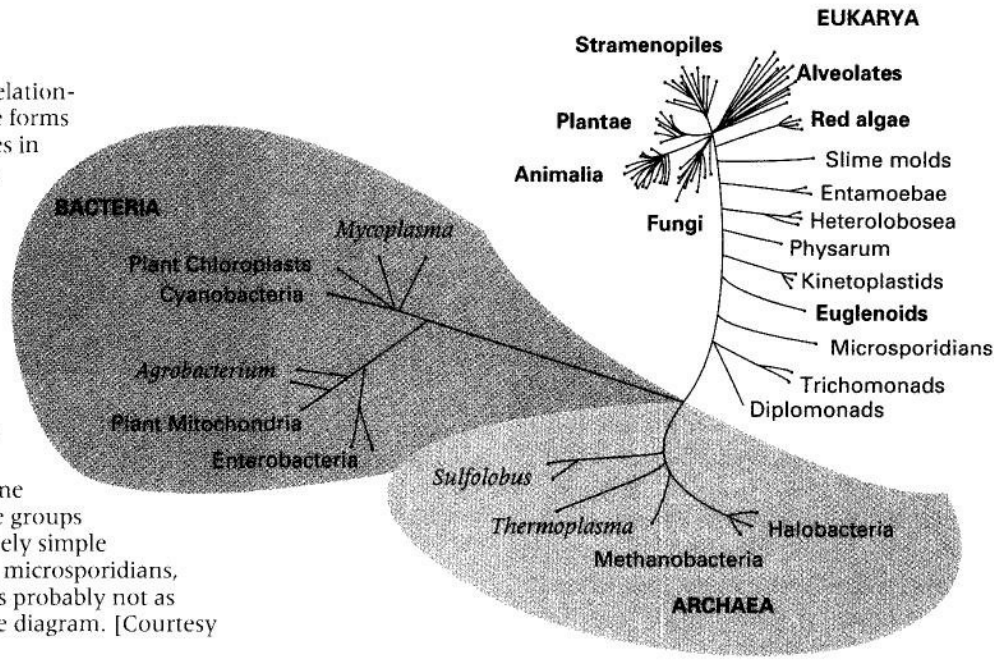


Figure 1.21 Evolutionary relationships among the major life forms as inferred from similarities in nucleotide sequence in an RNA molecule found in the small subunit of the ribosome. The three major kingdoms—Bacteria, Archaea, and Eukarya—are apparent. Plants, animals, and fungi are more closely related to each other than to members of either of the other kingdoms. Among eukaryotes, the time of branching of the diverse groups of undifferentiated, relatively simple organisms (trichomonads, microsporidians, euglenoids, and so forth) is probably not as ancient as suggested by the diagram. [Courtesy of Mitchell L. Sogin.]

rivals to the evolutionary scene of life on Earth. Animals came later still and primates very late indeed. What about human evolution? In the time scale of Earth history, human evolution is a matter of a few million years—barely a snap of the fingers.

If common ancestry is the source of the unity of life, what is the source of its diversity? Because differences among species are inherited, the original source of the differences must be mutation. However, mutations alone are not sufficient to explain why organisms are adapted to living in their environments—why ocean mammals have special adaptations for swimming and diving, why desert mammals have special adaptations that enable them to survive on minimal amounts of water. Mutations are chance events not directed toward any particular adaptive goal, such as longer fur among mammals living in the Arctic. The process that accounts for adaptation was described by Charles Darwin in his 1859 book *On the Origin of Species.* Darwin proposed that adaptation is the result of **natural selection:** Individual organisms carrying particular mutations or combinations of mutations that enable them to survive or reproduce more effectively in the prevailing environment leave more offspring than other organisms and so contribute their favorable genes disproportionately to future generations. If this process is repeated throughout the course of many generations, the entire species becomes genetically transformed—that is, it evolves—because a gradually increasing proportion of the population inherits the favorable mutations. Mutation, natural selection, and other features of the field of *population genetics* (also called *evolutionary genetics*) are discussed in Chapter 17.

Chapter Summary

Organisms of the same species have some traits (characteristics) in common but may differ from each other in innumerable other traits. Many of the differences between organisms result from genetic differences, the effects of the environment, or both. Genetics is the study of inherited traits, including those influenced in part by the environment. The elements of heredity consist of genes, which are transmitted from parents to offspring in reproduction. Although the sorting of genes in successive generations was first expressed numerically by Mendel, the chemical basis of genes was discovered by Miescher in the form of a weak acid, deoxyribonucleic acid (DNA). However, experimental proof that DNA is the genetic material did not come until about the middle of the twentieth century.

The first convincing evidence of the role of DNA in heredity came from the experiments of Avery, MacLeod, and McCarty, who showed that genetic characteristics in bacteria could be altered from one type to another by treatment with purified DNA. In studies of *Streptococcus pneumoniae,* they transformed mutant cells unable to cause pneumonia into cells that could do so by treating them with pure DNA from disease-causing forms. A second important line of evidence was the Hershey–Chase experiment. Hershey and Chase showed that the T2 bacterial virus injects primarily DNA into the host bacterium (*Escherichia coli*) and that a much higher proportion of parental DNA, compared with parental protein, is found among the progeny phage.

The three-dimensional structure of DNA, proposed in 1953 by Watson and Crick, gave many clues to the manner in which DNA functions as the genetic material. A molecule of DNA consists of two long chains of nucleotide subunits twisted around each other to form a right-handed helix. Each nucleotide subunit contains any one of four bases: A (adenine), T (thymine), G (guanine), or C (cytosine). The bases are paired in the two strands of a DNA molecule; wherever one strand has an A, the partner strand has a T, and wherever one strand has a G, the partner strand has a C. The base pairing means that the two paired strands in a DNA duplex molecule have complementary base sequences along their lengths. The structure of the DNA molecule suggested that genetic information could be coded in DNA in the sequence of bases. Mutations—changes in the genetic material—result from changes in the sequence of bases, such as the substitution of one nucleotide for another or the insertion or deletion of one or more nucleotides. The structure of DNA also suggested a mode of replication: The two strands of the parental DNA molecule separate, and each individual strand serves as a template for the synthesis of a new, complementary strand.

Most genes code for proteins. More precisely stated, most genes specify the sequence of amino acids in a polypeptide chain. The transfer of genetic information from DNA into protein is a multistep process that includes several types of RNA (ribonucleic acid). Structurally, an RNA strand is similar to a DNA strand except that the "backbone" contains a different sugar (ribose instead of deoxyribose) and RNA contains the base uracil (U) instead of thymine (T). Also, RNA is usually present in cells in the form of single, unpaired strands. The initial step in gene expression is transcription, in which a molecule of RNA is synthesized that is complementary in base sequence to whichever DNA strand is being transcribed.

In polypeptide synthesis, which takes place on a ribosome, the base sequence in the RNA transcript is translated in groups of three adjacent bases (codons). The codons are recognized by different types of transfer RNA (tRNA) through base pairing. Each type of tRNA is attached to a particular amino acid, and when a tRNA base-pairs with the proper codon on the ribosome, the growing end of the polypeptide chain is transferred to the amino acid on the tRNA. The table of all codons and the amino acids they specify is called the genetic code. Special codons specify the "start" (AUG, Met) and "stop" (UAA, UAG, and UGA) of polypeptide synthesis. The reason why various types of RNA are an intimate part of transcription and translation is probably that the earliest forms of life used RNA for both genetic information and enzyme catalysis.

A mutation that alters one or more codons in a gene can change the amino acid sequence of the resulting polypeptide chain synthesized in the cell. Often the altered protein is functionally defective, so an inborn error of metabolism results. One of the first inborn errors of metabolism studied was alkaptonuria; it results from the absence of an enzyme for cleaving homogentisic acid, which accumulates and is excreted in the urine, turning black upon oxidation. Phenylketonuria (PKU) is an inborn error of metabolism that affects the same metabolic pathway. The enzyme defect in PKU results in an inability to convert phenylalanine to tyrosine. Phenylalanine accumulation has catastrophic effects on the development of the brain. Children with the disease have severe mental deficits unless they are treated with a special diet low in phenylalanine.

Most visible traits of organisms result from many genes acting together in combination with environmental factors. The relationship between genes and traits is often complex because (1) every gene potentially affects many traits (pleiotropy), (2) every trait is potentially affected by many genes, and (3) many traits are significantly affected by environmental factors as well as by genes.

All living creatures are united by sharing many features of the genetic apparatus (for example, transcription and translation) and many aspects of metabolism. They share many similar genes in the genome and many similar proteins in the proteome. The unity of life results from all life being of common ancestry and provides evidence for evolution. There is also great diversity among living creatures. The three major kingdoms of organisms are the kingdoms Bacteria (whose members lack a membrane-bounded nucleus), Archaea (whose members share features with both bacteria and eukaryans but form a distinct group), and Eukarya (which includes all "higher" organisms whose cells have a membrane-bounded nucleus that contains DNA organized into discrete chromosomes). Members of the kingdoms Bacteria and Archaea are often collectively called prokaryotes. The ultimate source of diversity among organisms is mutation, but natural selection is the process by which mutations that enhance survival and reproduction are retained and mutations that are harmful are eliminated. Natural selection, first proposed by Darwin, is therefore the primary mechanism by which organisms become progressively better adapted to their environments.

Key Terms

adenine (A)
alkaptonuria
amino acid
Archaea
Bacteria
bacteriophage
base (in DNA or RNA)
biochemical pathway
block (in a biochemical pathway)
central dogma
chaperone
chromosome
codon
colony
complementary base sequence
cytosine (C)
deoxyribose
deoxyribonucleic acid (DNA)
double-stranded DNA
duplex DNA
enzyme
Eukarya
eukaryote
evolution
gene
genetic code
genetics
genome
guanine (G)
homogentisic acid
inborn error of metabolism
messenger RNA (mRNA)
metabolic pathway
metabolite
monomer
mutant
mutation
native conformation
natural selection
nucleotide
oligomerization domain
phage
phenylalanine hydroxylase (PAH)
phenylketonuria (PKU)
pleiotropic effect
pleiotropy
polarity (of DNA or RNA)
polypeptide chain
product molecule
prokaryote
protein folding
proteome
replication
ribonucleic acid (RNA)
ribose
ribosomal RNA (rRNA)
ribosome
single-stranded DNA
substrate molecule
template
tetramer
thymine (T)
transcript
transcription
transfer RNA (tRNA)
transformation
translation
uracil (U)
Watson–Crick base pairing

Review the Basics

- What were the key experiments showing that DNA is the genetic material?
- How did understanding the molecular structure of DNA give clues to its ability to replicate, to code for proteins, and to undergo mutation?
- Why is pairing of complementary bases a key feature of DNA replication?
- What is the process of transcription and in what ways does it differ from DNA replication?
- What three types of RNA participate in protein synthesis, and what is the role of each type of RNA?
- What is the "genetic code," and how is it relevant to the translation of a polypeptide chain from a molecule of messenger RNA?
- What is an inborn error of metabolism? How did this concept serve as a bridge between genetics and biochemistry?
- How does the "central dogma" explain Garrod's discovery that nonfunctional enzymes result from mutant genes?
- Explain why many mutant forms of phenylalanine hydroxylase have a simple amino acid replacement, yet the mutant polypeptide chains are absent or present in very small amounts.?
- What is a pleiotropic effect of a gene mutation? Give an example.
- What are some of the major differences in cellular organization among Bacteria, Archaea, and Eukarya?
- What process was Charles Darwin describing when he wrote the following statement? "As more individuals are produced than can possibly survive, there must in every case be a struggle for existence, either one individual with another of the same species, or with the individuals of distinct species, or with the physical conditions of life."

Guide to Problem Solving

Problem 1 In the human gene for the β chain of hemoglobin (the oxygen-carrying protein in the red blood cells), the first 30 nucleotides in the amino-acid–coding region have the sequence

3'-TACCACGTGGACTGAGGACTCCTCTTCAGA-5'

What is the sequence of the partner strand?

Answer The base pairing between the strands is A with T and G with C, but it is equally important that the strands in a DNA duplex have opposite polarity. The partner strand is therefore oriented with its 5' end at the left, and the base sequence is

5'-ATGGTGCACCTGACTCCTGAGGAGAAGTCT-3'

Problem 2 If the DNA duplex for the β chain of hemoglobin in Problem 1 were transcribed from left to right, deduce the base sequence of the RNA in this coding region.

Answer To deduce the RNA sequence, we must apply three concepts. First, in the transcription of RNA, the base pairing is such that an A, T, G, or C in the DNA template strand is transcribed as U, A, C, or G, respectively, in the RNA strand. Second, the RNA transcript and the DNA template strand have opposite polarity. Third (and critically for this problem), the RNA transcript is always transcribed in the 5'-to-3' direction, so the 5' end of the RNA is the end synthesized first. This being the case, and considering the opposite polarity, the 3' end of the template strand must be transcribed first. Because we are told that transcription takes place from left to right, we can deduce that the transcribed strand is that given in Problem 1. The RNA transcript therefore has the base sequence

5'-AUGGUGCACCUGACUCCUGAGGAGAAGUCU-3'

Problem 3 Given the RNA sequence coding for part of human β hemoglobin deduced in Problem 2, what is the amino acid sequence in this part of the β polypeptide chain?

Answer The polypeptide chain is translated in successive groups of three nucleotides (each group constituting a codon), starting at the 5' end of the coding sequence and moving in the 5'-to-3' direction. The amino acid corresponding to each codon can be found in the genetic code table. The first ten amino acids in the polypeptide chain are therefore

```
5'-AUG|GUG|CAC|CUG|ACU|CCU|GAG|GAG|AAG|UCU-3'
   Met|Val|His|Leu|Thr|Pro|Glu|Glu|Lys|Ser
```

Problem 4 A very important mutation in human hemoglobin occurs in the DNA sequence shown in Problem 1. In this mutation, the T at nucleotide position 20 is replaced with an A, where the numbering is from the left in the strand shown at the top in Problem 1. The mutant hemoglobin is called *sickle-cell hemoglobin*, and it is associated with a severe anemia known as *sickle-cell anemia*. Severe as the genetic disease is, the mutant gene is present at relatively high frequency in some human populations because carriers of the gene, who have only a mild anemia, are more resistant to falciparum malaria than are noncarriers. What is the nucleotide sequence of this region of the

GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- James D. Watson once said that he and Francis Crick had no doubt that their proposed DNA structure was essentially correct, because the structure was so beautiful it had to be true! At an internet site accessed by the keyword **DNA**, you can view a large collection of different types of models of DNA structure. Some models highlight the sugar–phosphate backbones, others the A—T and G—C base pairs, still others the helical structure of double-stranded DNA.

- How was **PKU** discovered? The story is told by the Norwegian physician Ivar Fölling: "The stage is set in 1934. A mother with two severely mentally retarded children came to see my father. . . . She had asked many doctors for help, which none had been able to give. But this woman was unusually persistent and would not accept the situation without explanation. She had also noticed that a peculiar smell always clung to her children. . . . This woman was advised to seek help from my father [who held a professorship of nutrient research at the University Hospital in Norway]. He of course had no real hope of being able to help her. But he did not want to reject her, and he agreed to examine the children. On clinical examination he found nothing of importance, except the [severe mental retardation], which was beyond doubt. [Urine analysis] was part of his thorough routine examination. On adding ferric chloride [to normal urine] the color normally stays brownish, [but in the case of these children] a deep green color developed. He had not seen this reaction before . . . He concluded that two mentally retarded children excreted a substance not found in normal urine. But which substance?" Consult this keyword site for more information on how Asbjörn Fölling finally tracked down the cause of the disease.

- With proper dietary control of blood phenylalanine, patients with PKU can develop normally and lead normal lives.

DNA duplex in sickle-cell hemoglobin (both strands) and that of the messenger RNA, and what is the amino acid replacement that results in sickle-cell hemoglobin?

Answer The mutation is already given as a T → A substitution at position 20. The sequence of the DNA duplex is obtained as in Problem 1, that of the RNA as in Problem 2, and that of the mutant polypeptide chain as in Problem 3, except that at each step there is one nucleotide (or one amino acid) that differs from the nonmutant. The DNA, RNA, and polypeptide in this region of sickle-cell hemoglobin are as follows, where the differences from the nonmutant gene are in red. The amino acid replacement is glutamic acid → valine.

DNA (transcribed strand)	3'-TAC CAC GTG GAC TGA GGA CAC CTC TTC AGA-5'
DNA (nontranscribed strand)	5'-ATG GTG CAC CTG ACT CCT GTG GAG AAG TCT-3'
	↓
RNA coding region	5'-AUG GUG CAC CUG ACU CCU GUG GAG AAG UCU-3'
Polypeptide chain	Met Val His Leu Thr Pro Val Glu Lys Ser

Problem 4 Answer—Nucleotide sequences

Analysis and Applications

1.1 Classify each of the following statements as true or false.

(a) Each gene is responsible for only one visible trait.
(b) Every trait is potentially affected by many genes.
(c) The sequence of nucleotides in a gene specifies the sequence of amino acids in a protein encoded by the gene.
(d) There is one-to-one correspondence between the set of codons in the genetic code and the set of amino acids encoded.

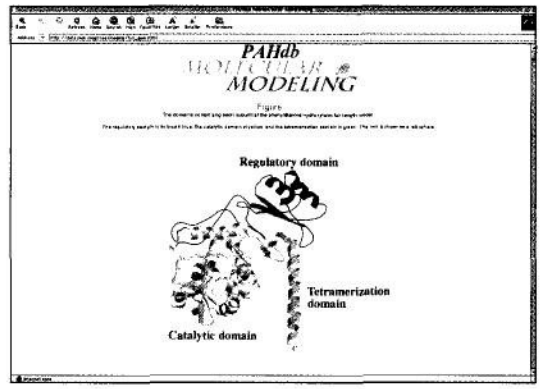
1.2 From their examination of the structure of DNA, what were Watson and Crick able to infer about the probable mechanisms of DNA replication, coding capability, and mutation?

When dietary control is relaxed, however, blood phenylalanine returns to high levels. This situation is extremely dangerous for a developing fetus, resulting in high risk of congenital heart disease, small head size, mental retardation, and slow growth. Affected children are said to have **maternal PKU**. They are affected, not because of their own inability to metabolize phenylalanine, but because of high levels of phenyalanine in their mothers' blood. The risk can be reduced, but not entirely eliminated, if PKU mothers plan their pregnancies and adhere to a strict dietary regimen prior to and during pregnancy. To learn more about this unanticipated consequence of dietary treatment of PKU, log onto the phenylalanine hydroxylase knowledge database at the keyword site.

- Perhaps surprisingly, the history of the bacteriophage T2 that figures so prominently in the experiments of Hershey and Chase is shrouded in mystery. Indeed, the time, place, and source material for the original isolation of phages T2, T4, and T6 (known as the "T-even phages") may never be known with certainty. Use the keyword **T2** to learn what is known about the origin of the bacteriophage and what sleuthing was required to discover what is known.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 1 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 1.



1.3 What does it mean to say that each strand of a duplex DNA molecule has a polarity? What does it mean to say that the paired strands in a duplex molecule have opposite polarity?

1.4 What is the end result of replication of a duplex DNA molecule?

1.5 What is the role of the messenger RNA in translation? What is the role of the ribosome? What is the role of transfer RNA? Is there more than one type of ribosome? Is there more than one type of transfer RNA?

1.6 What important observation about S and R strains of *Streptococcus pneumoniae* prompted Avery, MacLeod, and McCarty to study this organism?

1.7 In the transformation experiments of Avery, MacLeod, and McCarty, what was the strongest evidence that the substance responsible for the transformation was DNA rather than protein?

1.8 A chemical called phenol (carbolic acid) destroys proteins but not nucleic acids, and a strong alkali such as sodium hydroxide destroys both proteins and nucleic acids. In the transformation experiments with *Streptococcus pneumoniae*, what result would be expected if the S-strain extract had been treated with phenol? What would be expected if it had been treated with a strong alkali?

1.9 What feature of the physical organization of bacteriophage T2 made it suitable for use in the Hershey–Chase experiments?

1.10 Like DNA, molecules of RNA contain large amounts of phosphorus. When Hershey and Chase grew their T2 phage in bacterial cells that had grown in the presence of radioactive phosphorus, the RNA must also have incorporated the labeled phosphorus, and yet the experimental result was not compromised. Why not?

1.11 Although the Hershey–Chase experiments were widely accepted as proof that DNA is the genetic material, the results were not completely conclusive. Why not?

1.12 The DNA extracted from a bacteriophage contains 28 percent A, 28 percent T, 22 percent G, and 22 percent C. What can you conclude about the structure of this DNA molecule?

1.13 The DNA extracted from a bacteriophage consists of 24 percent A, 30 percent T, 20 percent G, and 26 percent C. What is unusual about this DNA? What can you conclude about its structure?

1.14 A double-stranded DNA molecule is separated into its constituent strands, and the strands are separated in an ultracentrifuge. In one of the strands the base composition is 24 percent A, 28 percent T, 22 percent G, and 26

percent C. What is the base composition of the other strand?

1.15 While studying sewage, you discover a new type of bacteriophage that infects *E. coli*. Chemical analysis reveals protein and RNA but no DNA. Is this possible?

1.16 One strand of a DNA duplex has the base sequence 5'-ATCGTATGCACTTTACCCGG-3'. What is the base sequence of the complementary strand?

1.17 A region along one strand of a double-stranded DNA molecule consists of tandem repeats of the trinucleotide 5'-TCG-3', so the sequence in this strand is

5'-TCGTCGTCGTCGTCG · · · -3'

What is the sequence in the other strand?

1.18 A duplex DNA molecule contains a random sequence of the four nucleotides with equal proportions of each. What is the average spacing between consecutive occurrences of the sequence 5'-GGCC-3'? Between consecutive occurrences of the sequence 5'-GAATTC-3'?

1.19 A region along a DNA strand that is transcribed contains no A. What base will be missing in the corresponding region of the RNA?

1.20 The duplex nucleic acid molecule shown here consists of a strand of DNA paired with a complementary strand of RNA. Is the RNA the top or the bottom strand? One of the base pairs is mismatched. Which pair is it?

5'-AUCGGUUACAUUCCGACUGA-3'
3'-TAGCCAATGTAAGGGTGACT-5'

1.21 The sequence of an RNA transcript that is initially synthesized is 5'-UAGCUAC-3', and successive nucleotides are added to the 3' end. This transcript is produced from a DNA strand with the sequence

3'-AAGTCGCATATCGATGCTAGCGCAACCT-5'

What is the sequence of the RNA transcript when synthesis is complete?

1.22 An RNA molecule folds back upon itself to form a "hairpin" structure held together by a region of base pairing. One segment of the molecule in the paired region has the base sequence 5'-AUACGAUA-3'. What is the base sequence with which this segment is paired?

1.23 A synthetic mRNA molecule consists of the repeating base sequence

5'-UUUUUUUUUUUU · · · -3'

When this molecule is translated *in vitro* using ribosomes, transfer RNAs, and other necessary constituents from *E. coli*, the result is a polypeptide chain consisting of the repeating amino acid Phe–Phe–Phe–Phe · · · . If you assume that the genetic code is a triplet code, what does this result imply about the codon for phenylalanine (Phe)?

1.24 A synthetic mRNA molecule consisting of the repeating base sequence

5'-UUUUUUUUUUU · · · -3'

is terminated by the addition, to the right-hand end, of a single nucleotide bearing A. When translated *in vitro*, the resulting polypeptide consists of a repeating sequence of phenylalanines terminated by a single leucine. What does this result imply about the codon for leucine?

1.25 With *in vitro* translation of an RNA into a polypeptide chain, the translation can begin anywhere along the RNA molecule. A synthetic RNA molecule has the sequence

5'-CGCUUACCACAUGUCGCGAACUCG-3'

How many reading frames are possible if this molecule is translated *in vitro*? How many reading frames are possible if this molecule is translated *in vivo*, in which translation starts with the codon AUG?

1.26 You have sequenced both strands of a double-stranded DNA molecule. To inspect the potential amino acid coding content of this molecule, you conceptually transcribe it into RNA and then conceptually translate the RNA into a polypeptide chain. How many reading frames will you have to examine?

1.27 A synthetic mRNA molecule consists of the repeating base sequence 5'-UCUCUCUCUCUCUC · · · -3'. When this molecule is translated *in vitro*, the result is a polypeptide chain consisting of the alternating amino acids Ser–Leu–Ser–Leu–Ser–Leu · · · . Why do the amino acids alternate? What does this result imply about the codons for serine (Ser) and leucine (Leu)?

1.28 A synthetic mRNA molecule consists of the repeating base sequence 5'-AUCAUCAUCAUCAUC · · · -3'. When this molecule is translated *in vitro*, the result is a mixture of three different polypeptide chains. One consists of repeating isoleucines (Ile–Ile–Ile–Ile · · ·), another of repeating serines (Ser–Ser–Ser–Ser · · ·), and the third of repeating histidines (His–His–His–His · · ·). What does this result imply about the manner in which an mRNA is translated?

1.29 How is it possible for a gene with a mutation in the coding region to encode a polypeptide with the same amino acid sequence as the nonmutant gene?

1.30 A polymer is made that has a random sequence consisting of 75 percent G's and 25 percent U's. Among the amino acids in the polypeptide chains resulting from *in vitro* translation, what is the expected frequency of Trp? of Val? of Phe?

Challenge Problems

1.31 The coding sequence in the messenger RNA for amino acids 1 through 10 of human phenylalanine hydroxylase is

5'-AUGUCCACUGCGGUCCUGAAAACCCAGGC-3'

(a) What are the first 10 amino acids?
(b) What sequence would result from a mutant RNA in which the red A was changed to G?
(c) What sequence would result from a mutant RNA in which the red C was changed to G?
(d) What sequence would result from a mutant RNA in which the red U was changed to C?
(e) What sequence would result from a mutant RNA in which the red G was changed to U?

1.32 A "frameshift" mutation is a mutation in which some number of base pairs, not a multiple of three, is inserted into or deleted from a coding region of DNA. The result is that, at the point of the frameshift mutation, the reading frame of protein translation is shifted with respect to the nonmutant gene. To see the consequence, consider that the coding sequence in the messenger RNA for the first 10 amino acids in human beta hemoglobin (part of the oxygen carrying protein in the blood) is

5'–AUGGUGCACCUGACUCCUGAGGAGAAGUCU · · · –3'

(a) What is the amino acid sequence in this part of the polypeptide chain?
(b) What would be the consequence of a frameshift mutation resulting in an RNA missing the red U?
(c) What would be the consequence of a frameshift mutation resulting in an RNA with a G inserted immediately in front of the red U?

1.33 With regard to the wildtype and mutant RNA molecules described in the previous problem, deduce the base sequence in both strands of the corresponding double-stranded DNA for:
(a) The wildtype sequence
(b) The single-base deletion
(c) The single-base insertion

Further Reading

Abedon, S. T. 2000. The murky origin of Snow White and her T-even dwarfs. *Genetics* 155: 481.

Bearn, A. G. 1994. Archibald Edward Garrod, the reluctant geneticist. *Genetics* 137: 1.

Calladine, C. R. 1997. *Understanding DNA: The Molecule and How It Works.* New York: Academic Press.

Ciechanover, A., A. Orian, and A. L. Schwartz. 2000. Ubiquitin-mediated proteolysis: Biological regulation via destruction. *Bioessays* 22: 442.

Doolittle, W. F. 2000. Uprooting the tree of life. *Scientific American,* February.

Gehrig, A., S. R. Schmidt, C. R. Muller, S. Srsen, K. Srsnova, and W. Kress. 1997. Molecular defects in alkaptonuria. *Cytogenetics & Cell Genetics* 76: 14.

Gould, S. J. 1994. The evolution of life on the earth. *Scientific American,* October.

Horowitz, N. H. 1996. The sixtieth anniversary of biochemical genetics. *Genetics* 143: 1.

Judson, H. F. 1996. *The Eighth Day of Creation: The Makers of the Revolution in Biology.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Mirsky, A. 1968. The discovery of DNA. *Scientific American,* June.

Olson, G. J., and C. R. Woese. 1997. Archaeal genomics: An overview. *Cell* 89: 991.

Radman, M., and R. Wagner. 1988. The high fidelity of DNA duplication. *Scientific American,* August.

Rennie, J. 1993. DNA's new twists. *Scientific American,* March.

Scazzocchio, C. 1997. Alkaptonuria: From humans to moulds and back. *Trends in Genetics* 13: 125.

Scriver, C. R., and P. J. Waters. 1999. Monogenic traits are not simple: Lessons from phenylketonuria. *Trends in Genetics* 15: 267.

Stadler, D. 1997. Ultraviolet-induced mutation and the chemical nature of the gene. *Genetics* 145: 863.

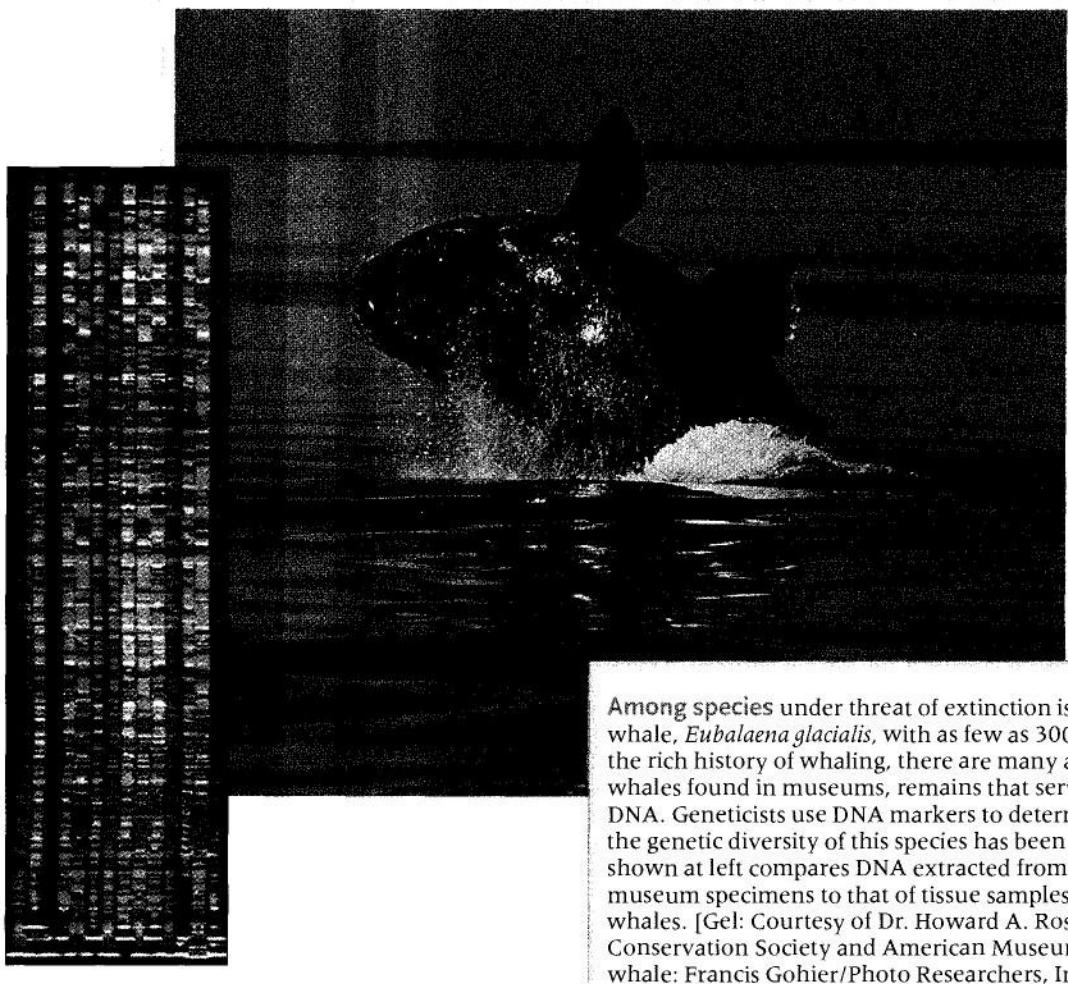
Susman, M. 1995. The Cold Spring Harbor phage course (1945–1970): A 50th anniversary remembrance. *Genetics* 139: 1101.

Watson, J. D. 1968. *The Double Helix.* New York: Atheneum.

Zhang, J. Z. 2000. Protein-length distributions for the three domains of life. *Trends in Genetics* 16: 107.

CHAPTER 2

DNA Structure and DNA Manipulation



Among species under threat of extinction is the Northern Right whale, *Eubalaena glacialis*, with as few as 300 remaining. Given the rich history of whaling, there are many artifacts of these whales found in museums, remains that serve as a source of DNA. Geneticists use DNA markers to determine to what extent the genetic diversity of this species has been reduced. The gel shown at left compares DNA extracted from whale baleen from museum specimens to that of tissue samples taken from living whales. [Gel: Courtesy of Dr. Howard A. Rosenbaum/Wildlife Conservation Society and American Museum of Natural History; whale: Francis Gohier/Photo Researchers, Inc.]

CHAPTER OUTLINE

2.1 **Genomes and Genetic Differences Among Individuals**
- DNA Markers as Landmarks in Chromosomes

2.2 **The Molecular Structure of DNA**
- Polynucleotide Chains
- Base Pairing and Base Stacking
- Antiparallel Strands
- DNA Structure as Related to Function

2.3 **The Separation and Identificatin of Genomic DNA Fragments**
- Restriction Enzymes and Site-Specific DNA Cleavage
- Gel Electrophoresis
- Nucleic Acid Hybridization
- The Southern Blot

2.4 **Selective Replication of Genomic DNA Fragments**
- Constraints on DNA Replication: Primers and 5'-to-3' Strand Elongation
- The Polymerase Chain Reaction

2.5 **The Terminology of Genetic Analysis**

2.6 **Types of DNA Markers Present in Genomic DNA**
- Single Nucleotide Polymorphisms (SNPs)
- Restriction Fragment Length Polymorphisms (RFLPs)
- Random Amplified Polymorphic DNA (RAPD)
- Amplified Fragment Length Polymorphisms (AFLPs)
- Simple Tandem Repeat Polymorphisms (STRPs)

2.7 **Applications of DNA Markers**
- Genetic Markers, Genetic Mapping, and "Disease Genes"
- Other Uses for DNA Markers

PRINCIPLES

- A DNA strand is a polymer of A, T, G and C deoxyribonucleotides joined 3'-to-5' by phosphodiester bonds.
- The two DNA strands in a duplex are held together by hydrogen bonding between the A–T and G–C base pairs and by base stacking of the paired bases.
- Each type of restriction endonuclease enzyme cleaves double-stranded DNA at a particular sequence of bases usually four or six nucleotides in length.
- The DNA fragments produced by a restriction enzyme can be separated by electrophoresis, isolated, sequenced, and manipulated in other ways.
- Separated strands of DNA or RNA that are complementary in nucleotide sequence can come together (hybridize) spontaneously to form duplexes.
- DNA replication takes place only by elongation of the growing strand in the 5'-to-3' direction through the addition of successive nucleotides to the 3' end.
- In the polymerase chain reaction, short oligonucleotide primers are used in successive cycles of DNA replication to amplify selectively a particular region of a DNA duplex.
- Genetic markers in DNA provide a large number of easily accessed sites in the genome that can be used to identify the chromosomal locations of disease genes, for DNA typing in individual identification, for the genetic improvement of cultivated plants and domesticated animals, and for many other applications.

CONNECTIONS

The Double Helix
James D. Watson and Francis H. C. Crick 1953
A Structure for Deoxyribose Nucleic Acid

Origin of the Human Genetic Linkage Map
David Botstein, Raymond L. White, Mark Skolnick, and Ronald W. Davis 1980
Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms

In Chapter 1, we reviewed the experimental evidence demonstrating that the genetic material is DNA. We saw how, through the unique structure of the DNA molecule, genetic information can be transcribed and translated into proteins that affect the inherited characteristics of organisms. When a mutant gene encodes a nonfunctional protein that results in some physical or physiological abnormality—for example, an "inborn error of metabolism"—the expression of that abnormality can be used to trace the transmission of the mutant gene from one generation to the next in a pedigree (family history). As a consequence, until recently, the first step in genetic analysis was the identification of organisms with abnormal traits, for example, peas with wrinkled seeds instead of round seeds and fruit flies with white eyes instead of red. These traits were studied by means of controlled crosses so that the parentage of each individual could be traced. Large-scale genetic studies were typically limited to one of a small number of model organisms especially favorable for isolating and identifying mutant genes, such as the budding yeast *Saccharomyces cerevisiae*, the nematode worm *Caenorhabditis elegans*, or the fruit fly *Drosophila melanogaster*.

Since the mid-1970s, studies in genetics have undergone a revolution based on the use of increasingly sophisticated ways to isolate and identify specific fragments of DNA. The culmination of these techniques was large-scale genomic sequencing—the ability to determine the correct sequence of the base pairs that make up the DNA in an entire genome and to identify the sequences associated with genes. Because many of the model organisms used in genetics have relatively small genomes, these sequences were completed first, in the late 1990s (**Figure 2.1**) The techniques used to sequence these simpler genomes were then scaled up to sequence the human genome. The initial "rough draft" of the human genome was announced in June 2000; this represents an important milestone in the Human Genome Project, whose goals in-

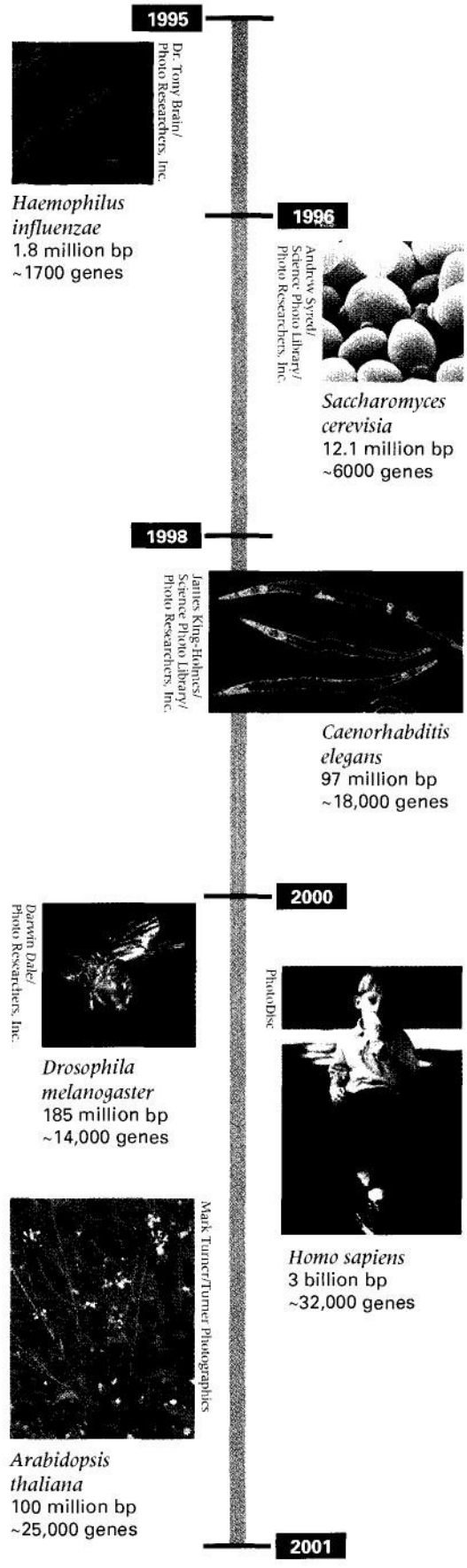


Figure 2.1 Timeline of large-scale genomic DNA sequencing.

clude determining the sequence, and identifying the function, of all human genes.

2.1 Genomes and Genetic Differences Among Individuals

The numbers associated with the genome of even a simple organism can be intimidating. The sequenced genomes of *D. melanogaster* and *C. elegans*, both approximately 100 million base pairs in length, encode 13,601 proteins and 18,424 proteins, respectively. The human genome is considerably larger. As found in a human reproductive cell, the human genome consists of 3 billion base pairs organized into 23 distinct chromosomes (each chromosome contains a single molecule of duplex DNA). A typical chromosome can contain several hundred to several thousand genes, arranged in linear order along the DNA molecule present in the chromosome. The sequences that make up the protein-coding part of these genes actually account for only about 4 percent of the entire genome. The other 96 percent of the sequences do not code for proteins. Some noncoding sequences are genetic "chaff" that gets separated from the protein-coding "wheat" when genes are transcribed and the RNA transcript is processed into messenger RNA. Other noncoding sequences are relatively short sequences that are found in hundreds of thousands of copies scattered throughout the genome. Still other noncoding sequences are remnants of genes called *pseudogenes*. As might be expected, identifying the protein-coding genes from among the large background of noncoding DNA in the human genome is a challenge in itself.

Geneticists often speak of the nucleotide sequence of "the" human genome because 99.9 percent of the DNA sequences in any two individuals are the same. This is our evolutionary legacy; it contains the genetic information that makes us human beings. In reality, however, there are many different human genomes. Geneticists have great interest in the 0.1 percent of the human DNA sequence—3 million base pairs—that differs from one genome to the next, because these differences include the mutations that are responsible for genetic diseases such as phenylketonuria and other inborn errors of metabolism, as well as the mutations that increase the risk of more complex diseases such as heart disease, breast cancer, and diabetes.

Fortunately, only a small proportion of all differences in DNA sequence are associated with disease. Some of the others are associated with inherited differences in height, weight, hair color, eye color, facial features, and other traits. Most of the genetic differences between people are completely harmless. Many have no detectable effects on appearance or health. Such differences can be studied only through direct examination of the DNA itself. These harmless differences are nevertheless important, because they serve as genetic markers.

DNA Markers as Landmarks in Chromosomes

In genetics, a **genetic marker** is any difference in DNA, no matter how it is detected, whose pattern of transmission from generation to generation can be tracked. Each individual who carries the marker also carries a length of chromosome on either side of it, so it *marks* a particular region of the genome. A mutant gene, or some portion of a mutant gene, can serve as a genetic marker. In the "classical" approach to genetics, it is the outward expression of a gene (or lack of expression) that forms the basis of genetic analysis. For example, a mutation causing wrinkled peas is a genetic marker, which can be identified through its effects on pea shape. In modern genetic analysis, *any* difference in DNA sequence between two individuals can serve as a genetic marker. And although these genetic markers are often harmless in themselves, they allow the positions of disease genes to be located and their DNA isolated, identified, and studied.

Genetic markers that are detected by direct analysis of the DNA are often called **DNA markers.** DNA markers are important in genetics because they serve as landmarks in long DNA molecules, such as those found in chromosomes, which allow genetic differences among individuals to be

tracked. They are like signposts along a highway. Using DNA markers as landmarks, the geneticist can identify the positions of normal genes, mutant genes, breaks in chromosomes, and other features important in genetic analysis (**Figure 2.2**).

The detection of DNA markers usually requires that the **genomic DNA** (the total DNA extracted from cells of an organism) be fragmented into pieces of manageable size (usually a few thousand nucleotide pairs) that can be manipulated in laboratory experiments. In the following sections, we examine some of the principal ways in which DNA is manipulated to reveal genetic differences among individuals, whether or not these differences find outward expression. Use of these methods broadens the scope of genetics, making it possible to carry out genetic analysis in *any* organism. This means that detailed genetic analysis is no longer restricted to human beings, domesticated animals, cultivated plants, and the relatively small number of model organisms favorable for genetic studies. Direct study of DNA eliminates the need for prior identification of genetic differences between individuals; it even eliminates the need for controlled crosses. The methods of molecular analysis discussed in this chapter have transformed genetics:

> The manipulation of DNA is the basic experimental operation in modern genetics.

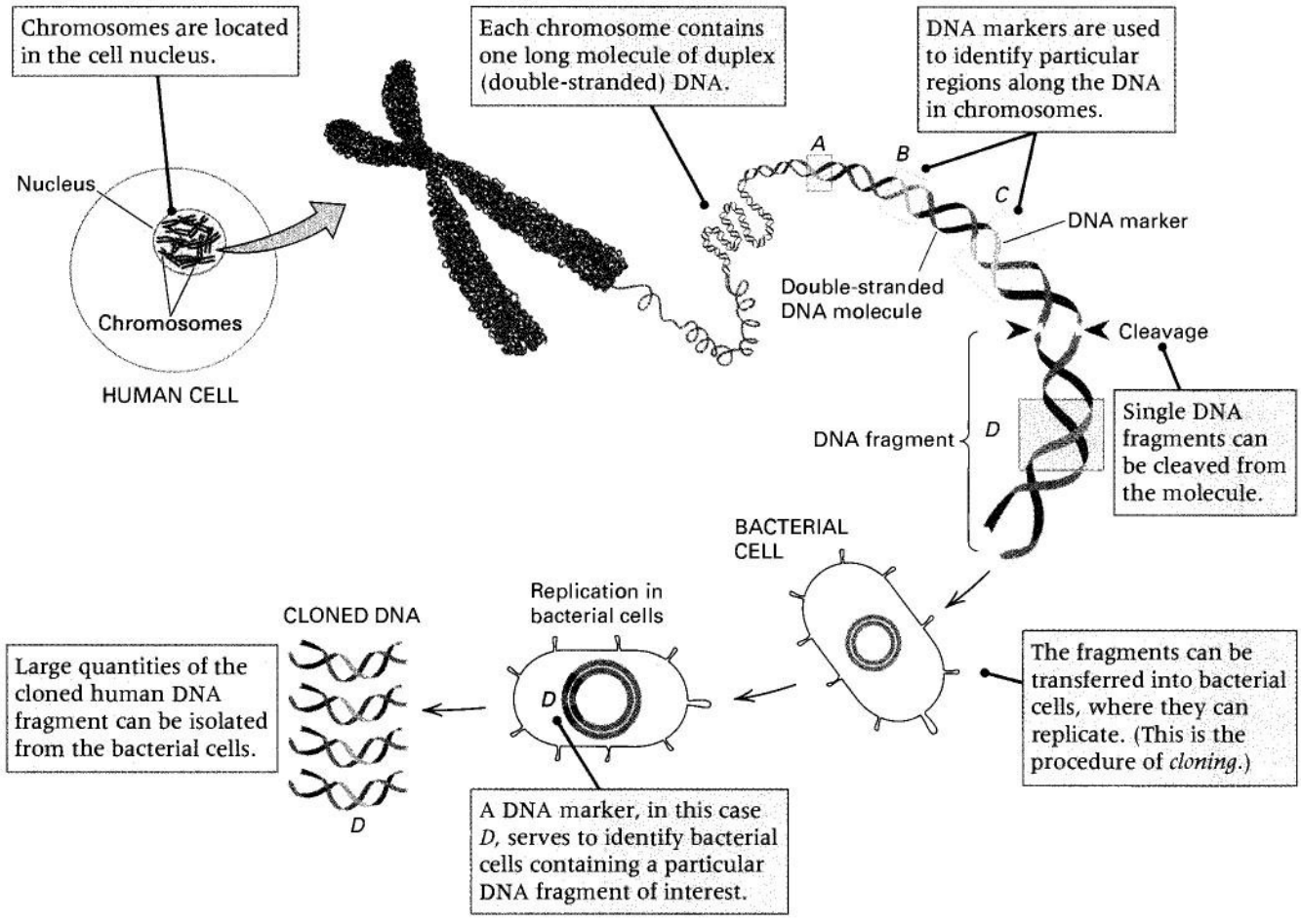


Figure 2.2 DNA markers serve as landmarks that identify physical positions along a DNA molecule, such as DNA from a chromosome. As shown above, a DNA marker can also be used to identify bacterial cells into which a particular fragment of DNA has been introduced. The procedure of DNA cloning is not quite as simple as indicated here; it is discussed further in Chapter 13.

These methods are the principal techniques used in virtually every modern genetics laboratory.



David Parker/Science Photo Library/Photo Researchers, Inc.

Much of the speed with which recent advances in genetics research have been made results from the use of high-throughput DNA sequencers coupled with computerized sequence acquisition, like these devices at the Sanger Centre near Cambridge, England. This technology has made it possible to determine the complete DNA sequence of the human genome.

2.2 The Molecular Structure of DNA

Modern experimental methods for the manipulation and analysis of DNA grew out of a detailed understanding of its molecular structure and replication. Therefore, to understand these methods, one needs to know something about the molecular structure of DNA. We saw in Chapter 1 that DNA is a helix of two paired, complementary strands, each composed of an ordered string of *nucleotides*, each bearing one of the bases A (adenine), T (thymine), G (guanine), or cytosine (C). Watson–Crick base pairing between A and T and between G and C in the complementary strands holds the strands together. The complementary strands also hold the key to replication, because each strand can serve as a template for the synthesis of a new complementary strand. We will now take a closer look at DNA structure and at the key features of its replication.

Polynucleotide Chains

In terms of biochemistry, a DNA strand is a polymer—a large molecule built from repeating units. The units in DNA are composed of 2′-deoxyribose (a five-carbon sugar), phosphoric acid, and the four nitrogen-containing bases denoted A, T, G, and C. The chemical structures of the bases are shown in **Figure 2.3**. Note that two of the bases have a double-ring structure; these are called **purines**. The other two bases have a single-ring structure; these are called **pyrimidines.**

- The purine bases are adenine (A) and guanine (G).
- The pyrimidine bases are thymine (T) and cytosine (C).

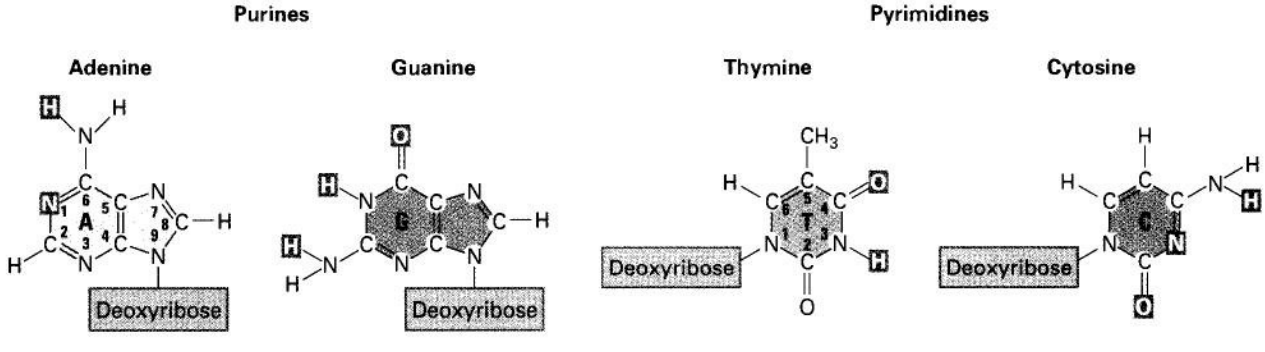


Figure 2.3 Chemical structures of the four nitrogen-containing bases in DNA: adenine, thymine, guanine, and cytosine. The nitrogen atom linked to the deoxyribose sugar is indicated. The atoms shown in red participate in hydrogen bonding between the DNA base pairs.

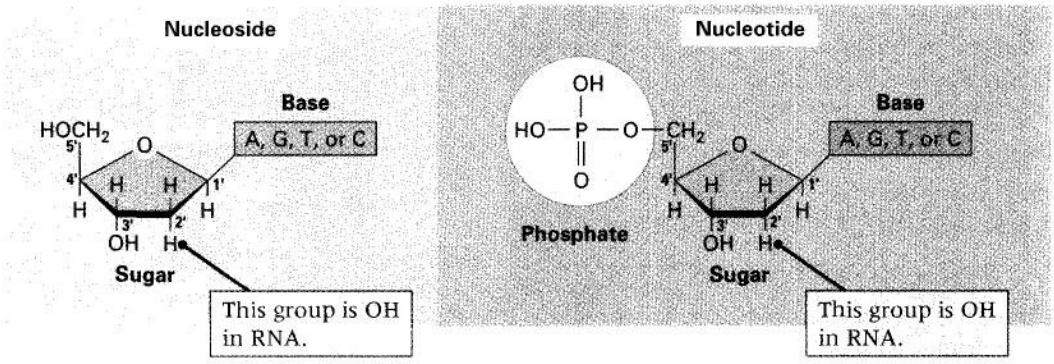


Figure 2.4 A typical nucleotide, showing the three major components (phosphate, sugar, and base), the difference between DNA and RNA, and the distinction between a nucleoside (no phosphate group) and a nucleotide (with phosphate). Nucleotides are monophosphates (with one phosphate group). Nucleoside diphosphates contain two phosphate groups, and nucleoside triphosphates contain three.

In DNA, each base is chemically linked to one molecule of the sugar deoxyribose, forming a compound called a **nucleoside.** When a phosphate group is also attached to the sugar, the nucleoside becomes a **nucleotide** (**Figure 2.4**). Thus a nucleotide is a nucleoside plus a phosphate. In the conventional numbering of the carbon atoms in the sugar in Figure 2.4, the carbon atom to which the base is attached is the 1' carbon. (The atoms in the sugar are given primed numbers to distinguish them from atoms in the bases.) The nomenclature of the nucleoside and nucleotide derivatives of the DNA bases is summarized in **Table 2.1**. Most of these terms are not needed in this book; they are included because they are likely to be encountered in further reading.

In nucleic acids, such as DNA and RNA, the nucleotides are joined to form a **polynucleotide chain,** in which the phosphate attached to the 5' carbon of one sugar is linked to the hydroxyl group attached to the 3' carbon of the next sugar in line (**Figure 2.5**). The chemical bonds by which the sugar components of adjacent nucleotides are linked through the phosphate groups are called **phosphodiester bonds.** The 5'–3'–5'–3' orientation of these linkages

Table 2.1 DNA nomenclature

Base	Nucleoside	Nucleotide
Adenine (A)	Deoxyadenosine	Deoxyadenosine-5' monophosphate (dAMP) diphosphate (dADP) triphosphate (dATP)
Guanine (G)	Deoxyguanosine	Deoxyguanosine-5' monophosphate (dGMP) diphosphate (dGDP) triphosphate (dGTP)
Thymine (T)	Deoxythymidine	Deoxythymidine-5' monophosphate (dTMP) diphosphate (dTDP) triphosphate (dTTP)
Cytosine (C)	Deoxycytidine	Deoxycytidine-5' monophosphate (dCMP) diphosphate (dCDP) triphosphate (dCTP)

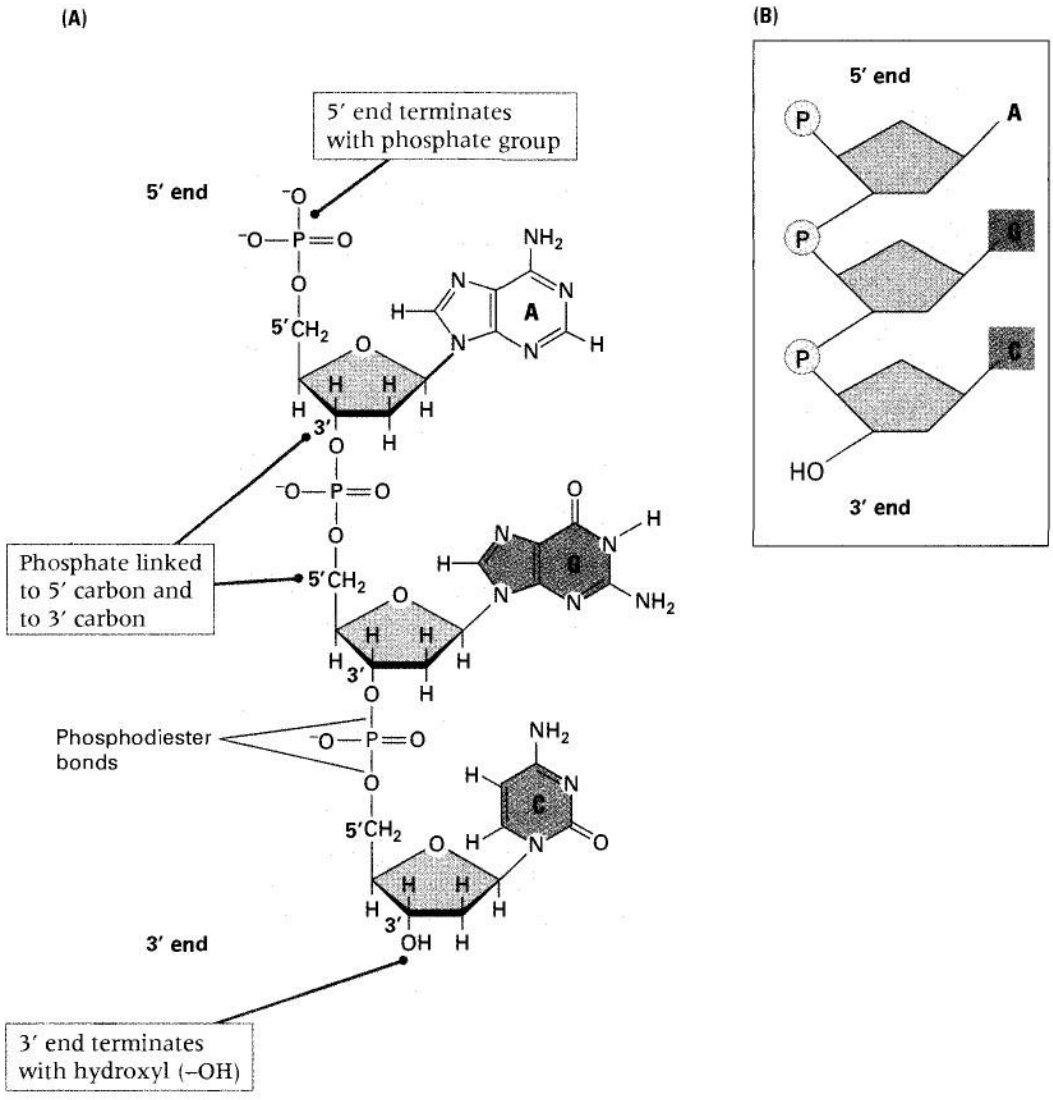


Figure 2.5 Three nucleotides at the 5' end of a single polynucleotide strand. (A) The chemical structure of the sugar–phosphate linkages, showing the 5'-to-3' orientation of the strand (the red numbers are those assigned to the carbon atoms). (B) A common schematic way to depict a polynucleotide strand.

continues throughout the chain, which typically consists of millions of nucleotides. Note that the terminal groups of each polynucleotide chain are a **5'-phosphate (5'-P) group** at one end and a **3'-hydroxyl (3'-OH) group** at the other. The asymmetry of the ends of a DNA strand implies that each strand has a **polarity** determined by which end bears the 5' phosphate and which end bears the 3' hydroxyl.

A few years before Watson and Crick proposed their essentially correct three-dimensional structure of DNA as a double helix, Erwin Chargaff developed a chemical technique to measure the amount of each base present in DNA. As we describe this technique, we will let the molar concentration of any base be represented by the symbol for the base in square brackets; for example, [A] denotes the molar concentration of adenine. Chargaff used his technique to measure the [A], [T], [G], and [C] content of the DNA from a variety of sources. He found that the **base composition** of the DNA, defined as the **percent G + C,** differs among species but is constant in all cells of an organism and within a species. Data on the base composition of

DNA from a variety of organisms are given in **Table 2.2**.

Chargaff also observed certain regular relationships among the molar concentrations of the different bases. These relationships are now called **Chargaff's rules**:

- The amount of adenine equals that of thymine: [A] = [T].
- The amount of guanine equals that of cytosine: [G] = [C].
- The amount of the purine bases equals that of the pyrimidine bases:

$$[A] + [G] = [T] + [C].$$

Although the chemical basis of these observations was not known at the time, one of the appealing features of the Watson–Crick structure of paired complementary strands was that it explained Chargaff's rules. Because A is always paired with T in double-stranded DNA, it must follow that [A] = [T]. Similarly, because G is paired with C, we know that [G] = [C]. The third rule follows by addition of the other two: [A] + [G] = [T] + [C]. In the next section, we examine the molecular basis of base pairing in more detail.

Base Pairing and Base Stacking

In the three-dimensional structure of the DNA molecule proposed in 1953 by Watson and Crick, the molecule consists of two polynucleotide chains twisted around one another to form a double-stranded helix in which adenine and thymine, and guanine and cytosine, are paired in opposite strands (**Figure 2.6**). In the standard structure, which is called the **B form of DNA,** each chain makes one complete turn every 34 Å. The helix is right-handed, which means that as one looks down the barrel, each chain follows a clockwise path as it progresses. The bases are spaced at 3.4 Å, so there are ten bases per helical turn in each strand and ten base pairs per turn of the double helix.

Table 2.2 Base composition of DNA from different organisms

Organism	Base (and percentage of total bases)				Base composition (percent G + C)
	Adenine	Thymine	Guanine	Cytosine	
Bacteriophage T7	26.0	26.0	24.0	24.0	48.0
Bacteria					
Clostridium perfringens	36.9	36.3	14.0	12.8	26.8
Streptococcus pneumoniae	30.2	29.5	21.6	18.7	40.3
Escherichia coli	24.7	23.6	26.0	25.7	51.7
Sarcina lutea	13.4	12.4	37.1	37.1	74.2
Fungi					
Saccharomyces cerevisiae	31.7	32.6	18.3	17.4	35.7
Neurospora crassa	23.0	22.3	27.1	27.6	54.7
Higher plants					
Wheat	27.3	27.2	22.7	22.8*	45.5
Maize	26.8	27.2	22.8	23.2*	46.0
Animals					
Drosophila melanagaster	30.8	29.4	19.6	20.2	39.8
Pig	29.4	29.6	20.5	20.5	41.0
Salmon	29.7	29.1	20.8	20.4	41.2
Human being	29.8	31.8	20.2	18.2	38.4

*Includes one-fourth 5-methylcytosine, a modified form of cytosine found in most plants more complex than algae and in many animals

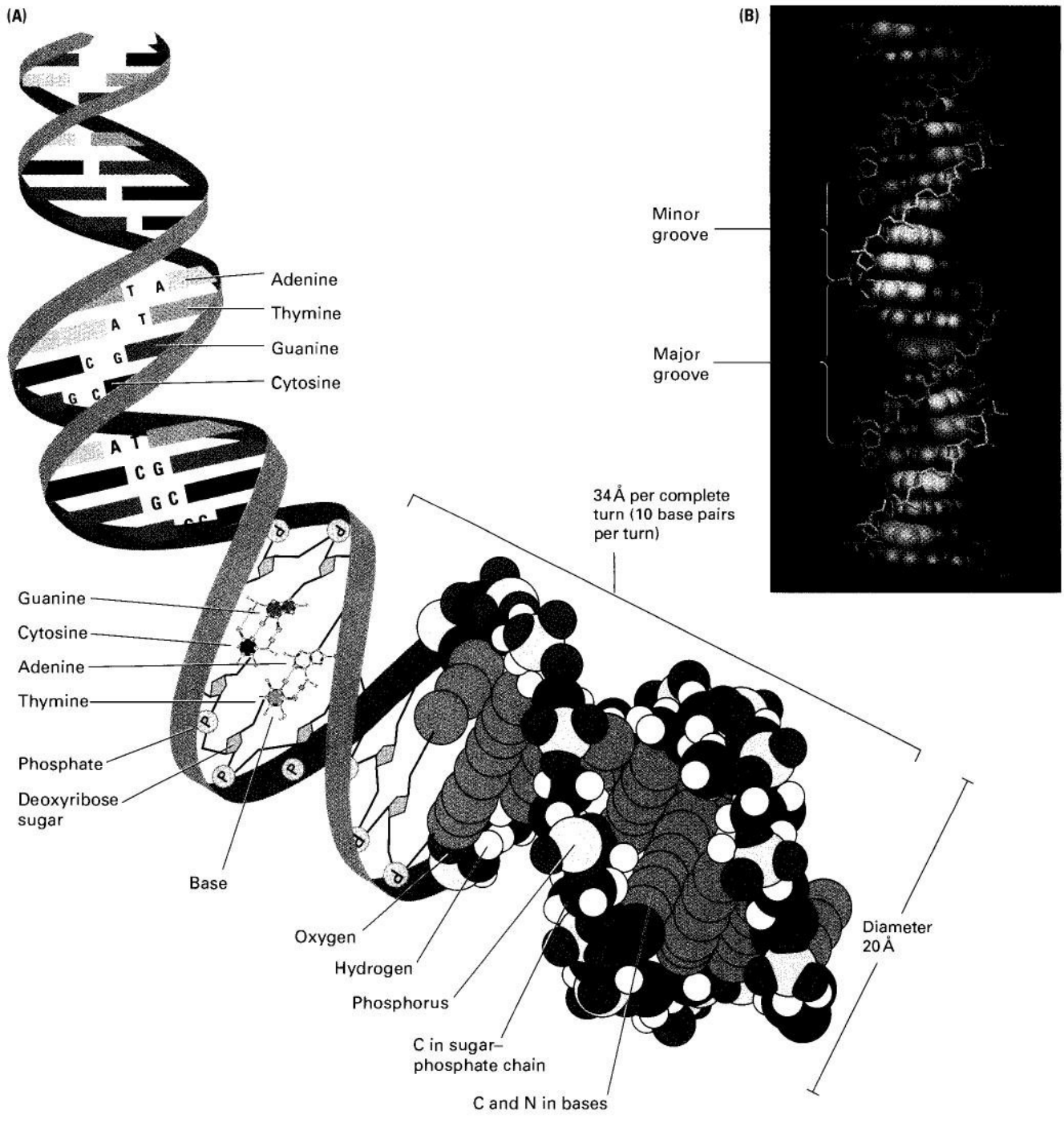


Figure 2.6 Two representations of DNA, illustrating the three-dimensional structure of the double helix. (A) In a ribbon diagram, the sugar–phosphate backbones are depicted as bands, with horizontal lines used to represent the base pairs. (B) A computer model of the B form of a DNA molecule. The stick figures are the sugar–phosphate chains winding around outside the stacked base pairs, forming a major groove and a minor groove. The color coding for the base pairs is as follows: A, red or pink; T, dark green or light green; G, dark brown or beige; C, dark blue or light blue. The bases depicted in dark colors are those attached to the blue sugar–phosphate backbone; the bases depicted in light colors are attached to the beige backbone. [B, courtesy of Antony M. Dean.]

The strands feature **base pairing,** in which each base is paired to a complementary base in the other strand by hydrogen bonds. (A **hydrogen bond** is a weak bond in which two participating atoms share a hydrogen atom between them.) The hydrogen bonds provide one type of force holding the strands together. In Watson–Crick base pairing, adenine (A) pairs with thymine (T), and guanine (G) pairs with cytosine (C). The hydrogen bonds that form in the adenine–thymine base pair and in the guanine–cytosine pair are illustrated in Figure 2.7. Note that an A–T pair (Figure 2.7A and B) has two hydrogen bonds and that a G–C pair (Figure 2.7C and D) has three hydrogen bonds. This means that the hydrogen bonding between G and C is stronger in the sense that it requires more energy to break; for example, the amount of heat required to separate the paired strands in a DNA duplex increases with the percent of G + C. Because nothing restricts the sequence of bases in a single strand, any sequence could be present along one strand. This explains Chargaff's observation that DNA from different organisms may differ in base composition. However, because the strands in duplex DNA are complementary, Chargaff's rules of [A] = [T] and [G] = [C] are true whatever the base composition.

In the B form of DNA, the paired bases are planar, parallel to one another, and perpendicular to the long axis of the double helix. This feature of double-stranded DNA is known as **base stacking.** The upper and lower faces of each nitrogenous base are relatively flat and nonpolar (uncharged). These surfaces are said to be **hydrophobic** because they bind poorly to water molecules, which are very polar. (The polarity refers to the asymmetrical distribution of charge across the V-shaped water molecule;

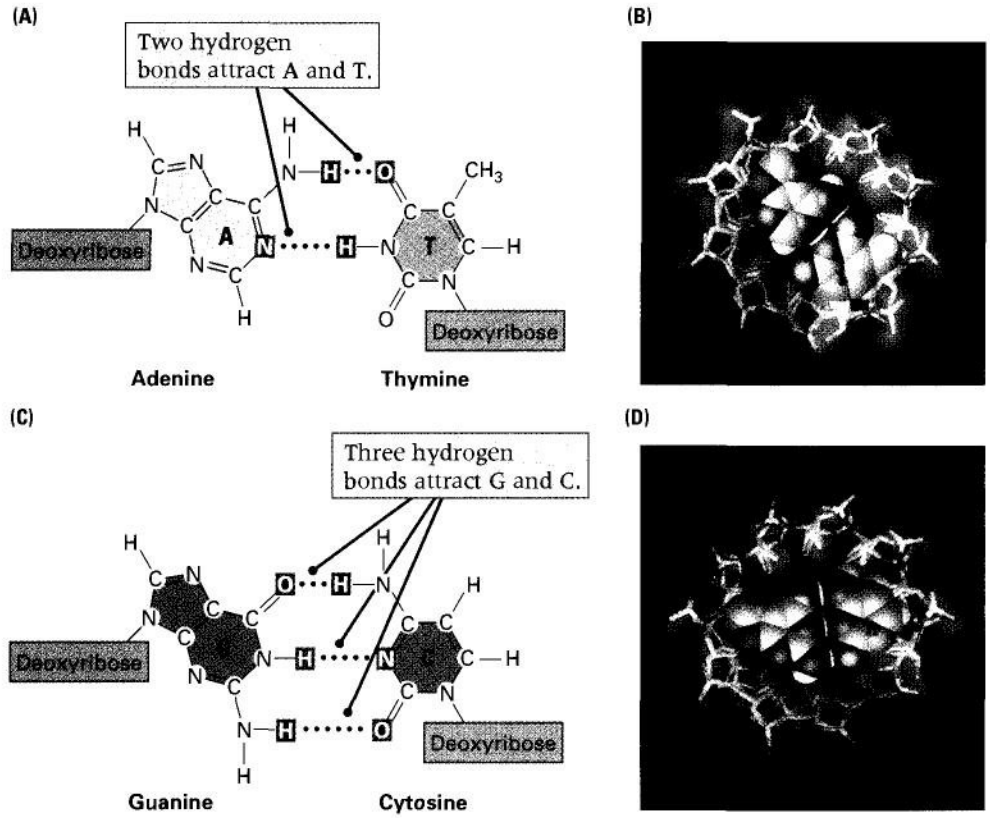


Figure 2.7 Normal base pairs in DNA. On the left, the hydrogen bonds (dotted lines) with the joined atoms are shown in red. (A and B) A–T base pairing. (C and D) G–C base pairing. In the space-filling models (B and D), the colors are as follows: C, gray; N, blue; O, red; and H (shown in the bases only), white. Each hydrogen bond is depicted as a white disk squeezed between the atoms sharing the hydrogen. The stick figures on the outside represent the backbones winding around the stacked base pairs. [Space-filling models courtesy of Antony M. Dean.]

the oxygen at the base of the V tends to be quite negative, whereas the hydrogens at the tips are quite positive). Owing to their repulsion of water molecules, the paired nitrogenous bases tend to stack on top of one another in such a way as to exclude the maximum amount of water from the interior of the double helix. Hence a double-stranded DNA molecule has a hydrophobic core composed of stacked bases, and it is the energy of base stacking that provides double-stranded DNA with much of its chemical stability.

When discussing a DNA molecule, molecular biologists frequently refer to the individual strands as single strands or as single-stranded DNA; they refer to the double helix as double-stranded DNA or as a *duplex* molecule. The two grooves spiraling along outside of the double helix are not symmetrical; one groove, called the **major groove,** is larger than the other, which is called the **minor groove.** Proteins that interact with double-stranded DNA often have regions that make contact with the base pairs by fitting into the major groove, into the minor groove, or into both grooves (Figure 2.6B).

Antiparallel Strands

Each backbone in a double helix consists of deoxyribose sugars alternating with phosphate groups that link the 3' carbon atom of one sugar to the 5' carbon of the next in line (Figure 2.5). The two polynucleotide strands of the double helix have opposite polarity in the sense that the 5' end of one strand is paired with the 3' end of the other strand. Strands with such an arrangement are said to be **antiparallel.** One implication of antiparallel strands in duplex DNA is that in each pair of bases, one base is attached to a sugar that lies above the plane of pairing, and the other base is attached to a sugar that lies below the plane of pairing. Another implication is that each terminus of the double helix possesses one 5'-P group (on one strand) and one 3'-OH group (on the other strand), as shown in **Figure 2.8.**

The diagram of the DNA duplex in Figure 2.6 is static and so somewhat misleading. DNA is a dynamic molecule, constantly in motion. In some regions, the strands can separate briefly and then come together again in the same conformation or in a different one. The right-handed double helix in Figure 2.6 is the standard B form, but depending on conditions, DNA can actually form more than 20 slightly different variants of a right-handed helix, and some regions can even form helices in which the strands twist to the left (called the **Z form of DNA**). If there are complementary stretches of nucleotides in the same strand, then a single strand, separated from its partner, can fold back upon itself like a hairpin. Even triple helices consisting of three strands can form in regions of DNA that contain suitable base sequences.

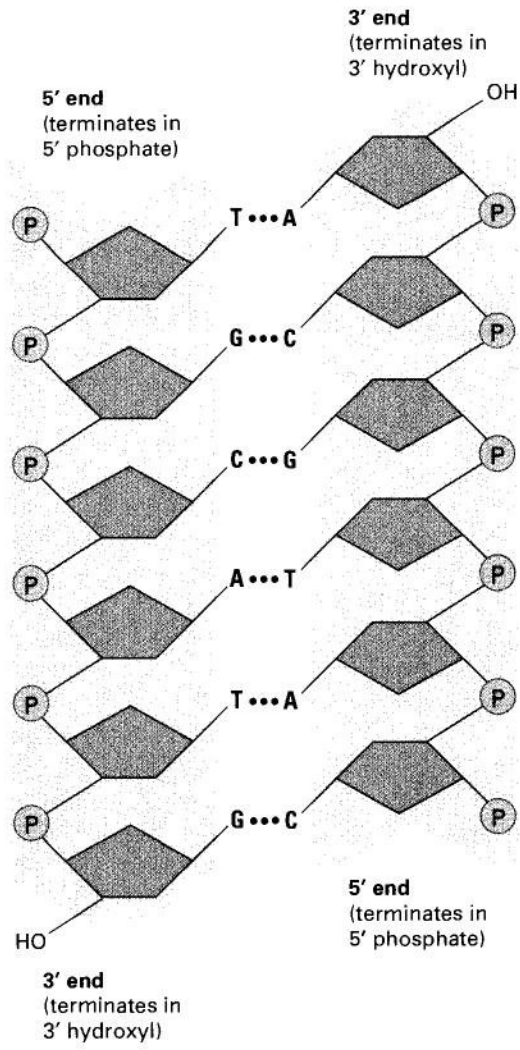


Figure 2.8 A segment of a DNA molecule, showing the antiparallel orientation of the complementary strands. The overlying blue arrows indicate the 5'-to-3' direction of each strand. The phosphates (P) join the 3' carbon atom of one deoxyribose (horizontal line) to the 5' carbon atom of the adjacent deoxyribose.

CoNnection

The Double Helix

James D. Watson and Francis H. C. Crick 1953
Cavendish Laboratory,
Cambridge, England
A Structure for Deoxyribose Nucleic Acid

This is one of the watershed papers of twentieth-century biology. After its publication, nothing in genetics was the same. Everything that was known, and everything still to be discovered, would now need to be interpreted in terms of the structure and function of DNA. The importance of the paper was recognized immediately, in no small part because of its lucid and concise description of the structure. Watson and Crick benefited tremendously in knowing that their structure was consistent with the unpublished structural studies of Maurice Wilkins and Rosalind Franklin. The same issue of Nature *that included the Watson and Crick paper also included, back to back, a paper from the Wilkins group and one from the Franklin group detailing their data and the consistency of their data with the proposed structure. It has been said that Franklin was poised a mere two half-steps from making the discovery herself, alone. In any event, Watson and Crick and Wilkins were awarded the 1962 Nobel Prize for their discovery of DNA structure. Rosalind Franklin, tragically, died of cancer in 1958 at the age of 38.*

We wish to suggest a structure for the salt of deoxyribose nucleic acid (DNA). . . . The structure has two helical chains each coiled round the same axis. . . . Both chains follow right-handed helices, but the two chains run in opposite directions. . . . The bases are on the inside of the helix and the phosphates on the outside. . . . There is a residue on each chain every 3.4 Å and the structure repeats after 10 residues. . . . The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fiber axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side. One of the pair must be a purine and the other a pyrimidine for bonding to occur. . . . Only specific pairs of bases can bond together. These pairs are adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine). In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined. . . . It has not escaped our notice that the specific pairing we have postulated immediately suggests a plausible copying mechanism for the genetic material. . . . We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. Maurice H. F. Wilkins, Dr. Rosalind Franklin and their co-workers at King's College, London.

If only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

Source: Nature 171: 737–738

DNA Structure as Related to Function

In the structure of the DNA molecule, we can see how three essential requirements of a genetic material are met.

1. Any genetic material must be able to be replicated accurately, so that the information it contains will be precisely replicated and inherited by daughter cells. The basis for exact duplication of a DNA molecule is the pairing of A with T and of G with C in the two polynucleotide chains. Unwinding and separation of the chains, with each free chain being copied, results in the formation of two identical double helices (see Figure 1.7).
2. A genetic material must also have the capacity to carry all of the information needed to direct the organization and metabolic activities of the cell. As we saw in Chapter 1, the product of most genes is a protein molecule—a polymer composed of repeating units of amino acids. The sequence

of amino acids in the protein determines its chemical and physical properties. A gene is expressed when its protein product is synthesized, and one requirement of the genetic material is that it direct the order in which amino acid units are added to the end of a growing protein molecule. In DNA, this is done by means of a genetic code in which groups of three bases specify amino acids. Because the four bases in a DNA molecule can be arranged in any sequence, and because the sequence can vary from one part of the molecule to another and from organism to organism, DNA can contain a great many unique regions, each of which can be a distinct gene. A long DNA chain can direct the synthesis of a variety of different protein molecules.

3. A genetic material must also be capable of undergoing occasional mutations in which the information it carries is altered. Furthermore, so that mutations will be heritable, the mutant molecules must be capable of being replicated as faithfully as the parental molecule. This feature is necessary to account for the evolution of diverse organisms through the slow accumulation of favorable mutations. Watson and Crick suggested that heritable mutations might be possible in DNA by rare mispairing of the bases, with the result that an incorrect nucleotide becomes incorporated into a replicating DNA strand.

2.3 The Separation and Identification of Genomic DNA Fragments

The following sections show how an understanding of DNA structure and replication has been put to practical use in the development of procedures for the separation and identification of particular DNA fragments. These methods are used primarily either to identify DNA markers or to aid in the isolation of particular DNA fragments that are of genetic interest. For example, consider a pedigree of familial breast cancer in which a particular DNA fragment serves as a marker for a bit of chromosome that also includes the mutant gene responsible for the increased risk; then the ability to identify the fragment is critically important in assessing the relative risk for each of the women in the pedigree who might carry the mutant gene. To take another example, suppose there is reason to believe that a mutation causing a genetic disease is present in a particular DNA fragment; then it is important to be able to pinpoint this fragment and isolate it from affected individuals to verify whether this hypothesis is true and, if so, to identify the nature of the mutation.

Most procedures for the separation and identification of DNA fragments can be grouped into two general categories:

1. Those that identify a specific DNA fragment present in genomic DNA by making use of the fact that complementary single-stranded DNA sequences can, under the proper conditions, form a duplex molecule. These procedures rely on *nucleic acid hybridization.*
2. Those that use prior knowledge of the sequence at the ends of a DNA fragment to specifically and repeatedly replicate this one fragment from genomic DNA. These procedures rely on selective DNA replication (*amplification*) by means of the *polymerase chain reaction.*

The major difference between these approaches is that the first (relying on nucleic acid hybridization) identifies fragments that are present in the genomic DNA itself, whereas the second (relying on DNA amplification) identifies experimentally manufactured *replicas* of fragments whose original templates (but not the replicas) were present in the genomic DNA. This difference has practical implications:

- Hybridization methods require a greater amount of genomic DNA for the experimental procedures, but relatively large fragments can be identified, and no prior knowledge of the DNA sequence is necessary.
- Amplification methods require extremely small amounts of genomic DNA for the experimental procedures, but the amplification is usually restricted to relatively small fragments, and some prior knowledge of DNA sequence is necessary.

The following sections discuss both types of approaches and give examples of how they are used. In methods that use nucleic acid hybridization to identify particular fragments present in genomic DNA, the first step is usually cutting the genomic DNA into fragments of experimentally manageable size. This procedure is discussed next.

Restriction Enzymes and Site-Specific DNA Cleavage

Procedures for chemical isolation of DNA, such as those developed by Avery, MacLeod, and McCarty (Chapter 1), usually lead to random breakage of double-stranded molecules into an average length of about 50,000 base pairs. This length is denoted 50 kb, where kb stands for **kilobases** (1 kb = 1000 base pairs). A length of 50 kb is close to the length of double-stranded DNA present in the bacteriophage λ that infects *E. coli*. The 50-kb fragments can be made shorter by vigorous shearing forces, such as occur in a kitchen blender, but one of the problems with breaking large DNA molecules into smaller fragments by random shearing is that the fragments containing a particular gene, or part of a gene, will be of different sizes. In other words, with random shearing, it is not possible to isolate and identify a *particular* DNA fragment on the basis of its size and sequence content, because each randomly sheared molecule that contains the desired sequence somewhere within it differs in size from all other molecules that contain the sequence. In this section we describe an important enzymatic technique that can be used for cleaving DNA molecules at specific sites. This method ensures that all DNA fragments that contain a particular sequence have the same size; furthermore, each fragment that contains the desired sequence has the sequence located at exactly the same position within the fragment.

The cleavage method makes use of an important class of DNA-cleaving enzymes isolated primarily from bacteria. The enzymes are called **restriction endonucleases** or **restriction enzymes,** and they are able to cleave DNA molecules at the positions at which particular, short sequences of bases are present. These naturally occurring enzymes serve to protect the bacterial cell by disabling the DNA of bacteriophages that attack it. Their discovery earned Werner Arber of Switzerland a Nobel Prize in 1978. Technically, the enzymes are known as *type II restriction endonucleases*. The restriction enzyme *Bam*HI is one example; it recognizes the double-stranded sequence

```
5'-GGATCC-3'
3'-CCTAGG-5'
```

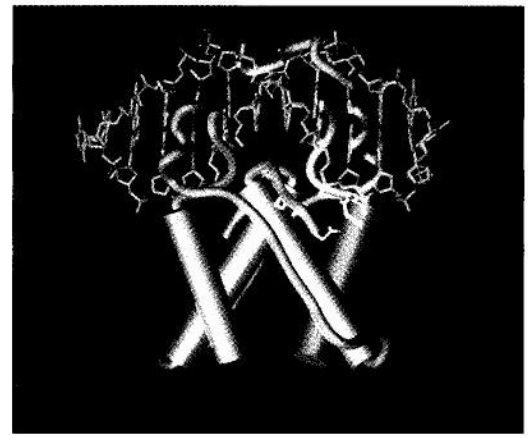


Figure 2.9 Structure of the part of the restriction enzyme *Bam*HI that comes into contact with its recognition site in the DNA (blue). The pink and green cylinders represent regions of the enzyme in which the amino acid chain is twisted in the form of a right-handed helix. [Courtesy of A. A. Aggarwal. Reprinted with permission from M. Newman, T. Strzelecka, L. F. Dorner, I. Schildkraut, and A. A. Aggarwal, *Science* 269: 656. Copyright 1995 American Association for the Advancement of Science.]

and cleaves each strand between the G-bearing nucleotides shown in red. **Figure 2.9** shows how the regions that make up the active site of *Bam*HI contact the recognition site (blue) just prior to cleavage, and the cleavage reaction is indicated in **Figure 2.10**.

Table 2.3 lists nine of the several hundred restriction enzymes that are known. Most restriction enzymes are named after the species in which they were found. *Bam*HI, for example, was isolated from the bacterium *Bacillus amyloliquefaciens* strain H, and it is the first (I) restriction enzyme isolated from this organism. Because the first three letters in the name of each restriction enzyme stand for the bacterial species of origin, these letters are printed in italics; the rest of the symbols in the name are not italicized. Most restriction enzymes recognize only one short base sequence, usually four or six nucleotide pairs. The enzyme binds with the DNA at these sites and makes a break in each strand of the DNA molecule, producing 3'-OH and 5'-P groups at each position. The nucleotide sequence recognized for cleavage by a restriction enzyme is called the **restriction site** of the enzyme. The examples in Table 2.3 show that some restriction enzymes cleave their restriction site asymmetrically (at different sites in the

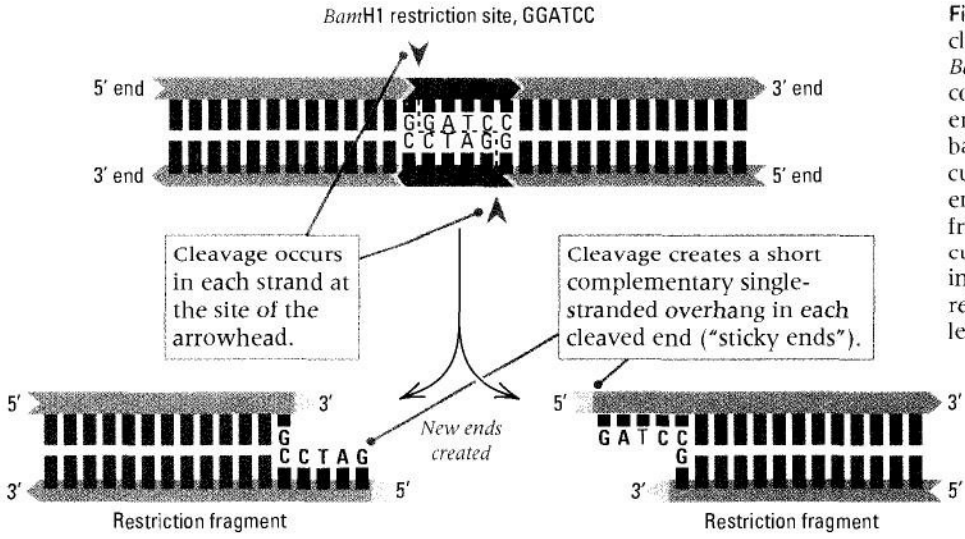


Figure 2.10 Mechanism of DNA cleavage by the restriction enzyme *Bam*HI. Wherever the duplex contains a *Bam*HI restriction site, the enzyme makes a single cut in the backbone of each DNA strand. Each cut creates a new 3′ end and a new 5′ end, separating the duplex into two fragments. In the case of *Bam*HI the cuts are staggered cuts, so the resulting ends terminate in single-stranded regions, each four base pairs in length.

Table 2.3 Some restriction endonucleases, their sources, and their cleavage sites

Enzyme (Microorganism)	Enzyme (Microorganism)	Enzyme (Microorganism)
*Eco*RI (*Escherichia coli*)	*Hind*III (*Haemophilus influenzae*)	*Alu*I (*Arthrobacter luteus*)
G A A T T C C T T A A G	A A G C T T T T C G A A	A G C T T C G A
*Bam*HI (*Bacillus amyloliquefaciens* H)	*Pst*I (*Providencia stuartii*)	*Rsa*I (*Rhodopseudomonas sphaeroides*)
G G A T C C C C T A G G	C T G C A G G A C G T C	G T A C C A T G
*Hae*II (*Haemophilus aegyptus*)	*Taq*I (*Thermus aquaticus*)	*Pvu*II (*Proteus vulgaris*)
Pu G C G C Py Py C G C G Pu	T C G A A G C T	C A G C T G G T C G A C

Target sequence and cleavage site; sticky ends

Target sequence and cleavage site; blunt ends

Note: The vertical dashed line indicates the axis of symmetry in each sequence. Red arrows indicate the sites of cutting. The enzyme *Taq*I yields cohesive ends consisting of two nucleotides, whereas the cohesive ends produced by the other enzymes contain four nucleotides. Pu and Py refer to any purine and pyrimidine, respectively.





two DNA strands), but other restriction enzymes cleave symmetrically (at the same site in both strands). The former leave **sticky ends** because each end of the cleaved site has a small, single-stranded overhang that is complementary in base sequence to the other end (Figure 2.10). In contrast, enzymes that have symmetrical cleavage sites yield DNA fragments that have **blunt ends.** In virtually all cases, the restriction site of a restriction enzyme reads the same on both strands, provided that the opposite polarity of the strands is taken into account; for example, each strand in the restriction site of *Bam*HI reads 5'-GGATCC-3' (Figure 2.10). A DNA sequence with this type of symmetry is called a **palindrome.** (In ordinary English, a palindrome is a word or phrase that reads the same forwards and backwards, such as "madam.")

Restriction enzymes have the following important characteristics:

- Most restriction enzymes recognize a single restriction site.
- The restriction site is recognized without regard to the source of the DNA.
- Because most restriction enzymes recognize a unique restriction site sequence, the number of cuts in the DNA from a particular organism is determined by the number of restriction sites present.

The DNA fragment produced by a pair of adjacent cuts in a DNA molecule is called a **restriction fragment.** A large DNA molecule will typically be cut into many restriction fragments of different sizes. For example, an *E. coli* DNA molecule, which contains 4.6×10^6 base pairs, is cut into several hundred to several thousand fragments, and mammalian genomic DNA is cut into more than a million fragments. Although these numbers are large, they are actually quite small relative to the number of sugar–phosphate bonds in the DNA of an organism.

Gel Electrophoresis

The DNA fragments produced by a restriction enzyme can be separated by size using the fact that DNA is negatively charged and moves in response to an electric field. If the terminals of an electrical power source are connected to the opposite ends of a horizontal tube containing a DNA solution, then the DNA molecules will move toward the positive end of the tube at a rate that depends on the electric field strength and on the shape and size of the molecules. The movement of charged molecules in an electric field is called *electrophoresis.*

The type of electrophoresis most commonly used in genetics is **gel electrophoresis.** An experimental arrangement for gel electrophoresis of DNA is shown in Figure 2.11. A thin slab of a gel, usually agarose or acrylamide, is prepared containing small slots (called *wells*) into which samples are placed. An electric field is applied, and the negatively charged DNA molecules penetrate and move through the gel toward the anode (the positively charged electrode). A gel is a complex molecular network that contains narrow, tortuous passages, so smaller DNA molecules pass through more easily; hence the rate of movement increases as the size of the DNA fragment decreases. Figure 2.12 shows the result of electrophoresis of a set of double-stranded DNA molecules in an agarose gel. Each discrete region containing DNA is called a **band.** The bands can be visualized under ultraviolet light after soaking the gel in the dye *ethidium bromide,* the molecules of which intercalate between the stacked bases in duplex DNA and render it fluorescent. In Figure 2.12, each band in the gel

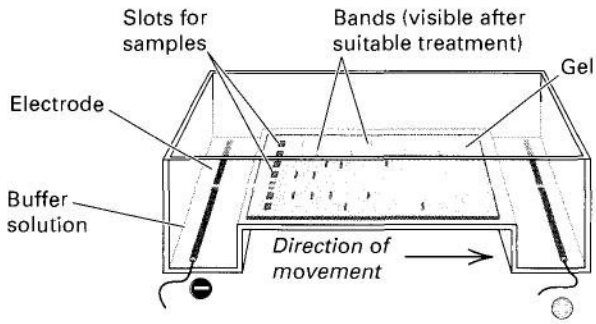


Figure 2.11 Apparatus for gel electrophoresis. Liquid gel is allowed to harden with an appropriately shaped mold in place to form "wells" for the samples (purple). After electrophoresis, the DNA fragments, located at various positions in the gel, are made visible by immersing the gel in a solution containing a reagent that binds to or reacts with DNA. The separated fragments in a sample appear as bands, which may be either visibly colored or fluorescent, depending on the particular reagent used. The region of a gel in which the fragments in one sample can move is called a *lane;* this gel has seven lanes.

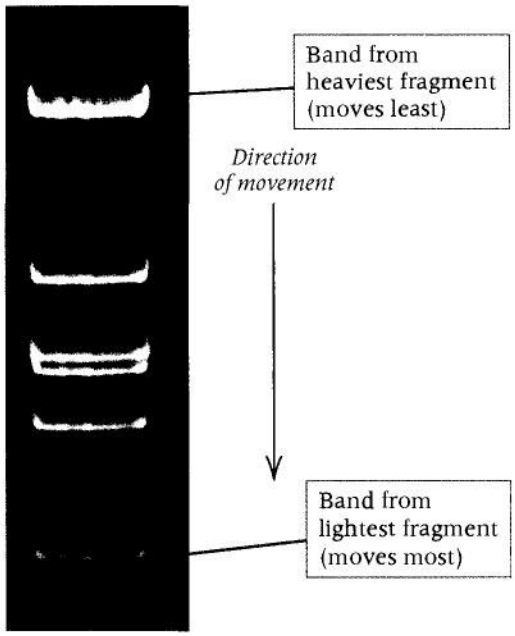


Figure 2.12 Gel electrophoresis of DNA. Fragments of different sizes were mixed and placed in a well. Electrophoresis was in the downward direction. The DNA has been made visible by the addition of a dye (ethidium bromide) that binds only to DNA and that fluoresces when the gel is illuminated with short-wavelength ultraviolet light.

results from the fact that all DNA fragments of a given size have migrated to the same position in the gel. To produce a visible band, a minimum of about 5×10^{-9} grams of DNA is required, which for a fragment of size 3 kb works out to about 10^9 molecules. The point is that a very large number of copies of any particular DNA fragment must be present in order to yield a visible band in an electrophoresis gel.

A linear double-stranded DNA fragment has an electrophoretic mobility that decreases in proportion to the logarithm of its length in base pairs—the longer the fragment, the slower it moves—but the proportionality constant depends on the agarose concentration, the composition of the buffering solution, and the electrophoretic conditions. This means that different concentrations of agarose allow efficient separation of different size ranges of DNA fragments (see **Figure 2.13**). Less dense gels, such as 0.6 percent agarose, are used to separate larger fragments; whereas more dense gels, such as 2 percent agarose, are used to separate smaller fragments. The inset in Figure 2.13 shows the dependence of electrophoretic mobility on the logarithm of fragment size. It also indicates that the linear relationship breaks down for the largest fragments that can be resolved under a given set of conditions.

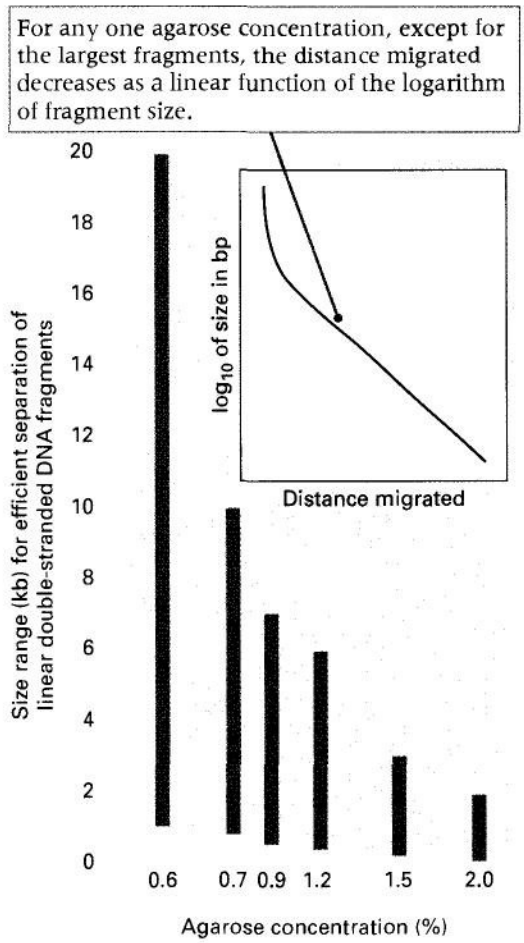


Figure 2.13 In agarose gels, the concentration of agarose is an important factor in determining the size range of DNA fragments that can be separated.

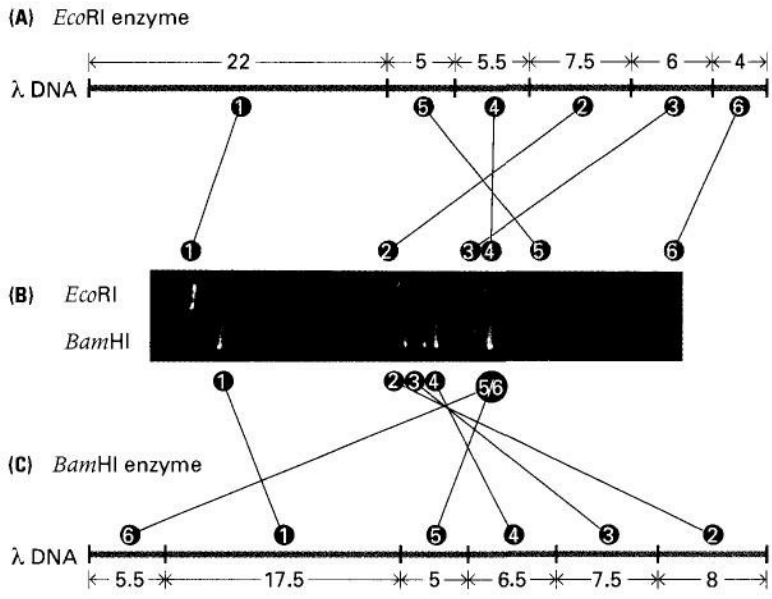


Figure 2.14 Restriction maps of λ DNA for the restriction enzymes (A) *Eco*RI and (C) *Bam*HI. The vertical bars indicate the sites of cutting. The numbers within the arrows are the approximate lengths of the fragments in kilobase pairs (kb). (B) An electrophoresis gel of *Bam*HI and *Eco*RI enzyme digests of λ DNA. Numbers indicate fragments in order from largest (1) to smallest (6); the circled numbers on the maps correspond to the numbers beside the gel. The DNA has not undergone electrophoresis long enough to separate bands 5 and 6 of the *Bam*HI digest.

Note: In Problem 2 at the end of this chapter (Guide to Problem Solving), we show how to use the results of a double digest to determine the particular order of fragments for a pair of restriction enzymes.

Because of the sequence specificity of cleavage, *a particular restriction enzyme produces a unique set of fragments for a particular DNA molecule.* Another enzyme will produce a different set of fragments from the same DNA molecule. In **Figure 2.14**, this principle is illustrated for the digestion of *E. coli* phage λ DNA by either *Eco*RI or *Bam*HI (see part B). The locations of the cleavage sites for these enzymes in λ DNA are shown in Figures 2.14A and C. A diagram showing sites of cleavage along a DNA molecule is called a **restriction map.** Particular DNA fragments can be isolated by cutting out the small region of the gel that contains the fragment and removing the DNA from the gel. One important use of isolated restriction fragments employs the enzyme *DNA ligase* to insert them into self-replicating molecules such as bacteriophage, plasmids, or even small artificial chromosomes (Figure 2.2). These procedures constitute **DNA cloning** and are the basis of one form of *genetic engineering,* discussed further in Chapter 13.

DNA fragments that have been cloned into organisms such as *E. coli* are widely used because the fragments can be isolated in large amounts and purified relatively easily. Among the uses of cloned DNA are:

- *DNA sequencing.* All current methods of DNA sequencing require cloned DNA fragments. These methods are discussed in Chapter 6.
- *Nucleic acid hybridization.* As we shall see below, an important application of cloned DNA fragments entails incorporating a radioactive or light-emitting "label" into them, after which the labeled material is used to "tag" DNA fragments containing similar sequences.
- *Storage and distribution.* Cloned DNA can be stored for long periods without risk of change and can easily be distributed to other researchers.

Nucleic Acid Hybridization

Most genomes are sufficiently large and complex that digestion with a restriction enzyme produces many bands that are the same or similar in size. Identifying a particular DNA fragment in a background of many other fragments of similar size presents a needle-in-a-haystack problem. Suppose, for example, that we are interested in a particular 3.0 *Bam*HI fragment from the human genome that serves as a marker indicating the presence of a genetic risk factor toward breast cancer among the individuals in a particular pedigree. This fragment of 3.0 kb is indistinguishable, on the basis of size alone, from fragments ranging from about 2.9 to 3.1 kb. How many fragments in this size range are expected? When human genomic DNA is cleaved with *Bam*HI, the average length of a restriction fragment is $4^6 = 4096$ base pairs, and the expected total number of *Bam*HI fragments is about 730,000; in the size range 2.9–3.1 kb, the expected number of fragments is about 17,000. What this means is that even though we know that the fragment we are interested in is 3 kb in length, it is only one of 17,000 fragments that are so similar in size that ours cannot be distinguished from the others by length alone.

This identification task is actually harder than finding a needle in a haystack because

haystacks are usually dry. A more accurate analogy would be looking for a needle in a haystack that had been pitched into a swimming pool full of water. This analogy is more relevant because gels, even though they contain a supporting matrix to make them semisolid, are primarily composed of water, and each DNA molecule within a gel is surrounded entirely by water. Clearly, we need some method by which the molecules in a gel can be immobilized and our *specific* fragment identified.

The DNA fragments in a gel are usually immobilized by transferring them onto a sheet of special filter paper consisting of nitrocellulose, to which DNA can be permanently (covalently) bound. How this is done is described in the next section. In this section we examine how the two strands in a double helix can be "unzipped" to form single strands and how, under the proper conditions, two single strands that are complementary or nearly complementary in sequence can be "zipped" together to form a different double helix. The "unzipping" is called **denaturation,** the "zipping" **renaturation.** The practical applications of denaturation and renaturation are many:

- A small part of a DNA fragment can be "zipped" with a much larger DNA fragment. This principle is used in identifying specific DNA fragments in a complex mixture, such as the 3-kb *Bam*HI marker for breast cancer that we have been considering. Applications of this type include the tracking of genetic markers in pedigrees and the isolation of fragments containing a particular mutant gene.
- A DNA fragment from one gene can be "zipped" with similar fragments from other genes in the same genome; this principle is used to identify different members of *families* of genes that are similar, but not identical, in sequence and that have related functions.
- A DNA fragment from one species can be "zipped" with similar sequences from other species. This allows the isolation of genes that have the same or related functions in multiple species. It is used to study aspects of molecular evolution, such as how differences in sequence are correlated with differences in function, and the patterns and rates of change in gene sequences as they evolve.

As we saw in Section 2.2, the double-stranded helical structure of DNA is maintained by base stacking and by hydrogen bonding between the complementary base pairs. When solutions containing DNA fragments are raised to temperatures in the range 85–100°C, or to the high pH of strong alkaline solutions, the paired strands begin to separate, or "unzip." Unwinding of the helix happens in less than a few minutes (the time depends on the length of the molecule). When the helical structure of DNA is disrupted, and the strands have become completely unzipped, the molecule is said to be **denatured.** A common way to detect denaturation is by measuring the capacity of DNA in solution to absorb ultraviolet light of wavelength 260 nm, because the absorption at 260 nm (A_{260}) of a solution of single-stranded molecules is 37 percent higher than the absorption of the double-stranded molecules at the same concentration. As shown in **Figure 2.15**, the progress of denaturation can be followed by slowly heating a solution of double-stranded DNA and recording the value of A_{260} at various temperatures. The temperature required for denaturation increases with G + C content, not only because G–C base pairs have three hydrogen bonds and A–T base pairs two, but because consecutive G–C base pairs have stronger base stacking.

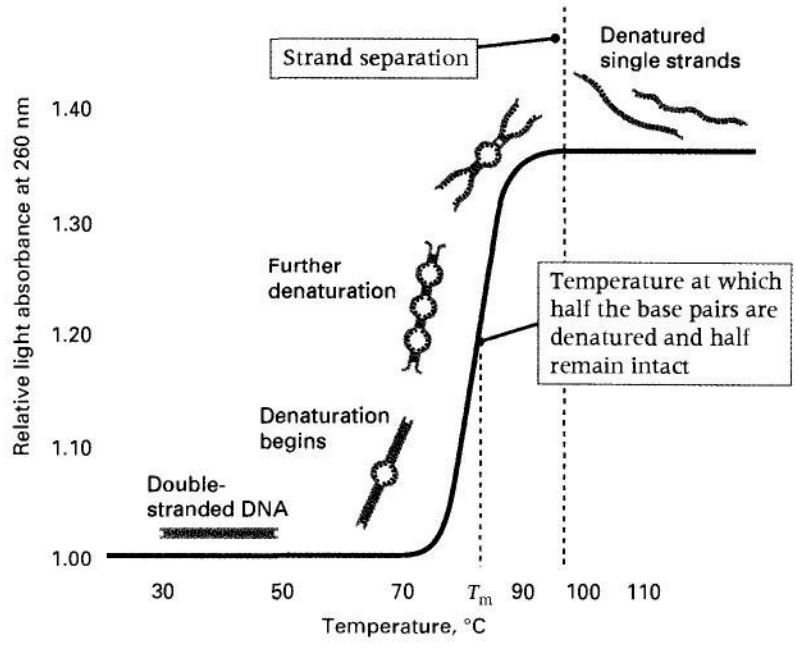
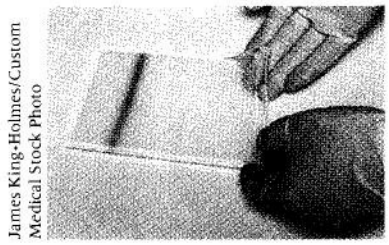


Figure 2.15 Mechanism of denaturation of DNA by heat. The temperature at which 50 percent of the base pairs are denatured is the *melting temperature*, symbolized T_m.

Fragments of DNA being placed in the wells of an agarose gel (right). The solution contains a purple dye that moves faster than any of the DNA fragments. When the dye has migrated toward the far edge of the gel (below), it indicates that the DNA fragments behind the dye have also been separated, so the gel can be removed from the electrophoresis apparatus in preparation for Southern blotting or other procedures.



James King-Holmes/Custom Medical Stock Photo



James Holmes/Cellmark Diagnostics/Science Photo Library/Photo Researchers, Inc.

Denatured DNA strands can, under certain conditions, form double-stranded DNA with other strands, provided that the strands are sufficiently complementary in sequence. This process of renaturation is called **nucleic acid hybridization** because the double-stranded molecules are "hybrid" in that each strand comes from a different source. For DNA strands to hybridize, two requirements must be met:

1. The salt concentration must be high (> 0.25M) to neutralize the negative charges of the phosphate groups, which would otherwise cause the complementary strands to repel one another.
2. The temperature must be high enough to disrupt hydrogen bonds that form at random between short sequences of bases within the same strand, but not so high that stable base pairs between the complementary strands are disrupted.

The initial phase of renaturation is a slow process because the rate is limited by the random chance that a region of two complementary strands will come together to form a short sequence of correct base pairs. This initial pairing step is followed by a rapid pairing of the remaining complementary bases and rewinding of the helix. Rewinding is accomplished in a matter of seconds, and its rate is independent of DNA concentration because the complementary strands have already found each other.

The example of nucleic acid hybridization in **Figure 2.16** will enable us to understand some of the molecular details and also to see how hybridization is used to "tag" and identify a particular DNA fragment. Shown in part A is a solution of denatured DNA, called the **probe,** in which each molecule has been labeled with either radioactive atoms or light-emitting molecules. Probe DNA is typically obtained from a clone, and the labeled probe usually contains denatured forms of both strands present in the original duplex molecule. (This has led to some confusing terminology. Geneticists say that probe DNA hybridizes with DNA fragments containing sequences that are *similar* to the probe, rather than *complementary*. What actually occurs is that one strand of the probe undergoes hybridization with a complementary sequence in the fragment. But because the probe usually contains both strands, hybridization takes place with any fragment that contains a similar sequence, each strand in the probe undergoing hybridization with the complementary sequence in the fragment.)

Part B in Figure 2.16 is a diagram of genomic DNA fragments that have been immobilized on a nitrocellulose filter. When the probe is mixed with the genomic fragments (part C), random collisions bring short, complementary stretches together. If the region of complementary sequence is short (part D), then random collision cannot initiate renaturation because the flanking sequences cannot pair; in this case the probe falls off almost immediately. If, however, a collision brings short sequences together in the correct register (part E), then this initiates renaturation, because the pairing proceeds zipperlike from the initial contact. The main point is that DNA fragments are able to hybridize only if the length of

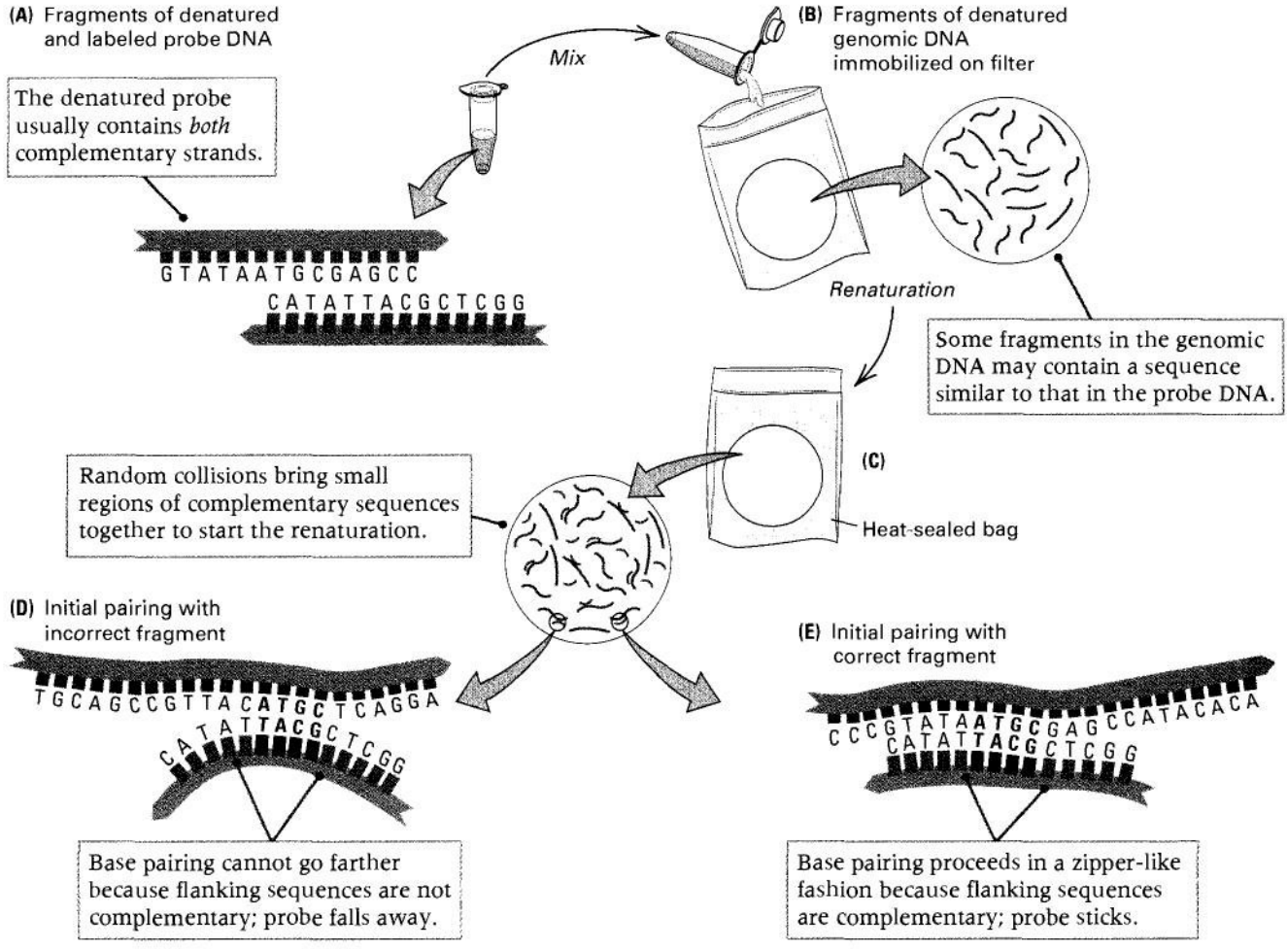


Figure 2.16 Nucleic acid hybridization. (A) Duplex molecules of probe DNA (obtained from a clone) are denatured and (B) placed in contact with a filter to which is attached denatured strands of genomic DNA. (C) Under the proper conditions of salt concentration and temperature, short complementary stretches come together by random collision. (D) If the sequences flanking the paired region are not complementary, then the pairing is unstable and the strands come apart again. (E) If the sequences flanking the paired region are complementary, then further base pairing stabilizes the renatured duplex.

the region in which they can pair is sufficiently long. Some mismatches in the paired region can be tolerated. How many mismatches are allowed is determined by the conditions of the experiment: The lower the temperature at which the hybridization is carried out, and the higher the salt concentration, the greater the proportion of mismatches that are tolerated.

The Southern Blot

The ability to renature DNA in the manner outlined in Figure 12.16 means that solution containing a small fragment of denatured DNA, if it is suitably labeled (for example, with radioactive ^{32}P), can be combined with a complex mixture of denatured DNA fragments, and upon renaturation the small fragment will "tag" with radioactivity any molecules in the complex mixture with which it can hybridize. The radioactive tag allows these molecules to be identified.

The methods of DNA cleavage, electrophoresis, transfer to nitrocellulose, and hybridization with a probe are all combined in the **Southern blot**, named after its inventor Edward Southern. In this procedure, a gel in which DNA molecules have been separated by electrophoresis is treated with

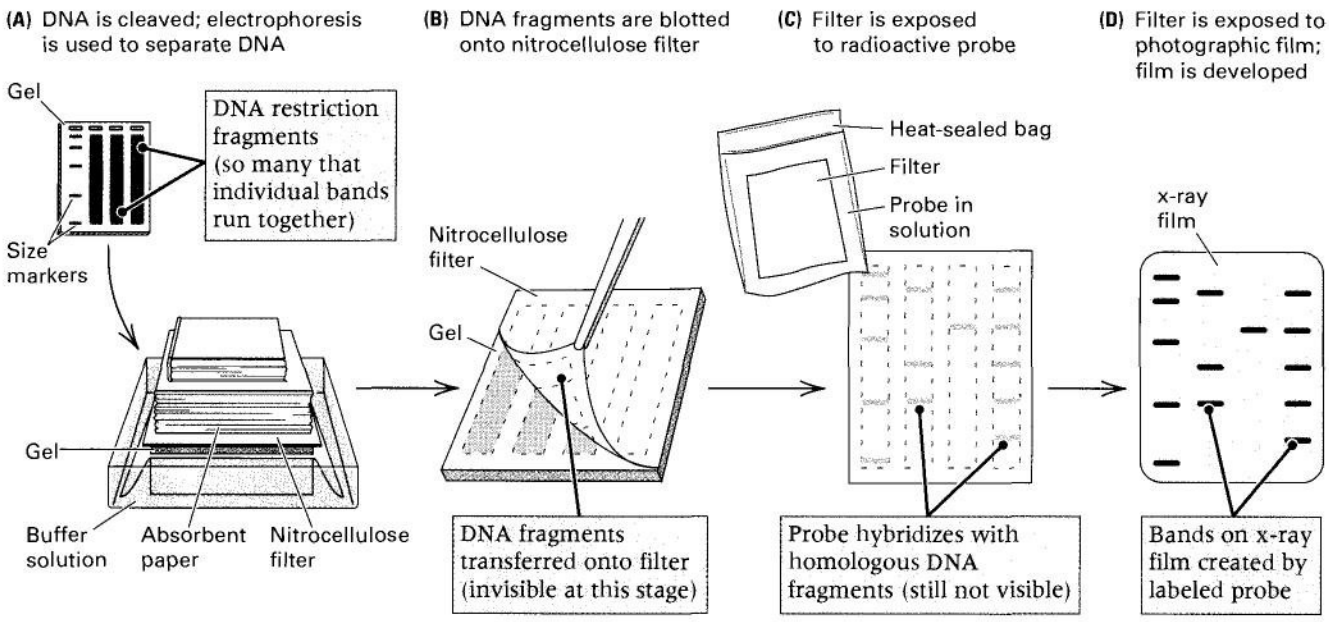


Figure 2.17 Southern blot. (A, B) DNA restriction fragments are separated by electrophoresis, blotted from the gel onto a nitrocellulose or nylon filter, and chemically attached by the use of ultraviolet light. (C) The strands are denatured and mixed with radioactive or light-sensitive probe DNA, which binds with complementary sequences present on the filter. The bound probe remains, whereas unbound probe washes off. (D) Bound probe is revealed by darkening of photographic film placed over the filter. The positions of the bands indicate which restriction fragments contain DNA sequences homologous to those in the probe.

alkali to denature the DNA and render it single-stranded (**Figure 2.17**). Then the DNA is transferred to a sheet of nitrocellulose filter in such a way that the relative positions of the DNA fragments are maintained. The transfer is accomplished by overlaying the nitrocellulose onto the gel and stacking many layers of absorbent paper on top; the absorbent paper sucks water molecules from the gel and through the nitrocellulose, to which the DNA fragments adhere. (This step is the "blot" component of the Southern blot, parts A and B.) Then the filter is treated so that the single-stranded DNA becomes permanently bound. The treated filter is mixed with a solution containing denatured probe (DNA or RNA) under conditions that allow complementary strands to hybridize to form duplex molecules (part C). Radioactive or other label present in the probe becomes stably bound to the filter, and therefore resistant to removal by washing, only at positions at which base sequences complementary to the probe are already present on the filter, so that the probe can form duplex molecules. The label is located by placing the paper in contact with x-ray film. After development of the film, blackened regions indicate positions of bands containing the radioactive or light-emitting label (part D).

The procedure in Figure 2.17 solves the wet-haystack problem by transferring and immobilizing the genomic DNA fragments to a filter and identifying, by hybridization, the ones that are of interest. Practical applications of Southern blotting center on identifying DNA fragments that contain sequences similar to the probe DNA or RNA, where the proportion of mismatched nucleotides allowed is determined by the conditions of hybridization. The advantages of the Southern blot are convenience and sensitivity. The sensitivity comes from the fact that both hybridization with a labeled probe and the use of photographic film amplify the signal; under typical conditions, a band can be observed on the film with only 5×10^{-12} grams of DNA—a thousand times less DNA than the amount required to produce a visible band in the gel itself.

2.4 Selective Replication of Genomic DNA Fragments

Although nucleic acid hybridization allows a particular DNA fragment to be identified when present in a complex mixture of fragments, it does not enable the fragment to be separated from the others and purified. Obtaining the fragment in purified form requires cloning, which is straightforward but time-consuming. (Cloning methods are discussed in Chapter 13.) However, if the fragment of interest is not too long, and if the nucleotide sequence at each end is known, then it becomes possible to obtain large quantities of the fragment merely by

selective replication. This process is called **amplification.** How would one know the sequence of the ends? Let us return to our example of the 3.0-kb *Bam*HI fragment that serves to mark a risk factor for breast cancer in certain pedigrees. Suppose that this fragment is cloned and sequenced from one affected individual, and it is found that, relative to the normal genomic sequence in this region, the *Bam*HI fragment is missing a region of 500 base pairs. At this point the sequences at the ends of the fragment are known, and we can also infer that amplification of genomic DNA from individuals with the risk factor will yield a band of 3.0 kb, whereas amplification from genomic DNA of noncarriers will yield a band of 3.5 kb. This difference allows every person in the pedigree to be diagnosed as a carrier or noncarrier merely by means of DNA amplification. To understand how amplification works, it is first necessary to examine a few key features of DNA replication.

Constraints on DNA Replication: Primers and 5'-to-3' Strand Elongation

As with most metabolic reactions in living cells, nucleic acids are synthesized in chemical reactions controlled by enzymes. An enzyme that forms the sugar–phosphate bond (the phosphodiester bond) between adjacent nucleotides in a nucleic acid chain is called a **DNA polymerase.** A variety of DNA polymerases have been purified, and for amplification of a DNA fragment, the DNA synthesis is carried out *in vitro* by combining purified cellular components in a test tube under precisely defined conditions. (*In vitro* means "without participation of living cells.")

In order for DNA polymerase to catalyze synthesis of a new DNA strand, preexisting single-stranded DNA must be present. Each single-stranded DNA molecule present in the reaction mix can serve as a template upon which a new partner strand is created by the DNA polymerase. For DNA replication to take place, the 5'-triphosphates of the four deoxynucleosides must also be present. This requirement is rather obvious, because the nucleoside triphosphates are the precursors from which new DNA strands are created. The triphosphates needed are the compounds denoted in Table 2.1 as dATP, dGTP, dTTP, and dCTP, which contain the bases adenine, guanine, thymine, and cytosine, respectively. Details of the structures of dCTP and dGTP are shown in **Figure 2.18**, in which the phosphate groups cleaved off during DNA synthesis are indicated. DNA synthesis requires all four nucleoside 5'-triphosphates and does not take place if any of them is omitted.

A feature found in all DNA polymerases is that

> A DNA polymerase can only *elongate* a DNA strand. It is not possible for DNA polymerase to *initiate* synthesis of a new strand, even when a template molecule is present.

One important implication of this principle is that DNA synthesis requires a preexisting segment of nucleic acid that is hydrogen-bonded to the template strand. This segment is called a **primer.** Because the primer molecule can be very short, it is an **oligonucleotide,** which literally means "few nucleotides." As we shall see in Chapter 6, in living cells the primer is a short segment of RNA, but in DNA amplification *in vitro*, the primer employed is usually DNA.

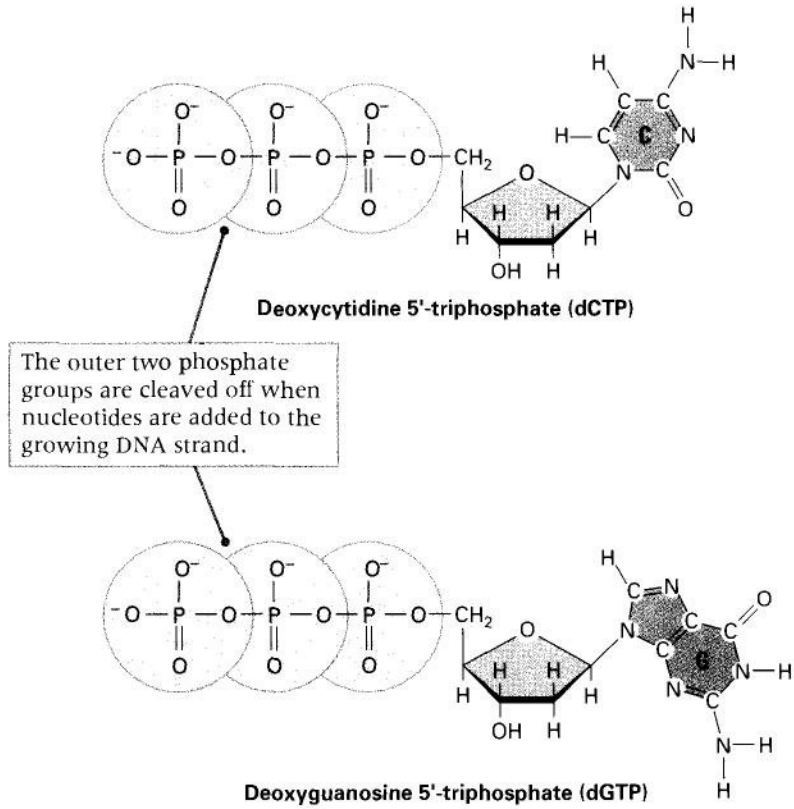


Figure 2.18 Two deoxynucleoside triphosphates used in DNA synthesis. The outer two phosphate groups are removed during synthesis.

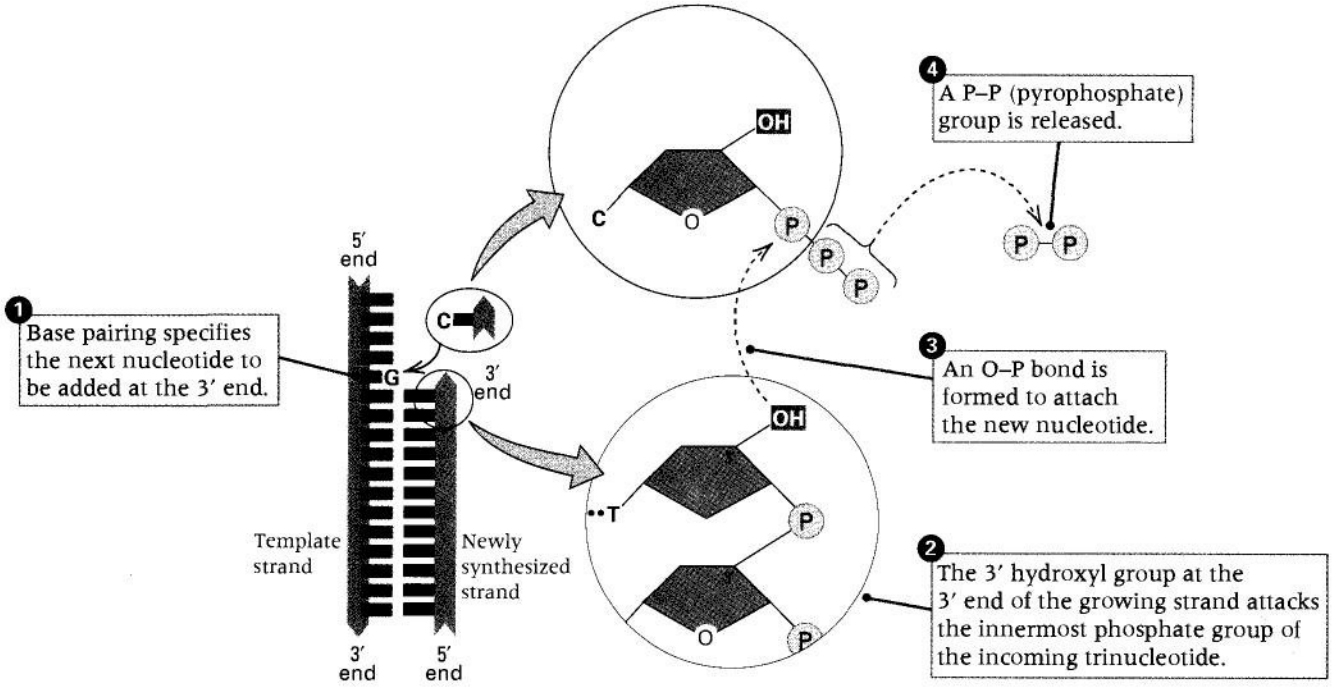


Figure 2.19 Addition of nucleotides to the 3'-OH terminus of a growing strand. The recognition step is shown as the formation of hydrogen bonds between the A and the T. The chemical reaction is that the 3'-OH group of the 3' end of the growing chain attacks the innermost phosphate group of the incoming trinucleotide.

It is the 3' end of the primer that is essential, because, as emphasized in Chapter 1,

> DNA synthesis proceeds only by addition of successive nucleotides to the 3' end of the growing strand. In other words, chain elongation always takes place in the 5'-to-3' direction (5' → 3').

The reason for the 5' → 3' direction of chain elongation is illustrated in **Figure 2.19**. It is a consequence of the fact that the reaction catalyzed by DNA polymerase is the formation of a phosphodiester bond between the free 3'-OH group of the chain being extended and the innermost phosphorus atom of the nucleoside triphosphate being incorporated at the 3' end. Recognition of the appropriate incoming nucleoside triphosphate in replication depends on base pairing with the opposite nucleotide in the template strand. DNA polymerase will usually catalyze the polymerization reaction that incorporates the new nucleotide at the primer terminus only when the correct base pair is present. The same DNA polymerase is used to add each of the four deoxynucleoside phosphates to the 3'-OH terminus of the growing strand.

The Polymerase Chain Reaction

The requirement for an oligonucleotide primer, and the constraint that chain elongation must always occur in the 5' → 3' direction, make it possible to obtain large quantities of a particular DNA sequence by selective amplification *in vitro*. The method for selective amplification is called the **polymerase chain reaction (PCR).** For its invention, Californian Kary B. Mullis was awarded a Nobel Prize in 1993. PCR amplification uses DNA polymerase and a *pair* of short, synthetic oligonucleotide primers, usually 18–22 nucleotides in length, that are complementary in sequence to the ends of the DNA sequence to be amplified. **Figure 2.20** gives an example in which the primer oligonucleotides (green) are 9-mers. These are too short for most practical purposes, but they will serve for illustration. The original duplex molecule (part A) is shown in blue. This duplex is mixed with a vast excess of primer molecules, DNA polymerase, and all four nucleoside triphosphates. When the temperature is raised, the strands of the duplex denature and become separated.

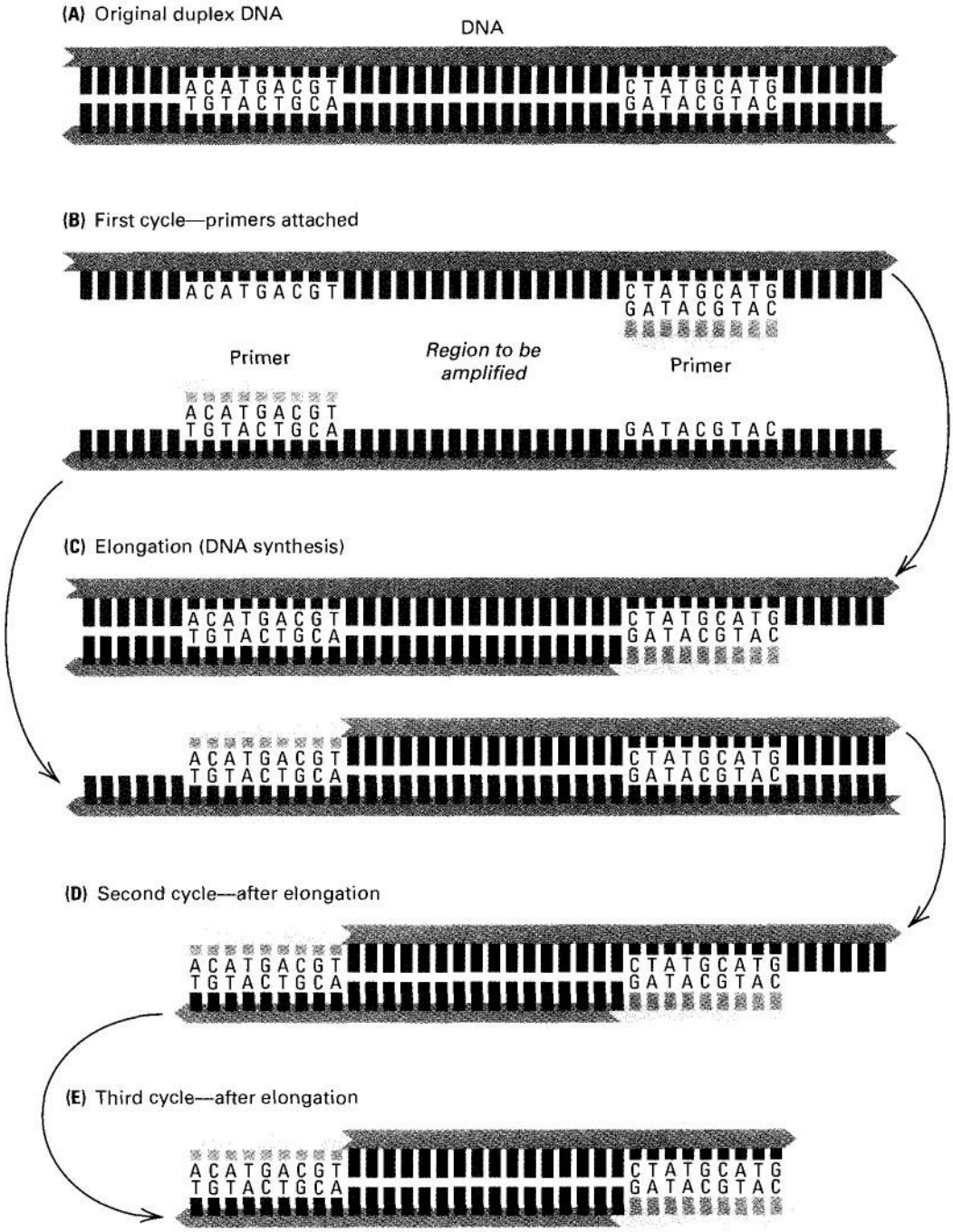


Figure 2.20 Role of primer sequences in PCR amplification. (A) Target DNA duplex (blue), showing sequences chosen as the primer-binding sites flanking the region to be amplified. (B) Primer (green) bound to denatured strands of target DNA. (C) First round of amplification. Newly synthesized DNA is shown in pink. Note that each primer is extended *beyond* the other primer site. (D) Second round of amplification (only one strand shown); in this round, the newly synthesized strand terminates at the opposite primer site. (E) Third round of amplification (only one strand shown); in this round, both strands are truncated at the primer sites. Primer sequences are normally about twice as long as shown here.

When the temperature is lowered again to allow renaturation, the primers, because they are in great excess, become annealed to the separated template strands (part B). Note that the primer sequences are different from each other but complementary to sequences present in opposite strands of the original DNA duplex and flanking the region to be amplified. The primers are oriented with their 3' ends pointing in the direction of the region to be amplified, because each DNA strand elongates only at the 3' end. After the primers have annealed, each is elongated by DNA polymerase using the original strand as a template, and the newly synthesized DNA strands (red) grow toward each other as synthesis proceeds (part C). Note that:

> A region of duplex DNA present in the original reaction mix can be PCR-amplified only if the region is flanked by the primer oligonucleotides.

To start a second cycle of PCR amplification, the temperature is raised again to denature the duplex DNA. Upon lowering of the temperature, the original parental strands anneal with the primers and are replicated as shown in Figure 2.20B and C. The daughter strands produced in the first round of amplification also anneal with primers and are replicated, as shown in part D. In this case, although the daughter duplex molecules are identical in sequence to the original parental molecule, they consist entirely of primer oligonucleotides and nonparental DNA that was synthesized in either the first or the second cycle of PCR. As successive cycles of denaturation, primer annealing, and elongation occur, the original parental strands are diluted out by the proliferation of new daughter strands until eventually, virtually every molecule produced in the PCR has the structure shown in part D.

The power of PCR amplification is that the number of copies of the template strand increases in exponential progression: 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, and so forth, doubling with each cycle of replication. Starting with a mixture containing as little as one molecule of the fragment of interest, repeated rounds of DNA replication increase the number of amplified molecules exponentially. For example, starting with a single molecule, 25 rounds of DNA replication will result in $2^{25} = 3.4 \times 10^7$ molecules. This number of molecules of the amplified fragment is so much greater than that of the other unamplified molecules in the original mixture that the amplified DNA can often be used without further purification. For example, a single fragment of 3 kb in *E. coli* accounts for only 0.06 percent of the DNA in this organism. However, if this single fragment were replicated through 25 rounds of replication, then 99.995 percent of the resulting mixture would consist of the amplified sequence. A 3-kb fragment of human DNA constitutes only 0.0001 percent of the total genome size. Amplification of a 3-kb fragment of human DNA to 99.995 percent purity would require about 34 cycles of PCR.

An overview of the polymerase chain reaction is shown in **Figure 2.21**. The DNA sequence to be amplified is again shown in blue and the oligonucleotide primers in green. The oligonucleotides anneal to the ends of the sequence to be amplified and become the substrates for chain elongation by DNA polymerase. In the first cycle of PCR amplification, the DNA is denatured to separate the strands. The denaturation temperature is usually around 95°C. Then the temperature is decreased to allow annealing in the presence of a vast excess of the primer oligonucleotides. The annealing temperature is typically in the range of 50°C–60°C, depending largely on the G + C content of the oligonucleotide primers. To complete the cycle, the temperature is raised slightly, to about 70°C, for the elongation of each primer. The steps of denaturation, renaturation, and replication are repeated from 20–30 times, and in each cycle the number of molecules of the amplified sequence is doubled.

Implementation of PCR with conventional DNA polymerases is not practical, because at the high temperature necessary for denaturation, the polymerase is itself irreversibly unfolded (denatured) and becomes inactive. However, DNA polymerase isolated from certain organisms is heat stable because the organisms normally live in hot springs at temperatures well above 90°C, such as are found in Yellowstone National Park. Such organisms are said to be **thermophiles**. The most widely used heat-stable DNA polymerase is called *Taq* polymerase, because it was originally isolated from the thermophilic bacterium *Thermus aquaticus*.

PCR amplification is very useful for generating large quantities of a specific DNA sequence. The principal limitation of the technique is that the DNA sequences at the ends of the region to be amplified must be known so that primer oligonucleotides can be synthesized. In addition, sequences longer than about 5000 base pairs cannot be replicated efficiently by conventional PCR procedures, although there are modifications of PCR that allow longer fragments to be amplified. On the other hand, many applications require amplification of relatively small fragments. The major advantage of PCR amplification is that it requires only trace amounts of template DNA. Theoretically only one template molecule is required, but in practice the amplification of a single molecule may fail because the molecule may, by chance, be broken or damaged. But amplification is usually reliable with as

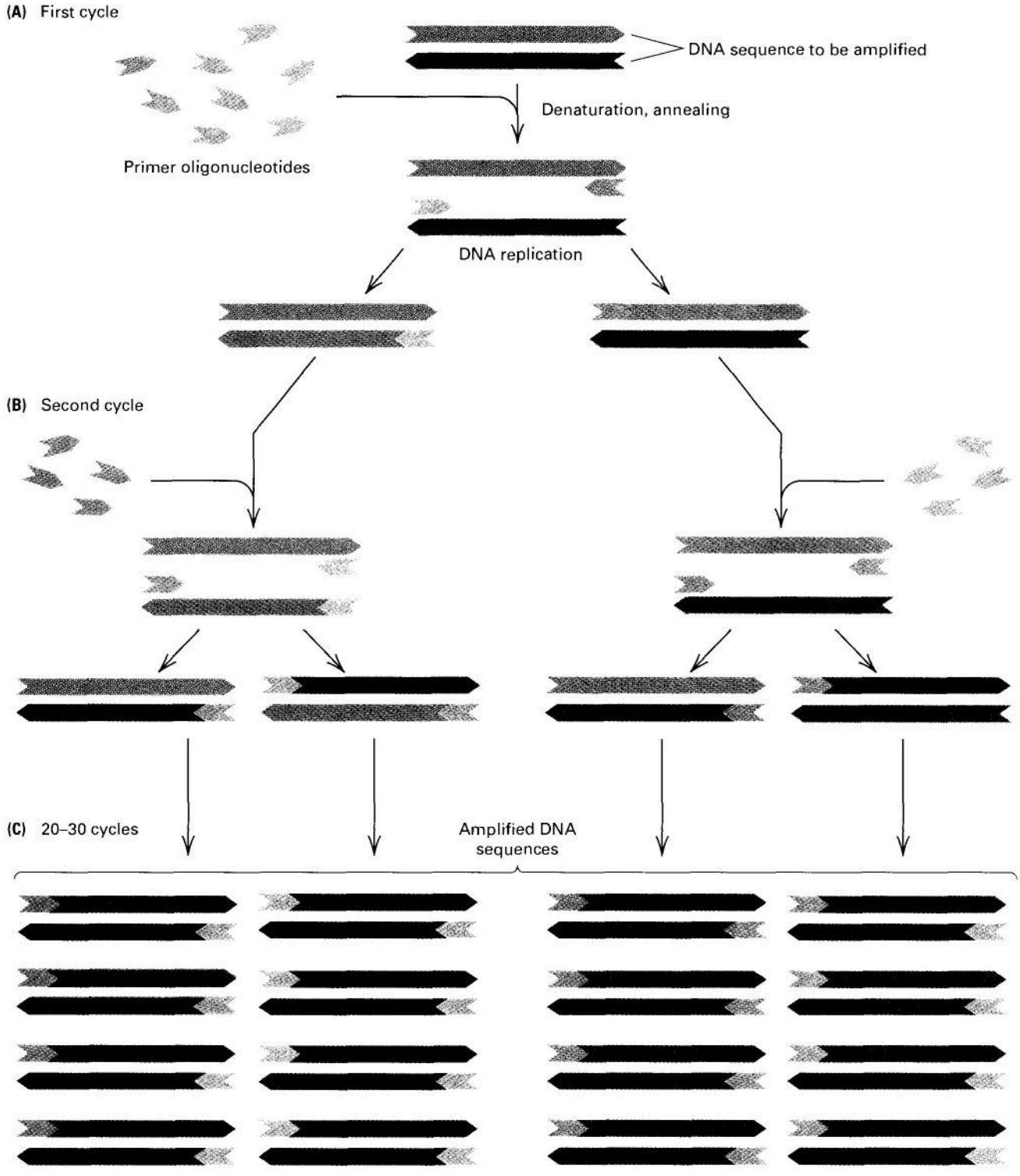


Figure 2.21 Polymerase chain reaction (PCR) for amplification of particular DNA sequences. Only the region to be amplified is shown. Oligonucleotide primers (green) that are complementary to the ends of the target sequence (blue) are used in repeated rounds of denaturation, annealing, and DNA replication. Newly replicated DNA is shown in pink. The number of copies of the target sequence doubles in each round of replication, eventually overwhelming any other sequences that may be present.

few as 10–100 template molecules, which makes PCR amplification 10,000–100,000 times more sensitive than detection via nucleic acid hybridization.

The exquisite sensitivity of PCR amplification has led to its use in DNA typing for criminal cases in which a minuscule amount of biological material has been left behind by the perpetrator (skin cells on a cigarette butt or hair-root cells on a single hair can yield enough template DNA for amplification). In research, PCR is widely used in the study of independent mutations in a gene whose sequence is known in order to identify the molecular basis of each mutation, to study DNA sequence variation among alternative forms of a gene that may be present in natural populations, or to examine differences among genes with the same function in different species. The PCR procedure has also come into widespread use in clinical laboratories for diagnosis. To take just one very important example, the presence of the human immunodeficiency virus (HIV), which causes acquired immune deficiency syndrome (AIDS), can be detected in trace quantities in blood banks via PCR by using primers complementary to sequences in the viral genetic material. These and other applications of PCR are facilitated by the fact that the procedure lends itself to automation by the use of mechanical robots to set up and run the reactions.

2.5 The Terminology of Genetic Analysis

In order to discuss the types of DNA markers that modern geneticists commonly use in genetic analysis, we must first introduce some key terms that provide the essential vocabulary of genetics. These terms can be understood with reference to **Figure 2.22**. In Chapter 1 we defined a **gene** as an element of heredity, transmitted from parents to offspring in reproduction, that influences one or more hereditary traits. Chemically, a gene is a sequence of nucleotides along a DNA molecule. In a population of organisms, not all copies of a gene may have exactly the same nucleotide sequence. For example, whereas one form of a gene has the codon GCA, which specifies alanine in the polypeptide chain that the gene encodes, another form of the same gene may have, at the same position, the codon GCG, which also specifies alanine. Hence the two forms of the gene encode the same sequence of amino acids yet differ in DNA sequence. The alternative forms of a gene are called **alleles** of the gene. Different alleles may also code for different amino acid sequences, sometimes with drastic effects. Recall the example of the *PAH* gene for phenyalanine hydroxylase in Chapter 1, in which a change in codon 408 from CGG (arginine) to TGG (tryptophan) results in an inactive enzyme that becomes expressed as the inborn error of metabolism phenylketonuria.

Within a cell, genes are arranged in linear order along microscopic thread-like bodies called **chromosomes,** which we will examine in detail in Chapters 4 and 8. Each human reproductive cell contains one complete set of 23 chromosomes containing 3×10^9 base pairs of DNA. A typical chromosome contains several hundred to several thousand genes. In humans the average is approximately 3500 genes per chromosome. Each chromosome contains a single molecule of duplex DNA along its length, complexed with proteins and very tightly coiled. The DNA in the average human chromosome, when fully extended, has relative dimensions comparable to those of a wet spaghetti noodle 25 miles long; when the DNA is coiled in the form of a chromosome, its physical compaction is comparable to that of the same noodle coiled and packed into an 18-foot canoe.

The physical position of a gene along a chromosome is called the **locus** of the gene. In most higher organisms, including human beings, each cell other than a sperm or egg contains two copies of each type of chromosome—one from the mother and one inherited from the father. Each member of such a pair of chromosomes is said to be **homologous** to the other. (The chromosomes that determine sex are an important exception, discussed in Chapter 4, that we will ignore for now.) At any locus, therefore, each individual carries two alleles, because one allele is present at a corresponding position in each of the homologous maternal and paternal chromosomes (Figure 2.22).

The genetic constitution of an individual is called its **genotype.** For a particular gene, if the two alleles at the locus in an individual are indistinguishable from each other, then for this gene the genotype of the individual is said to be **homozygous** for the allele that is present. If the two alleles at the locus are different from each other, then for this gene the genotype of the individual is said to be **heterozygous** for the alleles that are present. Typographically, genes are indicated in italics, and alleles are typically distinguished by uppercase or lowercase letters (*A* versus *a*), subscripts (A_1 versus A_2), superscripts (a^+ versus a^-), or sometimes just + and −. Using these symbols, homozygous genes would be portrayed by any of these formulas: *AA*, *aa*, A_1A_1, A_2A_2, a^+a^+, a^-a^-, +/+, or −/−. As in the last two examples, the slash is sometimes used to separate alleles present in homologous chromosomes to avoid ambiguity. Heterozygous genes would be portrayed by any of the formulas *Aa*, A_1A_2, a^+a^-, or +/−. In Figure 2.22, the genotype *Bb* is heterozygous because the *B* and *b* alleles are distinguishable (which is why they are assigned different symbols), whereas the genotype *CC* is homozygous. These genotypes could also be written as *B*/*b* and *C*/*C*, respectively.

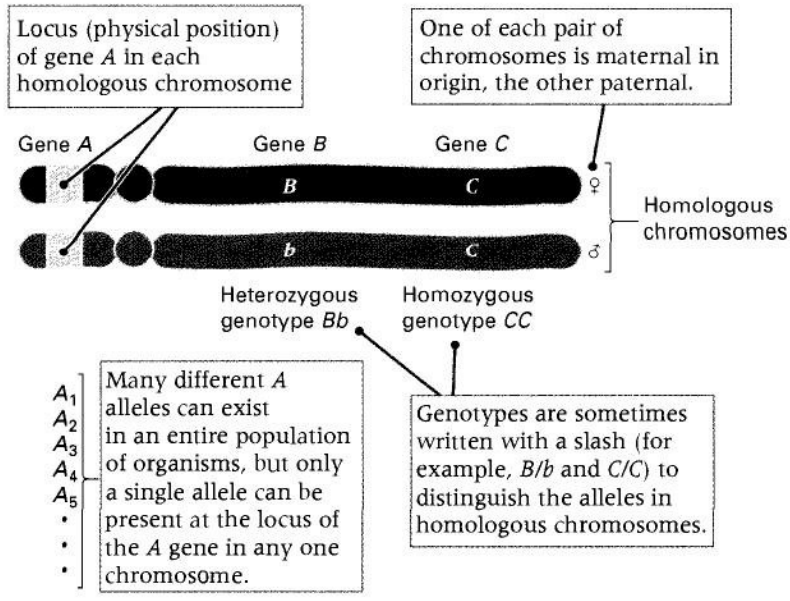


Figure 2.22 Key concepts and terms used in modern genetics. Note that a single gene can have any number of alleles in the population as a whole, but no more than two alleles can be present in any one individual.

Whereas the alleles that are present in an individual constitute its genotype, the physical or biochemical expression of the genotype is called the **phenotype.** To put it as simply as possible, the distinction is that the genotype of an individual is what is on the *inside* (the alleles in the DNA), whereas the phenotype is what is on the *outside* (the observable traits, including biochemical traits, behavioral traits, and so forth). The distinction between genotype and phenotype is critically important because there usually is not a one-to-one correspondence between genes and traits. Most complex traits, such as hair color, skin color, height, weight, behavior, life span, and reproductive fitness, are influenced by many genes. Most traits are also influenced more or less strongly by environment. This means that the same genotype can result in different phenotypes, depending on the environment. Compare, for example, two people with a genetic risk for lung cancer; if one smokes and the other does not, the smoker is much more likely to develop the disease. Environmental effects also imply that the same phenotype can result from more than one genotype; smoking again provides an example, because most smokers who are not genetically at risk can also develop lung cancer.

2.6 Types of DNA Markers Present in Genomic DNA

Genetic variation, in the form of multiple alleles of many genes, exists in most natural populations of organisms. We have called such genetic differences between individuals *DNA markers*; they are also called **DNA polymorphisms.** (The term *polymorphism* literally means "multiple forms.") The methods of DNA manipulation examined in Sections 2.3 and 2.4 can be used in a variety of combinations to detect differences among individuals. Anyone who reads the literature in modern genetics will encounter a bewildering variety of acronyms referring to different ways in which genetic polymorphisms are detected. The different approaches are in use because no single

method is ideal for all applications, each method has its own advantages and limitations, and new methods are continually being developed. In this section we examine some of the principal methods for detecting DNA polymorphisms among individuals.

Single-Nucleotide Polymorphisms (SNPs)

A **single-nucleotide polymorphism,** or **SNP** (pronounced "snip"), is present at a particular nucleotide site if the DNA molecules in the population frequently differ in the identity of the nucleotide pair that occupies the site. For example, some DNA molecules may have a T−A base pair at a particular nucleotide site, whereas other DNA molecules in the same population may have a C−G base pair at the same site. This difference constitutes a SNP. The SNP defines two "alleles" for which there could be three genotypes among individuals in the population: homozygous with T−A at the corresponding site in both homologous chromosomes, homozygous with C−G at the corresponding site in both homologous chromosomes, or heterozygous with T−A in one chromosome and C−G in the homologous chromosome. The word *allele* is in quotation marks above because the SNP need not be in a coding sequence, or even in a gene. In the human genome, any two randomly chosen DNA molecules are likely to differ at one SNP site about every 1000–3000 bp in protein-coding DNA and at about one SNP site every 500–1000 bp in noncoding DNA. Note, in the definition of a SNP, the stipulation that DNA molecules must differ at the nucleotide site "frequently." This provision excludes rare genetic variation of the sort found in less than 1 percent of the DNA molecules in a population. The reason for the exclusion is that genetic variants that are too rare are not generally as useful in genetic analysis as the more common variants. A catalog of SNPs is regarded as the ultimate compendium of DNA markers, because SNPs are the most common form of genetic differences among people and because they are distributed approximately uniformly along the chromosomes. By the middle of 2001, some 300,000 SNPs are expected to have been identified in human populations and their positions in the chromosomes located.

Restriction Fragment Length Polymorphisms (RFLPs)

Although most SNPs require DNA sequencing to be studied, those that happen to be located within a restriction site can be analyzed with a Southern blot. An example of this situation is shown in **Figure 2.23**, where the SNP consists of a T−A nucleotide pair in some molecules and a C−G pair in others. In this example, the polymorphic nucleotide

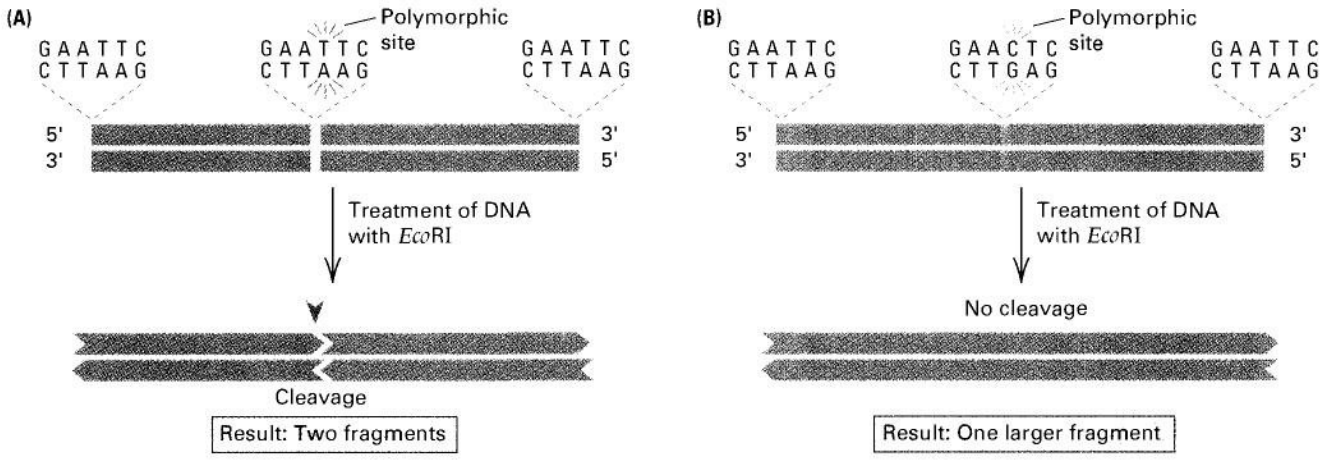


Figure 2.23 A minor difference in the DNA sequence of two molecules can be detected if the difference eliminates a restriction site. (A) This molecule contains three restriction sites for *Eco*RI, including one at each end. It is cleaved into two fragments by the enzyme. (B) This molecule has an altered *Eco*RI site in the middle, in which 5'-GAATTC-3' becomes 5'-GAACTC-3'. The altered site cannot be cleaved by *Eco*RI, so treatment of this molecule with *Eco*RI results in one larger fragment.

site is included in a cleavage site for the restriction enzyme *Eco*RI (5'-GAATTC-3'). The two nearest flanking *Eco*RI sites are also shown. In this kind of situation, DNA molecules with T—A at the SNP will be cleaved at both flanking sites and also at the middle site, yielding two *Eco*RI restriction fragments. Alternatively, DNA molecules with C—G at the SNP will be cleaved at both flanking sites but not at the middle site (because the presence of C—G destroys the *Eco*RI restriction site) and so will yield only one larger restriction fragment. A SNP that eliminates a restriction site is known as a **restriction fragment length polymorphism,** or **RFLP** (pronounced either as "riflip" or by spelling it out).

Because RFLPs change the number and size of DNA fragments produced by digestion with a restriction enzyme, they can be detected by the Southern blotting procedure discussed in Section 2.3. An example appears in **Figure 2.24**. In this case the labeled probe DNA hybridizes near the restriction site at the far left and identifies the position of this restriction fragment in the electrophoresis gel. The duplex molecule labeled "allele *A*" has a restriction site in the middle and, when cleaved and subjected to electrophoresis, yields a small band that

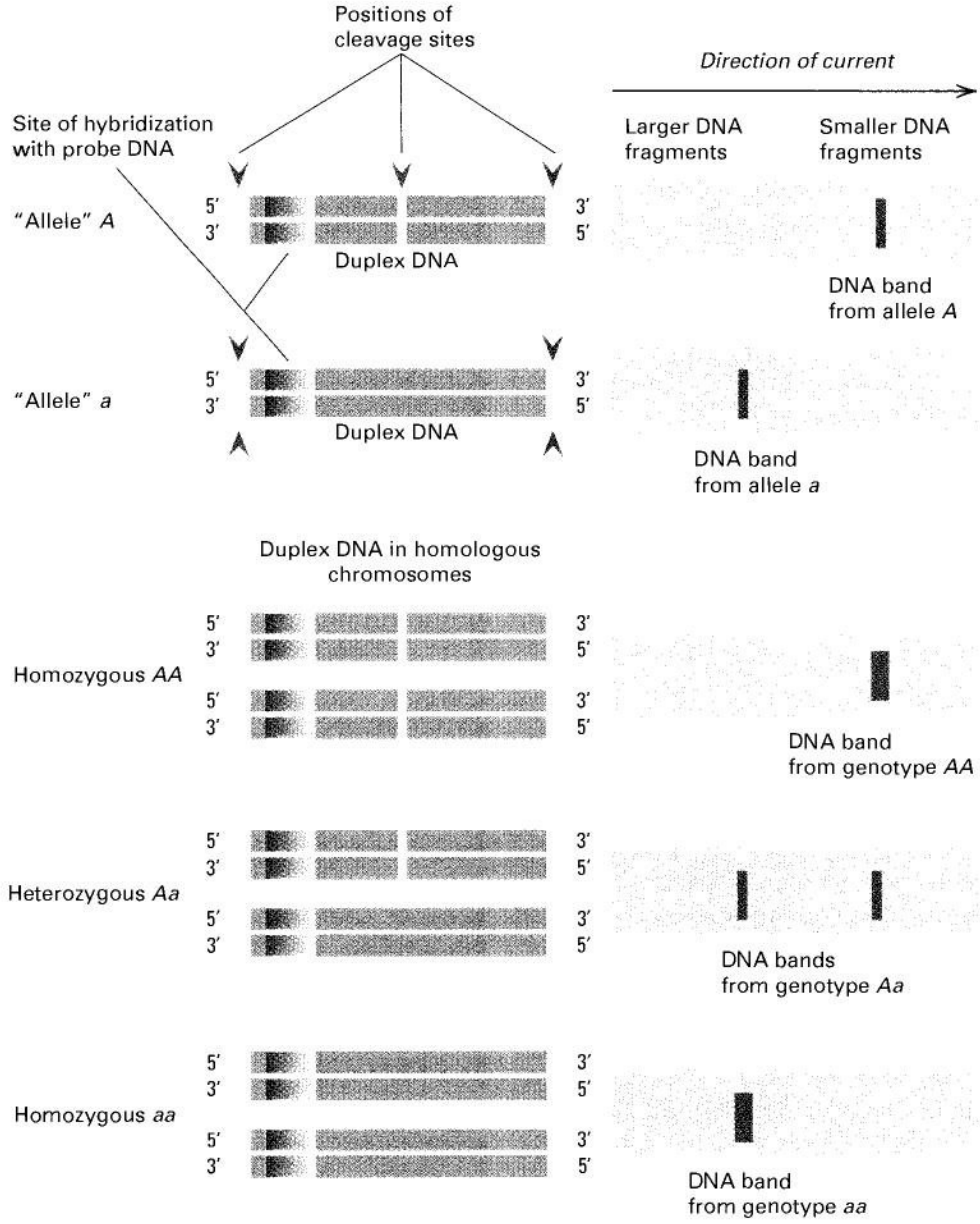


Figure 2.24 In a restriction fragment length polymorphism (RFLP), alleles may differ in the presence or absence of a cleavage site in the DNA. In this example, the *a* allele lacks a restriction site that is present in the DNA of the *A* allele. The difference in fragment length can be detected by Southern blotting. RFLP alleles are codominant, which means (as shown at the bottom) that DNA from the heterozygous *Aa* genotype yields each of the single bands observed in DNA from homozygous *AA* and *aa* genotypes.

CoNnection

Origin of the Human Genetic Linkage Map

David Botstein,[1] Raymond L. White,[2] Mark Skolnick,[3] and Ronald W. Davis[4] 1980

[1] Massachusetts Institute of Technology, Cambridge, Massachusetts
[2] University of Massachusetts Medical Center, Worcester, Massachusetts
[3] University of Utah, Salt Lake City, Utah
[4] Stanford University, Stanford, California

Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms

This historic paper stimulated a major international effort to establish a genetic linkage map of the human genome based on DNA polymorphisms. Pedigree studies using these genetic markers soon led to the chromosomal localization and identification of mutant genes for hundreds of human diseases. A more ambitious goal, still only partly achieved, is to understand the genetic and environmental interactions involved in complex traits such as heart disease and cancer. The "small set of large pedigrees" called for in the excerpt was soon established by the Centre d'Etude du Polymorphisme Humain (CEPH) in Paris, France, and made available to investigators worldwide for genetic linkage studies. Today the CEPH maintains a database on the individuals in these pedigrees that comprises approximately 12,000 polymorphic DNA markers and 2.5 million genotypes.

No method of systematically mapping human genes has been devised, largely because of the paucity of highly polymorphic marker loci. The advent of recombinant DNA technology has suggested a theoretically possible way to define an arbitrarily large number of arbitrarily polymorphic marker loci. . . . A subset of such polymorphisms can readily be detected as differences in the length of DNA fragments after digestion with DNA sequence-specific restriction endonucleases. These restriction fragment length polymorphisms (RFLPs) can be easily assayed in individuals, facilitating large population studies. . . . [Genetic mapping] of many DNA marker loci should allow the establishment of a set of well-spaced, highly polymorphic genetic markers covering the entire human genome [and enabling] any trait caused wholly or partially by a major locus segregating in a pedigree to be mapped. Such a procedure would not require any knowledge of the biochemical nature of the trait or of the nature of the alterations in the DNA responsible for the trait. . . . The most efficient procedure will be to study a small set of large pedigrees which have been genotyped for all known polymorphic markers. . . . The resolution of genetic and environmental components of disease . . . must involve unraveling the underlying genetic predisposition, understanding the environmental contributions, and understanding the variability of expression of the phenotype. In principle, linked marker loci can allow one to establish, with high certainty, the genotype of an individual and, consequently, assess much more precisely the contribution of modifying factors such as secondary genes, likelihood of expression of the phenotype, and environment.

In principle, linked marker loci can allow one to establish, with high certainty, the genotype of an individual.

Source: American Journal of Human Genetics 32: 314–331

contains sequences homologous to the probe DNA. The duplex molecule labeled "allele *a*" lacks the middle restriction site and yields a larger band. In this situation there can be three genotypes—*AA*, *Aa*, or *aa*, depending on which alleles are present in the homologous chromosomes—and all three genotypes can be distinguished as shown in the Figure 2.24. Homozygous *AA* yields only a small fragment, homozygous *aa* yields only a large fragment, and heterozygous *Aa* yields both a small and a large fragment. Because the presence of both the *A* and *a* alleles can be detected in heterozygous *Aa* genotypes, *A* and *a* are said to be **codominant.** In Figure 2.24, the bands from *AA* and *aa* have been shown as somewhat thicker than those from *Aa*, because each *AA* genotype has two copies of the *A* allele and each *aa* genotype has two copies of the *a* allele, compared with only one copy of each allele in the heterozygous genotype *Aa*.

Random Amplified Polymorphic DNA (RAPD)

For studying DNA markers, one limitation of Southern blotting is that it requires material (probes available in the form of cloned

DNA) and one limitation of PCR is that it requires sequence information (so primer oligonucleotides can be synthesized). These are not severe handicaps for organisms that are well studied (for example, human beings, domesticated animals and cultivated plants, and model genetic organisms such as yeast, fruit fly, nematode, or mouse), because research materials and sequence information are readily available. But for the vast majority of organisms that biologists study, there are neither research materials nor sequence information. Genetic analysis can still be carried out in these organisms by using an approach called **random amplified polymorphic DNA** or **RAPD** (pronounced *"rapid"*), described in this section.

RAPD analysis makes use of a set of PCR primers of 8–10 nucleotides whose sequence is essentially *random*. The random primers are tried individually or in pairs in PCR reactions to amplify fragments of genomic DNA from the organism of interest. Because the primers are so short, they often anneal to genomic DNA at multiple sites. Some primers anneal in the proper orientation and at a suitable distance from each other to support amplification of the unknown sequence between them. Among the set of amplified fragments are ones that can be amplified from some genomic DNA samples but not from others, which means that the presence or absence of the amplified fragment is polymorphic in the population of organisms.

In most organisms it is usually straightforward to identify a large number of RAPDs that can serve as genetic markers for many different kinds of genetic studies. An example of RAPD gel analysis is illustrated in **Figure 2.25**, where three pairs of primers (sets 1–3) are used to amplify genomic DNA from four individuals in a population. The fragments that amplify are then separated on an electrophoresis gel and visualized after straining with ethidium bromide. Many amplified bands are typically observed for each primer set, but only some of these are

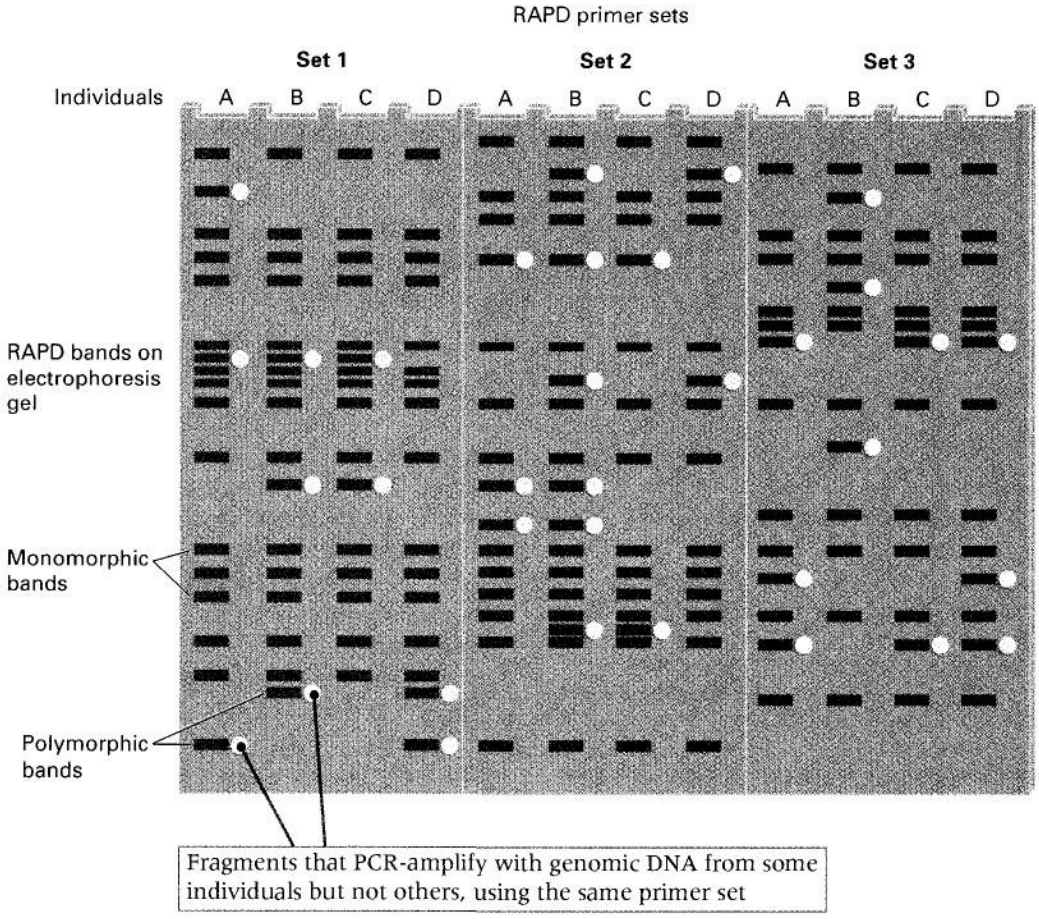


Figure 2.25 Random amplified polymorphic DNA (RAPD) is detected through the use of relatively short primer sequences that, by chance, match genomic DNA at multiple sites that are close enough together to support PCR amplification. Genomic DNA from a single individual typically yields many bands, only some of which are polymorphic in the population. Different sets of primers amplify different fragments of genomic DNA.

polymorphic. These are indicated in Figure 2.25 by the colored dots. The amplified bands that are not polymorphic are said to be **monomorphic** in the sample, which means that they are the same from one individual to the next. This example shows 17 RAPD polymorphisms. **Figure 2.26** shows an actual RAPD gel amplified from genomic DNA obtained from small tissue samples from a population of fish (*Campostoma anomalum*) in the Great Miami River Basin, Ohio. The fish were collected as part of a water quality assessment program to determine whether fish populations in stressful water environments progressively lose their genetic variation (that is, become increasingly monomorphic). Each pair of samples is flanked by a lane containing DNA size standards producing a "ladder" of fragments at 100-bp increments.

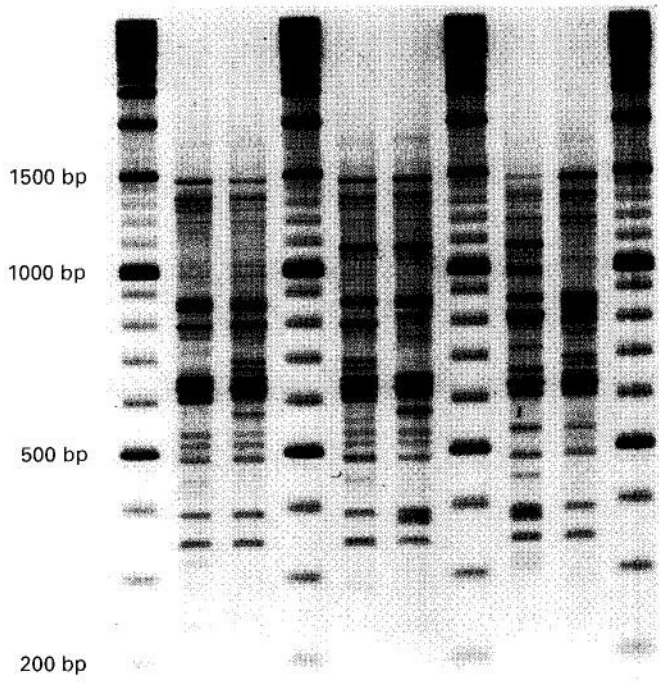


Figure 2.26 RAPD polymorphisms in the stoneroller fish (*Campostoma anomalum*) trapped in tributaries of the Great Miami River in Ohio. Each pair of samples is flanked by a lane containing DNA size standards; in these lanes, the smallest DNA fragment is 100 base pairs (bp), and each successively larger fragment increases in size by 100 bp. Fragments whose sizes are multiples of 500 bp are present in greater concentration and so yield darker bands. [Courtesy of Michael Simonich, Manju Garg, and Ana Braam (Pathology Associates International, Cincinnati, Ohio).]

Figures 2.24 through 2.26 illustrate an important point:

> In modern genetics, the phenotypes that are studied are very often bands in a gel rather than physical or physiological characteristics.

Figure 2.25 offers a good example. Each position at which a band is observed in one or more samples is a phenotype, whether or not the band is polymorphic. For example, primer set 1 yields a total of 19 bands, of which 5 are polymorphic and 14 are monomorphic in the sample. The phenotypes could be named in any convenient way, such as by indicating the primer set and the fragment length. For example, suppose that the smallest amplified fragment for primer set 1 is 125 bp, which is the polymorphic fragment at the bottom left in Figure 2.25. We could name this fragment unambiguously as 1-125 because it is a fragment of 125 bp amplified by primer set 1.

To understand why a DNA band is a *phenotype*, rather than a gene or a genotype, it is useful to assign different names to the "alleles" that do or do not support amplification. (The word *allele* is in quotation marks again, because the 1-125 fragment that is amplified need not be part of a gene.) We are talking only about the 1-125 fragment, so we could call the allele capable of supporting amplification the plus (+) allele and the allele not capable of supporting amplification the minus (−) allele. Then there are three possible genotypes with regard to the amplified fragment: +/+, +/−, and −/−. Using genomic DNA from these genotypes, the homozygous +/+ and heterozygous +/− will both support amplification of the 1-125 fragment, whereas the −/− genotype will not support amplification. Hence the presence of the 1-125 fragment is the *phenotype* observed in both +/+ and +/− genotypes. In other words, with regard to amplification, the + allele is **dominant** to the − allele, because the phenotype (presence of the 1-125 band) is present in both homozygous +/+ and heterozygous +/− genotypes. Therefore, on the basis of the phenotype for the 1-125 band in Figure 2.25, we could say that individuals A and D could have either a +/+ or +/− genotype but that individuals B and C must have genotype −/−.

Amplified Fragment Length Polymorphisms (AFLPs)

Because RAPD primers are small and may not match the template DNA perfectly, the amplified DNA bands often differ a great deal in how dark or light they appear. This variation creates a potential problem, because some exceptionally dark bands may actually result from two amplified DNA fragments of the same size, and some exceptionally light bands may be difficult to visualize consistently. To obtain amplified fragments that yield more uniform band intensities, double-stranded oligonucleotide sequences that match the primer sequences perfectly can be attached to genomic restriction fragments enzymatically prior to amplification. This method, which is outlined in **Figure 2.27**, yields a class of DNA polymorphisms known as **amplified fragment length polymorphisms,** or **AFLPs** (usually pronounced by spelling it out). The first step (part A) is to digest genomic DNA with a restriction enzyme; this example uses the enzyme *Eco*RI, whose restriction site is 5'-GAATTC-3'. Digestion yields a large number of restriction fragments flanked by what remains of an *Eco*RI site on each side. In the next step (part B), double-stranded oligonucleotides called **primer adapters,** with single-stranded overhangs complementary to those on the restriction fragments, are ligated onto the restriction fragments using the enzyme *DNA ligase*. The resulting fragments (C) are ready for amplification by means of PCR. Note that the same adapter is ligated onto each end, so a single primer sequence will anneal to both ends and support amplification.

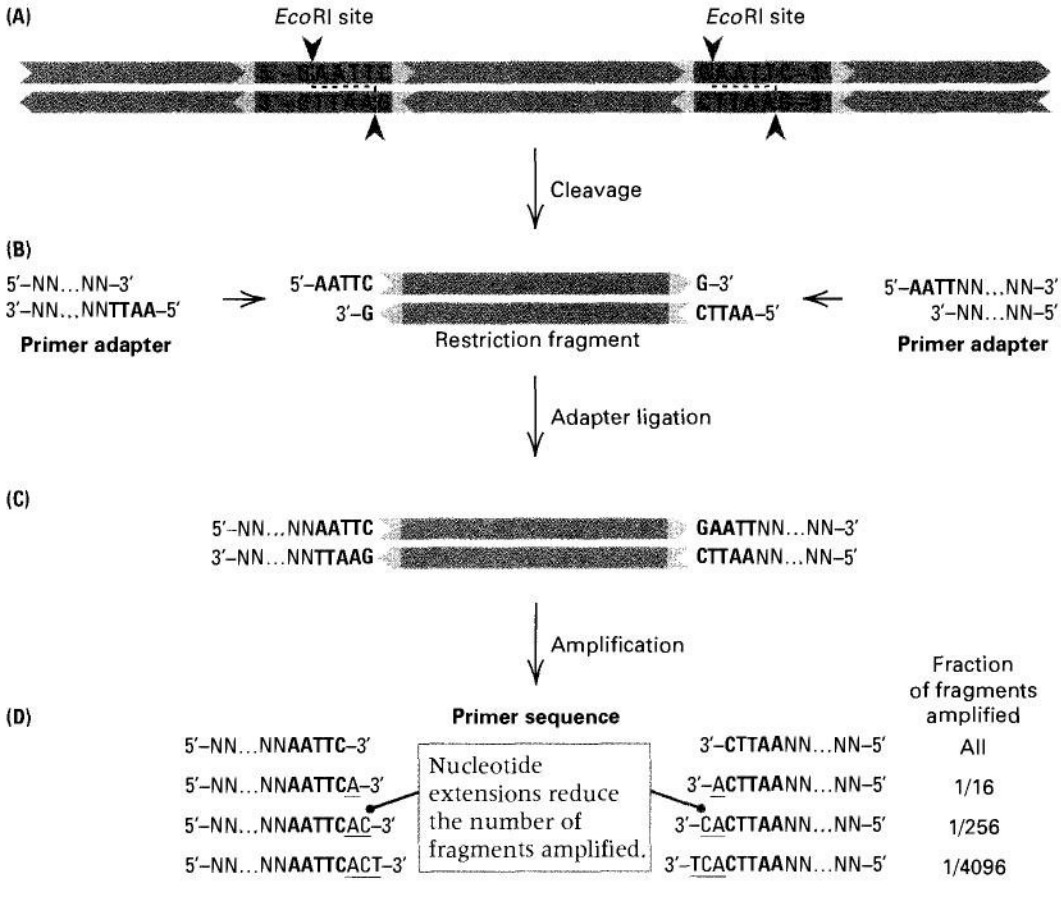


Figure 2.27 An amplified fragment length polymorphism (AFLP). (A and B) Genomic DNA is digested with one or more restriction enzymes (in this case, *Eco*RI). (C) Oligonucleotide adapters are ligated onto the fragments; note that the single-stranded overhang of the adaptors matches those of the genomic DNA fragments. (D) The resulting fragments are subjected to PCR using primers complementary to the adapters. The number of amplified fragments can be adjusted by manipulating the number of nucleotides in the adapters that are also present in the primers.

There are nevertheless a number of choices concerning the primer sequence. A primer that matches the adapters perfectly will amplify all fragments, but this often results in so many amplified fragments that they are not well separated in the gel. A PCR primer must match perfectly at its 3' end to be elongated. Thus, additional nucleotides added to the 3' end reduce the number of amplified fragments, because these primers will amplify only those fragments that, by chance, have a complementary nucleotide immediately adjacent to the *Eco*RI site. For example, the primer sequence in the second row has a single-nucleotide 3' extension; because only $1/4 \times 1/4$ of the fragments would be expected to have a complementary T immediately adjacent to the *Eco*RI site on both sides, this primer is expected to amplify 1/16 of all the restriction fragments. Similarly, the primer sequence in the third row has a two-nucleotide 3' extension, so this primer is expected to amplify $1/16 \times 1/16 = 1/256$ of all the fragments. One application of AFLP analysis is to organisms with large genomes, such as grasshoppers and crickets, for which RAPD analysis would yield an excessive number of amplified bands. How large is a "large" genome? Compared to the human genome, that of the brown mountain grasshopper *Podisma pedestris* is 7 times larger, and those of North American salamanders in the genus *Amphiuma* are 70 times larger! (Genome size is discussed in Chapter 8.)

Simple Tandem Repeat Polymorphisms (STRPs)

One more type of DNA polymorphism warrants consideration because it is useful in DNA typing for individual identification and for assessing the degree of genetic relatedness between individuals. This type of polymorphism is called a **simple tandem repeat polymorphism (STRP)** because the genetic differences among DNA molecules consist of the number of copies of a short DNA sequence that may be repeated many times in tandem at a particular locus in the genome. STRPs that are present at different loci may differ in the sequence and length of the repeating unit, as well as in the minimum and maximum number of tandem copies that occur in DNA molecules in the population. A STRP with a repeating unit of 2–9 bp is often called a **microsatellite** or a **simple sequence length polymorphism (SSLP),** whereas a STRP with a repeating unit of 10–60 bp is often called a **minisatellite** or a **variable number of tandem repeats (VNTR).**

Figure 2.28 shows an example of a STRP with a copy number ranging from 1 through 10. Because the number of copies determines the size of any restriction fragment that includes the STRP, each DNA molecule yields a single-size restriction fragment depending on the number of copies it contains. The STRP in Figure 2.28 has 10 different "alleles" (again, we use quotation marks because the STRP may not be in a gene), which could be distinguished either by Southern blotting using a probe to a unique (nonrepeating) sequence within the restriction fragment or by PCR amplification using primers to a unique sequence on either side of the tandem repeats. In this situation the locus is said to have **multiple alleles** in the population. Even with multiple alleles, however, any one chromosome can carry only one of the alleles, and any individual genotype can carry at most two different alleles. Nevertheless, a large number of alleles means an even larger number of genotypes, which is the feature that gives STRPs their utility in individual identification. For example, even with only 10 alleles in a population of organisms, there could be 10 different homozygous genotypes and 45 different heterozygous genotypes.

More generally, with n alleles there are n homozygous genotypes and $n(n - 1)/2$ heterozygous genotypes, or $n(n + 1)/2$ different genotypes altogether. With STRPs, not only are there a relatively large number of alleles, but no one allele is exceptionally common, so each of the many genotypes in the population has a relatively low frequency. If the genotypes at 6–8 STRP loci are considered simultaneously, then each possible multiple-locus genotype is exceedingly rare. Because of their high degree of variation among people, STRPs are widely used in **DNA typing** (sometimes called **DNA fingerprinting**) to establish individual identity for use in criminal investigations, parentage determinations, and so forth (Chapter 17).

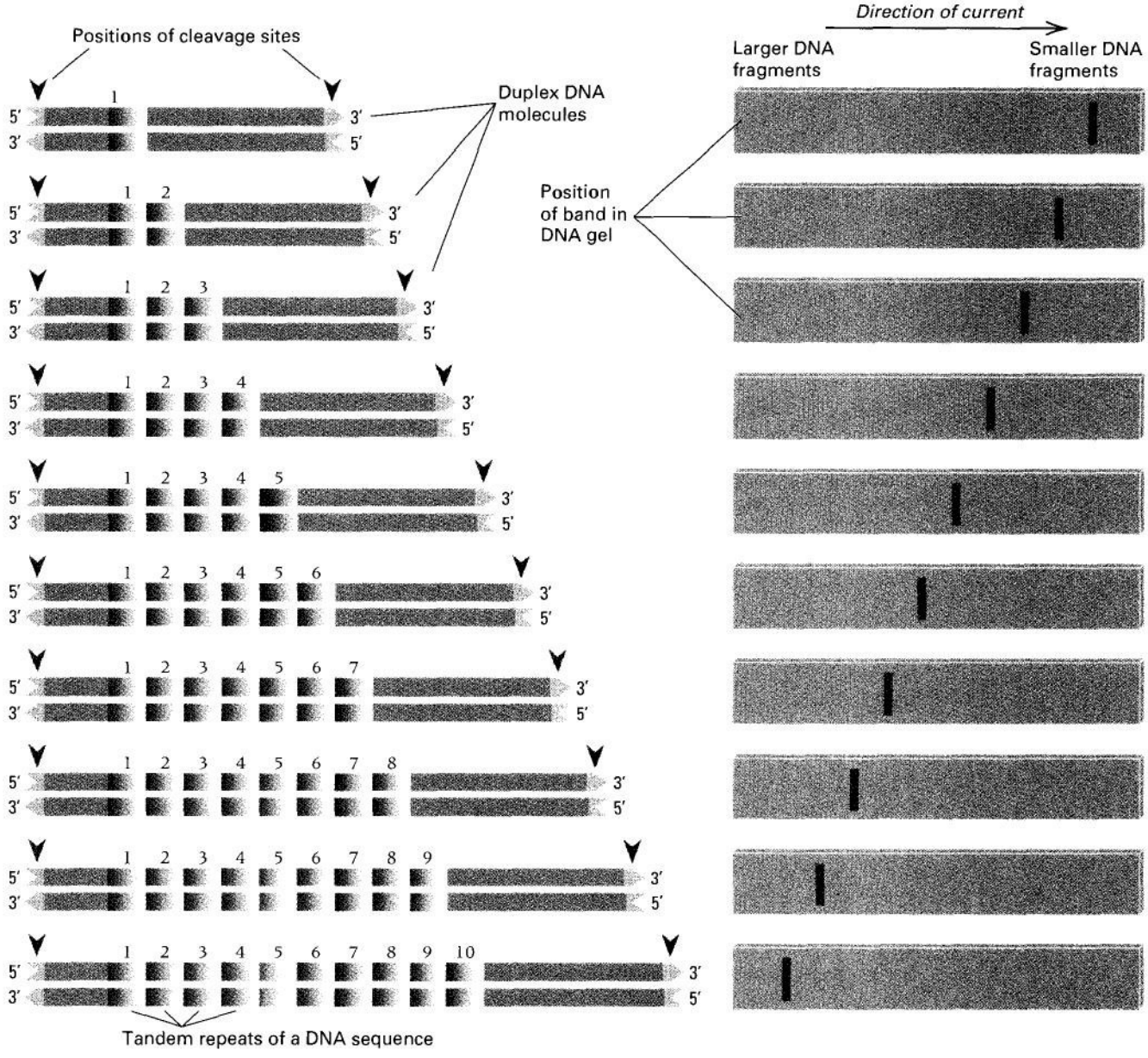


Figure 2.28 In a simple tandem repeat polymorphism (STRP), the alleles in a population differ in the number of copies of a short sequence (typically 2–60 bp) that is repeated in tandem along the DNA molecule. This example shows alleles in which the repeat number varies from 1 to 10. Cleavage at restriction sites flanking the STRP yields a unique fragment length for each allele. The alleles can also be distinguished by the size of the fragment amplified by PCR using primers that flank the STRP.

2.7 Applications of DNA Markers

Why are geneticists interested in DNA markers (DNA polymorphisms)? Their interest can be justified on any number of grounds. In this section we consider the reasons most often cited.

Genetic Markers, Genetic Mapping, and "Disease Genes"

Perhaps the key goal in studying DNA polymorphisms in human genetics is to identify the chromosomal location of mutant genes associated with hereditary diseases. In the context of disorders caused by the interaction of multiple genetic and environmental



Pam Francis/TimePix

Some people carry genetic risk factors and pass them on to their children—these three sisters have all had breast cancer. For breast cancer, several of the major genetic risk factors have now been identified. DNA studies can identify women at risk so that they can have frequent examinations to allow early detection. This greatly increases the chances for long-term survivial.

factors, such as heart disease, cancer, diabetes, depression, and so forth, it is important to think of a harmful allele as a **risk factor** for the disease, which increases the probability of occurrence of the disease, rather than as a sole causative agent. This needs to be emphasized, especially because genetic risk factors are often called **disease genes.** For example, the major "disease gene" for breast cancer in women is the gene *BRCA1*. For women who carry a mutant allele of *BRCA1*, the lifetime risk of breast cancer is about 36 percent, and hence most women with this genetic risk factor do *not* develop breast cancer. On the other hand, among women who are not carriers, the lifetime risk of breast cancer is about 12 percent, and hence many women without the genetic risk factor *do* develop breast cancer. Indeed, *BRCA1* mutations are found in only 16 percent of affected women who have a family history of breast cancer. The importance of a genetic risk factor can be expressed quantitatively as the **relative risk,** which equals the risk of the disease in persons who carry the risk factor as compared to the risk in persons who do not. The relative risk for *BRCA1* equals 3.0 (calculated as 36 percent/12 percent).

The utility of DNA polymorphisms in locating and identifying disease genes results from **genetic linkage,** the tendency for genes that are sufficiently close together in a chromosome to be inherited together. Genetic linkage will be discussed in detail in Chapter 5, but the key concepts are summarized in **Figure 2.29,** which shows the location of many DNA polymorphisms along a chromosome that also carries a genetic risk factor denoted *D* (for disease gene). Each DNA polymorphism serves as a genetic marker for its own location in the chromosome. The importance of genetic linkage is that DNA markers that are sufficiently close to the disease gene will tend to be inherited together with the disease gene in pedigrees—and the closer the markers, the stronger this association. Hence, the initial approach to the identification of a disease gene is to find DNA markers that are genetically linked with the disease gene in order to identify its chromosomal location, a procedure known as **genetic mapping.** Once the chromosomal position is known, other methods can be used to pinpoint the disease gene itself and to study its functions.

If genetic linkage seems a roundabout way to identify disease genes, consider the alternative. The human genome contains approximately 80,000 genes. If genetic linkage did not exist, then we would have to examine 80,000 DNA polymorphisms, one in each gene, in order to identify a disease gene. But the human genome has only 23 pairs of chromosomes, and because of genetic linkage and the power of genetic mapping, it actually requires only a few hundred DNA polymorphisms to identify the chromosome and approximate location of a genetic risk factor.

Other Uses for DNA Markers

DNA polymorphisms are widely used in all aspects of modern genetics because they provide a large number of easily accessed genetic markers for genetic mapping and other purposes. Among the other uses of DNA polymorphisms are the following.

Individual identification. We have already mentioned that DNA polymorphisms have application as a means of DNA typing (DNA fingerprinting) to identify different individuals in a population. DNA typing in other organisms is used to determine individual animals in endangered species and to identify the degree of genetic relatedness among individual organisms that live in packs or herds. For example, DNA typing in

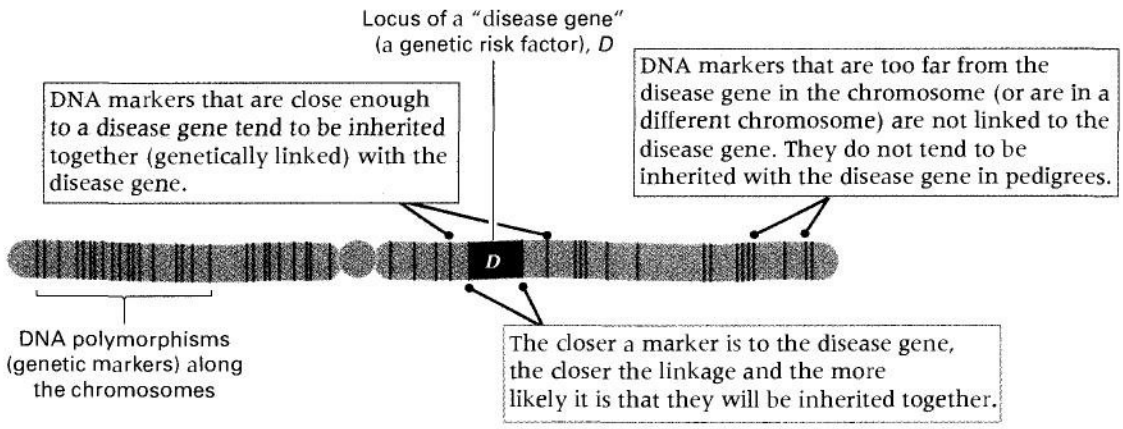


Figure 2.29 Concepts in genetic localization of genetic risk factors for disease. Polymorphic DNA markers (indicated by the vertical lines) that are close to a genetic risk factor (*D*) in the chromosome tend to be inherited together with the disease itself. The genomic location of the risk factor is determined by examining the known genomic locations of the DNA polymorphisms that are linked with it.

wild horses has shown that the wild stallion in charge of a harem of mares actually sires fewer than one-third of the foals.

Epidemiology and food safety science. DNA typing also has important applications in tracking the spread of viral and bacterial epidemic diseases, as well as in identifying the source of contamination in contaminated foods.

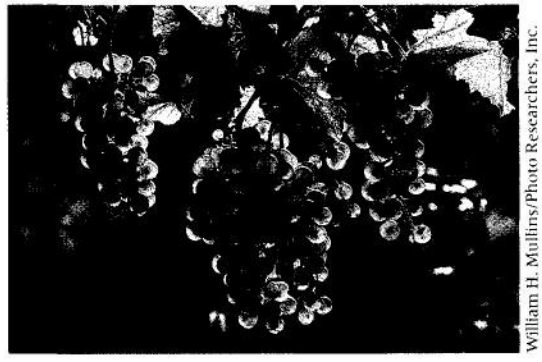
Human population history. DNA polymorphisms are widely used in anthropology to reconstruct the evolutionary origin, global expansion, and diversification of the human population.

Improvement of domesticated plants and animals. Plant and animal breeders have turned to DNA polymorphisms as genetic markers in pedigree studies to identify, by genetic mapping, genes that are associated with favorable traits in order to incorporate these genes into currently used varieties of plants and breeds of animals.

History of domestication. Plant and animal breeders also study genetic polymorphisms to identify the wild ancestors of cultivated plants and domesticated animals, as well as to infer the practices of artificial selection that led to genetic changes in these species during domestication.

DNA polymorphisms as ecological indicators. DNA polymorphisms are being evaluated as biological indicators of genetic diversity in key indicator species present in biological communities exposed to chemical, biological, or physical stress. They are also used to monitor genetic diversity in endangered species and species bred in captivity.

Evolutionary genetics. DNA polymorphisms are studied in an effort to describe the patterns in which different types of genetic variation occur throughout the genome, to infer the evolutionary mechanisms by which genetic variation is maintained, and to illuminate the processes by which genetic polymorphisms within species become transformed into genetic differences between species.



William H. Mullins/Photo Researchers, Inc.

A paternity test–for wines. Genetic researchers have shown that the notion of "purity of breed" thought to be so critical to the commercial value of certain wines is ill-founded. DNA tests indicate that wine grapes such as the cabernet sauvignon and chardonnay (shown here) are in fact hybrids.

Population studies. Population ecologists employ DNA polymorphisms to assess the level of genetic variation in diverse populations of organisms that differ in genetic organization (prokaryotes, eukaryotes, organelles), population size, breeding structure, or life-history characters, and they use genetic polymorphisms within subpopulations of a species as indicators of population history, patterns of migration, and so forth.

Evolutionary relationships among species. Differences in homologous DNA sequences between species is the basis of molecular systematics, in which the sequences are analyzed to determine the ancestral history (phylogeny) of the species and to trace the origin of morphological, behavioral, and other types of adaptations that have arisen in the course of evolution.

Chapter Summary

The sequence of bases in the human genome is 99.9% identical from one person to the next. The remaining 0.1%—comprising 3 million base pairs—differs among individuals. Included in these differences are many mutations that cause or increase the risk of disease, but the majority of the differences are harmless in themselves. Any of these differences between genomes can be used as a genetic marker. Genetic markers are widely employed in genetics to serve as positional landmarks along a chromosome or to identify particular cloned DNA fragments. The manipulation of DNA molecules to identify genetic markers is the basic experimental operation in modern genetics.

A DNA strand is a polymer of deoxyribonucleotides, each composed of a nitrogenous base, a deoxyribose sugar, and a phosphate. Sugars and phosphates alternate in forming a polynucleotide chain with one terminal 3'-OH group and one terminal 5'-P group. In double-stranded (duplex) DNA, the two strands are paired and antiparallel. Each end of the double helix carries a terminal 3'-OH group in one strand and a terminal 5'-P group in the other strand. The four bases found in DNA are the purines, adenine (A) and guanine (G), and the pyrimidines, cytosine (C) and thymine (T). Equal numbers of purines and pyrimidines are found in double-stranded DNA (Chargaff's rules), because the bases are paired as A–T pairs and G–C pairs. The hydrogen-bonded base pairs, along with hydrophobic base stacking of the nucleotide pairs in the core of the double helix, hold the two polynucleotide strands together in a double helix.

Duplex DNA can be cleaved into fragments of defined length by restriction enzymes, each of which cleaves DNA at a specific recognition sequence (restriction site) usually four or six nucleotide pairs in length. These fragments can be separated by electrophoresis. The positions of particular restriction fragments in a gel can be visualized by means of nucleic acid hybridization, in which strands of duplex DNA that have been separated (denatured) by heating are mixed and come together (renature) with strands having complementary nucleotide sequences. In a Southern blot, denatured and labeled probe DNA is mixed with denatured DNA made up of restriction fragments that have been transferred to a filter membrane after electrophoresis. The probe DNA anneals and forms stable duplexes with whatever fragments contain sufficiently complementary base sequences, and the positions of these duplexes can be determined by exposing the filter to x-ray film on which radioactive emission (or, in some procedures, light emission) produces an image of the band. Particular DNA sequences can also be amplified without cloning by means of the polymerase chain reaction (PCR), in which short, synthetic oligonucleotides are used as primers to replicate repeatedly and amplify the sequence between them. The primers must flank, and have their 3' ends oriented toward, the region to be amplified, because DNA polymerase can elongate the primers only by the addition of successive nucleotides to the 3' end of the growing chain. Each round of PCR amplification results in a doubling of the number of amplified fragments.

Most genes are present in pairs in the nonreproductive cells of most animals and higher plants. One member of each gene pair is in the chromosome inherited from the maternal parent, and the other member of the gene pair is at a corresponding location (locus) in the homologous chromosome inherited from the paternal parent. A gene can have different forms that correspond to differences in DNA sequence. The different forms of a gene are called alleles. The particular combination of alleles present in an organism constitutes its genotype. The observable characteristics of the organism constitute its phenotype. In an organism, if the two alleles of a gene pair are the same (for example, *AA* or *aa*), then the genotype is homozygous for the *A* or *a* allele; if the alleles are different (*Aa*), then the genotype is heterozygous. Even though each genotype can include at most two alleles, multiple alleles are often encountered among the individuals in natural populations.

DNA polymorphisms (DNA markers) are common in natural populations of most organisms. Among the most widely used DNA polymorphisms are single-nucleotide polymorphisms (SNPs), restriction fragment length polymorphisms (detected by Southern blots), and such PCR-based polymorphisms as random amplified polymorphic DNA (RAPD), amplified fragment length polymorphisms (AFLPs), and simple tandem repeat polymorphisms (STRPs). DNA polymorphisms are used in genetic map-

ping studies to identify DNA markers that are genetically linked to disease genes (genetic risk factors) in the chromosome in order to pinpoint their location. They are also used in DNA typing for identifying individuals, tracking the course of virus and bacterial epidemics, studying human population history, and improving cultivated plants and domesticated animals, as well as for the genetic monitoring of endangered species and for many other purposes.

Key Terms

allele
amplification
amplified fragment length polymorphism (AFLP)
antiparallel
band
base composition
base pairing
base stacking
B form of DNA
blunt end
Chargaff's rules
chromosome
codominant
denaturation
denatured DNA
disease gene
DNA cloning
DNA fingerprinting
DNA marker
DNA polymerase
DNA polymorphism
DNA typing
dominant
5'-P (phosphate) group
gel electrophoresis
gene
genetic linkage
genetic mapping
genetic marker
genomic DNA
genotype
heterozygous
homologous chromosomes
homozygous
hydrogen bond
hydrophobic
kilobase (kb)
locus
major groove
microsatellite
minisatellite
minor groove
monomorphic
multiple alleles
nucleic acid hybridization
nucleoside
nucleotide
oligonucleotide
palindrome
percent G + C
phenotype
phosphodiester bond
polarity
polymerase chain reaction (PCR)
polynucleotide chain
primer
primer adapter
probe
purine
pyrimidine
random amplified polymorphic DNA (RAPD)
relative risk
renaturation
restriction endonuclease
restriction enzyme
restriction fragment
restriction fragment length polymorphism (RFLP)
restriction map
restriction site
risk factor
simple sequence length polymorphism (SSLP)
simple tandem repeat polymorphism (STRP)
single-nucleotide polymorphism (SNP)
Southern blot
sticky end
thermophile
3'-OH (hydroxyl) group
variable number of tandem repeats (VNTR).
Z form of DNA

Review the Basics

- What four bases are commonly found in the nucleotides in DNA? Which form base pairs?
- Which chemical groups are present at the extreme 3′ and 5′ ends of a single polynucleotide strand?
- What does it mean to say that a single strand of DNA strand has a polarity? What does it mean to say that the DNA strands in a duplex molecule are antiparallel?
- What are restriction enzymes and why are they important in the study of particular DNA fragments? What does it mean to say that most restriction sites are palindromes?
- Describe how a Southern blot is carried out. Explain what it used for. What is the role of the probe?
- How does the polymerase chain reaction work? What is it used for? What information about the target sequence must be known in advance? What is the role of the oligonucleotide primers?
- What is a DNA marker? Explain how harmless DNA markers can serve as aids in identifying disease genes through genetic mapping.
- Define and given an example of each of the following key genetic terms: locus, allele, genotype, heterozygous, homozygous, phenotype.

GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- DNA is like **Coca-Cola®**? According to this keyword site, it is. It contains sugar, which is highly soluble in water; phosphate groups, which are of moderate solubility; and bases, which have extremely low solubility. (The base in the soft drink is caffeine, which is chemically similar to adenine and can sometimes be incorporated into DNA, causing a mutation.) As this keyword site emphasizes, the most important property contributing to the stability of double-stranded DNA is the stacking of the base pairs on top of one another as a result of hydrophobic interactions. For further discussion of this feature of DNA, and much else of interest regarding the discovery and analysis of this critical biological macromolecule, consult the keyword site.

- The concept of the polymerase chain reaction (**PCR**) occurred to Kary Mullis one night while cruising on Route 128 from San Francisco to Mendocino. He immediately realized that this approach would be unique in its ability to amplify, at an exponential rate, a specific nucleotide sequence present in a vanishingly small quantity amid a much larger background of total nucleic acid. Once its feasibility was demonstrated, PCR was quickly recognized as a major technical advance in molecular biology. The new technique earned Mullis the 1993 Nobel Prize in chemistry, and today it is the basis of a large number of experimental and diagnostic procedures. At this keyword site you can learn more about the

Guide to Problem Solving

Problem 1 Distinguish between base pairing and base stacking in double-stranded DNA.

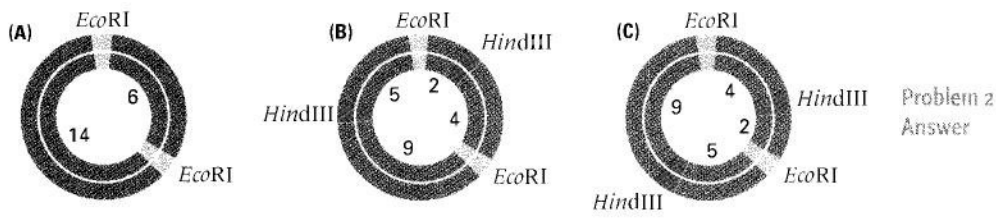
Answer Base pairing is the hydrogen bonding between corresponding bases in opposite strands of duplex DNA; A (adenine) is paired with T (thymine), and G (guanine) is paired with C (cytosine). Base stacking refers to the hydrophobic (water-hating) interaction between consecutive base pairs along a DNA duplex, which promotes the formation of a "stack" of base pairs with the sugar–phosphate backbones of the strands running along outside.

Problem 2 The restriction enzyme *Eco*RI cleaves double-stranded DNA at the sequence 5'-GAATTC-3', and the restriction enzyme *Hind*III cleaves at 5'-AAGCTT-3'. A 20-kilobase (kb) circular plasmid is digested with each enzyme individually and then in combination, and the resulting fragment sizes are determined by means of electrophoresis. The results are as follows:

*Eco*RI alone	fragments of 6 kb and 14 kb
*Hind*III alone	fragments of 7 kb and 13 kb
*Eco*RI and *Hind*III	fragments of 2 kb, 4 kb, 5 kb and 9 kb

How many possible restriction maps are compatible with these data? For each possible restriction map, make a diagram of the circular molecule and indicate the relative positions of the *Eco*RI and *Hind*III restriction sites.

Answer Because the single-enzyme digests give two bands each, there must be two restriction sites for each enzyme in the molecule. Furthermore, because digestion with *Hind*III makes both the 6-kb and the 14-kb restriction fragments disappear, each of these fragments must contain one *Hind*III site. Considering the sizes of the fragments in the double digest, the 6-kb *Eco*RI fragment must be cleaved into 2-kb and 4-kb fragments, and the 14-kb *Eco*RI fragment must be cleaved into 5-kb and 9-kb fragments. Two restriction maps are compatible with the data, depending on which end of the 6-kb *Eco*RI fragment the *Hind*III site is nearest. The position of the remaining *Hind*III site is determined by the fact that the 2-kb and 5-kb fragments in the double digest must be adjacent in the intact molecule in order for a 13-kb fragment to be produced by *Hind*III digestion alone. The accompanying figure shows the relative positions of the



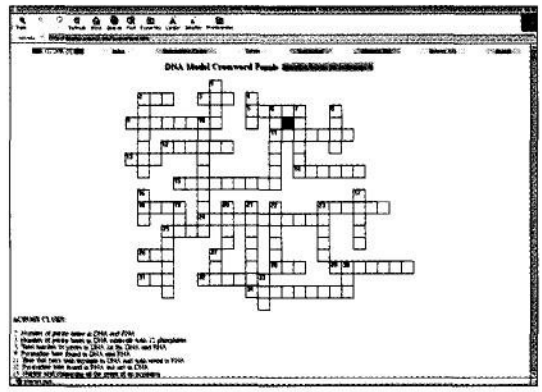
Problem 2 Answer

development of PCR from Mullis's original conception, including two major innovations that were necessary to perfect the process.

- Human beings rely on plants for food, shelter, and medicines. Although at least 5000 species are cultivated, modern agricultural research emphasizes a few widely cultivated crops while largely ignoring plants such as Bambara groundnut (*Vigna subterranea*), breadfruit (*Artocarpus altilis*), carob (*Ceratonia siliqua*), coriander (*Coriandrum sativum*), emmer wheat (*Triticum dicoccum*), oca (*Oxalis tuberosa*), and ulluco (*Ullucus tuberosus*). To learn more about these minor crops and the use of **molecular markers** for characterizing and preserving their genetic diversity, consult this keyword site.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 2 and you will be linked automatically.

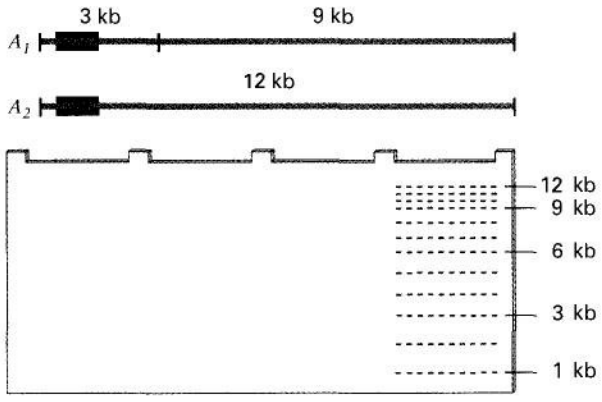
- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 2.



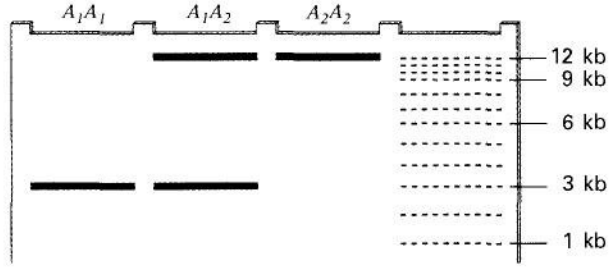
[Courtesy of Wayne's Word and Palomar College.]

*Eco*RI sites (part A). Parts B and C are the two possible restriction maps, which differ according to whether the *Eco*RI site at the top generates the 2-kb or the 4-kb fragment in the double digest.

Problem 3 The accompanying diagram shows the positions of restriction sites (tick marks) for a particular restriction enzyme that can be present in the DNA at a locus in a human chromosome. The DNA present in any particular chromosome may be that shown at the top or that shown at the bottom. A probe DNA binds to the fragments at the position shown by the rectangle. With respect to an RFLP based on these fragments, three genotypes are possible. What are they? Use the symbol A_1 to refer to the allele that yields the upper DNA fragment, and use A_2 to refer to the allele that yields the lower DNA fragment. In the accompanying gel diagram, indicate the genotypes across the top and the phenotype (band position or positions) expected for each genotype. The scale on the right shows the expected positions of fragments ranging in size from 1 to 12 kb.



Answer After cleavage with the restriction enzyme, the A_1-type DNA yields a 3-kb fragment that binds with the probe (yielding a 3-kb band) and a 9-kb fragment that does not bind with the probe (yielding no visible band), whereas the A_2-type DNA yields a 12-kb fragment that binds with the probe (yielding a 12-kb band). A particular chromosome may carry allele A_1 or allele A_2. Because individuals have two copies of each chromoosme (except for the sex chromosomes), any individual may carry A_1A_1, A_1A_2, or A_2A_2. DNA from homozygous A_1A_1 genotypes yields a 3-kb band, that from heterozygous A_1A_2 genotypes yields both a 3-kb and a 12-kb band, and that from homozygous A_2A_2 genotypes yields a 12-kb band. The expected phenotypes are illustrated here.



5'-GTACGGGCAATGGTAATTTTTCAGGAACCAGGGCCCTTAAGCCGTC-3'
3'-CATGCCCGTTACCATTAAAAACTCCTTGGTCCCGGGAATTCGGCAG-5'

Problem 4

Problem 4 A geneticist plans to use the polymerase chain reaction (PCR) to amplify part of the DNA sequence shown above, using oligonucleotide primers that are hexamers matching the regions shown in red. (In practice, hexamers are too short for most purposes.) State the sequence of the primer oligonucleotides that should be used, including the polarity, and give the sequence of the DNA molecule that results from amplification.

Answer The primers must be able to base-pair with the chosen primer sites and must be oriented with their 3' ends facing one another. Thus the "forward primer" (the one that is elongated in a left-to-right direction) should have the sequence 5'-GCAATG-3' and the "reverse primer" (the one that is elongated in a right-to-left direction) should have the sequence 3'-TTCGGC-5'. The resulting amplified sequence is shown below.

5'-GCAATGGTAATTTTTCAGGAACCAGGGCCCTTAAGCCG-3'
3'-CGTTACCATTAAAAACTCCTTGGTCCCGGGAATTCGGC-5'

Analysis and Applications

2.1 Many restriction enzymes produce restriction fragments that have "sticky ends." What does this mean?

2.2 Which of the following sequences are palindromes and which are not? Explain your answer. Symbols such as (A/T) mean that the site may be occupied by (in this case) either A or T, and N stands for any nucleotide.

(a) 5'-AATT-3'
(b) 5'-AAAA-3'
(c) 5'-AANTT-3'
(d) 5'-AA(A/T)AA-3'
(e) 5'-AA(G/C)TT-3'

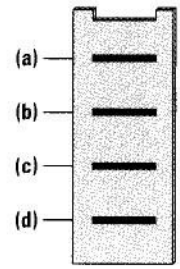
2.3 The following list gives half of each of a set of palindromic restriction sites. What is the complete sequence of each restriction site? (N stands for any nucleotide.)

(a) 5'-AA??-3'
(b) 5'-ATG???-3'
(c) 5''-GGN??-3'
(d) 5'-ATNN??-3'

2.4 Apart from the base sequence, what is different about the ends of restriction fragments produced by the following restriction enzymes? (The downward arrow represents the site of cleavage in each strand.)

(a) *Hae*III (5'-GG ↓ CC-3')
(b) *Mae*I (5'-C ↓ TAG-3')
(c) *Cfo*I (5'-GCG ↓ C-3')

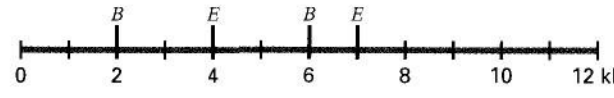
2.5 A solution contains double-stranded DNA fragments of size 3 kb, 6 kb, 9 kb, and 12 kb. They are separated in an electrophoresis gel. In the diagram of the gel at the right, match the fragment sizes with the correct bands.

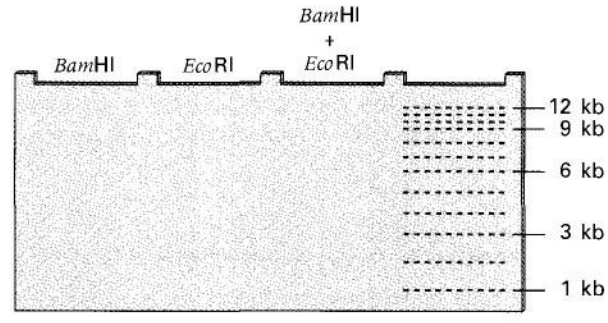


2.6 The linear DNA fragment shown here has cleavage sites for *Bam*HI (*B*) and *Eco*RI (*E*). In the accompanying diagram of an electrophoresis gel, indicate the positions at which bands would be found after digestion with:

(a) *Bam*HI alone
(b) *Eco*RI alone
(c) *Bam*HI and *Eco*RI together

The dashed lines on the right indicate the positions to which bands of 1–12 kb would migrate.

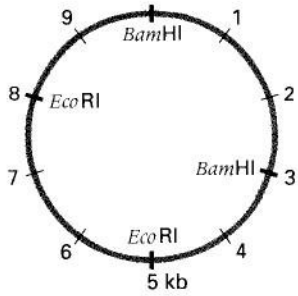


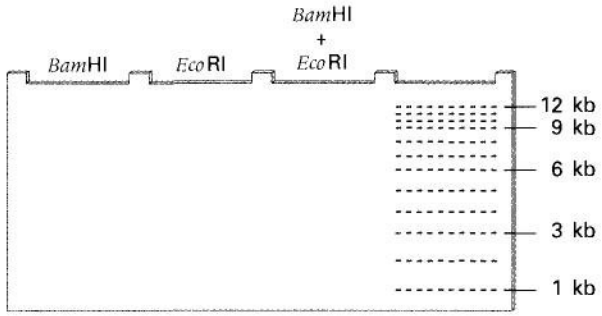


2.7 The circular DNA molecule shown at the top of page 81 has cleavage sites for *Bam*HI and *Eco*RI. In the accompanying diagram of an electrophoresis gel, indicate the positions at which bands would be found after digestion with:

(a) *Bam*HI alone
(b) *Eco*RI alone
(c) *Bam*HI and *Eco*RI together

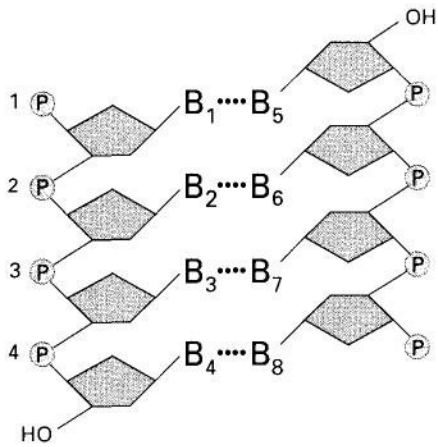
The dashed lines on the right indicate the positions to which bands of 1–12 kb would migrate.





2.8 Consider the accompanying diagram of a region of duplex DNA, in which the B's represent bases in Watson–Crick pairs. Specify as precisely as possible the identity of:

(a) B_5, assuming that B_1 = A
(b) B_6, assuming that B_2 = C
(c) B_7, assuming that B_3 = purine
(d) B_8, assuming that B_4 = A or T



2.9 Refer to the DNA molecule diagrammed in Problem 2.8. In the precursor nucleotides of this molecule, which base was each of the phosphate groups 1–4 associated with?

2.10 In a random sequence consisting of equal proportions of all four nucleotides, what is the probability that a particular short sequence of nucleotides matches a restriction site for:

(a) A restriction enzyme with a 4-base cleavage site?
(b) A restriction enzyme with a 6-base cleavage site?
(c) A restriction enzyme with an 8-base cleavage site?

2.11 In a random sequence consisting of equal proportions of all four nucleotides, what is the average distance between restriction sites for:

(a) A restriction enzyme with a 4-base cleavage site?
(b) A restriction enzyme with a 6-base cleavage site?
(c) A restriction enzyme with an 8-base cleavage site?

2.12 If human DNA were essentially a random sequence of 3×10^9 bp with equal proportions of all four nucleotides (this is an oversimplification), approximately how many restriction fragments would be expected from cleavage with

(a) A "4-cutter" restriction enzyme?
(b) A "6-cutter" restriction enzyme?
(c) An "8-cutter" restriction enzyme?

2.13 Consider the restriction enzymes *Bam*HI (cleavage site 5'-G ↓ GATCC-3') and *Sau*3A (cleavage site 5'-↓ GATC-3'), where the downward arrow denotes the site of cleavage in each strand. Is every *Bam*HI site a *Sau*3A site? Is every *Sau*3A site a *Bam*HI site? Explain your answer.

2.14 A DNA duplex with the sequence shown here is cleaved with *Bam*HI (cleavage site 5'-G ↓ GATCC-3'), where the arrow denotes the site of cleavage in each strand. If the resulting fragments were brought together in the right order, and the breaks in the backbones repaired, what possible DNA duplexes would be expected?

5'-ATTGGATCCAAACCCCAAAGGATCCTTA-3'
3'-TAACCTAGGTTTGGGGTTTCCTAGGAAT-5'

2.15 The restriction enzymes *Pst*I, *Pvu*II, and *Mlu*I have the following restriction sites, where the arrow indicates the site of cleavage in each strand.

*Pst*I 5'-CTGCA ↓ G-3'
*Pvu*II 5'-CAG ↓ CTG-3'
*Mlu*I 5'-A ↓ CGCGT-3'

A DNA duplex with the sequence shown below is digested. What fragments would result from cleavage with:

(a) *Pst*I?
(b) *Pvu*II?
(c) *Mlu*I?

2.16 With regard to the restriction enzymes and the DNA duplex in Problem 2.15, what fragments would result from digestion with

(a) *Pst*I and *Mlu*I?
(b) *Pvu*II and *Mlu*I?

5'-ATGCCCTGCAGTACCATGACGCGTTACGCAGCTGATCGAAACGCGTATATATGCC-3'
3'-TACGGGACGTCATGGTACTGCGCAATGCGTCGACTAGCTTTGCGCATATATACGG-5'

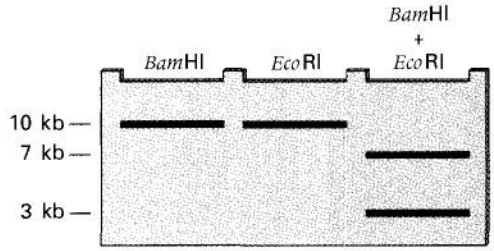
Problems 2.15, 2.16

2.17 Consider the sequence:

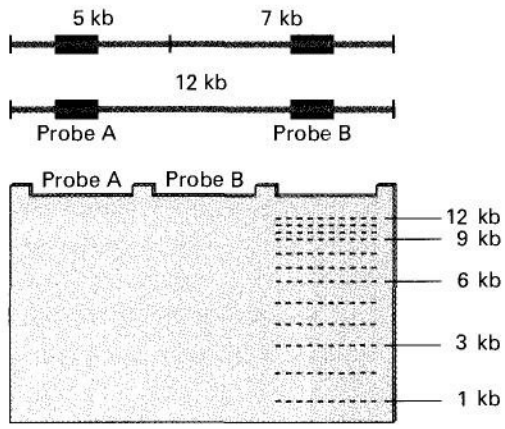
5'-CTGCAGGTG-3'
3'-GACGTCCAC-5'

If this sequence were cleaved with *Pst*I (5'-CTGCA ↓ G-3'), could it still be cleaved with *Pvu*II (5'-CAG ↓ CTG-3')? If it were cleaved with *Pvu*II, could it still be cleaved with *Pst*I? Explain your answer.

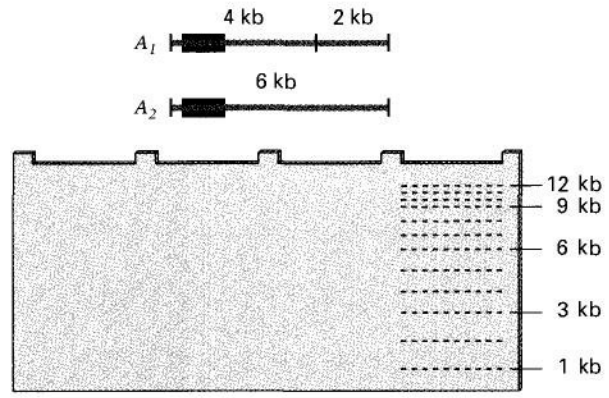
2.18 A circular DNA molecule is cleaved with *Bam*HI, *Eco*RI, or the two restriction enzymes together. The accompanying diagram shows the resulting electrophoresis gel, with the band sizes indicated. Draw a diagram of the circular DNA, showing the relative positions of the *Bam*HI and *Eco*RI sites.



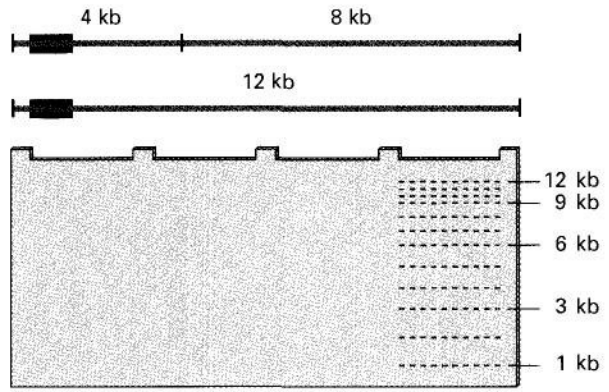
2.19 In the diagrams of DNA fragments shown here, the tick marks indicate the positions of restriction sites for a particular restriction enzyme. A mixture of the two types of molecules is digested and analyzed with a Southern blot using either probe A or probe B, which hybridizes to the fragments where shown by the rectangles. In the accompanying gel diagram, indicate the bands that would result from the use of each of these probes. (The scale on the right shows the expected positions of fragments from 1 to 12 kb.)



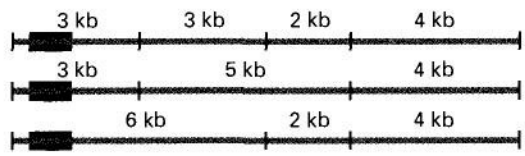
2.20 In the accompanying diagram, the tick marks indicate the positions of restriction sites in two alternative DNA fragments that can be present at the *A* locus in a human chromosome. An RFLP analysis is carried out, using probe DNA that binds to the fragments at the position shown by the rectangle. With respect to this RFLP, how many genotypes are possible? (Use the symbol A_1 to refer to the allele that yields the upper DNA fragment, and use A_2 to refer to the allele that yields the lower DNA fragment.) In the accompanying gel diagram, indicate the genotypes across the top and the phenotype (band position or positions) expected for each genotype. (The scale on the right shows the expected positions of fragments from 1 to 12 kb.)



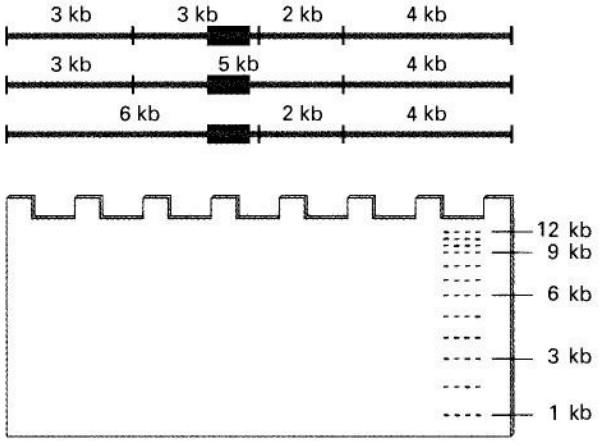
2.21 The accompanying diagram shows the DNA fragments associated with an RFLP revealed by a probe that hybridizes where shown by the rectangle. The tick marks are cleavage sites for the restriction enzyme used in the RFLP analysis. How many alleles does this RFLP have? How many genotypes are possible? In the accompanying gel diagram, indicate the phenotype (pattern of bands) expected of each genotype. (The scale on the right shows the expected positions of fragments from 1 to 12 kb.)



2.22 The thick horizontal lines shown below represent alternative DNA molecules at a particular locus in a human chromosome. The tick marks indicate the positions of restriction sites for a particular restriction enzyme. Genomic DNA from a sample of people is digested and analyzed by a Southern blot using a probe DNA that hybridizes at the position shown by the rectangle. How many possible RFLP alleles would be observed in the sample? How many genotypes?



2.23 The RFLPs described in Problem 2.22 are analyzed with the same restriction enzyme but a different probe, which hybridizes at the site indicated here by the rectangle. How many RFLP alleles would be found? How many genotypes? (Use the symbols A_1, A_2, . . . to indicate the alleles.) In the accompanying gel diagram, indicate the genotypes across the top and the phenotype (band position or positions) expected for each genotype. (The scale on the right shows the expected positions of fragments from 1 to 12 kb.)



2.24 If hexamers were long enough oligonucleotides to serve as specific primers for PCR (for most purposes they are too short), what DNA fragment would be amplified using the "forward" primer pair 5'-AATGCC-3' and the "reverse" primer 3'-GCATGT-5' on the double-stranded DNA molecule shown at the bottom of this page?

2.25 Would the primer pairs 3'-AATGCC-5' and 5'-GCATGT-3' amplify the same fragment described in the previous problem? Explain your answer.

2.26 A human DNA fragment of 3 kb is to be amplified by PCR. The total genome size is 3×10^9 bp.

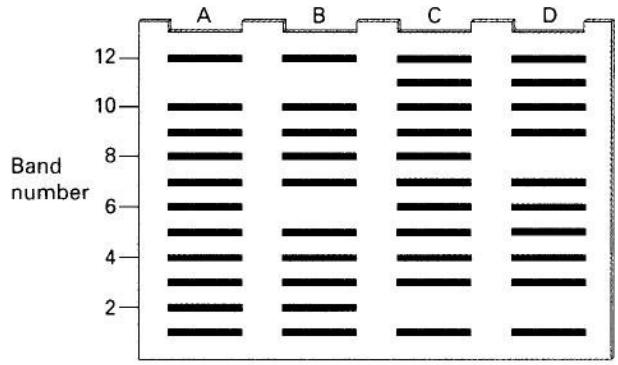
(a) Prior to amplification, what fraction of the total DNA does the target sequence constitute?

(b) What fraction does it constitute after 10 cycles of PCR?

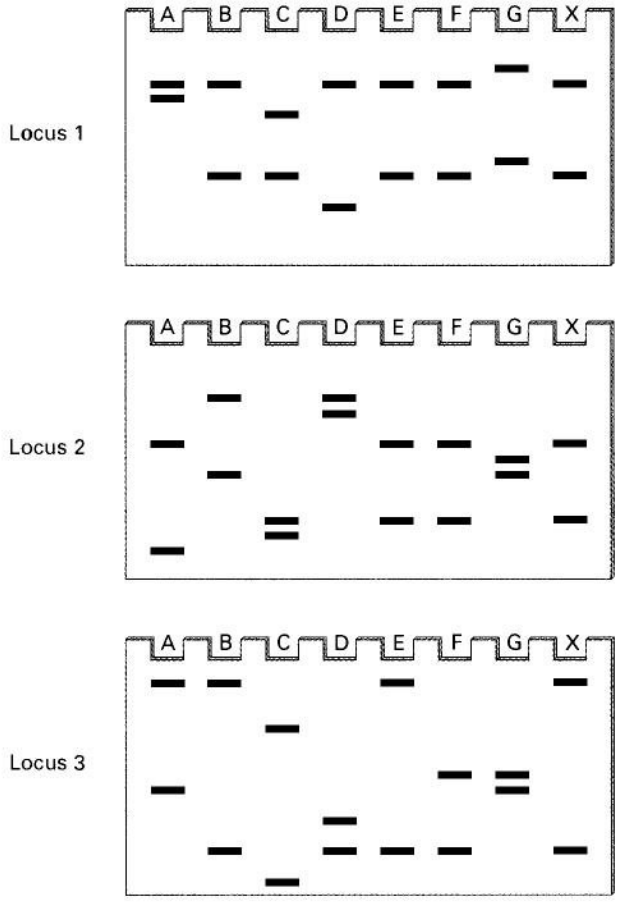
(c) After 20 cycles of PCR?

(d) After 30 cycles of PCR?

2.27 RAPD analysis is carried out using genomic DNA from four individuals (A–D) sampled from a natural population of Hawaiian crickets. The gel shown at the top of the page resulted from PCR with one of the primer pairs tested. Which bands are the RAPD polymorphisms?



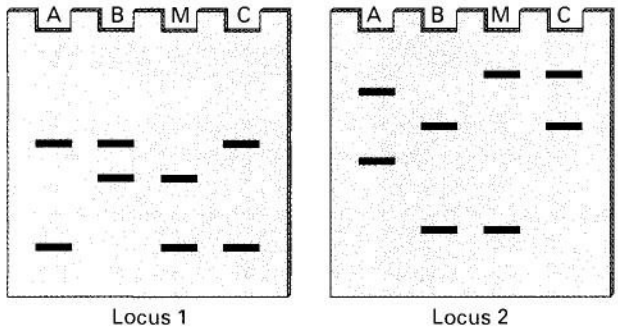
2.28 A cigarette butt found at the scene of a robbery is found to have a sufficient number of epithelial cells stuck to the paper for the DNA to be extracted and typed. Shown below are the results of typing for three probes (locus 1, locus 2, and locus 3) of the evidence (X) and 7 suspects (A through G). Which of the suspects can be excluded? Which cannot be excluded? Can you identify the robber? Explain your reasoning.



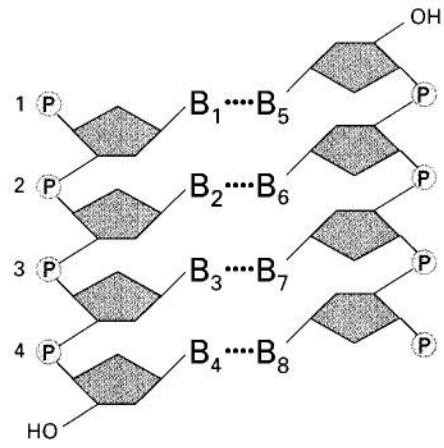
5'-GATTACCGGTAAATGCCGGATTAACCCGGGTTATCAGGCCACGTACAACTGGAGTCC-3'
3'-CTAATGGCCATTTACGGCCTAATTGGGCCCAATAGTCCGGTGCATGTTGACCTCAGG-5'

Problem 2.24

2.29 A woman is uncertain which of two men is the father of her child. DNA typing is carried out on blood from the child (C), the mother (M), and each of the two males (A and B), using probes for a highly polymorphic DNA marker on two different chromosomes ("locus 1" and "locus 2"). The result is shown in the accompanying diagram. Can either male be excluded as the possible father? Explain your reasoning.



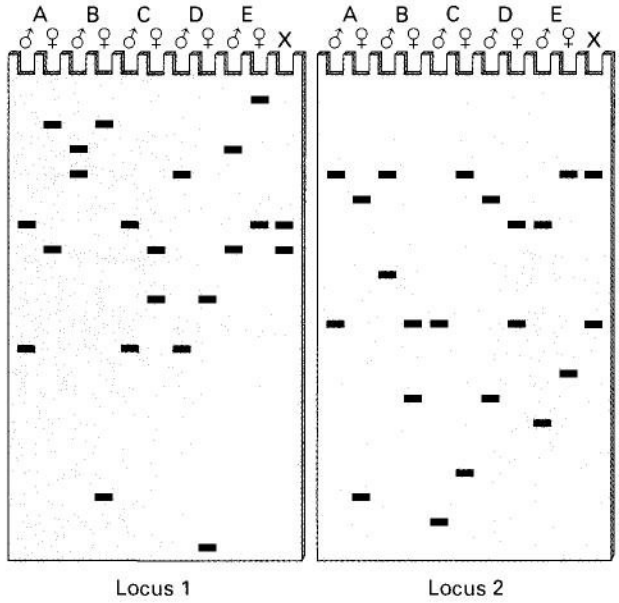
2.30 Snake venom phosphodiesterase cleaves the chemical bonds shown in red in the accompanying diagram, leaving mononucleotides that are phosphorylated in the 3' position. If the phosphates numbered 2 and 4 are radioactive, which mononucleotides will be radioactive after cleavage with snake venom phosphodiesterase?



Challenge Problems

2.31 The genome of *Drosophila melanogaster* is 180×10^6 bp, and a fragment of size 1.8 kb is to be amplified by PCR. How many cycles of PCR are necessary for the amplified target sequence to constitute at least 99 percent of the total DNA?

2.32 A murder victim is found in an advanced state of decomposition and cannot be identified. Police suspect that the victim is one of five persons reported by their parents as missing. DNA typing is carried out on tissues from the victim (X) and on the five sets of parents (A through E), using probes for a highly polymorphic DNA marker on two different chromosomes ("locus 1" and "locus 2"). The result is shown in the diagram at the right. How do you interpret the fact that genomic DNA from each individual yields two bands? Can you identify the parents of the victim? Explain your reasoning.



Problem 2.32

2.33 The snake venom phosphodiesterase enzyme described in Problem 2.30 was originally used in a procedure called "nearest neighbor" analysis. In this procedure, a DNA strand is synthesized in the presence of all four trinucleotides, one of which carries a radioactive phosphate in the α (innermost) position. Then the DNA is digested to completion with snake venom phosphodiesterase, and the resulting mononucleotides are separated and assayed for radioactivity. Examine the diagram in Problem 2.30, and then answer the following questions.

(a) How does this procedure reveal the "nearest neighbors" of the radioactive nucleotide?

(b) Is the "nearest neighbor" on the 5' or the 3' side of the labeled nucleotide?

Further Reading

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314.

Calladine, C. R., and H. Drew. 1997. *Understanding DNA: The Molecule and How it Works*. 2d ed. San Diego: Academic Press.

Cruzan, M. B. 1998. Genetic markers in plant evolutionary ecology. *Ecology* 79: 400.

Danna, K., and D. Nathans. 1971. Specific cleavage of Simian Virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proceedings of the National Academy of Sciences*, USA 68: 2913.

DePamphilis, M. L., ed. 1996. *DNA Replication in Eukaryotic Cells*. Cold Spring Harbor, NY: Cold Spring Harbor Press.

Eeles, R. A. and A. C. Stamps. 1993. *Polymerase Chain Reaction (PCR): The Technique and Its Application*. Austin TX: R. G. Landes.

Frank-Kamenetskii, M. D. 1997. *Unraveling DNA: The Most Important Molecule of Life*. Tr. by L. Liapin. Reading MA: Addison-Wesley.

Hartl, D. L. 2000. *A Primer of Populaton Genetics*. 3d ed. Sunderland, MA: Sinauer.

Jorde, L. B., M. Bamshad, and A. R. Rogers. 1998. Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *Bioessays* 20: 126.

Kumar, L. S. 1999. DNA markers in plant improvement: An overview. *Biotechnology Advances* 17: 143.

Loxdale, H. D., and G. Lushai. 1998. Molecular markers in entomology. *Bulletin of Entomological Research* 88: 577–600.

Mitton, J. B. 1994. Molecular approaches to population biology. *Annual Review of Ecology & Systematics* 25: 45.

Mullis, K. B. 1990. The unusual origin of the polymerase chain reaction. *Scientific American*, April.

Olby, R. C. 1994. *The Path to the Double Helix: The Discovery of DNA*. New York: Dover.

Pena, S. D. J., V. F. Prado, and J. T. Epplen. 1995. DNA diagnosis of human genetic individuality. *Journal of Molecular Medicine* 73: 555.

Sayre, A. 1975. *Rosalind Franklin and DNA*. New York: Norton.

Schafer, A. J., and J. R. Hawkins. 1998. DNA variation and the future of human genetics. *Nature Biotechnology* 16: 33.

Smithies, O. 1995. Early days of electrophoresis. *Genetics* 139: 1.

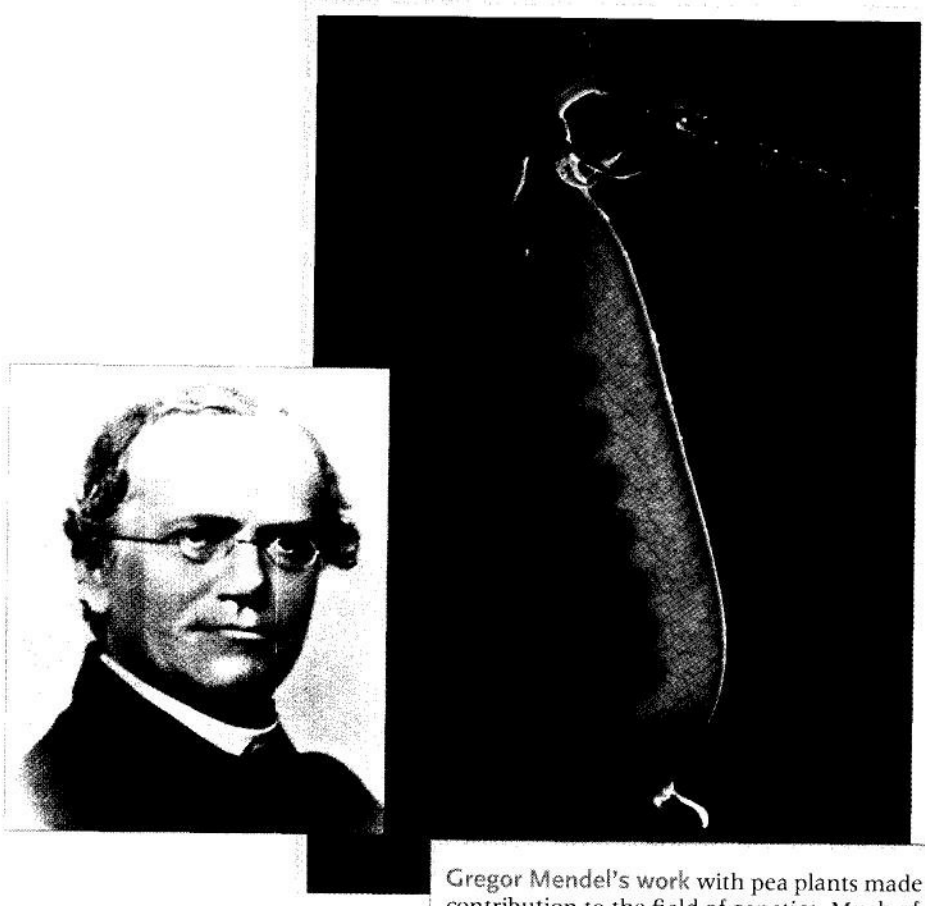
Thomson, G., and M. S. Esposito. 1999. The genetics of complex diseases. *Trends in Genetics* 15: M17.

Wang, D. G., J.-B. Fan, C.-J. Siao, A. Berno, P. Young, R. Sapolsky *et al.* 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280: 1077.

Watson, J. D. 1968. *The Double Helix*. New York: Atheneum.

CHAPTER 3

Transmission Genetics: The Principle of Segregation



Gregor Mendel's work with pea plants made a significant contribution to the field of genetics. Much of that work was carried out during the years 1856–1863, 25 years before scientists discovered a link between heredity and the process of cell division. In 1868 his fellow monks elected him Abbot, and thereafter he had little time to pursue experimental work. [Mendal: Leslie Holzer/Photo Researchers, Inc./Science Source. Pea pod: Richard Gross/Biological Photo.]

CHAPTER OUTLINE

3.1 Morphological and Molecular Phenotypes

3.2 Segregation of a Single Gene
Phenotypic Ratios in the F_2 Generation
The Principle of Segregation
Verification of Segregation
The Testcross and the Backcross

3.3 Segregation of Two or More Genes
The Principle of Independent Assortment
The Testcross with Unlinked Genes
The Big Experiment

3.4 Human Pedigree Analysis
Characteristics of Dominant and Recessive Inheritance
Molecular Markers in Human Pedigrees

3.5 Pedigrees and Probability
Mutually Exclusive Possibilities
Independent Possibilities

3.6 Incomplete Dominance and Epistasis
Multiple Alleles
Human ABO Blood Groups
Epistasis

3.7 Genetic Analysis: Mutant Screens and the Complementation Test
The Complementation Test in Gene Identification
Why Does the Complementation Test Work?

PRINCIPLES

- Inherited traits are determined by the genes present in the reproductive cells united in fertilization.
- Genes are usually inherited in pairs—one from the mother and one from the father.
- The genes in a pair may differ in DNA sequence and in their effect on the expression of a particular inherited trait.
- The maternally and paternally inherited genes are not changed by being together in the same organism.
- In the formation of reproductive cells, the paired genes separate again into different cells.
- Random combinations of reproductive cells containing different genes result in Mendel's ratios of traits appearing among the progeny.
- The ratios actually observed for any trait are determined by the types of dominance and gene interaction.
- In genetic analysis, the complementation test is used to determine whether two recessive mutations that cause a similar phenotype are alleles of the same gene. The mutant parents are crossed, and the phenotype of the progeny is examined. If the progeny phenotype is nonmutant (complementation occurs), the mutations are in different genes; if the progeny phenotype is mutant (complementation does not occur), the mutations are in the same gene.

CONNECTIONS

What Did Gregor Mendel Think He Discovered?
Gregor Mendel 1866
Experiments on Plant Hybrids

Troubadour
The Huntington's Disease Collaborative Research Group 1993
A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes

In Chapter 2 we emphasized that in modern genetics, a typical phenotype consists of a band present at a particular position in a DNA electrophoresis gel. In a method of genetic analysis such as RAPD (random amplified polymorphic DNA), genomic DNA from a single individual can yield 30 or more bands (see Figure 2.26). Each of these bands represents a phenotype and results from the PCR (polymerase chain reaction) amplification of a single region of DNA in the genome of the individual.

In this chapter we consider how genes are transmitted from parents to offspring and how this determines the distribution of genotypes and phenotypes among related individuals. The study of the inheritance of traits constitutes **transmission genetics.** This subject is also called **Mendelian genetics** because the underlying principles were first deduced from experiments in garden peas (*Pisum sativum*) carried out in the years 1856–1863 by Gregor Mendel, a monk at the monastery of St. Thomas in the town of Brno (Brünn), in the Czech Republic. He reported his experiments to a local natural history society, published the results and his interpretation in its scientific journal in 1866, and began exchanging letters with one of the leading botanists of the time. His experiments were careful and exceptionally well documented, and his paper contains the first clear exposition of the statistical rules governing the transmission of genes from generation to generation. Nevertheless, Mendel's paper was ignored for 34 years until its significance was finally appreciated.



In this small garden plot adjacent to the monastery of St. Thomas, Gregor Mendel grew more than 33,500 pea plants over a period of eight years, including more than 6400 plants in one year alone. He received some help from two fellow monks who assisted in the experiments.

3.1 Morphological and Molecular Phenotypes

Geneticists study traits in which one organism differs from another in order to discover the genetic basis of the difference. Until the advent of molecular genetics, geneticists dealt mainly with morphological traits, in which the differences between organisms can be expressed in terms of color, shape, or size. Mendel studied seven morphological traits contrasting in seed shape, seed color, flower color, pod shape, and so forth (**Figure 3.1**). Perhaps the most widely known example of a contrasting Mendelian trait is round versus wrinkled seeds. As pea seeds dry, they lose water and shrink. Round seeds are round because they shrink uniformly; wrinkled seeds are wrinkled because they shrink irregularly. The wrinkled phenotype is due to the absence of a branched-chain form of starch known as *amylopectin*, which is not synthesized in wrinkled seeds owing to a defect in the enzyme starch-branching enzyme I (SBEI).

The nonmutant, or **wildtype,** allele of the gene for SBEI is designated *W* and the mutant allele *w*. Seeds that are heterozygous *Ww* have only half as much SBEI enzyme as wildtype homozygous *WW* seeds, but this half the normal amount of enzyme produces enough amylopectin for the heterozygous *Ww* seeds to shrink uniformly and remain phenotypically round. Hence, with respect to seed shape, the wildtype *W* allele is dominant over the mutant *w* allele. Because the *w* allele is recessive, only the homozygous *ww* seeds become wrinkled.

The molecular basis of the wrinkled mutation is that the SBEI gene has become interrupted by the insertion, into the gene, of a DNA sequence called a **transposable element.** Such DNA sequences are capable of moving (*transposition*) from one location to another within a chromosome or between chromosomes. The molecular mechanism of transposition is discussed in Chapter 7, but for our present purposes it is necessary to know only that transposable elements are present in most genomes, especially the large genomes of eukaryotes, and that many spontaneous mutations result from the insertion of transposable elements into a gene.

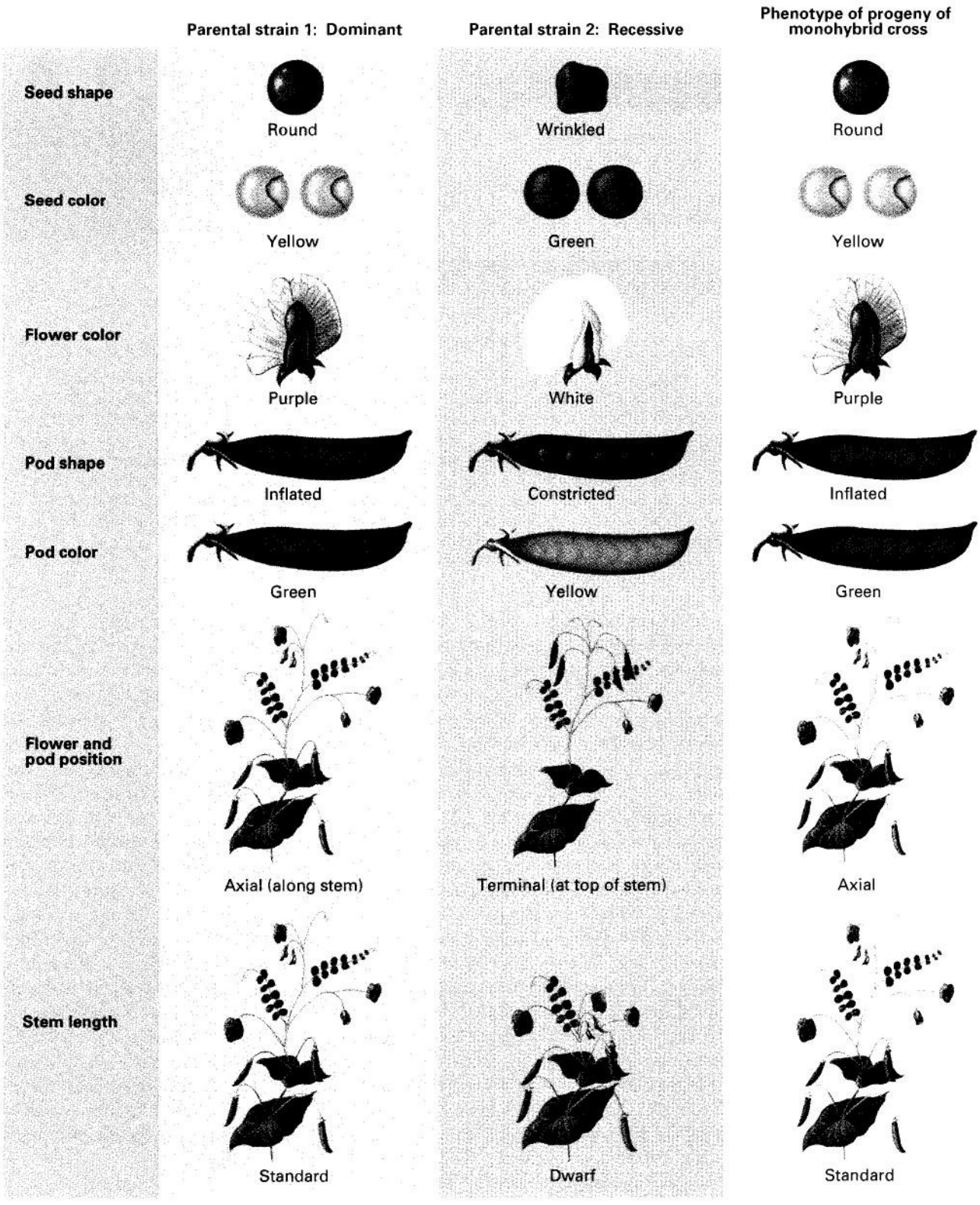


Figure 3.1 The seven different traits studied in peas by Mendel. The phenotype shown at the far right is the dominant trait, which appears in the hybrid produced by crossing.

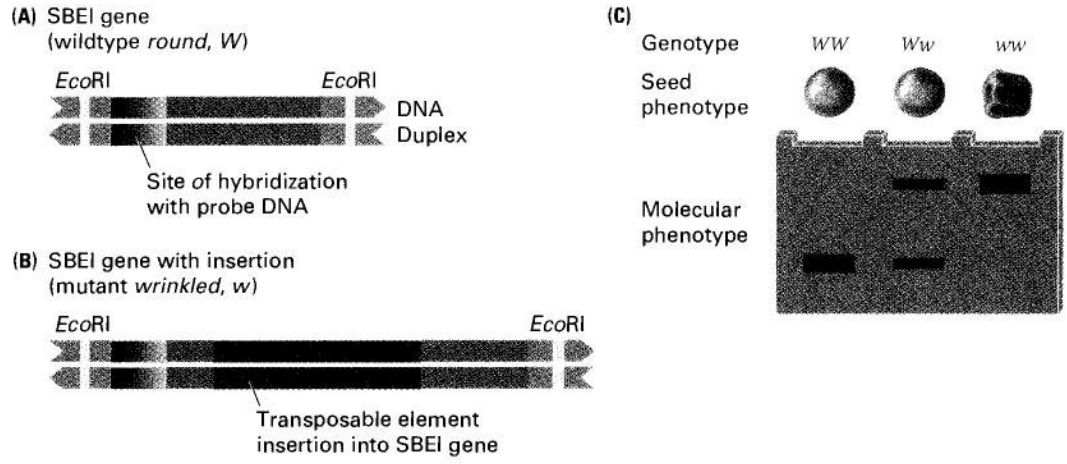


Figure 3.2 (A) *W* (round) is an allele of a gene that specifies the amino acid sequence of starch branching enzyme I (SBEI). (B) *w* (wrinkled) is an allele that encodes an inactive form of the enzyme because its DNA sequence is interrupted by the insertion of a transposable element. (C) At the level of the morphological phenotype, *W* is dominant to *w:* Genotype *WW* and *Ww* have round seeds, whereas genotype *ww* has wrinkled seeds. The molecular difference between the alleles can be detected as a restriction fragment length polymorphism (RFLP) using the enzyme *Eco*RI and a probe that hybridizes at the site shown. At the molecular level, the alleles are codominant: DNA from each genotype yields a different molecular phenotype—a single band differing in size for homozygous *WW* and *ww,* and both bands for heterozygous *Ww.*

Figure 3.2 includes a diagram of the DNA structure of the wildtype *W* and mutant *w* alleles and shows the DNA insertion that interrupts the *w* allele. Highlighted are two *Eco*RI restriction sites, present in both alleles, that flank the site of the insertion. The diagram in part C indicates what pattern of bands would be expected if one were to carry out a Southern blot (Section 2.3) in which genomic DNA was digested to completion with *Eco*RI and then the resulting fragments were separated by electrophoresis and hybridized with a labeled probe complementary to a region shared between the *W* and *w* alleles. The *Eco*RI fragment from the *W* allele would be smaller than that of the *w* allele, because of the inserted DNA in the *w* allele, and thus it would migrate faster than the corresponding fragment from the *w* allele and move to a position closer to the bottom of the gel. Genomic DNA from homozygous *WW* would yield a single, faster-migrating band; that from homozygous *ww* a single, slower-migrating band; and that from heterozygous *Ww* two bands with the same electrophoretic mobility as those observed in the homozygous genotypes. In Figure 3.2C, the band in the homozygous genotypes is shown as somewhat thicker than those from the heterozygous genotype, because the single band in each homozygous genotype comes from the two copies of whichever allele is homozygous, and thus it contains more DNA than the corresponding DNA in the heterozygous genotype, in which only one copy of each allele is present.

Hence, as illustrated in Figure 3.2C, the RFLP analysis clearly distinguishes between the genotypes *WW, Ww,* and *ww,* because the heterozygous *Ww* genotype exhibits both of the bands observed in the homozygous genotypes. This situation is described by saying that the *W* and *w* alleles are **codominant** with respect to the molecular phenotype. However, as indicated by the seed shapes in Figure 3.2C, *W* is dominant over *w* with respect to morphological phenotype. In the discussion that follows, we use the RFLP analysis of the *W* and *w* alleles to emphasize the importance of molecular phenotypes in modern genetics and to demonstrate experimentally the principles of genetic transmission.

3.2 Segregation of a Single Gene

Mendel selected peas for his experiments for two primary reasons. First, he had access to varieties that differed in contrasting traits, such as round versus wrinkled seeds

and yellow versus green seeds. Second, his earlier studies had indicated that peas usually reproduce by self-pollination, in which pollen produced in a flower is used to fertilize the eggs in the same flower. To produce hybrids by cross-pollination (**outcrossing**), all he needed to do was open the keel petal (enclosing the reproductive structures), remove the immature anthers (the pollen-producing structures) before they shed pollen, and dust the stigma (the female structure) with pollen taken from a flower on another plant (**Figure 3.3**). The relatively small space needed to grow each plant, and the relatively large number of progeny that could be obtained, gave him the opportunity, as he says in his paper, to "determine the number of different forms in which hybrid progeny appear" and to "ascertain their numerical interrelationships."

The fact that garden peas are normally self-fertilizing means that in the absence of deliberate outcrossing, plants with contrasting traits are usually homozygous for alternative alleles of a gene affecting the trait; for example, plants with round seeds have genotype *WW* and those with wrinkled seeds genotype *ww*. The homozygous genotypes are indicated experimentally by the observation that hereditary traits in each

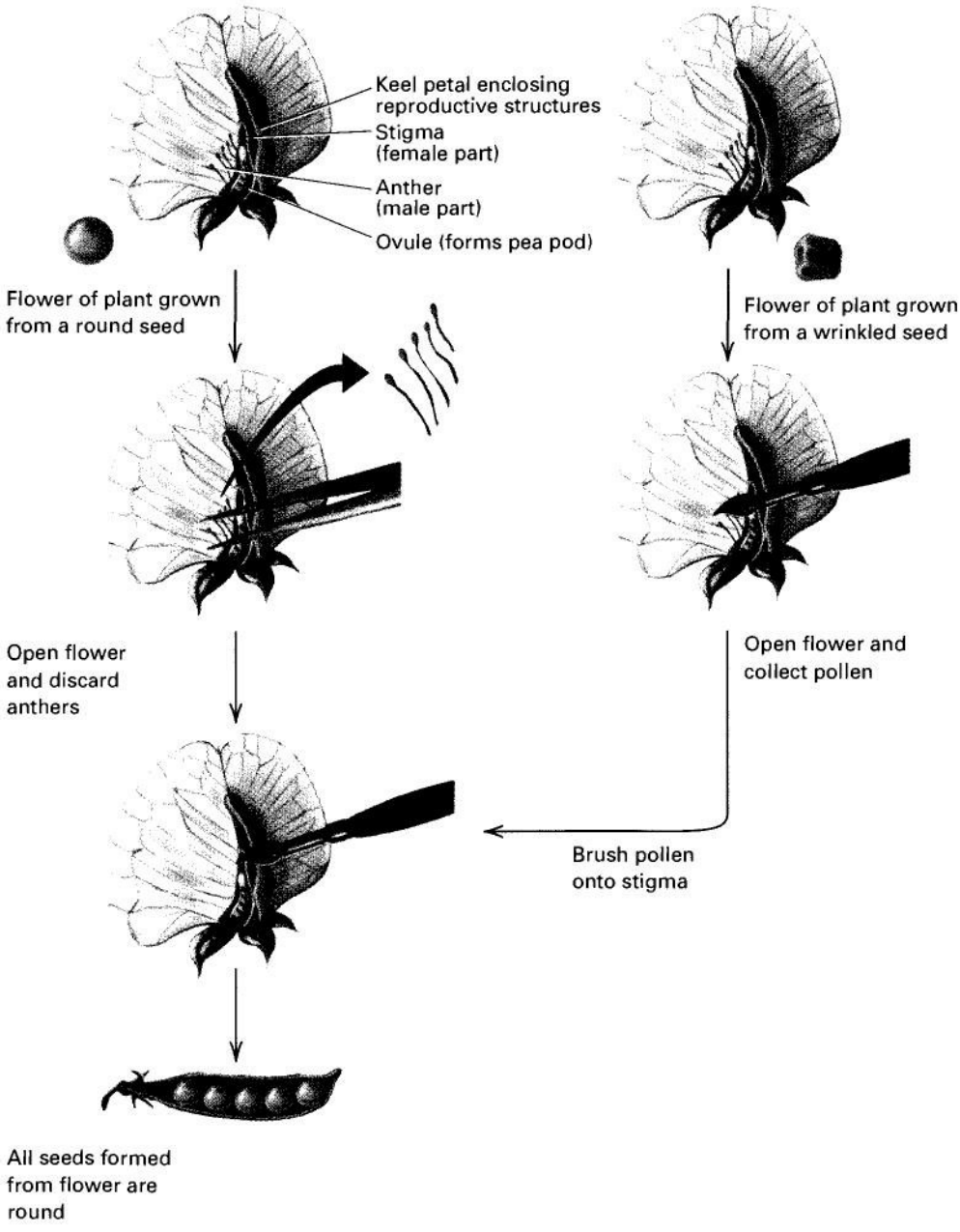


Figure 3.3 Crossing pea plants requires some minor surgery in which the anthers of a flower are removed before they produce pollen. The stigma, or female part of the flower, is not removed. It is fertilized by brushing with mature pollen grains taken from another plant.

variety are **true-breeding,** which means that plants produce only progeny like themselves when allowed to self-pollinate normally. Outcrossing between plants that differ in one or more traits creates a **hybrid.** If the parents differ in one, two, or three traits, the hybrid is a *monohybrid, dihybrid,* or *trihybrid.* In keeping track of parents and their hybrid progeny, we say that the parents constitute the **P_1 generation** and their hybrid progeny the **F_1 generation.**

Because garden peas are sexual organisms, each cross can be performed in two ways, depending on which phenotype is present in the female parent and which in the male parent. With round versus wrinkled, for example, the female parent can be round and the male wrinkled, or the reverse. These are called **reciprocal crosses.** Mendel was the first to demonstrate the following important principle:

> The outcome of a genetic cross does not depend on which trait is present in the male and which is present in the female; reciprocal crosses yield the same result.

This principle is illustrated for round and wrinkled seeds in **Figure 3.4.** The gel icons show the RFLP bands that genomic DNA from each type of seed in these crosses would yield. The genotypes of the crosses and their progeny are as follows:

Cross A: *WW* ♀ × *ww* ♂ → *Ww*
Cross B: *ww* ♀ × *WW* ♂ → *Ww*

In both reciprocal crosses, the progeny have the morphological phenotype of round seeds, but as shown by the RFLP diagrams on the right, all progeny genotypes are actually heterozygous *Ww* and therefore different from either parent. The genetic equivalence of reciprocal crosses illustrated in Figure 3.4 is a principle that is quite general in its applicability, but there is an important exception, having to do with sex chromosomes, that will be discussed in Chapter 4.

In the following section we examine a few of Mendel's original experiments in the context of RFLP analysis in order to relate the morphological phenotypes and their

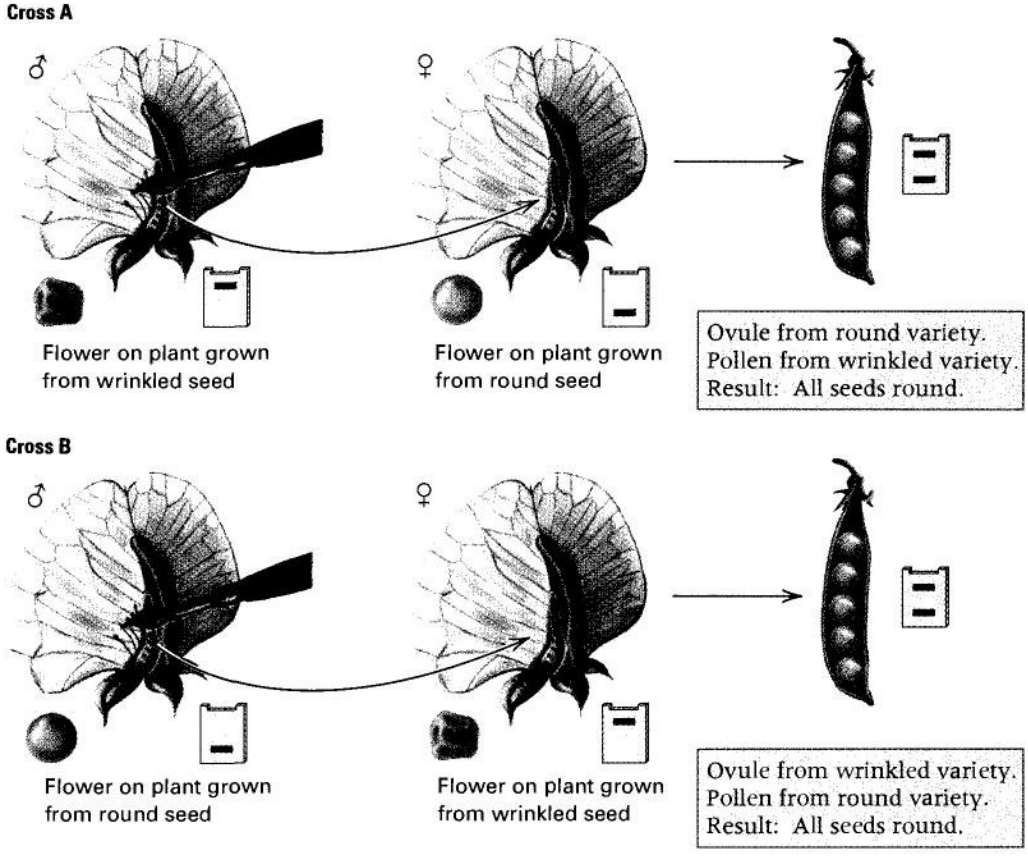
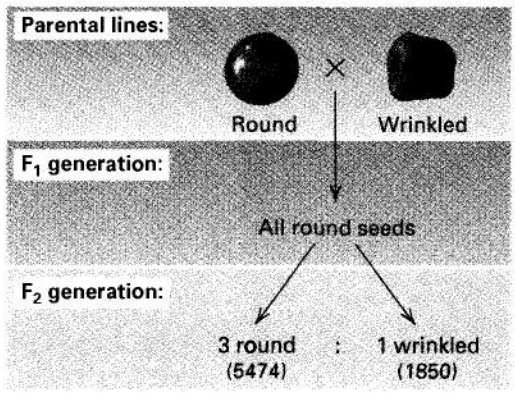


Figure 3.4 Morphological and molecular phenotypes showing the equivalence of reciprocal crosses. In this example, the hybrid seeds are round and yield an RFLP pattern with two bands, irrespective of the direction of the cross.

ratios to the molecular phenotypes that would be expected.

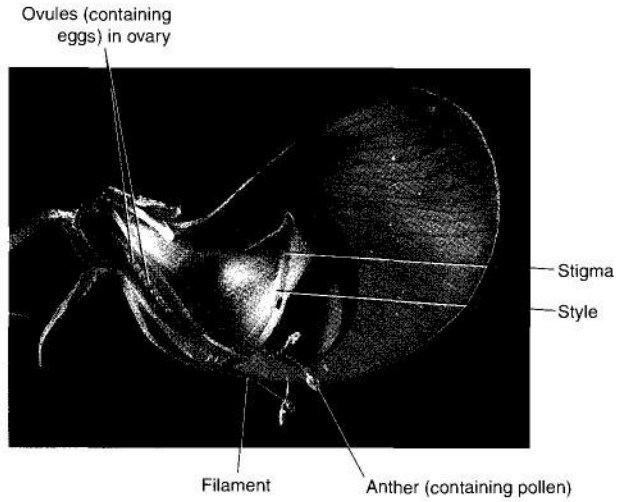
Phenotypic Ratios in the F_2 Generation

Although the progeny of the crosses in Figure 3.4 have the dominant phenotype of round seeds, the RFLP analysis shows that they are actually heterozygous. The *w* allele is hidden with respect to the morphological phenotype because *w* is recessive to *W*. Nevertheless, the wrinkled phenotype reappears in the next generation when the hybrid progeny are allowed to undergo self-fertilization. For example, if the round F_1 seeds from Cross A in Figure 3.4 were grown into sexually mature plants and allowed to undergo self-fertilization, some of the resulting seeds would be round and others wrinkled. The progeny seeds produced by self-fertilization of the F_1 generation constitute the **F_2 generation.** When Mendel carried out this experiment, in the F_2 generation he observed the results shown in the following diagram:



Note that the ratio 5474 : 1850 is approximately 3 : 1.

A 3 : 1 ratio of dominant : recessive forms in the F_2 progeny is characteristic of simple Mendelian inheritance. Mendel's data demonstrating this point are shown in Table 3.1. Note that the first two traits in the table (round versus wrinkled seeds and yellow versus green seeds) have many more observations than any of the others. The reason is that these traits can be classified directly in the seeds, whereas the others can be classified only in the mature plants, and Mendel could analyze many more seeds than he could mature plants. The



The anatomy of a pea plant makes it easy to work with. Mendal chose the pea plant because it is small, easy to grow, and has a short generation time. [Jean-Michel Labat/Ardea.]

principal observations from the data in Table 3.1 are

- The F_1 hybrids express only the dominant trait (because the F_1 progeny are heterozygous—for example, *Ww*)
- In the F_2 generation, plants with either the dominant or the recessive trait are present (which means that some F_2 progeny are homozygous—for example, *ww*).
- In the F_2 generation, there are approximately three times as many plants with the dominant phenotype as plants with the recessive phenotype.

The 3 : 1 ratio observed in the F_2 generation is the key to understanding the mechanism of genetic transmission. In the next section we use RFLP analysis of the *W* and *w* alleles to explain why this ratio is produced.

Table 3.1 Results of Mendel's monohybrid experiments

Parental traits	F_1 trait	Number of F_2 progeny	F_2 ratio
Round × wrinkled (seeds)	Round	5474 round, 1850 wrinkled	2.96 : 1
Yellow × green (seeds)	Yellow	6022 yellow, 2001 green	3.01 : 1
Purple × white (flowers)	Purple	705 purple, 224 white	3.15 : 1
Inflated × constricted (pods)	Inflated	882 inflated, 299 constricted	2.95 : 1
Green × yellow (unripe pods)	Green	428 green, 152 yellow	2.82 : 1
Axial × terminal (flower position)	Axial	651 axial, 207 terminal	3.14 : 1
Long × short (stems)	Long	787 long, 277 short	2.84 : 1

The Principle of Segregation

The 3 : 1 ratio can be explained with reference to **Figure 3.5**. This is the heart of Mendelian genetics. You should master it and be able to use it to deduce the progeny types produced in crosses. The diagram illustrates these key features of single-gene inheritance:

1. Genes come in pairs, which means that a cell or individual has *two* copies (alleles) of each gene.
2. For each pair of genes, the alleles may be identical (homozygous *WW* or homozygous *ww*), or they may be different (heterozygous *Ww*).
3. Each reproductive cell (**gamete**) produced by an individual contains only *one* allele of each gene (that is, either *W* or *w*).
4. In the formation of gametes, any particular gamete is equally likely to include either allele (hence, from a heterozygous *Ww* genotype, half the gametes contain *W* and the other half contain *w*).
5. The union of male and female reproductive cells is a random process that reunites the alleles in pairs.

The essential feature of transmission genetics is the separation, technically called **segregation,** in unaltered form, of the two alleles in an individual during the formation of its reproductive cells. Segregation

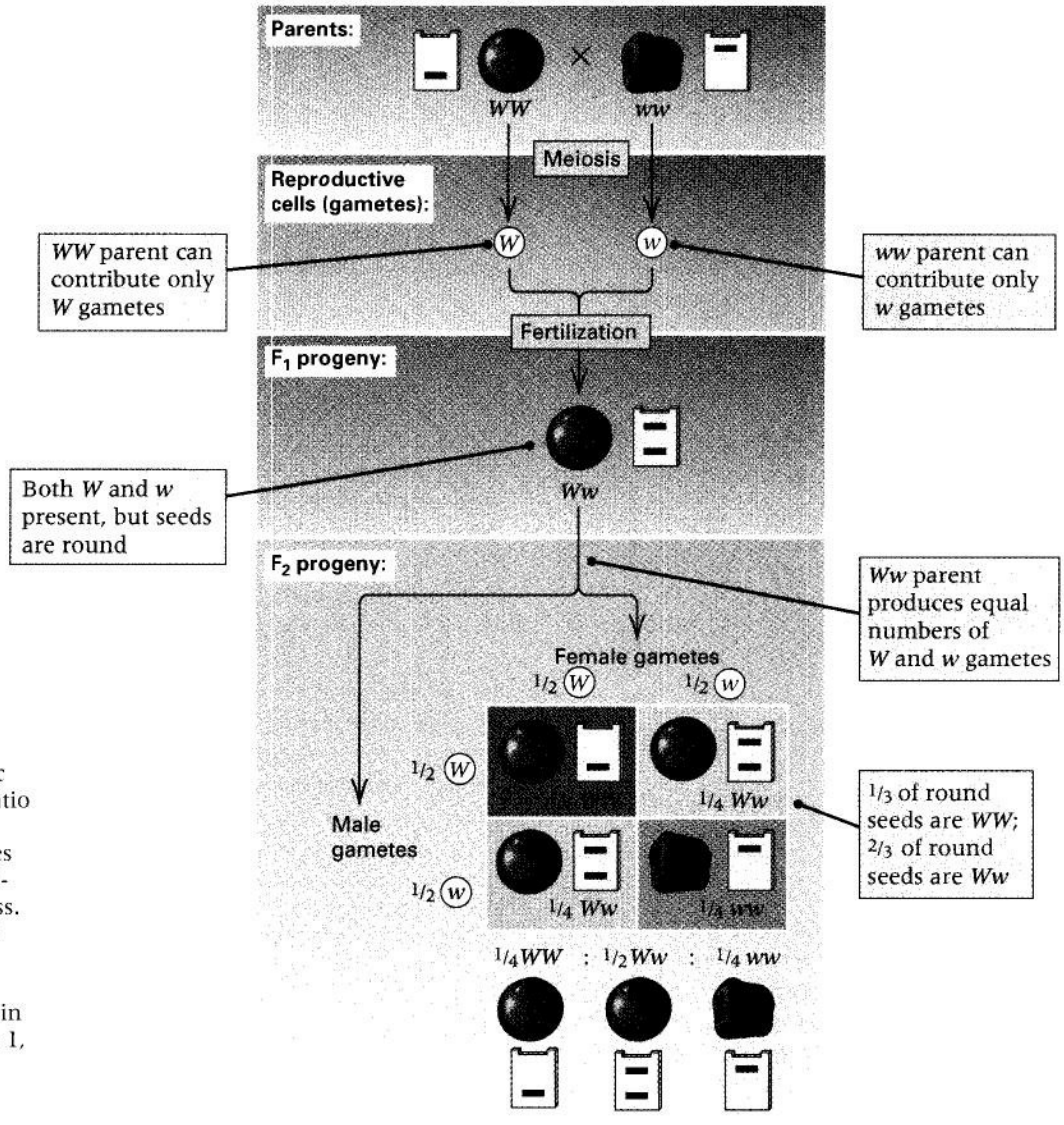


Figure 3.5 A diagrammatic explanation of the 3 : 1 ratio of dominant : recessive morphological phenotypes observed in the F_2 generation of a monohybrid cross. The 3 : 1 ratio is observed because of dominance. Note that the ratio of *WW* : *Ww* : *ww* genotypes in the F_2 generation is 1 : 2 : 1, as can be seen from the restriction fragment phenotypes.

corresponds to points 3 and 4 in the foregoing list. The principle of segregation is sometimes called *Mendel's first law.*

> **The Principle of Segregation:** In the formation of gametes, the paired hereditary determinants separate (segregate) in such a way that each gamete is equally likely to contain either member of the pair.

Another key feature of transmission genetics is that the hereditary determinants are present as pairs in both the parental organisms and the progeny organisms but as single copies in the reproductive cells. This feature corresponds to points 1 and 5 in the foregoing list.

Figure 3.5 illustrates the biological mechanism underlying the important Mendelian ratios in the F_2 generation of 3 : 1 for phenotypes and 1 : 2 : 1 for genotypes. To understand these ratios, consider first the parental generation in which the original cross is *WW* × *ww*. The sex of the parents is not stated because reciprocal crosses yield identical results. (There is, however, a convention in genetics that unless otherwise specified, crosses are given with the female parent listed first.) In the original cross, the *WW* parent produces only *W*-containing gametes, whereas the *ww* parent produces only *w*-containing gametes. Segregation still takes place in the homozygous genotypes as well as in the heterozygous genotype, even though all of the gametes carry the same type of allele (*W* from homozygous *WW* and *w* from homozygous *ww*). When the *W*-bearing and *w*-bearing gametes come together in fertilization, the hybrid genotype is heterozygous *Ww*, which is shown by the bands in the gel icon next to the F_1 progeny. With regard to seed shape, the hybrid *Ww* seeds are round because *W* is dominant over *w*.

When the heterozygous F_1 progeny form gametes, segregation implies that half the gametes will contain the *W* allele and the other half will contain the *w* allele. These gametes come together at random when an F_1 individual is self-fertilized or when two F_1 individuals are crossed. The result of random fertilization can be deduced from the sort of cross-multiplication square shown at the bottom of the figure, in which the female gametes and their frequencies are arrayed across the top margin and the male gametes and their frequencies along the left-hand margin. This calculating device is widely used in genetics and is called a **Punnett square** after its inventor Reginald C. Punnett (1875–1967). The Punnett square in Figure 3.5 shows that random combinations of the F_1 gametes result in an F_2 generation with the genotypic composition 1/4 *WW*, 1/2 *Ww*, and 1/4 *ww*. This can be confirmed by the RFLP banding patterns in the gel icons because of the codominance of *W* and *w* with respect to the molecular phenotype. But because *W* is dominant over *w* with respect to the morphological phenotype, the *WW* and *Ww* genotypes have round seeds and the *ww* genotypes have wrinkled seeds, yielding the phenotypic ratio of round : wrinkled seeds of 3 : 1. Hence it is a combination of segregation, random union of gametes, and dominance that results in the 3 : 1 ratio.

The ratio of F_2 genotypes is as important as the ratio of F_2 phenotypes. The Punnett square in Figure 3.5 also shows that the ratio of *WW* : *Ww* : *ww* genotypes is 1 : 2 : 1, which can be confirmed directly by the RFLP analysis.

Verification of Segregation

The round seeds in Figure 3.5 conceal a genotypic ratio of 1 *WW* : 2 *Ww*. To say the same thing in another way, among the F_2 seeds that are round (or, more generally, among organisms that show the dominant morphological phenotype), 1/3 are homozygous (in this example, *WW*) and 2/3 are heterozygous (in this example, *Ww*). This conclusion is obvious from the RFLP patterns in Figure 3.5, but it is not at all obvious from the morphological phenotypes. Unless you knew something about genetics already, it would be a very bold hypothesis, because it implies that two organisms with the same morphological phenotype (in this case round seeds) might nevertheless differ in molecular phenotype and in genotype.

Yet this is exactly what Mendel proposed. But how could this hypothesis be tested experimentally? He realized that it could be tested via self-fertilization of the F_2 plants. With self-fertilization, plants grown

CoNnection

What Did Gregor Mendel Think He Discovered?

Gregor Mendel 1866
Monastery of St. Thomas, Brno
[then Brünn], Czech Republic
Experiments on Plant Hybrids
(original in German)

Mendel's paper is remarkable for its precision and clarity. It is worth reading in its entirety for this reason alone. Although the most important discovery attributed to Mendel is segregation, he never uses this term. His description of segregation is found in the first passage in italics in the excerpt. (All of the italics are reproduced from the original.) In his description of the process, he takes us carefully through the separation of A *and* a *in gametes and their coming together again at random in fertilization. One flaw in the description is Mendel's occasional confusion between genotype and phenotype, which is illustrated by his writing* A *instead of* AA *and* a *instead of* aa *in the display toward the end of the passage. Most early geneticists made no consistent distinction between genotype and phenotype until 1909, when the terms themselves were coined.*

Artificial fertilization undertaken on ornamental plants to obtain new color variants initiated the experiments reported here. The striking regularity with which the same hybrid forms always reappeared whenever fertilization between like species took place suggested further experiments whose task it was to follow that development of hybrids in their progeny. . . . This paper discusses the attempt at such a detailed experiment. . . . Whether the plan by which the individual experiments were set up and carried out was adequate to the assigned task should be decided by a benevolent judgment. . . . [Here the experimental results are described in detail.] Thus experimentation also justifies the assumption *that pea hybrids form germinal and pollen cells that in their composition correspond in equal numbers to all the constant forms resulting from the combination of traits united through fertilization.* The difference of forms among the progeny of hybrids, as well as the ratios in which they are observed, find an adequate explanation in the principle [of segregation] just deduced. The simplest case is given by the series for *one pair of differing traits.* It is shown that this series is described by the expression: $A + 2Aa + a$, in which A and a signify the forms with constant differing traits, and Aa the form hybrid for both. The series contains four individuals in three different terms. In their production, pollen and germinal

Whether the plan by which the individual experiments were set up and carried out was adequate to the assigned task should be decided by a benevolent judgment.

from the homozygous *WW* genotypes should be true-breeding for round seeds, whereas those from the heterozygous *Ww* genotypes should yield round and wrinkled seeds in the ratio of 3 : 1. On the other hand, the plants grown from wrinkled seeds should be true-breeding for wrinkled because these plants are homozygous *ww*. The results Mendel obtained are summarized in **Figure 3.6**. As predicted from the genetic hypothesis, plants grown from F_2 wrinkled seeds were true-breeding for wrinkled seeds, yielding only wrinkled seeds in the F_3 generation. But some of the plants grown from round seeds showed evidence of segregation. Among 565 plants grown from round F_2 seeds, 372 plants produced both round and wrinkled seeds in a proportion very close to 3 : 1, whereas the remaining 193 plants produced only round seeds in the F_3 generation. The ratio 193 : 372 equals 1 : 1.93, which is very close to the ratio 1 : 2 of *WW* : *Ww* genotypes predicted from the genetic hypothesis.

An important feature of the homozygous round and homozygous wrinkled seeds produced in the F_2 and F_3 generations is that the phenotypes are exactly the same as those observed in the original parents in the P_1 generation. This makes sense in terms of DNA, because the DNA of each allele remains unaltered unless a new mutation happens to occur. Mendel described this result in a letter by saying that in the progeny of crosses, "the two parental traits appear, separated and unchanged, and there is nothing to indicate that one of them has either inherited or taken over anything from the other." From this finding, he concluded that the hereditary determinants for

cells of form *A* and *a* participate, on the average, equally in fertilization; therefore each form manifests itself twice, since four individuals are produced. Participating in fertilization are thus:

Pollen cells $A + A + a + a$
Germinal cells $A + A + a + a$

It is entirely a matter of chance which of the two kinds of pollen combines with each single germinal cell. However, according to the laws of probability, in an average of many cases it will always happen that every pollen form *A* and *a* will unite equally often with every germinal-cell form *A* and *a;* therefore, in fertilization, one of the two pollen cells *A* will meet a germinal cell *A*, the other a germinal cell *a*, and equally, one pollen cell *a* will become associated with a germinal cell *A*, and the other *a*.

Pollen cells A A a a
Germinal cells A A a a

The result of fertilization can be visualized by writing the designations for associated germinal and pollen cells in the form of fractions, pollen cells above the line, germinal cells below. In the case under discussion one obtains

$$\frac{A}{A} + \frac{A}{a} + \frac{a}{A} + \frac{a}{a}$$

In the first and fourth terms germinal and pollen cells are alike; therefore the products of their association must be constant, namely *A* and *a;* in the second and third, however, a union of the two differing parental traits takes place again, therefore the forms arising from such fertilizations are absolutely identical with the hybrid from which they derive. *Thus, repeated hybridization takes place.* The striking phenomenon, that hybrids are able to produce, in addition to the two parental types, progeny that resemble themselves is thus explained: *Aa* and *aA* both give the same association, *Aa*, since, as mentioned earlier, it makes no difference to the consequence of fertilization which of the two traits belongs to the pollen and which to the germinal cell. Therefore

$$\frac{A}{A} + \frac{A}{a} + \frac{a}{A} + \frac{a}{a} = A + 2Aa + a$$

This represents the *average* course of self-fertilization of hybrids when two differing traits are associated in them. In individual flowers and individual plants, however, the ratio in which the members of the series are formed may be subject to not insignificant deviations. . . . Thus it was proven experimentally that, in *Pisum*, hybrids form *different kinds* of germinal and pollen cells and that this is the reason for the variability of their offspring.

Source: Verhandlungen des naturforschenden den Vereines in Brünn 4: 3–47

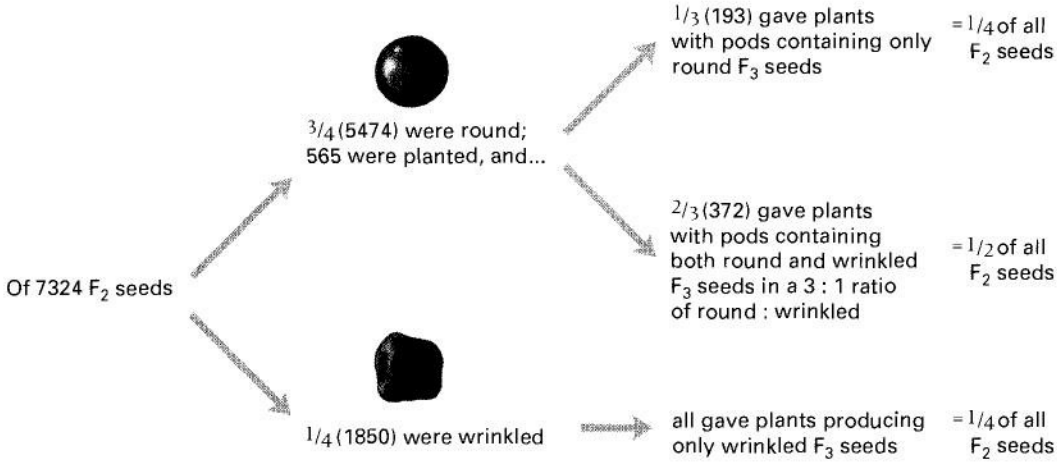


Figure 3.6 Summary of F_2 phenotypes and the progeny produced by self-fertilization.

the traits in the parental lines were transmitted as two different elements that retain their purity in the hybrids. In other words, the hereditary determinants do not "mix" or "contaminate" each other. In modern terminology, this means that, with rare but important exceptions, genes are transmitted unchanged from generation to generation.

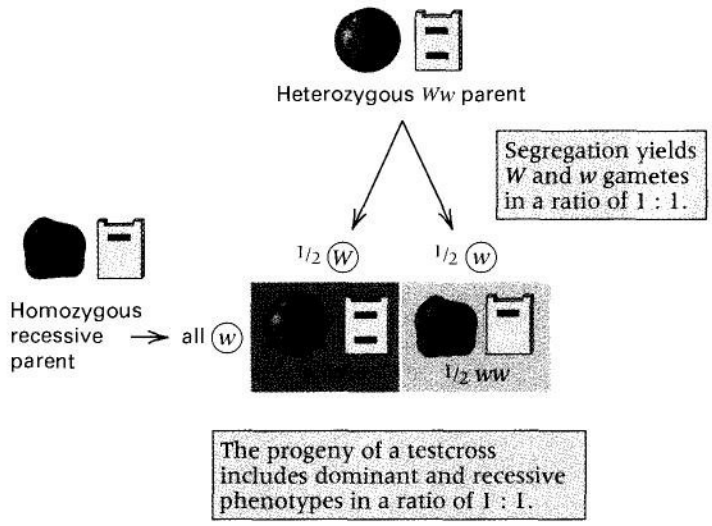


Figure 3.7 In a testcross of a *Ww* heterozygous parent with a *ww* homozygous recessive, the progeny are *Ww* and *ww* in the ratio of 1 : 1. A testcross shows the result of segregation.

The Testcross and the Backcross

Another straightforward way of testing the genetic hypothesis in Figure 3.5 is by means of a **testcross,** a cross between an organism that is heterozygous for one or more genes (for example, *Ww*) and an organism that is homozygous for the recessive alleles (for example, *ww*). The result of such a testcross is shown in **Figure 3.7**. Because the heterozygous parent is expected to produce *W* and *w* gametes in equal numbers, whereas the homozygous recessive produces only *w* gametes, the expected progeny are 1/2 with the genotype *Ww* and 1/2 with the genotype *ww*. The former have the dominant phenotype because *W* is dominant over *w*, and the latter have the recessive phenotype. A testcross is often extremely useful in genetic analysis:

> In a testcross, the phenotypes of the progeny reveal the relative frequencies of the different gametes produced by the heterozygous parent, because the recessive parent contributes only recessive alleles.

Mendel carried out a series of testcrosses with various traits. The results are shown in **Table 3.2**. In all cases, the ratio of phenotypes among the testcross progeny is very close to the 1 : 1 ratio expected from segregation of the alleles in the heterozygous parent.

Another valuable type of cross is a **backcross,** in which hybrid organisms are crossed with one of the parental genotypes. Backcrosses are commonly used by geneticists and by plant and animal breeders, as we will see in later chapters. Note that the testcrosses in Table 3.2 are also backcrosses, because in each case, the F_1 heterozygous parent came from a cross between the homozygous dominant and the homozygous recessive.

Table 3.2 Mendel's testcross results

Testcross (F_1 heterozygote × homozygous recessive)	Progeny from testcross	Ratio
Round × wrinkled seeds	193 round, 192 wrinkled	1.01 : 1
Yellow × green seeds	196 yellow, 189 green	1.04 : 1
Purple × white flowers	85 purple, 81 white	1.05 : 1
Long × short stems	87 long, 79 short	1.10 : 1

3-3 Segregation of Two or More Genes

The results of many genetic crosses depend on the segregation of the alleles of two or more genes. The genes may be in different chromosomes or in the same chromosome. Although in this section we consider the case of genes that are in two different chromosomes, the same principles apply to genes that are in the same chromosome but are so far distant from each other that they segregate independently. The case of *linkage* of genes in the same chromosome is examined in Chapter 5.

To illustrate the principles, we consider again a cross between homozygous genotypes, but in this case homozygous for the alleles of two genes. A specific example is a true-breeding variety of garden peas with seeds that are wrinkled and green (genotype *ww gg*) versus a variety with seeds that are round and yellow (genotype *WW GG*). As suggested by the use of uppercase and lowercase symbols for the alleles, the dominant alleles are *W* and *G*, the recessive alleles *w* and *g*. Crossing these strains yields F_1 seeds with the genotype *Ww Gg*, which are phenotypically round and yellow because of the dominance relations. When the F_1 seeds are grown into mature plants and self-fertilized, the F_2 progeny show the result of simultaneous segregation of the *W*, *w* allele pair and the *G*, *g* allele pair. When

Mendel perfomed this cross, he obtained the following numbers of F2 seeds:

Round, yellow	315
Round, green	108
Wrinkled, yellow	101
Wrinkled, green	32
Total	556

In these data, the first thing to be noted is the expected 3 : 1 ratio for each trait considered separately. The ratio of round : wrinkled (pooling across yellow and green) is

$$(315 + 108) : (101 + 32)$$
$$= 423 : 133$$
$$= 3.18 : 1$$

And the ratio of yellow : green (pooling across round and wrinkled) is

$$(315 + 101) : (108 + 32)$$
$$= 416 : 140$$
$$= 2.97 : 1$$

Both of these ratios are in satisfactory agreement with 3 : 1. (Testing for goodness of fit to a predicted ratio is described in Chapter 4.) Furthermore, in the F_2 progeny of the dihybrid cross, the separate 3 : 1 ratios for the two traits were combined at random. With random combinations, as shown in **Figure 3.8**, among the 3/4 of the progeny that are round, 3/4 will be yellow and 1/4 green; similarly, among the 1/4 of the progeny that are wrinkled, 3/4 will be yellow and 1/4 green. The overall proportions of round yellow to round green to wrinkled yellow to wrinkled green are therefore expected to be

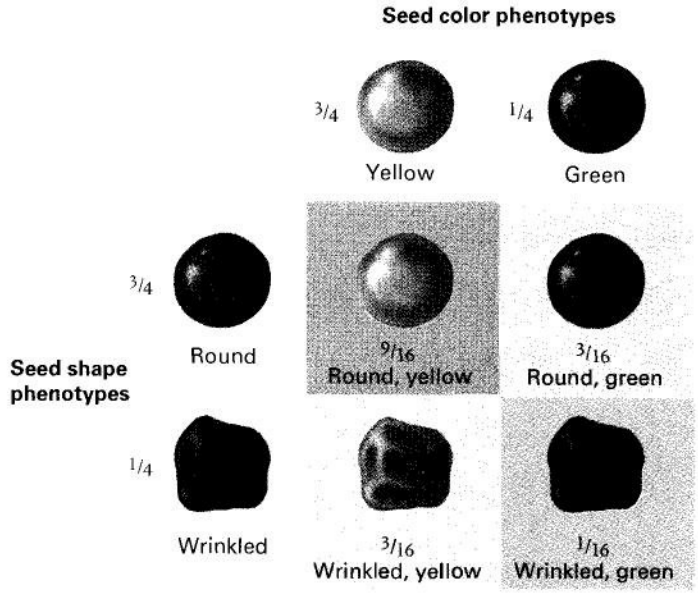
$$3/4 \times 3/4 : 3/4 \times 1/4 : 1/4 \times 3/4 : 1/4 \times 1/4$$
$$= 9/16 : 3/16 : 3/16 : 1/16$$

The observed ratio of 315 : 108 : 101 : 32 equals 9.84 : 3.38 : 3.16 : 1, which is a satisfactory fit to the 9 : 3 : 3 : 1 ratio expected from the Punnett square in Figure 3.8.

The Principle of Independent Assortment

The independent segregation of the *W*, *w* and *G*, *g* allele pairs is illustrated in **Figure 3.9**. What independence means is that if a gamete contains *W*, it is equally likely to contain *G* or *g*; and if a gamete contains *w*, it is equally likely to contain *G* or *g*. The implication is that the four gametes are formed in equal frequencies:

1/4 *WG* 1/4 *Wg* 1/4 *wG* 1/4 *wg*



Ratio of phenotypes in the F_2 progeny of a dihybrid cross is 9 : 3 : 3 : 1.

Figure 3.8 The 3 : 1 ratio of round : wrinkled, when combined at random with the 3 : 1 ratio of yellow : green, yields the 9 : 3 : 3 : 1 ratio observed in the F_2 progeny of the dihybrid cross.

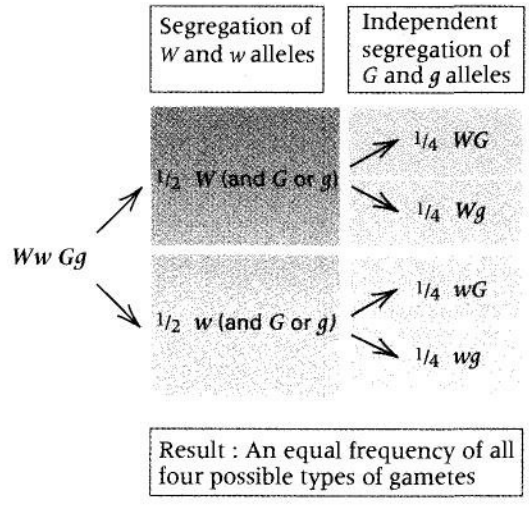
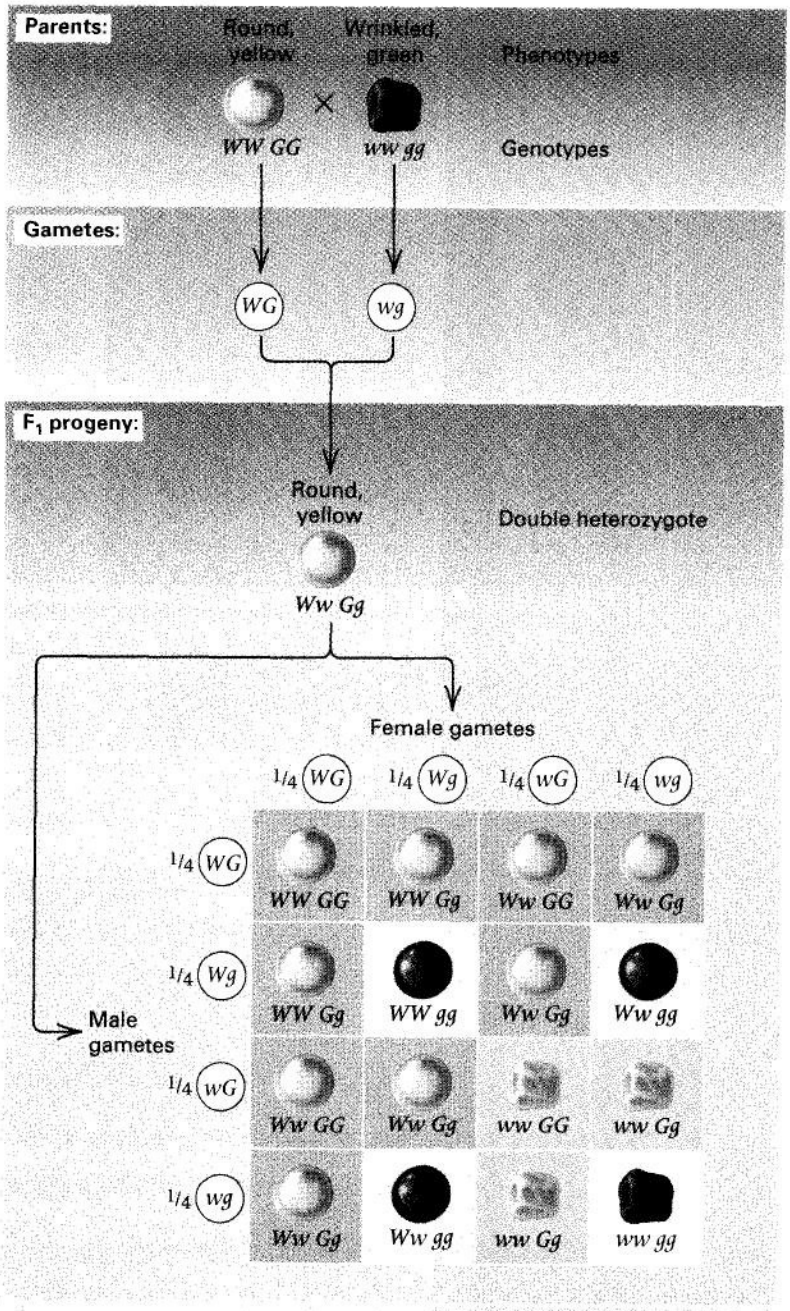


Figure 3.9 Independent segregation of the *W*, *w* and *G*, *g* allele pairs means that among each of the *W* and *w* gametic classes, the ratio of *G* : *g* is 1 : 1. Likewise, among each of the *G* and *g* gametic classes, the ratio of *W* : *w* is 1 : 1.



F_2 progeny:

Genotypes	Phenotypes
$^1/_{16}$ *WW GG* + $^2/_{16}$ *WW Gg* + $^2/_{16}$ *Ww GG* + $^4/_{16}$ *Ww Gg*	= $^9/_{16}$ round, yellow
$^1/_{16}$ *ww GG* + $^2/_{16}$ *ww Gg*	= $^3/_{16}$ wrinkled, yellow
$^1/_{16}$ *WW gg* + $^2/_{16}$ *Ww gg*	= $^3/_{16}$ round, green
$^1/_{16}$ *ww gg*	= $^1/_{16}$ wrinkled, green

Figure 3.10 Independent assortment is the biological basis for the 9 : 3 : 3 : 1 ratio of F_2 phenotypes resulting from a dihybrid cross.

The result of independent assortment when the four types of gametes combine at random to form the zygotes of the next generation is shown in **Figure 3.10**. Note that the expected ratio of phenotypes among the F_2 progeny is

$$9 : 3 : 3 : 1$$

However, as the Punnett square also shows, the ratio of genotypes in the F_2 generation is more complex; it is

$$1 : 2 : 1 : 2 : 4 : 2 : 1 : 2 : 1$$

The reason for this ratio is shown in **Figure 3.11**. Among seeds that have the *WW* genotype, the ratio of

$$GG : Gg : gg \text{ equals } 1 : 2 : 1$$

Among seeds that have the *Ww* genotype, the ratio is

$$2 : 4 : 2$$

(which is a 1 : 2 : 1 ratio multiplied by 2 because there are twice as many *Ww* genotypes as either *WW* or *ww*). And among seeds that have the *ww* genotype, the ratio of

$$GG : Gg : gg \quad \text{equals} \quad 1 : 2 : 1$$

The phenotypes of the seeds are shown beneath the genotypes, and the combined phenotypic ratio is

$$9 : 3 : 3 : 1$$

Figure 3.11 also shows that among seeds that are *GG*, the ratio of

$$WW : Ww : ww \quad \text{equals} \quad 1 : 2 : 1$$

among seeds that are *Gg*, it is

$$2 : 4 : 2$$

and among seeds that are *gg*, it is

$$1 : 2 : 1$$

Therefore, the independent segregation means that among each of the possible genotypes formed by one pair of alleles, the ratio of homozygous dominant to heterozygous to homozygous recessive is 1 : 2 : 1 for the other independent pair of alleles.

The principle of independent segregation of two pairs of alleles in different chromosomes (or located sufficiently far apart in the same chromosome) has come to be known as the principle of independent as-

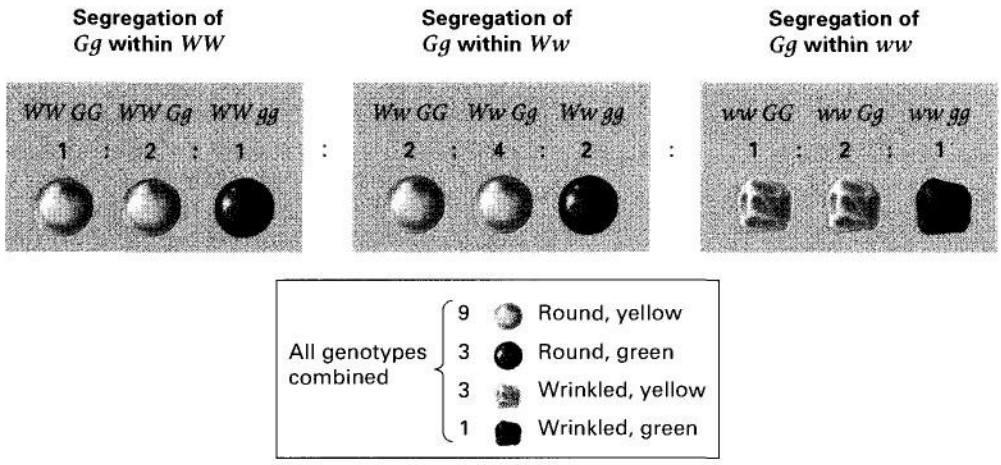


Figure 3.11 Genotypes and phenotypes of the F_2 progeny of the dihybrid cross for seed shape and seed color.

sortment. It is also sometimes referred to as *Mendel's second law.*

> **The Principle of Independent Assortment:** Segregation of the members of any pair of alleles is independent of the segregation of other pairs in the formation of reproductive cells.

Although the principle of independent assortment is fundamental in Mendelian genetics, the phenomenon of linkage, caused by proximity of genes in the same chromosome, is an important exception.

The Testcross with Unlinked Genes

Genes that show independent assortment are said to be **unlinked.** The hypothesis of independent assortment can be tested directly in a testcross with the double homozygous recessive:

$$Ww\ Gg \times ww\ gg$$

The result of the testcross is shown in **Figure 3.12.** Because plants with doubly heterozygous genotypes produce four types of gametes—*W G, W g, w G, and w g*—in equal frequencies, whereas the plants with *ww gg* genotypes produce only *w g* gametes, the possible progeny genotypes are *Ww Gg, Ww gg, ww Gg,* and *ww gg,* and these are expected in equal frequencies. Because of the dominance relations—*W* over *w* and *G* over *g*—the progeny phenotypes are expected to be round yellow, round green, wrinkled yellow, and wrinkled green in a ratio of

$$1 : 1 : 1 : 1$$

As with the one-gene testcross, in a two-gene testcross the ratio of progeny phenotypes is a direct demonstration of the ratio of gametes produced by the doubly heterozygous parent. In the actual cross, Mendel obtained 55 round yellow, 51 round green, 49 wrinkled yellow, and 53 wrinkled green, which is in good agreement with the predicted 1 : 1 : 1 : 1 ratio.

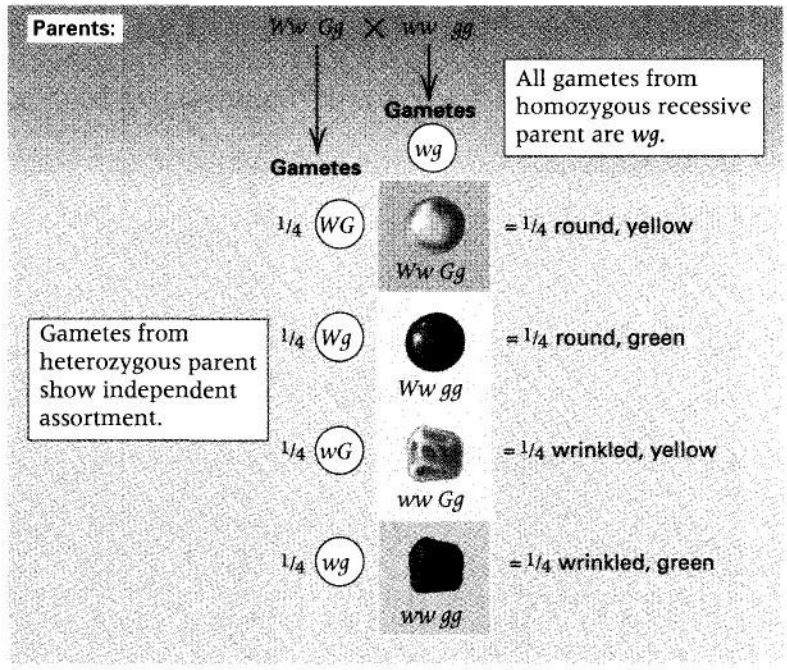


Figure 3.12 Genotypes and phenotypes resulting from a testcross of the *Ww Gg* double heterozygote.

The Big Experiment

By analogy with the independent segregation of two genes illustrated in Figure 3.8, one might expect that independent segregation of three genes would produce F_2 progeny in which combinations of the phenotypes are given by successive terms in the multiplication of $[(3/4) + (1/4)]^3$, which, when multiplied out, yields 27 : 9 : 9 : 9 : 3 : 3 : 3 : 1. A specific example is shown in **Figure 3.13**. The allele pairs are *W, w* for round versus wrinkled seeds; *G, g* for yellow versus green seeds; and *P, p* for purple versus white flowers. (The alleles indicated with the uppercase letters are dominant.) With independent segregation, among the F_2 progeny the most frequent phenotype (27/64) has the dominant form of all three traits, the next most frequent (9/64) has the dominant form of two of the traits, the next most frequent (3/64) has the dominant form of only one trait, and the least frequent (1/64) is the triple recessive. Observe that if you consider any one of the traits and ignore the other two, then the ratio of phenotypes is 3 : 1; and if you consider any two of the traits, then the ratio of phenotypes is 9 : 3 : 3 : 1. This means that all of the possible one- and two-gene independent segregations are present in the overall three-gene independent segregation.

Mendel tested this hypothesis, too, but by this time he was complaining of the amount of work the experiment entailed, noting that "of all the experiments, it required the most time and effort." The result of the experiment is shown in **Figure 3.14**. From top to bottom, the three Punnett squares show the segregation of *W* and *w* from *G* and *g* in the genotypes *PP, Pp,* and *pp*. In each box, the number in red is the expected number of plants of each genotype, assuming independent assortment, and the number in black is the observed number of each genotype of plant. The excellent agreement confirmed what Mendel regarded as the main conclusion of all of his experiments:

> "Pea hybrids form germinal and pollen cells that in their composition correspond in equal numbers to all the constant forms resulting from the combination of traits united through fertilization."

In this admittedly somewhat turgid sentence, Mendel incorporated both segregation and independent assortment. In modern terms, what he means is that with independent segregation, the gametes produced by any hybrid plant consist of equal numbers of all possible combinations of the alleles that are heterozygous. For example,

$(\frac{3}{4}W- + \frac{1}{4}ww) \times (\frac{3}{4}G- + \frac{1}{4}gg) \times (\frac{3}{4}P- + \frac{1}{4}pp)$

					Observed number	Expected number
27/64	*W*−	*G*−	*P*−	Round, yellow, purple	269	270
9/64	*W*−	*G*−	*pp*	Round, yellow, white	98	90
9/64	*W*−	*gg*	*P*−	Round, green, purple	86	90
9/64	*ww*	*G*−	*P*−	Wrinkled, yellow, purple	88	90
3/64	*W*−	*gg*	*pp*	Round, green, white	27	30
3/64	*ww*	*G*−	*pp*	Wrinkled, yellow, white	34	30
3/64	*ww*	*gg*	*P*−	Wrinkled, green, purple	30	30
1/64	*ww*	*gg*	*pp*	Wrinkled, green, white	7	10

- For any one gene, the ratio of phenotypes is 48 : 16 = 3 : 1
- For any pair of genes, the ratio of phenotypes is 36 : 12 : 12 : 4 = 9 : 3 : 3 : 1

Figure 3.13 With independent assortment, the expected ratio of phenotypes in a trihybrid cross is obtained by multiplying the three independent 3 : 1 ratios of the dominant and recessive phenotypes. A dash used in a genotype symbol indicates that either the dominant or the recessive allele is present; for example, *W*− refers collectively to the genotypes *WW* and *Ww*. (The expected numbers total 640 rather than 639 because of round-off error.)

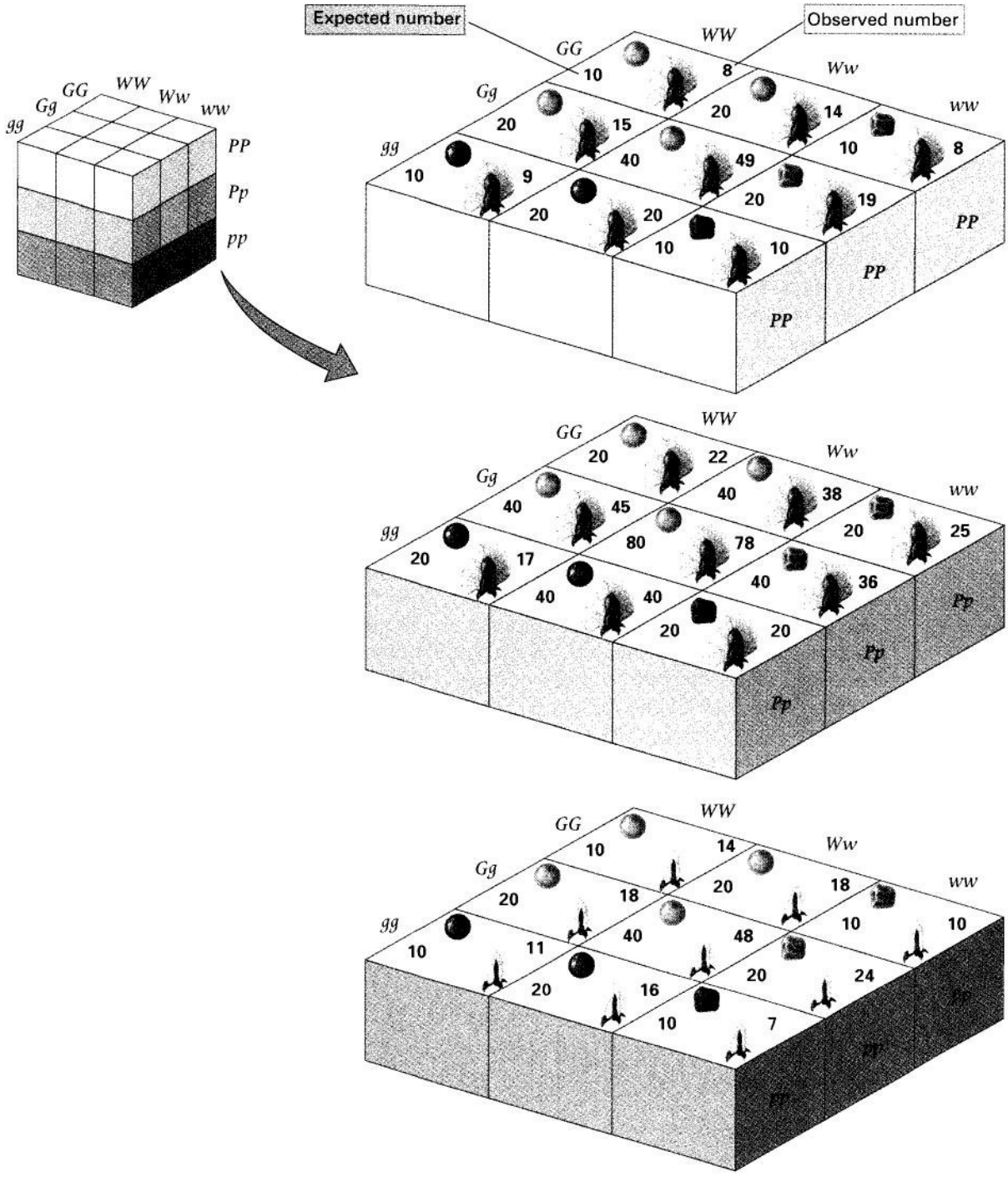


Figure 3.14 Progeny observed in the F_2 generation of a trihybrid cross with the allelic pairs *W, w* and *G, g* and *P, p*. In each box, the red entry is the expected number and the black entry is the observed. Note that each gene, by itself, yields a 1 : 2 : 1 ratio of genotypes and that each pair of genes yields a 1 : 2 : 1 : 2 : 4 : 2 : 1 : 2 : 1 ratio of genotypes.

the cross *WW gg* × *ww GG* produces F_1 progeny of genotype *Ww Gg*, which yields the gametes *W G*, *W g*, *w G*, and *w g* in equal numbers. Segregation is illustrated by the 1 : 1 ratio of *W* : *w* and *G* : *g* gametes, and independent assortment is illustrated by the equal numbers of *W G*, *W g*, *w G*, and *w g* gametes.

3.4 Human Pedigree Analysis

Large deviations from expected genetic ratios are often found in individual human families and in domesticated large animals because of the relatively small number of progeny. The effects of segregation are nevertheless evident upon examination of the phenotypes among several generations of related individuals. A diagram of a family tree showing the phenotype of each individual among a group of relatives is a **pedigree.** In this section we introduce basic concepts in pedigree analysis.

Characteristics of Dominant and Recessive Inheritance

Figure 3.15 defines the standard symbols used in depicting human pedigrees. Females are represented by circles and males by squares. (A diamond is used if the sex is unknown—as, for example, in a miscarriage.) Persons with the phenotype of interest are indicated by colored or shaded symbols. For recessive alleles, heterozygous carriers are depicted with half-filled symbols. A mating between a female and a male is indicated by joining their symbols with a horizontal line, which is connected vertically to a second horizontal line, below, that connects the symbols for their offspring. The offspring within a sibship, called **siblings** or **sibs,** are represented from left to right in order of their birth.

A pedigree for the trait *Huntington disease,* caused by a dominant mutation, is shown in Figure 3.16. The numbers in the pedigree are for convenience in referring to particular persons. The successive generations are designated by Roman numerals. Within any generation, all of the persons are numbered consecutively from left to right. The pedigree starts with the woman I-1 and the man I-2. The man has Huntington disease, which is a progressive nerve degeneration that usually begins about middle age. It results in severe physical and mental disability and then death. The dominant allele, *HD,* that causes Huntington disease is rare. All affected persons in the pedigree have the heterozygous genotype *HD hd,* whereas nonaffected persons have the homozygous normal genotype *hd hd.* The disease has complete penetrance. The **penetrance** of a genetic disorder is the proportion of individuals with the at-risk genotype who actually express the trait; *complete penetrance* means the trait is expressed in 100 percent of persons with that genotype. The pedigree demonstrates the following characteristic features of inheri-

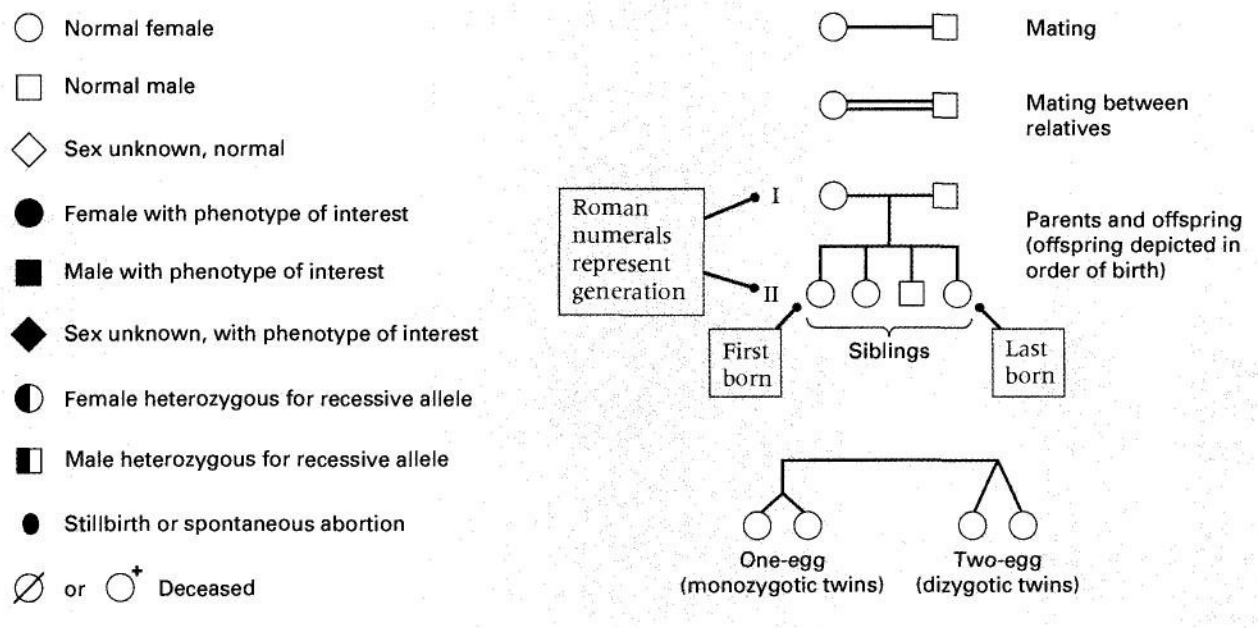


Figure 3.15 Conventional symbols used in depicting human pedigrees.

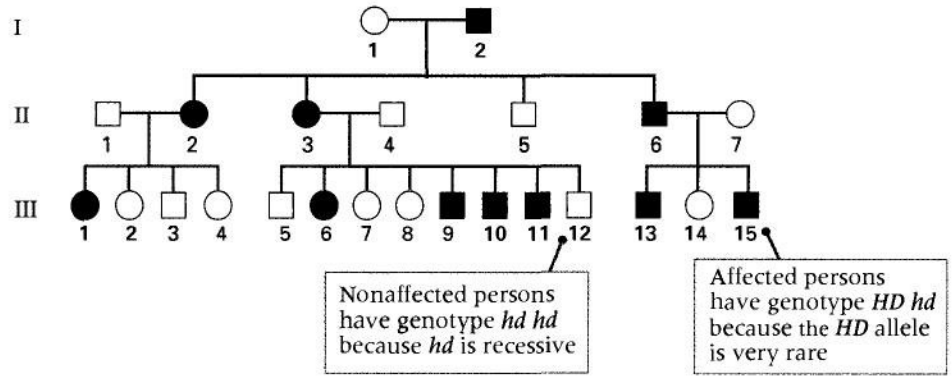


Figure 3.16 Pedigree of a human family showing the inheritance of the dominant gene for Huntington disease. Females and males are represented by circles and squares, respectively. Red symbols indicate persons affected with the disease.

tance due to a rare dominant allele with complete penetrance.

1. Females and males are equally likely to be affected.
2. Affected offspring have one affected parent (except for rare new mutations), and the affected parent is equally likely to be the mother or the father.
3. On average, half of the individuals in sibships with an affected parent are affected.

A pedigree for a trait due to a homozygous recessive allele is shown in **Figure 3.17**. The trait is *albinism*, absence of pigment in the skin, hair, and iris of the eyes. This pedigree illustrates characteristics of inheritance due to a rare recessive allele with complete penetrance:

1. Females and males are equally likely to be affected.
2. Affected individuals, if they reproduce, usually have unaffected progeny.
3. Most affected individuals have unaffected parents.
4. The parents of affected individuals are often relatives.
5. Among siblings of affected individuals, the proportion affected is approximately 25 percent.

With rare recessive inheritance, the mates of homozygous affected persons are usually homozygous for the normal allele, so all of the offspring will be heterozygous and not affected. Heterozygous **carriers** of the mutant allele are considerably more common than homozygous affected individuals, because it is more likely that a person will inherit only one copy of a rare mutant allele than two copies. Most homozygous recessive genotypes therefore result from



Albinism, a recessive genetic disorder, is characterized by a lack of the pigment melanin. This 37-year-old woman with albinism has very light skin and white (actually colorless) hair. [Dr. P. Marazzi/Science Photo Library/Photo Researchers, Inc.]

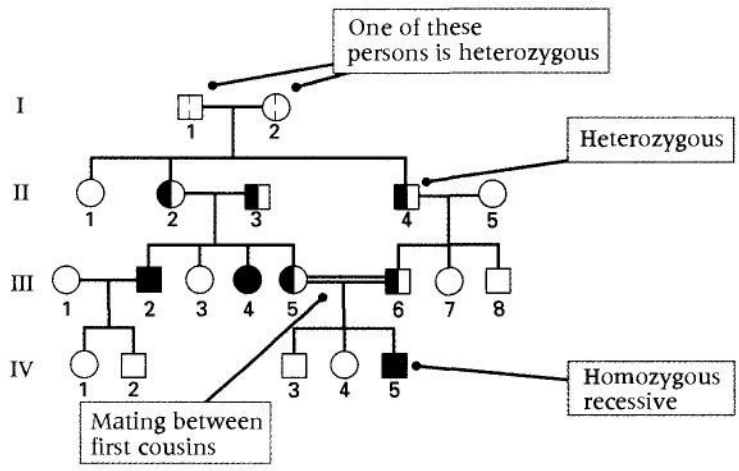


Figure 3.17 Pedigree of albinism. With recessive inheritance, affected persons (filled symbols) often have unaffected parents. The double horizontal line indicates a mating between relatives—in this case, first cousins.

matings between carriers (heterozygous × heterozygous), in which each offspring has a 1/4 chance of being affected. Another important feature of rare recessive inheritance is that the parents of affected individuals are often related (**consanguineous**). A mating between relatives is indicated with a double line connecting the partners, as for the first-cousin mating in Figure 3.17. Mating between relatives is important for recessive alleles to become homozygous, because when a recessive allele is rare, it is more likely to become homozygous through inheritance from a common ancestor than from parents who are completely unrelated. The reason is that the carrier of a rare allele may have many descendants who are also carriers. If two of these carriers should mate with each other (for example, in a first-cousin mating), then the hidden recessive allele can become homozygous with a probability of 1/4. Mating between relatives constitutes *inbreeding*, and the consequences of inbreeding are discussed further in Chapter 17. Because an affected individual indicates that the parents are heterozygous carriers, the expected proportion of affected individuals among the siblings is approximately 25 percent, but the exact value depends on the details of how affected individuals are identified and included in the database.

Molecular Markers in Human Pedigrees

Before the advent of molecular methods, there were many practical obstacles to the study of human genetics.

- Most genes that cause genetic diseases are rare, so they are observed in only a small number of families.
- Many genes of interest in human genetics are recessive, so they are not detected in heterozygous genotypes.
- The number of offspring per human family is relatively small, so segregation cannot usually be detected in single sibships.
- The human geneticist cannot perform testcrosses or backcrosses, because human matings are not manipulated by an experimenter.

Because techniques for manipulating DNA allow direct access to the DNA, modern genetic studies of human pedigrees are carried out primarily using genetic markers present in the DNA itself, rather than through the phenotypes produced by mutant genes. The human genome includes one single-nucleotide polymorphism per 500–3000 bp, depending on the region being studied. This means that two randomly chosen human genomes differ at 1–6 million nucleotide positions.

Various types of DNA polymorphisms were discussed in Chapter 2, along with the methods by which they are detected and studied. An example of a DNA polymorphism segregating in a three-generation human pedigree is shown in **Figure 3.18.** The type of polymorphism is a *simple tandem repeat polymorphism (STRP),* in which each allele differs in size according to the number of copies it contains of a short DNA sequence repeated in tandem. The differences in size are detected by electrophoresis after amplification of the region by PCR. STRP markers usually have as many as 20 codominant alleles, and the majority of in-

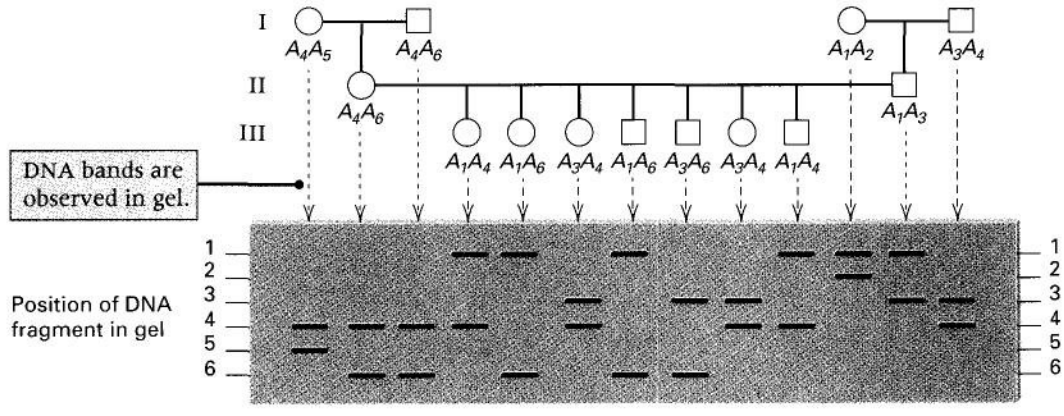


Figure 3.18 Human pedigree showing segregation of VNTR alleles. Six alleles (A_1–A_6) are present in the pedigree, but any one person can have only one allele (if homozygous) or two alleles (if heterozygous).

dividuals are heterozygous for two different alleles. In the example in Figure 3.18, each of the parents is heterozygous, as are all of the children. More than 5000 genetic markers of this type have been identified in the human genome, each heterozygous in an average of 70 percent of individual genotypes.

Six alleles are depicted in Figure 3.18, denoted by A_1 through A_6. In the gel, the numbers of the bands correspond to the subscripts of the alleles. The mating in generation II is between two heterozygous genotypes: $A_4A_6 \times A_1A_3$. Because of segregation in each parent, four genotypes are possible among the offspring (A_4A_1, A_4A_3, A_6A_1, and A_6A_3); these would conventionally be written with the smaller subscript first, as A_1A_4, A_3A_4, A_1A_6, and A_3A_6. With random fertilization the offspring genotypes are equally likely, as may be verified from a Punnett square for the mating. Figure 3.18 illustrates some of the principal advantages of multiple, codominant alleles for human pedigree analysis: (1) Heterozygous genotypes can be distinguished from homozygous genotypes. (2) Many individuals in the population are heterozygous, and so many matings are informative in regard to segregation. (3) Each segregating genetic marker yields up to four distinguishable offspring genotypes.

3.5 Pedigrees and Probability

Genetic analysis benefits from large numbers of progeny because these numbers determine the degree to which observed ratios of genotypes and phenotypes fit their expected values. Greater statistical variation occurs in smaller numbers—and therefore potentially greater deviations from the expected values. For example, among the seeds of individual pea plants in which a 3 : 1 ratio was expected, Mendel observed ratios ranging from 1.85 : 1 to 4.85 : 1. The large variation results from the relatively small number of seeds per plant, which averaged about 34 in these experiments. Taken together, the seeds yielded a phenotypic ratio of 3.08 : 1, which not only fits 3 : 1 but is actually a better fit than that observed in any of the individual plants. Because of statistical variation in small numbers, a working knowledge of probability is basic to understanding genetic transmission. In the first place, each event of fertilization represents a chance combination of alleles present in the parental gametes. In the second place, the proportions of the different types of offspring obtained from a cross are the cumulative result of numerous independent events of fertilization.

In the analysis of genetic crosses, the probability of a particular outcome of a fertilization event may be considered as equivalent to the proportion of times that this outcome is expected to be realized in numerous repeated trials. The reverse is also true: The proportion of times that an outcome is expected to be realized in numerous repeated trials is equivalent to the probability that it is realized in a single trial. To take a specific example using the pedigree in Figure 3.18, the mating at the upper left is $A_4A_5 \times A_4A_6$. Among a large number of offspring from such a mating, the expected proportion of A_4A_6 genotypes is $1/4$. Equivalently, we could say that any offspring whose genotype is unknown has a probability of genotype A_4A_6 equal to $1/4$. However, the genotype of the female offspring shown is already given as A_4A_6, so relative to this individual, the situation has become a certainty: The probability that her genotype is A_4A_6 equals 1, and the probability that her genotype is anything else equals 0.

To evaluate the probability of a genetic event usually requires an understanding of the mechanism of inheritance and knowledge of the particular cross. For example, in asserting that the mating I-1 $\times$ I-2 in Figure 3.18 has a probability of yielding an A_4A_6 offspring equal to $1/4$, we needed to take the parental genotypes into account. If the mating were $A_4A_6 \times A_4A_6$, then the probability of an A_4A_6 offspring would be $1/2$ rather than $1/4$; if the mating were $A_4A_4 \times A_6A_6$, then the probability of an A_4A_6 offspring would be 1.

In many genetic crosses, the possible outcomes of fertilization are equally likely. Suppose that there are n possible outcomes, each as likely as any other, and that in m of these, a particular outcome of interest is realized; then the probability of the outcome of interest is m/n. In the language of probability, a possible outcome of interest is typically called an *event*. As an example,

CoNnection Troubadour

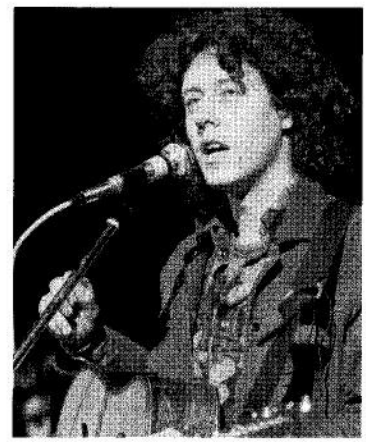
The Huntington's Disease Collaborative Research Group
1993
Comprising 58 authors
among 9 institutions
A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes

Modern genetic research is sometimes carried out by large collaborative groups in a number of research institutions scattered across several countries. This approach is exemplified by the search for the gene responsible for Huntington disease. The search was highly publicized because of the severity of the disease, the late age of onset, and the dominant inheritance. Famed folk singer Woody Guthrie, who wrote "This Land Is Your Land" and other well-known tunes, died of the disease in 1967. When the gene was identified, it turned out to encode a protein (now called huntingtin) of unknown function that is expressed in many cell types throughout the body and not, as expected, exclusively in nervous tissue. Within the coding sequence of this gene is a trinucleotide repeat (5'-CAG-3') that is repeated in tandem a number of times according to the general formula $(5'\text{-CAG-}3')_n$. Among normal alleles, the number n of repeats ranges from 11 to 34 with an average of 18; among mutant alleles, the number of repeats ranges from 40 to 86. This tandem repeat is genetically unstable in that it can, by some unknown mechanism, increase in copy number ("expand"). In two cases in which a new mutant allele was analyzed, one had increased in repeat number from 36 to 44 and the other from 33 to 49. This is a mutational mechanism that is quite common in some human genetic diseases. The excerpt cites several other examples. The authors also emphasize that their discovery raises important ethical issues, including genetic testing, confidentiality, and informed consent.

Huntington disease (HD) is a progressive neurodegenerative disorder characterized by motor disturbance, cognitive loss, and psychiatric manifestations. It is inherited in an autosomal dominant fashion and affects approximately 1 in 10,000 individuals in most populations of European origin. The hallmark of HD is a distinctive choreic [jerky] movement disorder that typically has a subtle, insidious onset in the fourth to fifth decade of life and gradually worsens over a course of 10 to 20 years until death. . . . The

Huntington disease, a dominant genetic disorder, took the life of folksinger–songwriter Woody Guthrie (left) in 1967. The condition does not exhibit itself until individuals are 30 years of age, or older—after they may have had children. Given the severity of the disease, it raises important ethical issues for those children. Woody's son Arlo (right) is also a folksinger and songwriter.


AP/Wide World Photos


Corbis/Henry Diltz

consider again the mating $A_4A_5 \times A_4A_6$. If we were concerned with the exact specification of each genotype, then we would distinguish among all of the four possible genotypes of progeny (A_4A_4, A_4A_6, A_4A_5, and A_5A_6) and assert that each is equally likely because it has a probability of 1/4. On the other hand, if we were concerned only with homozygosity or heterozygosity, then there would be only two possible outcomes, homozygous (A_4A_4) and heterozygous (A_4A_6, A_4A_5, and A_5A_6), which have probabilities 1/4 and 3/4, respectively. Or again, if we were interested only in whether or not an offspring carries the A_5 allele, then the probabilities would be 1/2 for "yes" and 1/2 for "no." These examples illustrate the point that offspring of a genetic cross can be classified or grouped together in different ways, depending on the particular question

genetic defect causing HD was assigned to chromosome 4 in one of the first successful linkage analyses using DNA markers in humans. Since that time, we have pursued an approach to isolating and characterizing the HD gene based on progressively refining its localization. . . . [We have found that a] 500-kb segment is the most likely site of the genetic defect. [The abbreviation kb stands for kilobase pairs; 1 kb equals 1000 base pairs.] Within this region, we have identified a large gene, spanning approximately 210 kb, that encodes a previously undescribed protein. The reading frame contains a polymorphic $(CAG)_n$ trinucleotide repeat with at least 17 alleles in the normal population, varying from 11 to 34 CAG copies. On HD chromosomes, the length of the trinucleotide repeat is substantially increased. . . . Elongation of a trinucleotide repeat sequence has been implicated previously as the cause of three quite different human disorders, the fragile-X syndrome, myotonic dystrophy, and spino-bulbar muscular atrophy. . . . It can be expected that the capacity to monitor directly the size of the trinucleotide repeat in individuals "at risk" for HD will revolutionize testing for the disorder. . . . We consider it of the utmost importance that the current internationally accepted guidelines and counseling protocols for testing people at risk continue to be observed, and that samples from unaffected relatives should not be tested inadvertently or without full consent. . . . With the mystery of the genetic basis of HD apparently solved, [it opens] the next challenges in the effort to understand and to treat this devastating disorder.

We consider it of the utmost importance that the current internationally accepted guidelines and counseling protocols for testing people at risk continue to be observed, and that samples from unaffected relatives should not be tested inadvertently or without full consent.

Source: Cell 72: 971–983

of interest, and different types of groupings may have different probabilities associated with them.

Mutually Exclusive Possibilities

Sometimes an outcome of interest includes two or more different possibilities. The offspring of the mating $A_4A_5 \times A_4A_6$ again provides a convenient example. As we have seen, if the offspring are classified as either homozygous or heterozygous, then the outcome "homozygous" consists of just one possibility (A_4A_4) and the outcome "heterozygous" consists of three possibilities (A_4A_6, A_4A_5, and A_5A_6). Furthermore, if an individual genotype is any one of the three heterozygous genotypes, it cannot at the same time be any of the other heterozygous genotypes. In other words, only one genotype can be realized in any one organism, so the realization of one possibility precludes the realization of another in the same organism. Events that exclude each other in this manner are said to be *mutually exclusive*. When events are mutually exclusive, their probabilities are combined according to the addition rule.

> **Addition Rule:** The probability of the realization of one or the other of two mutually exclusive events, A or B, is the sum of their separate probabilities.

In symbols, if we use *Prob* to mean *probability*, then the addition rule for two possible outcomes is written

$$Prob\{\text{A or B}\} = Prob\{\text{A}\} + Prob\{\text{B}\}$$

Application of the addition rule is straightforward in the above examples of homozygous versus heterozygous genotypes, but there are three possible outcomes instead of two. Because the offspring genotypes are equally likely, the probability of a heterozygous offspring equals $Prob\{A_4A_6, A_4A_5, \text{or } A_5A_6\} = Prob\{A_4A_6\} + Prob\{A_4A_5\} + Prob\{A_5A_6\} = 1/4 + 1/4 + 1/4 = 3/4$. Because 3/4 is the probability of an individual offspring being heterozygous, it is also the expected proportion of heterozygous genotypes among a large number of progeny.

Independent Possibilities

Events that are not mutually exclusive may be *independent*, which means that the realization of one outcome has no influence on the possible realization of any others. We have already seen an example in the independent segregation of round versus wrinkled as against yellow versus green peas (Figure 3.9). The principle is that when the possible outcomes of an experiment or observation are independent, the probability that they are realized together is obtained by multiplication.

Successive offspring from a cross are also independent events, which means that the genotypes of early progeny have no influence on the probabilities in later progeny (**Figure 3.19**). The independence of successive offspring contradicts the widespread belief that in each human family, the ratio of girls to boys must "even out" at approximately 1 : 1. According to this reasoning, a family with four girls would be more likely to have a boy the next time around. But this belief is supported neither by theory nor by actual data on the sex ratio in human sibships. The data indicate that human families are equally likely to have a girl or a boy on any birth, irrespective of the sex distribution in previous births. Although statistics guarantees that the sex ratio will balance out when averaged over a very large number of sibships, this does not imply that it will equalize in any individual sibship. To be concrete, among families in which there are five children, sibships consisting of five boys balance those consisting of five girls, yielding an overall sex ratio of 1 : 1; nevertheless, both of these sibships have an unusual sex ratio.

When events are independent (such as independent traits or successive offspring from a cross), the probabilities are combined by means of the multiplication rule.

Multiplication Rule: The probability of two independent events, A and B, being realized simultaneously is given by the product of their separate probabilities.

In symbols, the multiplication rule is written

$$Prob\{\text{A and B}\} = Prob\{\text{A}\} \cdot Prob\{\text{B}\}$$

The multiplication rule can be used to answer questions such as this: For two offspring from the mating $A_4A_5 \times A_4A_6$, what is the probability of one homozygous genotype and one heterozygous genotype? The birth order homozygous–heterozygous has probability $1/4 \times 3/4$, and the birth order heterozygous–homozygous has probability $3/4 \times 1/4$. These possibilities are mutually exclusive, so the overall probability is

$$1/4 \times 3/4 + 3/4 \times 1/4 = 2(1/4)(3/4) = 3/8$$

Note that the addition rule was used twice in solving this problem, once to calculate the probability of a heterozygous offspring and again to calculate the probability of both birth orders. The multiplication rule can also be used to calculate the probability of a specific genotype among the progeny of a complex cross. For example, if a quadruple heterozygous genotype *Aa Bb Cc Dd* is self-fertilized, what is the probability of a quadruple heterozygous offspring, *Aa Bb Cc Dd*? Assuming independent assortment, the answer is $(1/2)(1/2)(1/2)(1/2) = (1/2)^4 = 1/16$, because each of the inde-

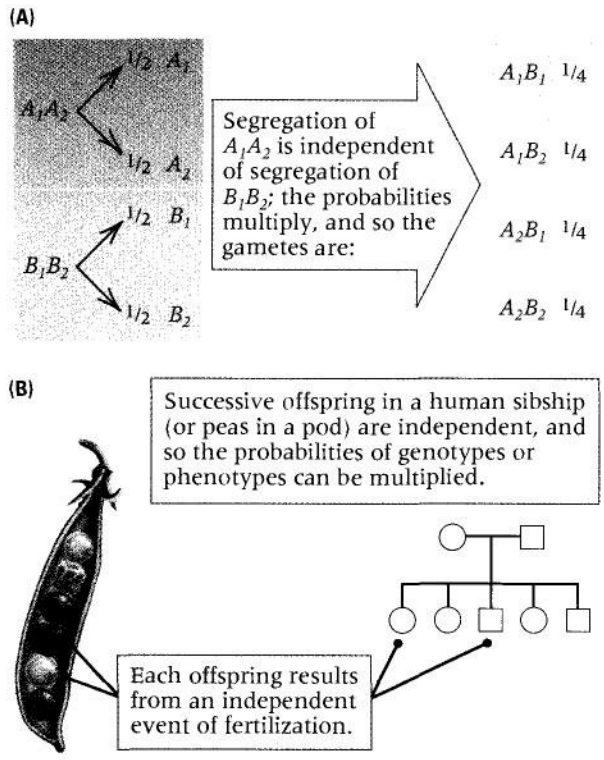


Figure 3.19 In genetics, two important types of independence are (A) independent segregation of alleles that show independent assortment and (B) independent fertilizations resulting in successive offspring. In these cases, the probabilities of the individual outcomes of segregation or fertilization are multiplied to obtain the overall probability.

pendent genes yields a heterozygous offspring with probability 1/2.

3.6 Incomplete Dominance and Epistasis

Dominance and codominance are not the only possibilities for pairs of alleles. There are situations of **incomplete dominance,** in which the phenotype of the heterozygous genotype is intermediate between the phenotypes of the homozygous genotypes. A classic example of incomplete dominance concerns flower color in the snapdragon *Antirrhinum majus* (**Figure 3.20**). In wildtype flowers, a red type of anthocyanin pigment is formed by a sequence of enzymatic reactions. A wildtype enzyme, encoded by the *I* allele, is limiting to the rate of the overall reaction, so the amount of red pigment is determined by the amount of enzyme that the *I* allele produces. The alternative *i* allele codes for an inactive enzyme, and *ii* flowers are ivory in color. Because the amount of the critical enzyme is reduced in *Ii* heterozygotes, the amount of red pigment in the flowers is reduced also, and the effect of the dilution is to make the flowers pink. A cross between plants differing in flower color therefore gives direct phenotypic evidence of segregation (Figure 3.20). The cross *II* (red) × *ii* (ivory) yields F_1 plants with genotype *Ii* and pink flowers. In the F_2 progeny obtained by self-pollination of the F_1 hybrids, one experiment resulted in 22 plants with red flowers, 52 with pink flowers, and 23 with ivory flowers, which fits the expected ratio of 1 : 2 : 1.

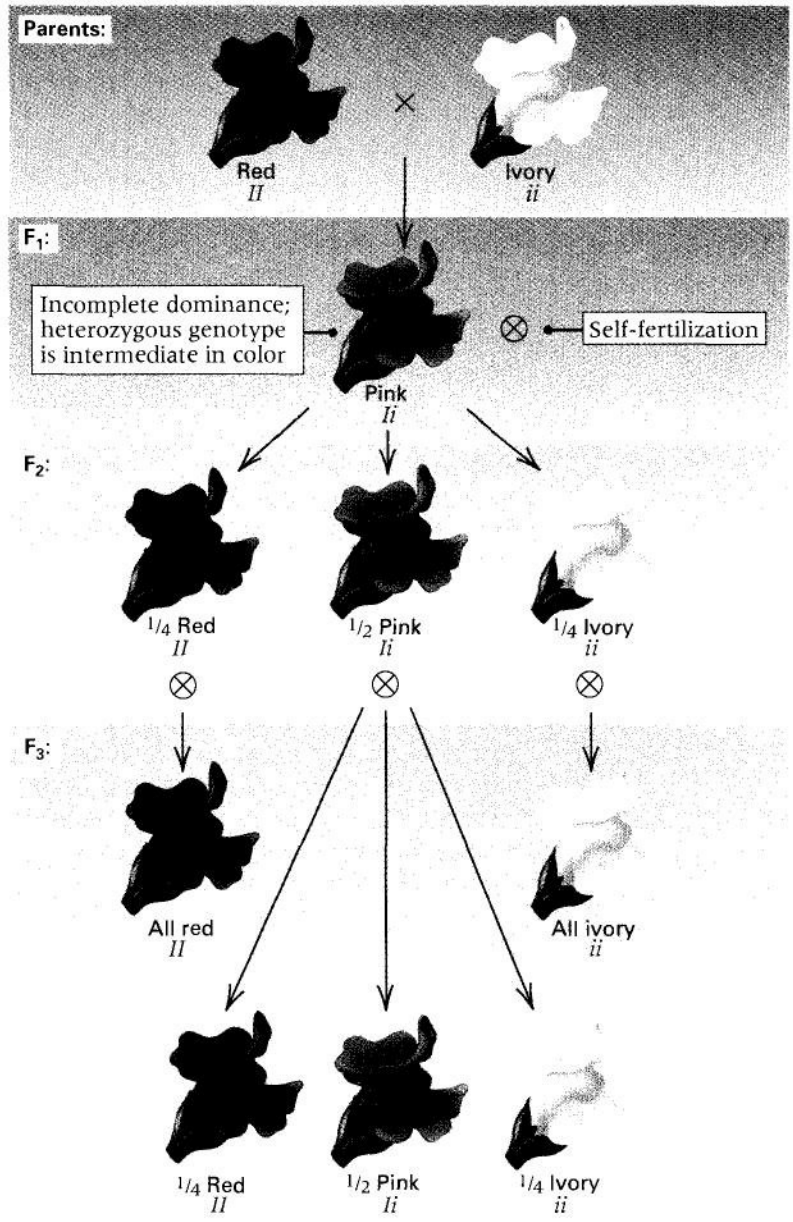


Figure 3.20 Red versus white flower color in snapdragons shows no dominance.

Multiple Alleles

The occurrence of **multiple alleles** is exemplified by the alleles A_1–A_6 of the STRP marker in the human pedigree in Figure 3.18. Multiple alleles are relatively common in natural populations and, as in this example, can be detected most easily by molecular methods. In the DNA of a gene, each nucleotide can be A, T, G, or C, so a gene of n nucleotides can theoretically mutate at any of the positions to any of the three other nucleotides. The number of possible single-nucleotide differences in a gene of length n is therefore $3 \times n$. If n = 5000, for example, there are potentially 15,000 alleles (not counting any of the possibilities with more than one nucleotide substitution). Most of the potential alleles do not actually exist at any one time. Some are absent because they did not occur, others did occur but were eliminated by chance or because they were harmful, and still others are present but at such a low frequency that they remain undetected. Nevertheless, at the level of DNA sequence, most genes in

most natural populations have multiple alleles, all of which can be considered "wild-type." Multiple wildtype alleles are useful in such applications as DNA typing because two unrelated people are unlikely to have the same genotype, especially if several different loci, each with multiple alleles, are examined. Many harmful mutations also exist in multiple forms. Recall from Chapter 1 that more than 400 mutant forms of the phenylalanine hydroxylase gene have been identified in patients with phenylketonuria. The alleles A_1–A_6 in Figure 3.18 also illustrate that although a *population* of organisms may contain any number of alleles, any particular organism or cell may carry no more than two, and any gamete may carry no more than one.

In some cases, the multiple alleles of a gene exist merely by chance and reflect the history of mutations that have taken place in the population and the dissemination of these mutations among population subgroups by migration and interbreeding. In other cases, there are biological mechanisms that favor the maintenance of a large number of alleles. For example, genes that control self-sterility in certain flowering plants can have large numbers of allelic types. This type of self-sterility is found in species of red clover that grow wild in many pastures. The self-sterility genes prevent self-fertilization because a pollen grain can undergo pollen tube growth and fertilization only if it contains a self-sterility allele different from either of the alleles present in the flower on which it lands. In other words, a pollen grain containing an allele already present in a flower will not function on that flower. Because all pollen grains produced by a plant must contain one of the self-sterility alleles present in the plant, pollen cannot function on the same plant that produced it, and self-fertilization cannot take place. Under these conditions, any plant with a new allele has a selective advantage, because pollen that contains the new allele can fertilize all flowers except those on the same plant. Through evolution, populations of red clover have accumulated hundreds of alleles of the self-sterility gene, many of which have been isolated and their DNA sequences determined. Many of the alleles differ at multiple nucleotide sites, which implies that the alleles in the population are very old.

Human ABO Blood Groups

In a multiple allelic series, there may be different dominance relationships between different pairs of alleles. An example is found in the human ABO blood groups, which are determined by three alleles denoted I^A, I^B, and I^O. (Actually, there are two slightly different variants of the I^A allele.) The blood group of any person may be A, B, AB, or O, depending on the type of polysaccharide (polymer of sugars) present on the surface of red blood cells. One of two different polysaccharides, A or B, can be formed from a precursor molecule that is modified by the enzyme product of the I^A or the I^B allele. The gene product is a glycosyl transferase enzyme that attaches a sugar unit to the precursor (**Figure 3.21**). The I^A or the I^B alleles encode different forms of the enzyme with replacements at four amino acid sites; these alter the substrate specificity so that each enzyme attaches a different sugar. People of genotype I^AI^A produce red blood cells having only the A polysaccharide and are said to have blood type A. Those of genotype I^BI^B have red blood cells with only the B polysaccharide and have blood type B. Heterozygous I^AI^B people have red cells with both the A and the B polysaccharides and have blood type AB. The third allele, I^O, encodes an enzymatically inactive protein that leaves the precursor unchanged; neither the A nor the B type of polysaccharide is produced. Homozygous I^OI^O persons therefore lack both the A and the B polysaccharides and are said to have blood type O.

In this multiple allelic series, the I^A and I^B alleles are codominant: The heterozygous genotype has the characteristics of both homozygous genotypes—the presence of both the A and the B carbohydrate on the red blood cells. On the other hand, the I^O allele is recessive to both I^A and I^B. Hence, heterozygous I^AI^O genotypes produce the A polysaccharide and have blood type A, and heterozygous I^BI^O genotypes produce the B polysaccharide and have blood type B. The genotypes and phenotypes of the ABO blood group system are summarized in the first three columns of **Table 3.3**.

ABO blood groups are important in medicine because of the need for blood transfusions. A crucial feature of the ABO system is that most human blood contains

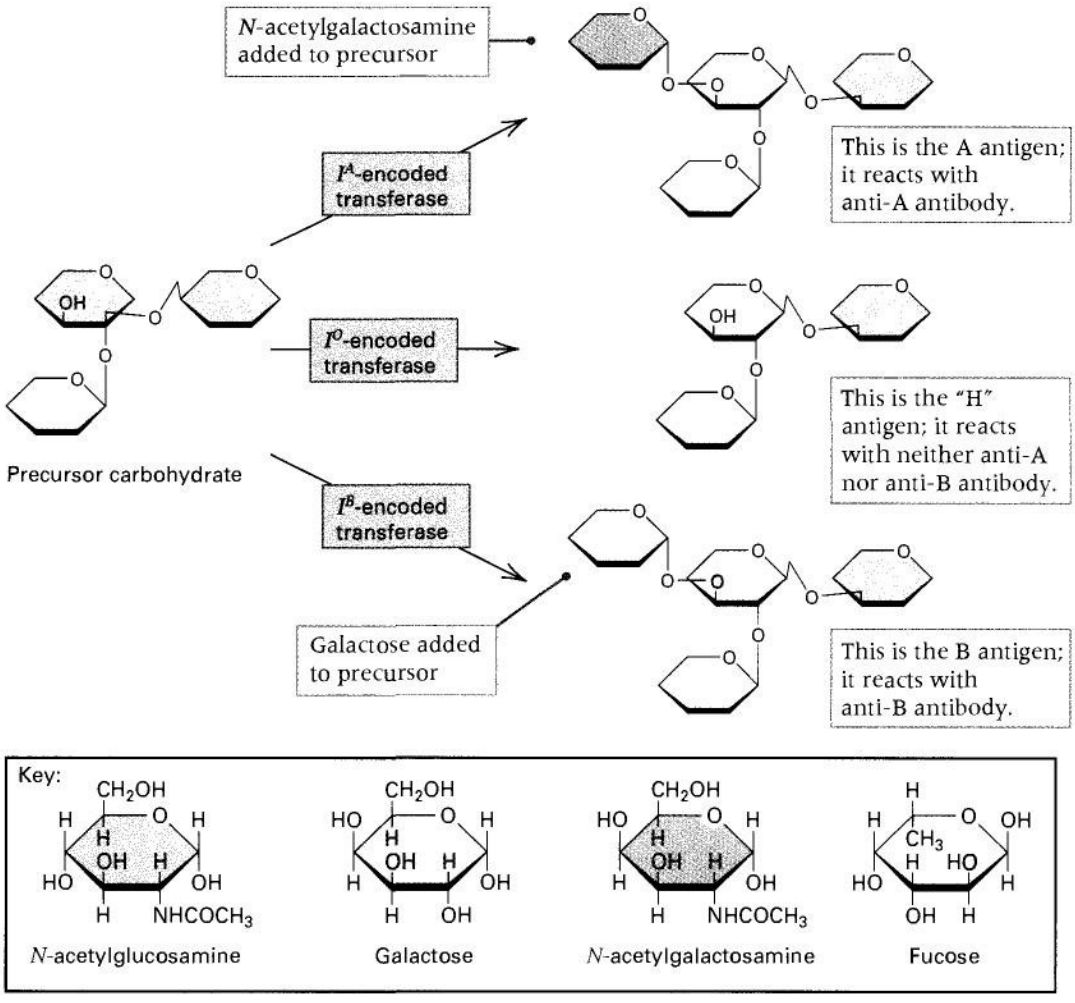


Figure 3.21 The ABO antigens on the surface of human red blood cells are carbohydrates. They are formed from a precursor carbohydrate by the action of transferase enzymes encoded by alleles of the *I* gene. Allele I^O codes for an inactive enzyme and leaves the precursor (called the H substance) unmodified. The I^A allele encodes an enzyme that adds *N*-acetylgalactosamine (purple) to the precursor. The I^B allele encodes an enzyme that adds galactose (green) to the precursor. The other colored sugar units are *N*-acetylglucosamine (orange) and fucose (yellow).

antibodies to either the A or the B polysaccharide. An **antibody** is a protein made by the immune system in response to a stimulating molecule called an **antigen** and is capable of binding to the antigen. An antibody is usually specific in that it recognizes only one antigen. Some antibodies combine with antigen and form large molecular aggregates that may precipitate.

Antibodies act to defend against invading viruses and bacteria. Although antibodies do not normally form without prior

Table 3.3 Genetic control of the human ABO blood groups

Genotype	Antigens present on red blood cells	ABO blood goup phenotype	Antibodies present in blood fluid	Blood types that can be tolerated in transfusion	Blood types that can accept blood for transfusion
I^AI^A	A	Type A	Anti-B	A & O	A & AB
I^AI^O	A	Type A	Anti-B	A & O	A & AB
I^BI^B	B	Type B	Anti-A	B & O	B & AB
I^BI^O	B	Type B	Anti-A	B & O	B & AB
I^AI^B	A & B	Type AB	Neither anti-A nor anti-B	A, B, AB & O	AB only
I^OI^O	Neither A nor B	Type O	Anti-A & anti-B	O only	A, B, AB & O

stimulation by the antigen, people capable of producing anti-A and anti-B antibodies do produce them. Production of these antibodies may be stimulated by antigens that are similar to polysaccharides A and B and that are present on the surfaces of many common bacteria. However, a mechanism called *tolerance* prevents an organism from producing antibodies against its own antigens. This mechanism ensures that A antigen or B antigen elicits antibody production only in people whose own red blood cells do not contain A or B, respectively. The end result:

> People of blood type O make both anti-A and anti-B antibodies, those of blood type A make anti-B antibodies, those of blood type B make anti-A antibodies, and those of blood type AB make neither type of antibody.

The antibodies found in the blood fluid of people with each of the ABO blood types are shown in the fourth column in Table 3.3. The clinical significance of the ABO blood groups is that transfusion of blood containing A or B red-cell antigens into persons who make antibodies against them results in an agglutination reaction in which the donor red blood cells are clumped. In this reaction, the anti-A antibody agglutinates red blood cells of either blood type A or blood type AB, because both carry the A antigen (**Figure 3.22**). Similarly, anti-B antibody agglutinates red blood cells of either blood type B or blood type AB. When the blood cells agglutinate, many blood vessels are blocked, and the recipient of the transfusion goes into shock and may die. Incompatibility in the other

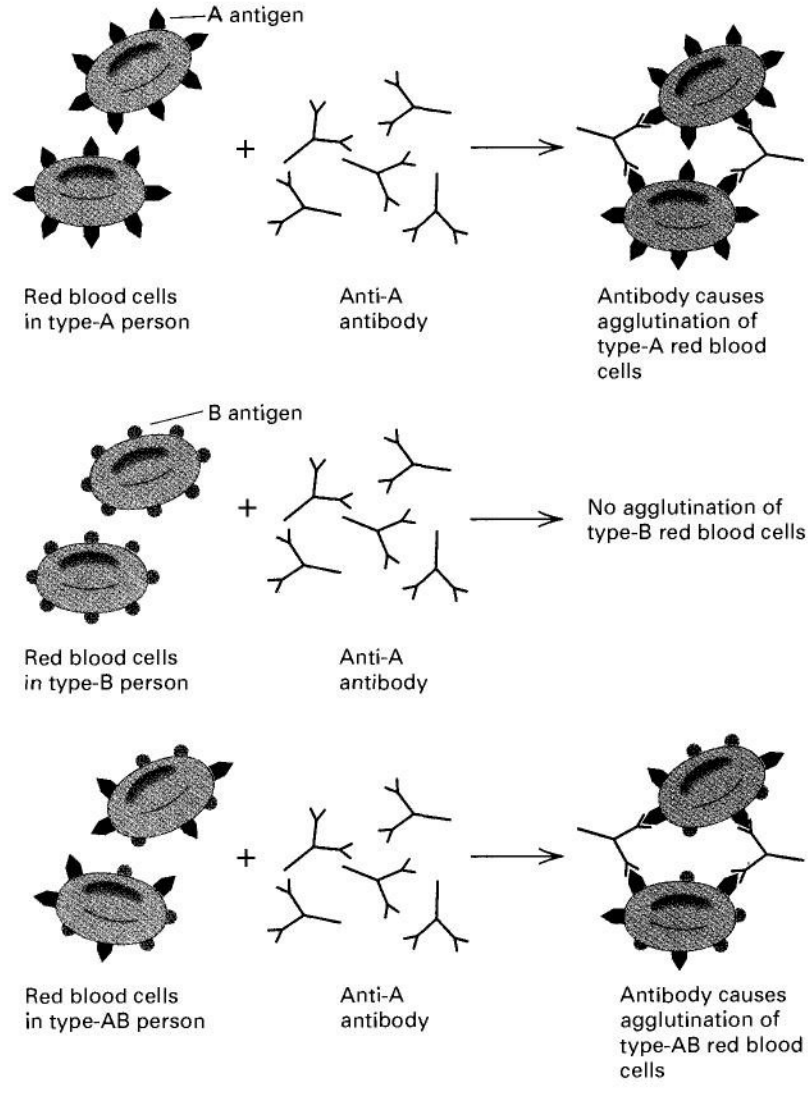


Figure 3.22 Antibody against type-A antigen agglutinates red blood cells that carry the type-A antigen, whether or not they also carry the type-B antigen. Blood fluid containing anti-A antibody agglutinates red blood cells of type A and type AB, but not red blood cells of type B or type O.

direction, in which the donor blood contains antibodies against the recipient's red blood cells, is usually acceptable because the donor's antibodies are diluted so rapidly that clumping is avoided. The types of compatible blood transfusions are shown in the last two columns of Table 3.3. Note that a person of blood type AB can receive blood from a person of any other ABO type; type AB is called a *universal recipient.* Conversely, a person of blood type O can donate blood to a person of any ABO type; type O is called a *universal donor.*

Epistasis

In Chapter 1 we saw how the products of several genes may be necessary to carry out all the steps in a biochemical pathway. In genetic crosses in which two mutations that affect different steps in a single pathway are both segregating, the typical F_2 ratio of 9 : 3 : 3 : 1 is not observed. Gene interaction that perturbs the normal Mendelian ratios is known as **epistasis.** One type of epistasis is illustrated by the interaction of the *C, c* and *P, p* allele pairs affecting flower coloration in peas. These genes encode enzymes in the biochemical pathway for the synthesis of anthocyanin pigment, and the production of anthocyanin requires the presence of at least one wildtype dominant allele of each gene. The proper way to represent this situation genetically is to write the required genotype as

$$C-\ P-$$

where each dash is a "blank" that may be filled with either allele of the gene. Hence *C*— comprises the genotypes *CC* and *Cc*, and likewise *P*— comprises the genotypes *PP* and *Pp*. All four genotypes included in the symbol *C*— *P*—, and only these genotypes, have purple flowers.

Figure 3.23 shows a cross between the homozygous recessive genotypes *pp* and *cc*. The phenotype of the flowers in the F_1 generation is purple because the genotype is *Cc Pp*. Self-fertilization of the F_1 plants (indicated by the encircled cross sign) results in the F_2 progeny genotypes shown in the Punnett square. Because only the *C*— *P*— progeny have purple flowers, the ratio of purple flowers to white flowers in the F_2 generation is 9 : 7. The epistasis does not change the result of independent segregation, it merely conceals the fact that the underlying ratio of the genotypes *C*— *P*— : *C*— *pp* : *cc P*— : *cc pp* is 9 : 3 : 3 : 1.

For a trait determined by the interaction of two genes, each with a dominant allele,

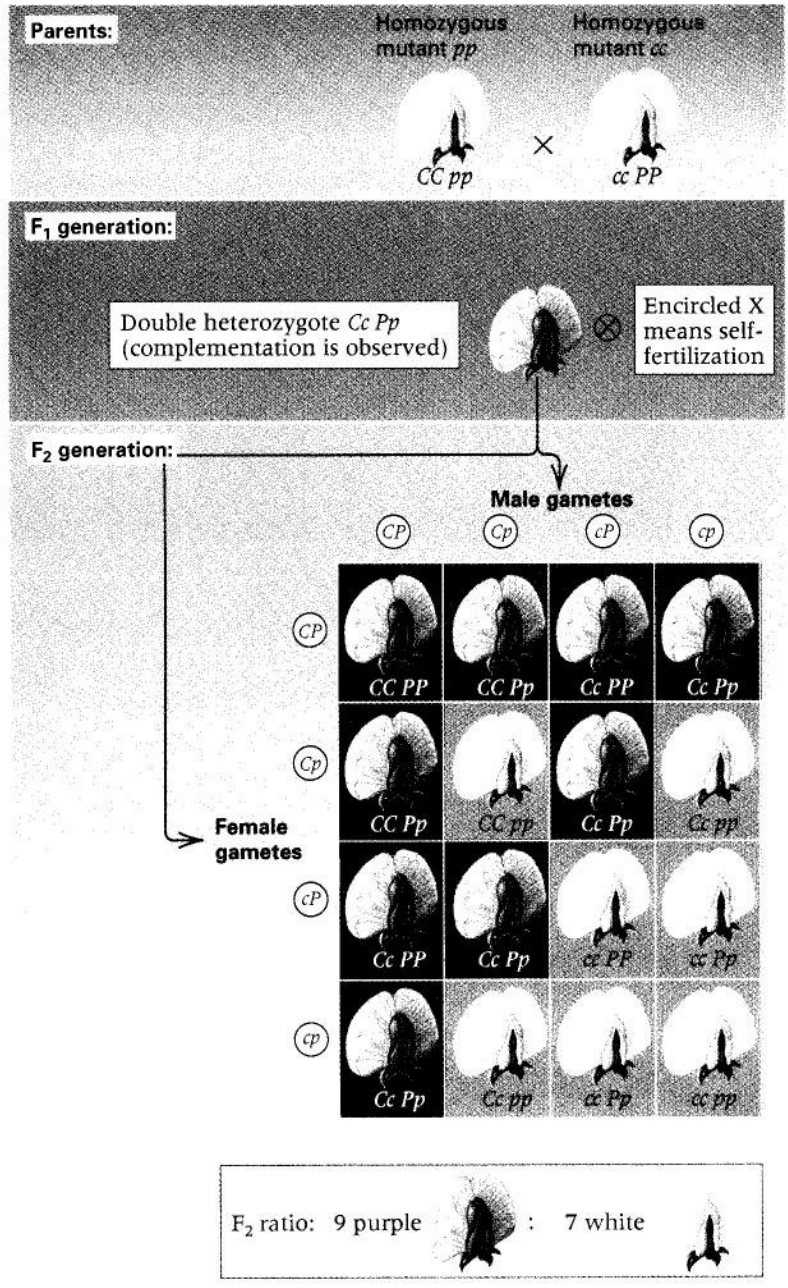


Figure 3.23 Epistasis in the determination of flower color in peas. Formation of the purple pigment requires the dominant allele of both the *C* and *P* genes. With this type of epistasis, the dihybrid F_2 ratio is modified to 9 purple : 7 white.

there are only a limited number of ways in which the 9 : 3 : 3 : 1 dihybrid ratio can be modified. The possibilities are illustrated in **Figure 3.24**. In part A are the genotypes produced in the F_2 generation by independent segregation. In the absence of epistasis, the F_2 ratio of phenotypes is 9 : 3 : 3 : 1. The possible modified ratios are shown in part B. In each row, the color coding indicates phenotypes that are indistinguishable because of epistasis, and the resulting modified ratio is given. For example, in the modified ratio at the bottom, the phenotypes of the "3 : 3 : 1" classes are indistinguishable, resulting in a 9 : 7 ratio. This is the ratio observed in the segregation of the *C, c* and *P, p* alleles in Figure 3.23, with its 9 : 7 ratio of purple to white flowers. Taking all the possible modified ratios in Figure 3.24B together, there are nine possible dihybrid ratios when both genes show complete dominance. Examples of each of the modified ratios are known. The most frequently encountered modified ratios are 9 : 7, 12 : 3 : 1, 13 : 3 , 9 : 4 : 3, and 9 : 6 : 1. The types of epistasis that result in these modified ratios are illustrated in the following examples, which are taken from a variety of organisms. Other examples can be found in the problems at the end of the chapter.

9 : 7 is observed when a homozygous recessive mutation in either or both of two different genes results in the same mutant phenotype, as in Figure 3.23.

12 : 3 : 1 results when the presence of a dominant allele at one locus masks the

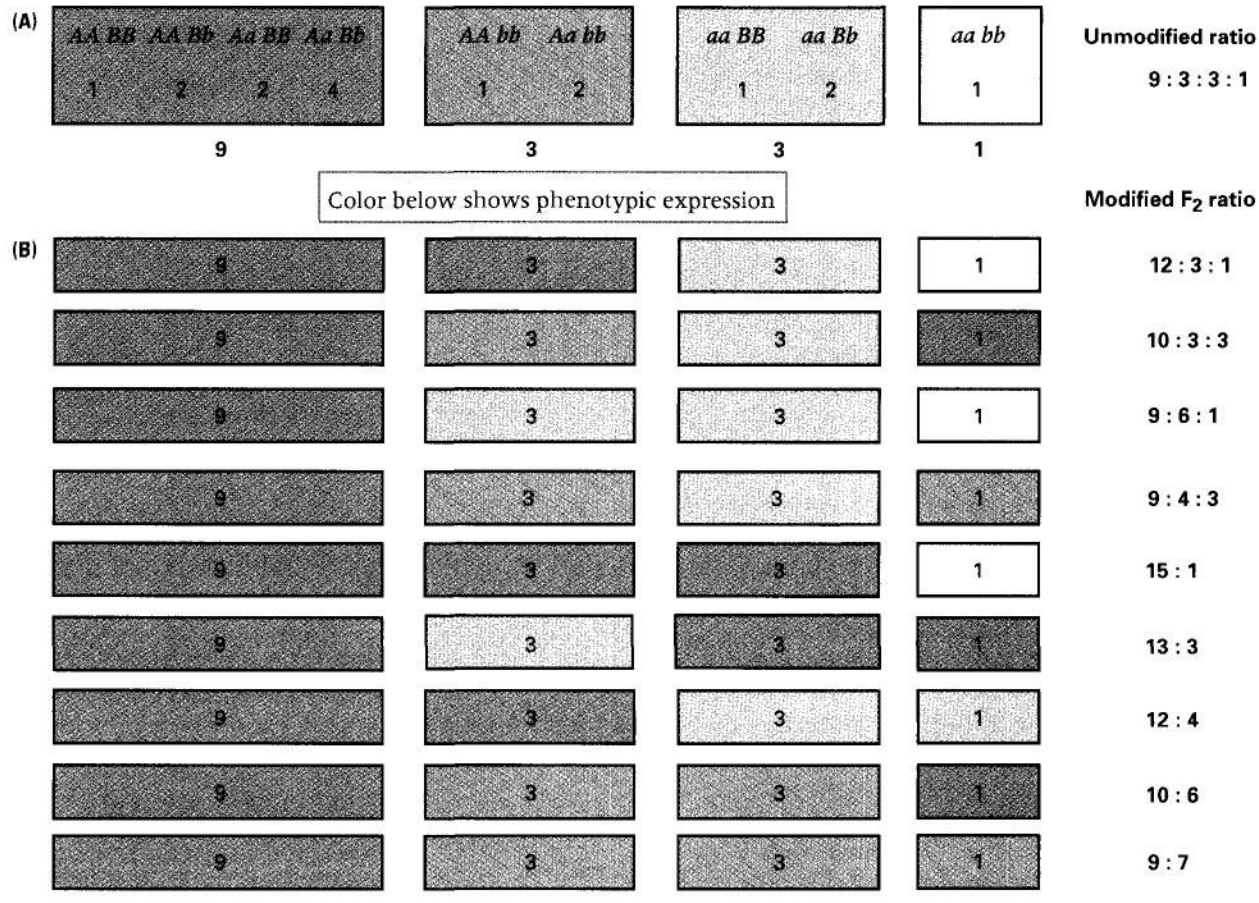
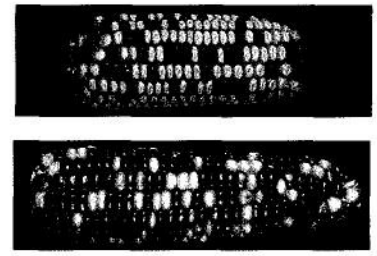


Figure 3.24 Modified F_2 dihybrid ratios. (A) The F_2 genotypes of two independently assorting genes with complete dominance result in a 9 : 3 : 3 : 1 ratio of phenotypes, provided there is no interaction (epistasis) between the genes. (B) If there is epistasis that renders two or more of the phenotypes indistinguishable (indicated by the colors), then the F_2 ratio is modified.

genotype at a different locus, such as the *A*— genotype rendering the *B*— and *bb* genotypes indistinguishable. For example, in genetic study of the color of the hull in oat seeds, a variety with white hulls was crossed with a variety with black hulls. The F_1 hybrid seeds had black hulls. Among 560 progeny in the F_2 generation, the hull phenotypes observed were 418 black, 106 gray, and 36 white. The ratio of phenotypes is 11.6 : 3.9 : 1, or very nearly 12 : 3 : 1. A genetic hypothesis to explain these results is that the black-hull phenotype is due to the presence of a dominant allele *A* and the gray-hull phenotype is due to another dominant allele *B* whose effect is apparent only in *aa* genotypes. On the basis of this hypothesis, the original varieties had genotypes *aa bb* (white) and *AA BB* (black). The F_1 has genotype *Aa Bb* (black). If the *A*, *a* allele pair and the *B*, *b* allele pair undergo independent assortment, then the F_2 generation is expected to have the genotypic and phenotypic composition 9/16 *A*— *B*— (black hull), 3/16 *A*— *bb* (black hull), 3/16 *aa B*— (gray hull), 1/16 *aa bb* (white hull), or 12 : 3 : 1.

13 : 3 is illustrated by the difference between White Leghorn chickens (genotype *CC II*) and White Wyandotte chickens (genotype *cc ii*). Both breeds have white feathers because the *C* allele is necessary for colored feathers, but the *I* allele in White Leghorns is a dominant inhibitor of feather coloration. The F_1 generation of a dihybrid cross between these breeds has the genotype *Cc Ii*, which is expressed as white feathers because of the inhibitory effects of the *I* allele. In the F_2 generation, only the *C*— *ii* genotype has colored feathers, so there is a 13 : 3 ratio of white : colored.

9 : 4 : 3 is observed when homozygosity for a recessive allele with respect to one gene masks the expression of the genotype of a different gene. For example, if the *aa* genotype has the same phenotype regardless of whether the genotype is *B*— or *bb*, then the 9 : 4 : 3 ratio results. As an example, in mice the grayish "agouti" coat color results from a horizontal band of yellow pigment just beneath the tip of each hair. The agouti pattern is due to a dominant allele *A*, and in *aa* animals the coat color is black. A second dominant allele, *C*, is necessary for the formation of hair pigments of any kind, and *cc* animals are albino (white). In a cross of *AA CC* (agouti) × *aa cc* (albino), the F_1 progeny are *Aa Cc* and phenotypically agouti. Crosses between F_1 males and females produce F_2 progeny in the proportions 9/16 *A*— *C*— (agouti), 3/16 *A*— *cc* (albino), 3/16 *aa C*— (black), 1/16 *aa cc* (albino), or 9 agouti : 4 albino : 3 black.



Ratios of phenotype in a dihybrid cross are quite readily seen in the color of kernals in an ear of corn: top, a 13:3 ratio; bottom, a 9:4:3 ratio. [Dan Guravich/Photo Researchers, Inc.]

9 : 6 : 1 implies that homozygosity for either of two recessive alleles yields the same phenotype but that the phenotype of the double homozygote is different. In Duroc–Jersey pigs, red coat color requires the presence of two dominant alleles *R* and *S*. Pigs of genotype *R*— *ss* and *rr S*— have sandy-colored coats, and *rr ss* pigs are white. The F_2 ratio is therefore 9/16 *R*— *S*— (red), 3/16 *R*— *ss* (sandy), 3/16 *rr S*— (sandy), 1/16 *rr ss* (white), or 9 red : 6 sandy : 1 white.

3.7 Genetic Analysis: Mutant Screens and the Complementation Test

When a geneticist makes a statement such as "Flower color in peas is determined by a single gene with dominant allele *P* and recessive allele *p*," this does not mean that only one gene is needed for flower color. Implicit in the statement is the existence of two strains or varieties, one with colored flowers and one with white flowers, in which the difference is caused by one strain having genotype *PP* and the other *pp*. Many genes beside *P* are also necessary for purple flower coloration. Among them are other genes in the biochemical pathway for the synthesis of anthocyanin, as well as different genes needed for expression of the biochemical pathway in developing flowers. A third variety of peas that has white flowers may not have the genotype *pp*. In this variety, the white flowers could be caused by a mutation in one of the other genes needed for flower color. We have already seen an example in Figure 3.23, in

which the white flower at the upper right results from genotype *cc*. Furthermore, a geneticist interested in understanding the genetic basis of flower color would rarely be satisfied with having identified the *P* gene alone or both the *P* gene and the *C* gene. The ultimate goal of a genetic analysis of flower color would be, by isolating new mutations, to identify every gene necessary for purple coloration and then, through further study of the mutant phenotypes, to determine the normal function of each of the genes that affect the trait.

The Complementation Test in Gene Identification

In a genetic analysis of flower color, a geneticist would begin by isolating many new mutants with white flowers. Although mutations are usually very rare, their frequency can be increased by treatment with radiation or certain chemicals. The isolation of a set of mutants, all of which show the same type of phenotype, is called a **mutant screen.** Among the mutants that are isolated, some will contain mutations in genes already identified. For example, a genetic analysis of flower color in peas might yield one or more new mutations that changed the wildtype *P* allele into a recessive allele that blocks the formation of the purple pigment. Each of these mutations might differ in DNA sequence, forming a set of multiple alleles that are all mutant forms of the wildtype *P* allele. On the other hand, a mutant screen should also yield mutations in genes not previously identified. Each of the new genes might be also represented by multiple mutant alleles.



White pea flowers (*Pisum sativum*). Peas are true-breeding. Self fertilization takes place within an individual plant if it is left undisturbed. To determine whether the mutant white coloring is the product of one or more genes, geneticists use the complementation test. [Nigel Cattlin/Holt Studios International/Photo Researchers, Inc.]

In a mutant screen for flower color, most of the new mutations will be recessive. This is because most of the new mutant alleles will encode an inactive protein of some kind, and most protein-coding genes are expressed at such a level that one copy of the wildtype allele is sufficient to yield a normal morphological phenotype. (The molecular phenotype is often intermediate; the heterozygotes have half as much protein as the wildtype homozygotes.) New recessive alleles are therefore identified in homozygous genotypes, because homozygous recessive plants have white flowers. Among all the new mutant strains that are isolated, some will have a recessive mutation in one gene, others in a second gene, still others in a third gene, and so on. Complicating the issue are multiple alleles, because two or more independently isolated mutations may be alleles of the same wildtype gene. How can the strains be sorted out? How can the geneticist determine which mutations are alleles of the same gene and which are mutations in different genes?

Because the new mutations are recessive, all one has to do is cross the homozygous genotypes. The phenotype of the F_1 progeny yields the answer. This principle is illustrated in **Figure 3.25.** In part A we suppose that the mutations are recessive alleles of the same gene. Let us call them a_1 and a_2. Then the parental mutant strains have genotypes a_1a_1 and a_2a_2, and the F_1 progeny have the genotype a_1a_2. Because there is no wildtype allele at the locus, the phenotype of the F_1 is white (mutant). This result is called **noncomplementation,** and it means that the parental strains are homozygous for recessive alleles of the *same* gene.

In part B we suppose the parental mutant strains are homozygous for recessive alleles of different genes, say a_1a_1 and b_1b_1. With respect to the *B* locus, the genotype of the a_1a_1 strain is *BB*, and with respect to the *A* locus, the genotype of the b_1b_1 strain is

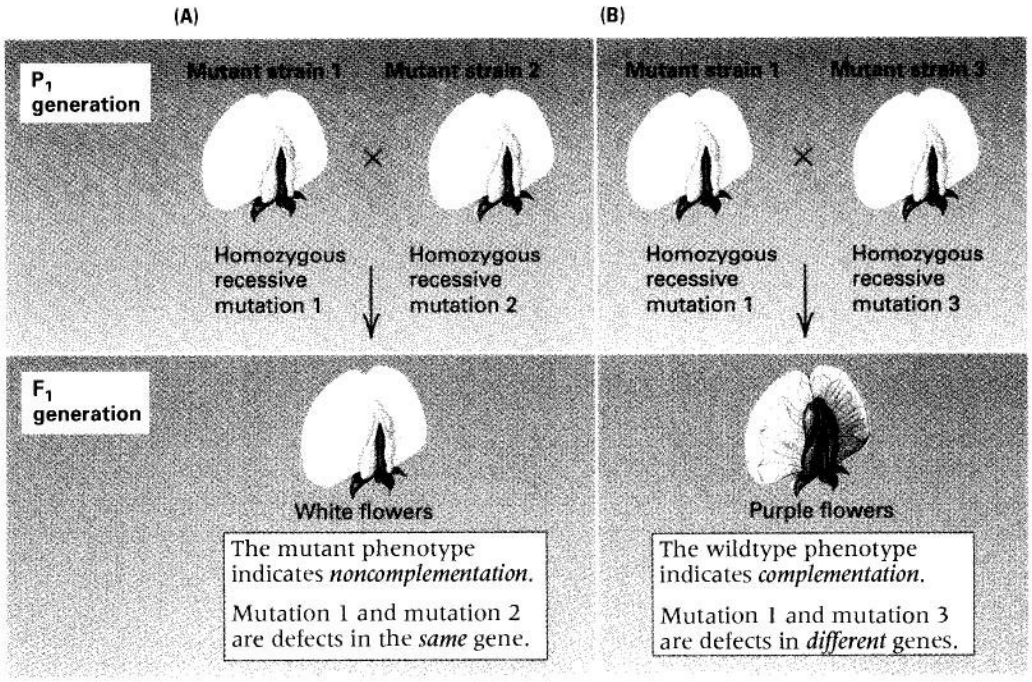


Figure 3.25 The complementation test reveals whether two recessive mutations are alleles of the same gene. In the complementation test, homozygous recessive genotypes are crossed. If the phenotype of the F_1 progeny is mutant (A), it means that the mutations in the parental strains are alleles of same gene. If the phenotype of the F_1 progeny is nonmutant (B), it means that the mutations in the parental strains are alleles of different genes.

AA; hence the genotypes of the mutant strains could be written more completely as a_1a_1 *BB* and *AA* b_1b_1. In this case, the F_1 progeny have the genotype Aa_1 Bb_1. The *A* allele masks the mutation a_1 and the *B* allele masks the mutation b_1; hence the phenotype is purple (wildtype). This result is called **complementation,** and it means that the parental strains are homozygous for recessive alleles of *different* genes.

The kind of cross illustrated in Figure 3.25 is a **complementation test.** As we have seen, it is used to determine whether recessive mutations in each of two different strains are alleles of the same gene. Because the result indicates the presence or absence of allelism, the complementation test is one of the key experimental operations in genetics. To illustrate the application of the test in practice, suppose a mutant screen were carried out to isolate more mutations for white flowers. Starting with a true-breeding strain with purple flowers, we treat pollen with x rays and use the irradiated pollen to fertilize ovules to obtain seeds. The F_1 seeds are grown and the resulting plants allowed to self-fertilize, after which the F_2 plants are grown. A few of the F_1 seeds may contain a new recessive mutation, but in this generation the genotype is heterozygous. Self-fertilization of such plants results in F_2 progeny in a ratio of 3 purple : 1 white. Because mutations resulting in a particular phenotype are quite rare, even when induced by radiation, only a few among many thousands of self-fertilized plants will be found to have a new white-flower mutation. Let us suppose that we were lucky enough to obtain six new mutations.

How are we going to name these mutations? We can make no assumptions about the number of genes represented. As far as we know at this stage, they could all be alleles of the same gene, or they could all be alleles of different genes. This is what we need to establish using the complementation test. For the moment, then, let us assign the mutations arbitrary names in series—*x1, x2, x3,* and so forth—with no implication about which may be alleles of each other.

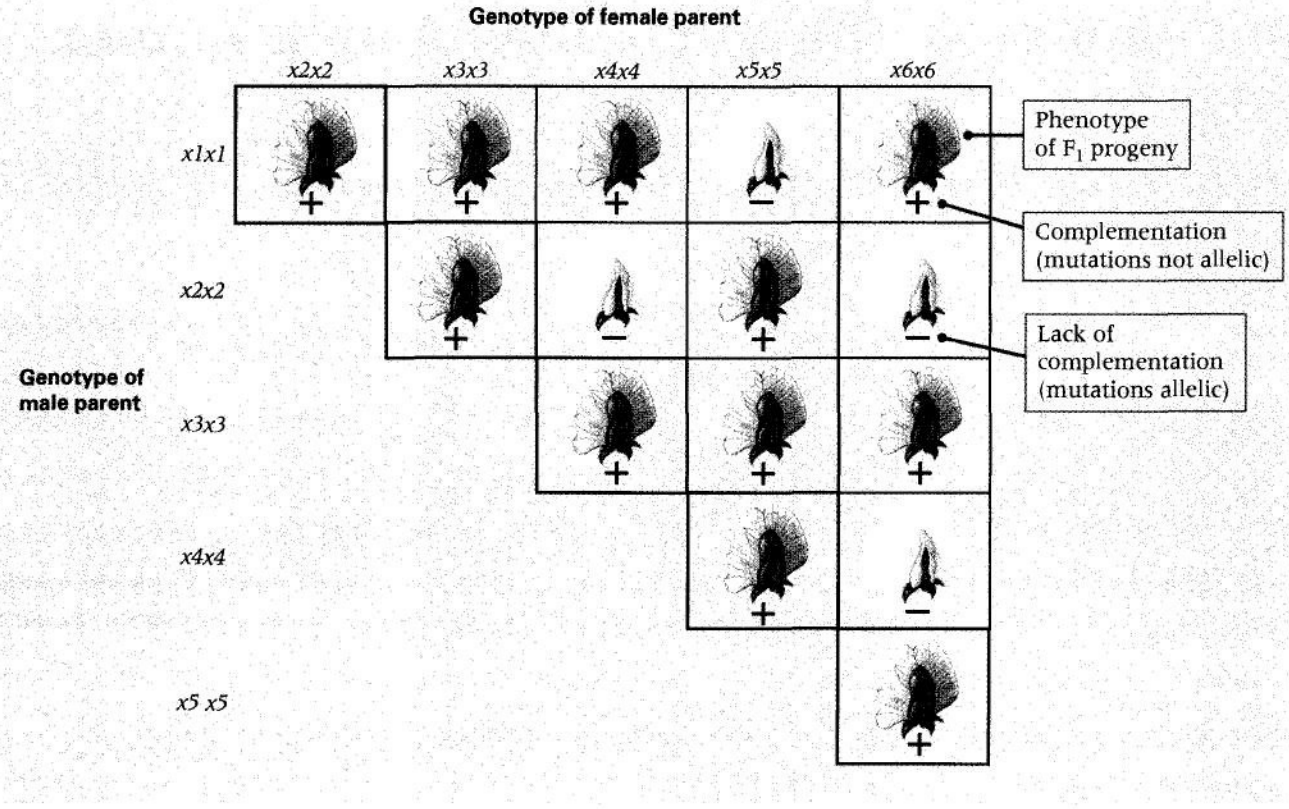


Figure 3.26 Results of complementation tests among six mutant strains of peas, each homozygous for a recessive allele resulting in white flowers. Each box gives the phenotype of the F_1 progeny of a cross between the male parent whose genotype is indicated in the far left column and the female parent whose genotype is indicated in the top row.

The next step is to classify the mutations into groups using the complementation test. **Figure 3.26** shows the results, conventionally reported in a triangular array of + and − signs. The crosses that yield F_1 progeny with the wildtype phenotype (in this case, purple flowers) are denoted with a + sign in the corresponding box, whereas those that yield F_1 progeny with the mutant phenotype (white flowers) are denoted with a − sign. The + signs indicate complementation between the mutant alleles in the parents, and the − signs indicate noncomplementation. (The bottom half of the triangle is unnecessary because reciprocal crosses yield the same results, and the diagonal elements are unnecessary because each strain is true-breeding within itself and so yields mutant progeny.) As we saw in Figure 3.25, complementation in a cross means that the parental strains have mutations in different genes. Lack of complementation means that the parental mutations are in the same gene. The following principle underlies the complementation test.

The Principle of Complementation:
If two recessive mutations are alleles of the same gene, then a cross between homozygous strains yields F_1 progeny that are mutant (noncomplementation); if they are alleles of different genes, then the F_1 progeny are wildtype (complementation).

In interpreting complementation data such as those in Figure 3.26, we actually apply the principle the other way around. We examine the phenotype of the F_1 progeny of each possible cross and then infer whether or not the parental strains have mutant alleles of the same gene.

In a complementation test, if the combination of two recessive mutations results in a mutant phenotype, then the mutations are regarded as alleles of the same gene; if the combination results in a wildtype phenotype, then the mutations are regarded as alleles of different genes.

A convenient way to analyze the data in Figure 3.26 is to arrange the alleles in a circle as shown in **Figure 3.27**A. Then, for each possible pair of mutations, connect the pair by a straight line if the mutations *fail* to complement (Figure 3.27B). According to the principle of complementation, the lines must connect mutations that are alleles of each other, because in a complementation test, lack of complementation means that the mutations are alleles. Each of the groups of noncomplementing mutations is called a **complementation group.** As we have seen, each complementation group defines a gene, so the complementation test provides the geneticist's technical definition:

> A gene is defined experimentally as a set of mutations that make up a single complementation group. Any pair of mutations within the group fail to complement one another, and the complementation test yields organisms with a mutant phenotype.

The mutations in Figure 3.27 therefore represent three genes, mutation of any one of which results in white flowers. Mutations *x1* and *x5* are alleles of one gene; *x3* is an allele of a different gene; and *x2*, *x4*, and *x6* are alleles of a third gene.

What about the genes *P* and *C* (Figure 3.23 on page 115) that also affect flower color? The complementation test tells us nothing about these, because we have not yet included them in any of the crosses. To test allelism with *P* and *C*, we would have to cross one mutant strain from each complementation group with strains of genotype *pp* and *cc*. If the F_1 progeny are mutant, it identifies the complementation group with an already known mutation. For example, suppose we cross each of the strains *x1x1*, *x3x3*, and *x4x4* with *pp* and *cc*. Suppose further that the progeny are all wildtype except in two crosses—*x1x1* × *pp* and *x3x3* × *cc*—in which the progeny are mutant. This result implies that the complementation group consisting of *x1* and *x5* also includes *p* and that the complementation group consisting of *x3* also includes *c*.

At this point in the genetic analysis, it is advisable to rename the mutations to indicate which ones are true alleles. (There is an old Chinese saying that the correct naming of things is the beginning of wisdom, and this is certainly true in the case of genes.) Because the *p* allele already had its name before the mutation screen was carried out, the new alleles of *p* (that is, *x1* and *x5*) should be renamed to reflect their allelism with *p*. Old names have priority, and so we must use the symbol *p* embellished with some sort of identifier. We might, for example, rename the *x1* and *x5* mutations p_2 and p_3 to indicate that they were the second and third mutations of *P* to be discovered and to convey their independent origins. Similarly, we might rename *x3* (the new allele of *c*) as c_2. The remaining complementation group consisting of *x2*, *x4*, and *x6* is a new one, and we can name it arbitrarily. For example, we might call the locus *albus* (Latin for "white") with the gene symbol *alb* and call the alleles alb_1, alb_2, and alb_3. The wildtype dominant allele of *alb*, which is necessary for purple coloration, would then be represented as *Alb* or as alb^+. The procedure of sorting new mutations into complementation groups and renaming them according to their allelism is an example of how geneticists identify genes and name alleles. Such renaming of alleles is the typical manner in which genetic terminology evolves as knowledge advances.

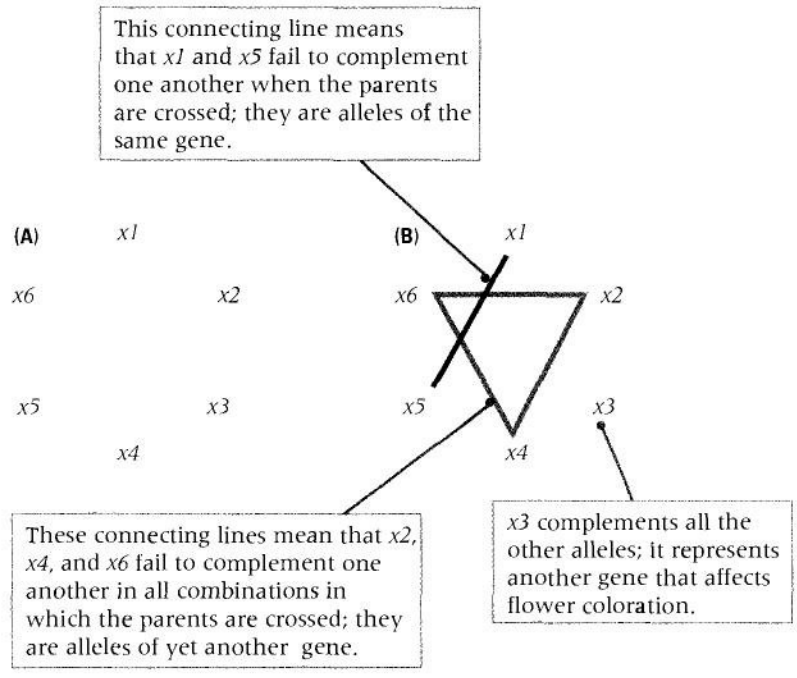


Figure 3.27 A method for interpreting the results of complementation tests. (A) Arrange the mutations in a circle. (B) Connect by a straight line any pair of mutations that fail to complement (that yield a mutant phenotype); any pair of mutations so connected are alleles of the same gene. In this example, there are three complementation groups, each of which represents a single gene needed for purple flower coloration.

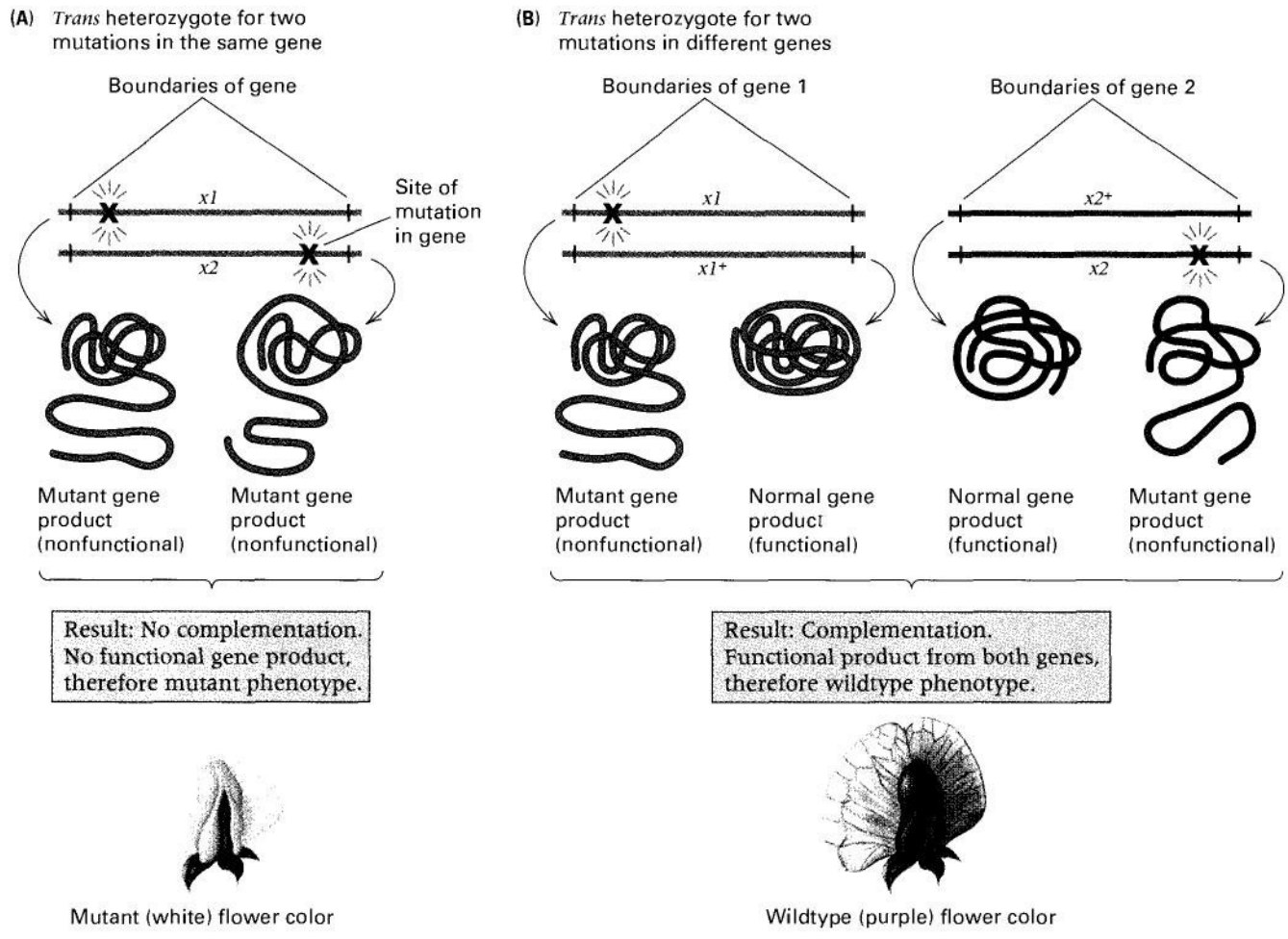


Figure 3.28 Molecular interpretation of a complementation test used in determining whether two mutations are alleles of the same gene (A) or alleles of different genes (B).

Why Does the Complementation Test Work?

The molecular basis of complementation is illustrated in **Figure 3.28**. Part A depicts the situation when two recessive mutations, *x1* and *x2*, each resulting in white flowers, are different mutations in the same gene. (The site of each mutation is indicated by a "burst.") The cross $x1x1 \times x2x2$ produces an F_1 hybrid in which *x1* and *x2* are in homologous DNA molecules; *x1* encodes a protein with one type of defect, and *x2* encodes a protein with a different type of defect, but both types of protein are nonfunctional. Hence alleles in the same gene yield a mutant phenotype (white flowers), because neither mutation encodes a wildtype form of the protein.

When the mutations are alleles of different genes, the situation is as depicted in Figure 3.28B. Because the mutations are in different genes, the homozygous *x1* strain is also homozygous for the wildtype allele $x2^+$ of the second locus; likewise, the homozygous *x2* strain is also homozygous for the wildtype allele $x1^+$ of the first locus. Hence, the same cross that yields the genotype *x1x2* in the case of allelic mutations (Figure 3.28A) yields $x1^+x1\ x2^+x2$ in the case of different genes. This is a double heterozygous genotype. Because the mutations are both recessive and in different genes, they do complement each other and yield an organism with a wildtype phenotype (purple flowers). With respect to the protein rendered defective by *x1*, there is a functional form encoded by the

wildtype allele brought in from the *x2x2* parent. With respect to the protein rendered defective by *x2*, there is again a functional form encoded by the wildtype allele brought in from the *x1x1* parent. Because a functional form of both proteins is produced, the result is a normal phenotype, or complementation.

Chapter Summary

Mendelian genetics deals with the hereditary transmission of genes from one generation to the next. One key principle is segregation, in which the two alleles in an individual separate during the formation of gametes so that each gamete is equally likely to contain either member of the pair. In a cross such as $AA \times aa$, in which only one gene is considered, the genotype of the offspring (constituting the F_1 generation) is heterozygous *Aa*. The phenotype of the F_1 progeny depends on the dominance relationships among the alleles. For many morphological traits, the wildtype allele, here denoted *A*, is dominant, and the phenotype of heterozygous *Aa* is indistinguishable from that of homozygous *AA*. In contrast, the alleles of molecular genetic markers are often codominant, and the phenotypes of *AA*, *Aa*, and *aa* are all distinct. For an RFLP marker, for example, the homozygous *AA* and *aa* genotypes have a phenotype consisting of a single band differing in electrophoretic mobility, whereas the heterozygous *Aa* genotype has a phenotype consisting of both bands.

In the formation of gametes, an *Aa* genotype produces *A*-bearing and *a*-bearing gametes in equal proportions. Hence in the $F_1 \times F_1$ cross $Aa \times Aa$, assuming random union of gametes in fertilization, the progeny (the F_2 generation) are expected to consist of genotypes $AA : Aa : aa$ in the proportions 1 : 2 : 1. The distribution of phenotypes in the F_2 generation again depends on the dominance relationships. If *A* is dominant to *a*, then the F_2 ratio of dominant : recessive phenotypes is expected to be 3 : 1. With codominance all three genotypes are distinguishable, and the ratio of F_2 phenotypes is 1 : 2 : 1.

For crosses involving two genes—for example, $AA\ BB \times aa\ bb$—the F_1 genotype is *Aa Bb*. Segregation of each gene implies that the ratios of $A : a$ and of $B : b$ gametes are both 1 : 1. If the genes are unlinked, they undergo independent segregation (independent assortment), and the gametic types $A\ B : A\ b : a\ B : a\ b$ are formed in the ratio 1 : 1 : 1 : 1. Hence the F_2 generation formed from the cross $F_1 \times F_1$ is expected to have genotypes given by the product of the expression $(1/4\ AA + 1/2\ Aa + 1/4\ aa) \times (1/4\ BB + 1/2\ Bb + 1/4\ bb)$. Using a dash to represent an allele of unspecified type, we can write the F_2 genotypes as $9\ A-\ B- : 3\ A-\ bb : 3\ aa\ B- : 1\ aa\ bb$, and if both *A* and *B* are dominant, the phenotypic ratio in the F_2 generation is 9 : 3 : 3 : 1. This ratio can be modified in various ways by interaction between the genes (epistasis). Different types of epistasis may result in dihybrid ratios such as 9 : 7 or 12 : 3 : 1 or 13 : 3 or 9 : 4 : 3.

The rules of probability provide the basis for predicting the outcomes of genetic crosses based on the principles of segregation and independent assortment. Two basic rules for combining probabilities are the addition rule and the multiplication rule. The addition rule applies to mutually exclusive events; it states that the probability of the realization of either one or the other of two events equals the sum of the respective probabilities. The multiplication rule applies to independent events; it states that the probability of the simultaneous realization of both of two events is equal to the product of the respective probabilities.

In some organisms, including human beings, it is not possible to perform controlled crosses, and genetic analysis is accomplished through the study of pedigrees through two or more generations. Pedigree analysis is the determination of the possible genotypes of the family members in a pedigree and of the probability that an individual member has a particular genotype. The goal of pedigree analysis is often to infer the genetic basis of an inherited disease or other condition—for example, to determine whether it may be due to a simple dominant or recessive allele. Although most morphological traits do not show simple Mendelian patterns of inheritance in pedigrees, molecular markers usually do. Prominent among these are SNPs (single-nucleotide polymorphisms), RFLPs (restriction fragment length polymorphisms), and STRPs (simple tandem repeat polymorphisms).

Multiple alleles are often encountered in natural populations or as a result of mutant screens. Genes with multiple alleles may have as few as three, such as those for the ABO blood groups, to as many as a hundred or more. Examples of large numbers of alleles include the genes used in DNA typing and the self-sterility alleles in some flowering plants. Although there may be multiple alleles in a population, each gamete can carry only one allele of each gene, and each organism can carry at most two different alleles of each gene.

The complementation test is the functional definition of a gene. If a cross between two homozygous recessive genotypes results in nonmutant progeny, the mutations are said to *complement* one another. Complementation is evidence that the alleles are mutations in different genes. On the other hand, if a cross between two homozygous recessives results in mutant progeny, then the alleles show *noncomplementation* (they fail to complement). Noncomplementation is evidence that the mutations are alleles of the *same* gene. For any group of recessive mutations, a complete complementation test entails crossing the homozygous recessives in all pairwise combinations. At the molecular level, lack of complementation implies that allelic mutations impair the function of the same protein molecule.

Key Terms

addition rule
antibody
antigen
backcross
carrier
codominance
complementation
complementation group
complementation test
consanguineous mating
epistasis
F_1 generation
F_2 generation
gamete
hybrid
incomplete dominance
independent assortment
Mendelian genetics
multiple alleles
multiplication rule
mutant screen
noncomplementation
outcrossing
P_1 generation
pedigree
penetrance
Punnett square
reciprocal cross
segregation
sib
sibling
testcross
transmission genetics
transposable element
true-breeding
wildtype
unlinked genes

Review the Basics

- What is the principle of segregation, and how is this principle demonstrated in the results of a single-gene (monohybrid) cross?
- What is the principle of independent assortment, and how is this principle demonstrated in the results of a two-gene (dihybrid) cross?
- Explain why random union of male and female gametes is necessary for Mendelian segregation and independent assortment to be observed in the progeny of a cross.
- What is the difference between mutually exclusive events and independent events? How are the probabilities of these two types of events combined? Give two examples of genetic events that are mutually exclusive and two examples of genetic events that are independent.
- When two pairs of alleles show independent assortment, under what conditions will a 9 : 3 : 3 : 1 ratio of phenotypes in the F_2 generation *not* be observed?
- Explain the following statement: "Among the F_2 progeny of a dihybrid cross, the ratio of genotypes is 1 : 2 : 1, but among the progeny that express the dominant phenotype, the ratio of genotypes is 1 : 2."
- What are the principal features of human pedigrees in which a rare dominant allele is segregating? In which a rare recessive allele is segregating?
- What is a mutant screen and how is it used in genetic analysis?
- Explain the statement: "In genetics, a gene is identified experimentally by a set of mutant alleles that fail to show complementation." What is complementation? How does a complementation test enable a geneticist to determine whether two different mutations are or are not mutations in the same gene?
- What does it mean to say that epistasis results in a "modified dihybrid F_2 ratio?" Give two examples of a modified dihybrid F_2 ratio, and explain the gene interactions that result in the modified ratio.

Guide to Problem Solving

Problem 1 Explain what each of the following ratios represents in the F_2 progeny of a single-gene (monohybrid) cross. Which are ratios of genotypes and which are ratios of phenotypes?
(a) 3 : 1
(b) 1 : 2 : 1
(c) 2 : 1

Answer **(a)** 3 : 1 is the expected ratio of phenotypes when one of the alleles is dominant. **(b)** 1 : 2 : 1 is the expected ratio of genotypes *AA* : *Aa* : *aa*. **(c)** 2 : 1 is the expected ratio of heterozygous *Aa* genotypes to homozygous *AA* genotypes.

Problem 2 Complete the table by inserting 0, 1, 1/2, or 1/4 for the probability of each genotype of progeny from each type of mating. For which mating are the parents identical in genotype but the progeny as variable in genotype as they can be for a single locus? For which mating are the parents as different as they can be for a single locus but the progeny identical to each other and different from either parent?

Mating	Progeny genotypes		
	AA	*Aa*	*aa*
AA × *AA*			
AA × *Aa*			
AA × *aa*			
Aa × *Aa*			
Aa × *aa*			
aa × *aa*			

Answer This table is to transmission genetics what the multiplication table is to arithmetic. It is fundamental to being able to solve almost any type of quantitative problem in transmission genetics. It needs to be thoroughly understood—memorized, if necessary!

Mating	Progeny genotypes		
	AA	*Aa*	*aa*
AA × *AA*	1	0	0
AA × *Aa*	1/2	1/2	0
AA × *aa*	0	1	0
Aa × *Aa*	1/4	1/2	1/4
Aa × *aa*	0	1/2	1/2
aa × *aa*	0	0	1

The mating for which the parents are identical in genotype, but the progeny are as variable in genotype as they can be for a single locus, is *Aa* × *Aa*. The mating for which the parents are as different as they can be for a single locus, but the progeny are identical to each other and different from either parent, is *AA* × *aa*.

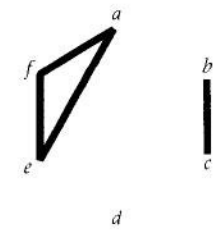
Problem 3 The data in the accompanying complementation matrix summarize the result of crosses between mutants designated *a* through *f*.

(a) Configure the allele symbols in the form of a circle, and use straight lines to connect the alleles that are in the same complementation group.

(b) State the conclusion in words: How many complementation groups are indicated, and which mutants are in each complementation group?

	a	*b*	*c*	*d*	*e*	*f*
a	−	+	+	+	−	−
b		−	−	+	+	+
c			−	+	+	+
d				−	+	+
e					−	−
f						−

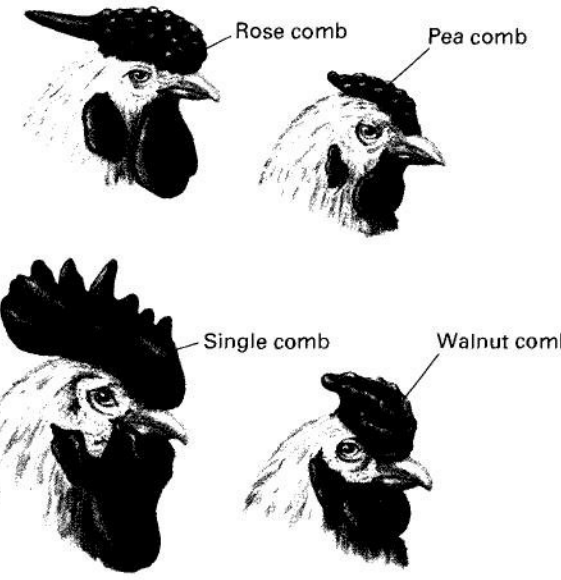
Answer **(a)** Follow the strategy outlined in the text, in which the alleles are placed in the shape of a circle and pairs of alleles that fail to show complementation are connected by a line. The resulting pattern is a visual representation of the complementation groups.



(b) The mutants define three complementation groups ("genes"). One complementation group consists of the alleles *a*, *e*, and *f*; another of the alleles *b* and *c*; and the other of the allele *d* only.

Problem 4 The accompanying illustration shows four alternative types of combs in chickens; they are called rose, pea, single, and walnut. The following data summarize the results of crosses. The rose and pea strains used in crosses 1, 2, and 5 are true breeding.

1. rose × single → rose
2. pea × single → pea
3. (rose × single) F_1 × (rose × single) F_1 → 3 rose : 1 single
4. (pea × single) F_1 × (pea × single) F_1 → 3 pea : 1 single
5. rose × pea → walnut
6. (rose × pea) F_1 × (rose × pea) F_1 → 9 walnut : 3 rose : 3 pea : 1 single



(a) What genetic hypothesis can explain these results?

(b) What are the genotypes of parents and progeny in each of the crosses?

(c) What are the genotypes of true-breeding strains of rose, pea, single, and walnut?

Answer **(a)** Cross 6 gives the Mendelian ratios expected when two genes are segregating, so a genetic hypothesis with two genes is necessary. Crosses 1 and 3 give the results expected if rose comb were due to a dominant allele (say, *R*). Crosses 2 and 4 give the results expected if pea comb were due to a dominant allele (say, *P*). Cross 5 indicates that walnut comb results from the interaction of *R* and *P*. The segregation in cross 6 means that *R* and *P* are not alleles of the same gene.

(b) 1. $RR\ pp \times rr\ pp \rightarrow Rr\ pp$
2. $rr\ PP \times rr\ pp \rightarrow rr\ Pp$
3. $Rr\ pp \times Rr\ pp \rightarrow 3/4\ R-\ pp : 1/4\ rr\ pp$
4. $rr\ Pp \times rr\ Pp \rightarrow 3/4\ rr\ P- : 1/4\ rr\ pp$
5. $RR\ pp \times rr\ PP \rightarrow Rr\ Pp$
6. $Rr\ Pp \times Rr\ Pp \rightarrow$
$9/16\ R-\ P- : 3/16\ R-\ pp : 3/16\ rr\ P- : 1/16\ rr\ pp$

(c) The true-breeding genotypes are *RR pp* (rose), *rr PP* (pea), *rr pp* (single), and *RR PP* (walnut).

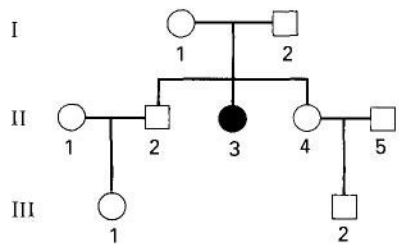
Analysis and Applications

3.1 What gametes can be formed by an individual organism of genotype *Aa*? Of genotype *Bb*? Of genotype *Aa Bb*?

3.2 How many different gametes can be formed by an organism with genotype *AA Bb Cc Dd Ee* and, in general, by an organism that is heterozygous for *m* genes and homozygous for *n* genes?

3.3 Round pea seeds are planted that were obtained from the F_2 generation of a cross between a true-breeding strain with round seeds and a true-breeding strain with wrinkled seeds. The pollen was collected and used *en masse* to fertilize plants from the true-breeding wrinkled strain. What fraction of the progeny is expected to have wrinkled seeds?

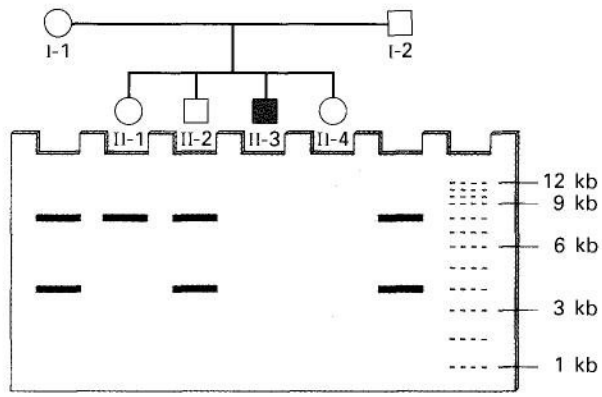
3.4 Phenylketonuria is a recessive inborn error of metabolism of the amino acid phenylalanine that results in severe mental retardation of affected children. The female II-3 (red circle) in the pedigree shown here is affected. If persons III-1 and III-2 (they are first cousins) mate, what is the probability that their offspring will be affected? (Assume that persons II-1 and II-5 are homozygous for the normal allele.)



3.5 Shown above right are a pedigree and gel diagram indicating the clinical phenotypes with respect to phenylketonuria and the molecular phenotypes with respect to an RFLP that overlaps the *PAH* gene for phenylalanine hydroxylase. Individual II-3 is affected.

(a) Indicate the expected molecular phenotype of II-3.

(b) Indicate the possible molecular phenotypes of II-4.



3.6 Assuming equal numbers of boys and girls, if a mating has already produced a girl, what is the probability that the next child will be a boy? If a mating has already produced two girls, what is the probability that the next child will be a boy? On what type of probability argument do you base your answers?

3.7 Assuming equal sex ratios, what is the probability that a sibship of four children consists entirely of boys? Of all boys or all girls? Of equal numbers of boys and girls?

3.8 In the following questions, you are asked to deduce the genotype of certain parents in a pedigree. The phenotypes are determined by dominant and recessive alleles of a single gene.

(a) A homozygous recessive results from the mating of a heterozygote and a parent with the dominant phenotype? What does this tell you about the genotype of the parent with the dominant phenotype?

(b) Two parents with the dominant phenotype produce nine offspring. Two have the recessive phenotype. What does this tell you about the genotype of the parents?

(c) One parent has a dominant phenotype, and the other has a recessive phenotype. Two offspring result, and both have the dominant phenotype. What genotypes are possible for the parent with the dominant phenotype?

GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

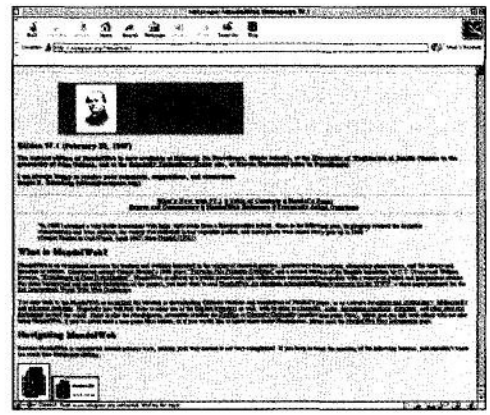
- Mendel's paper is one of the few nineteenth century scientific papers that reads almost as clearly as if it had been written today. It is important reading for every aspiring geneticist. You can access a conveniently annotated text using the keyword **Mendel**. Although modern geneticists make a clear distinction between genotype and phenotype, Mendel made no clear distinction between these concepts. At this keyword site you will find a treasure trove of information about Mendel, including his famous paper, essays, commentary, and a collection of images—all richly linked to additional Internet resources.

- **Huntington disease** is a devastating degeneration of the brain that begins in middle life. It affects about 30,000 Americans, and, because of its dominant mode of inheritance and complete penetrance, each of their 150,000 siblings and children has a 50–50 chance of developing the disease. Named after George Huntington, a Long Island physician who first described it in 1872, the disease's principal symptom is an involuntary, jerky motion of the head, trunk, and limbs called *chorea*, after the Greek word for "dance." At this keyword site you can learn about genetic testing programs, early symptoms, and the time course of the disease.

- The red and purple colors of flowers, as well as of autumn leaves, result from members of a class of pigments called anthocyanins. The biochemical pathway for **anthocyanin** synthesis in the snapdragon, *Antirrhinum majus*, can be found at this keyword site. The enzyme responsible for the first step in the pathway (chalcone synthase) limits the amount of pigment formed, which explains why red and white flowers in *Antirrhinum* show incomplete dominance.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 3 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 3.



3.9 Pedigree analysis tells you that a particular parent may have the genotype *AA BB* or *AA Bb*, each with the same probability. Assuming independent assortment, what is the probability of this parent's producing an *Ab* gamete? What is the probability of the parent's producing an *AB* gamete?

3.10 Assume that the trihybrid cross *AA BB rr* × *aa bb RR* is made in a plant species in which *A* and *B* are dominant but there is no dominance between *R* and *r*. Consider the F_2 progeny from this cross, and assume independent assortment.

(a) How many phenotypic classes are expected?

(b) What is the probability of the parental *aa bb RR* genotype?

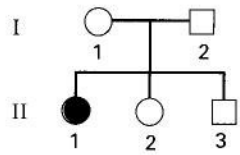
(c) What proportion would be expected to be homozygous for all three genes?

3.11 In the cross *Aa Bb Cc Dd* × *Aa Bb Cc Dd*, in which all genes undergo independent assortment, what proportion of offspring are expected to be heterozygous for all four genes?

3.12 The pattern of coat coloration in dogs is determined by the alleles of a single gene, with *S* (solid) being dominant over *s* (spotted). Black coat color is determined by the dominant allele *A* of a second gene, tan by homozygosity for the recessive allele *a*. A female having a solid tan coat is mated with a male having a solid black coat and produces a litter of six pups. The phenotypes of the pups are 2 solid tan, 2 solid black, 1 spotted tan, and 1 spotted black. What are the genotypes of the parents?

3.13 In the human pedigree shown here, the daughter indicated by the red circle (II-1) has a form of deafness determined by a recessive allele. What is the probability

that the phenotypically normal son (II-3) is heterozygous for the gene?



3.14 Huntington disease is a rare neurodegenerative human disease determined by a dominant allele, *HD*. The disorder is usually manifested after the age of forty-five. A young man has learned that his father has developed the disease.

(a) What is the probability that the young man will later develop the disorder?

(b) What is the probability that a child of the young man carries the *HD* allele?

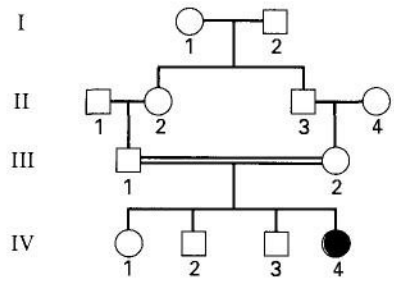
3.15 Assume that the trait in the accompanying pedigree is due to simple Mendelian inheritance.

(a) Is it likely to be due to a dominant allele or a recessive allele? Explain.

(b) What is the meaning of the double horizontal line connecting III-1 with III-2?

(c) What is the biological relationship between III-1 and III-2?

(d) If the allele responsible for the condition is rare, what are the most likely genotypes of all of the persons in the pedigree in generations I, II, and III? (Use *A* and *a* for the dominant and recessive alleles, respectively.)

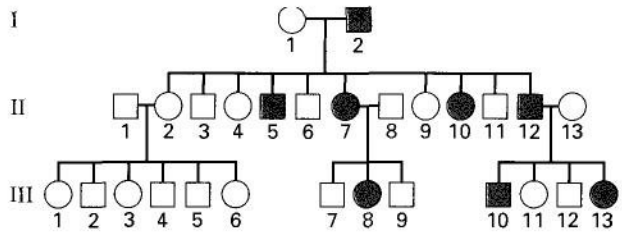


3.16 The Hopi, Zuni, and some other Southwest American Indians have a relatively high frequency of albinism (absence of skin pigment) resulting from homozygosity for a recessive allele, *a*. A normally pigmented man and woman, each of whom has an albino parent, have two children. What is the probability that both children are albino? What is the probability that at least one of the children is albino?

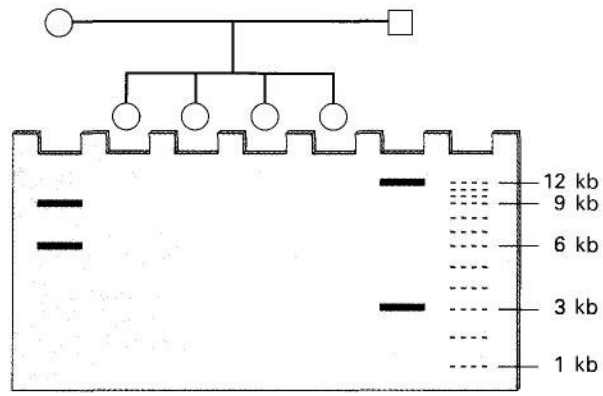
3.17 Say the trait in the indicated pedigree is due to simple Mendelian inheritance.

(a) Is it likely to be due to a dominant allele or a recessive allele? Explain.

(b) What are the most likely genotypes of all of the persons in the pedigree? (Use *A* and *a* for the dominant and recessive alleles.)



3.18 The accompanying pedigree and gel diagram show the phenotypes of the parents for an RFLP that has multiple alleles. What are the possible phenotypes of the progeny, and in what proportions are they expected?



3.19 Red kernel color in wheat results from the presence of at least one dominant allele of each of two independently segregating genes (in other words, *R*− *B*− genotypes have red kernels). Kernels on *rr bb* plants are white, and the genotypes *R*− *bb* and *rr B*− result in brown kernel color. Suppose that plants of a variety that is true breeding for red kernels are crossed with plants true breeding for white kernels.

(a) What is the expected phenotype of the F_1 plants?

(b) What are the expected phenotypic classes in the F_2 progeny and their relative proportions?

3.20 Heterozygous *Cp cp* chickens express a condition called creeper, in which the leg and wing bones are shorter than normal (*cp cp*). The dominant *Cp* allele is lethal when homozygous. Two alleles of an independently segregating gene determine white (*W*−) versus yellow (*ww*) skin color. From matings between chickens heterozygous for both of these genes, what phenotypic classes will be represented among the viable progeny, and what are their expected relative frequencies?

3.21 White Leghorn chickens are homozygous for a dominant allele, *C*, of a gene responsible for colored feathers, and also for a dominant allele, *I*, of an independently segregating gene that prevents the expression of *C*. The White Wyandotte breed is homozygous recessive for both genes *cc ii*. What proportion of the F_2 progeny obtained from mating White Leghorn × White Wyandotte F_1 hybrids would be expected to have colored feathers?

3.22 The F_2 progeny from a particular cross exhibit a modified dihybrid ratio of 9 : 7 (instead of 9 : 3 : 3 : 1). What phenotypic ratio would be expected from a testcross of the F_1?

3.23 Black hair in rabbits is determined by a dominant allele, *B*, and white hair by homozygosity for a recessive allele, *b*. Two heterozygotes mate and produce a litter of three offpring.

(a) What is the probability that the offspring are born in the order white-black-white? What is the probability that the offspring are born in either the order white-black-white or the order black-white-black?

(b) What is the probability that exactly two of the three offspring will be white?

3.24 Consider sibships consisting of 6 children, and assume a sex ratio of 1 : 1.

(a) What is the proportion with no girls?

(b) What is the proportion with exactly 1 girl?

(c) What is the proportion with exactly 2 girls?

(d) What is the proportion with exactly 3 girls?

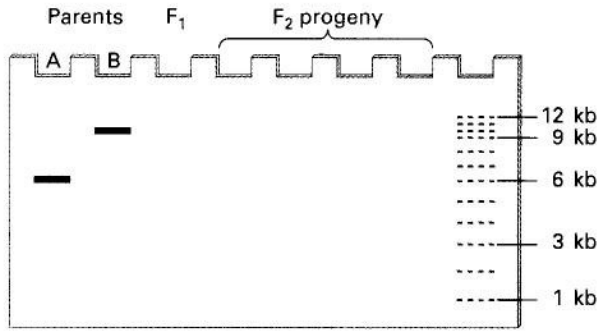
(e) What is the proportion with 3 or more boys?

3.25 Andalusian fowls are colored black, splashed white (resulting from an uneven sprinkling of black pigment through the feathers), or slate blue. Black and splashed white are true breeding, and slate blue is a hybrid that segregates in the ratio 1 black : 2 slate blue : 1 splashed white. If a pair of blue Andalusians is mated and the hen lays three eggs, what is the probability that the chicks hatched from these eggs will be one black, one blue, and one splashed white?

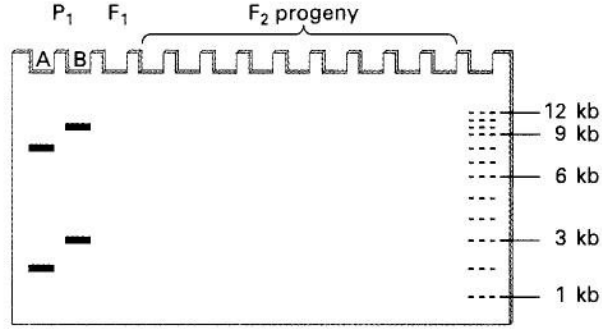
3.26 In the mating $Aa \times Aa$, what is the smallest number of offspring, *n*, for which the probability of at least one *aa* offspring exceeds 95 percent?

3.27 From the F_2 generation of a cross between mouse genotypes $AA \times aa$, one male progeny of genotype $A-$ was chosen and mated with an *aa* female. All of the progeny in the resulting litter were $A-$. From this result you would like to conclude that the sire's genotype is *AA*. How much confidence could you have in this conclusion for each litter size from 1 to 15? (In other words what is the probability that the sire's genotype is *AA*, given that the a priori probability is 1/3 and that a litter of *n* pups resulted in all $A-$ progeny?)

3.28 The accompanying gel diagram includes the phenotype of two parents (A and B) with respect to two different RAPD polymorphisms. Each parent is homozygous for an allele associated with a band defining a RAPD polymorphism. The two RAPD polymorphisms are at different loci and undergo independent assortment. In the gel diagram, indicate the expected phenotype of the F_1 progeny as well as all possible phenotypes of the F_2 progeny, along with their expected proportions.



3.29 The gel diagram shown below shows the phenotype of two parents (A and B), each homozygous for two RFLPs that undergo independent assortment. Parent A has genotype $A_1A_1\,B_1B_1$, where the A_1 allele yields a band of 2 kb and the B_1 allele yields a band of 8 kb. Parent B has genotype $A_2A_2\,B_2B_2$, where the A_2 allele yields a band of 3 kb and the B_2 allele yields a band of 10 kb. Show the expected phenotype of the F_1 progeny as well as all possible phenotypes of the F_2 progeny, along with their expected proportions.

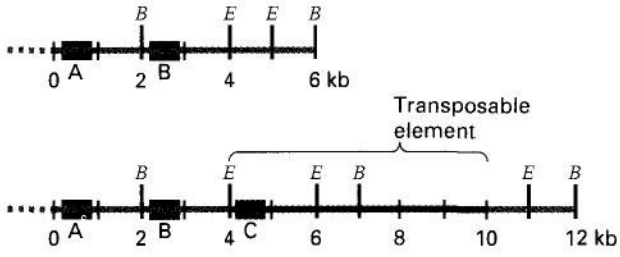


3.30 Complementation tests of the recessive mutant genes *a* through *f* produced the data in the accompanying matrix. The circles represent missing data. Assuming that all of the missing mutant combinations would yield data consistent with the entries that are known, complete the table by filling each circle with a + or − as needed.

	a	*b*	*c*	*d*	*e*	*f*
a	○	+	−	○	+	○
b		○	○	○	○	−
c			○	−	○	○
d				○	○	○
e					○	+
f						○

Challenge Problems

3.31 Diagrammed here is DNA from a wildtype gene (top) and a mutant allele (bottom) that has an insertion of a transposable element that inactivates the gene. The transposable element is present in many copies scattered throughout the genome. The symbols *B* and *E* represent the positions of restriction sites for *Bam*HI and *Eco*RI, respectively, and the rectangles show sites of hybridization with each of three probes (A, B, and C) that are available. The dots at the left indicate that the nearest site of either *Bam*HI or *Eco*RI cleavage is very far to the left of the region shown. Explain which probe and which single restriction enzyme you would use for RFLP analysis to identify both alleles. Also explain why any other choices would be unsuitable.



3.32 Meiotic drive is an unusual phenomenon in which two alleles do not show Mendelian segregation from the heterozygous genotype. Examples are known from mammals, insects, fungi, and other organisms. The usual mechanism is one in which both types of gametes are formed, but one of them fails to function normally. The excess of the driving allele over the other can range from a small amount to nearly 100 percent. Suppose that *D* is an allele showing meiotic drive against its alternative allele *d*, and suppose that *Dd* heterozygotes produce functional *D*-bearing and *d*-bearing gametes in the proportions 3/4 : 1/4. In the mating *Dd* × *Dd*,

(a) What are the expected proportions of *DD*, *Dd*, and *dd* genotypes?

(b) If *D* is dominant, what are the expected proportions of *D*− and *dd* phenotypes?

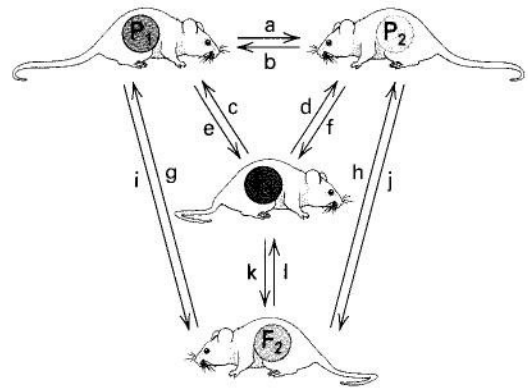
(c) Among the *D*− phenotypes, what is the ratio of *DD* : *Dd*?

(d) Answer parts (a) through (c), assuming that the meiotic drive takes place in only one sex.

3.33 The accompanying table summarizes the effect of inherited tissue antigens on the acceptance or rejection of transplanted tissues, such as skin grafts, in mammals. The tissue antigens are determined in a codominant fashion, so that tissue taken from a donor of genotype *Aa* carries both the A and the a antigen. In the table, the + sign means that a graft of donor tissue is accepted by the recipient, and the − sign means that a graft of donor tissue is rejected by the recipient. The rule is that *any graft will be rejected whenever the donor tissue contains an antigen not present in the recipient.* In other words, any transplant will be accepted if, and only if, the donor tissue does *not* contain an antigen different from any already present in the recipient.

		Donor		
		AA	*Aa*	*aa*
Recipient	*AA*	+		
	Aa	+	+	+
	aa			+

The diagram illustrated below shows all possible skin grafts between inbred (homozygous) strains of mice (P_1 and P_2) and their F_1 and F_2 progeny. Assume that the inbred lines P_1 and P_2 differ in only one tissue-compatibility gene. For each of the arrows, what is the probability of acceptance of a graft in which the donor is an animal chosen at random from the population shown at the base of the arrow and the recipient is an animal chosen at random from the population indicated by the arrowhead?



Further Reading

Bowler, P. J. 1989. *The Mendelian Revolution.* Baltimore, MD: Johns Hopkins University Press.

Carlson, E. A. 1987. *The Gene: A Critical History.* 2d ed. Philadelphia: Saunders.

Dunn, L. C. 1965. *A Short History of Genetics.* New York: McGraw-Hill.

Hartl, D. L., and V. Orel. 1992. What did Gregor Mendel think he discovered? *Genetics* 131: 245.

Hawley, R. S., and C. A. Mori. 1999. *The Human Genome: A User's Guide.* San Diego: Academic.

Henig, R. M. 2000. *The Monk in the Garden.* Boston, MA: Houghton Mifflin.

Judson, H. F. 1996. *The Eighth Day of Creation: The Makers of the Revolution in Biology.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Lewis, R. 2001. *Human Genetics: Concepts and Applications.* 4th ed. Dubuque, IA: McGraw-Hill.

Lewontin, R. C. 2000. *It Ain't Necessarily So: The Dream of the Human Genome and Other Illusions.* New York: New York Review of Books.

Maroni, G. 2000. *Molecular and Genetic Analysis of Human Traits.* Malden, MA: Blackwell.

Mendel, G. 1866. Experiments in plant hybridization. (Translation.) In *The Origins of Genetics: A Mendel Source Book,* ed. C. Stern and E. Sherwood. 1966. New York: Freeman.

Olby, R. C. 1966. *Origins of Mendelism.* London: Constable.

Orel, V. 1996. *Gregor Mendel: The First Geneticist.* Oxford, England: Oxford University Press.

Orel, V., and D. L. Hartl. 1994. Controversies in the interpretation of Mendel's discovery. *History and Philosophy of the Life Sciences* 16: 423.

Sandler, I. 2000. Development: Mendel's legacy to genetics. *Genetics* 154: 7.

Stern, C., and E. Sherwood, eds. 1966. *The Origins of Genetics: A Mendel Source Book.* New York: Freeman.

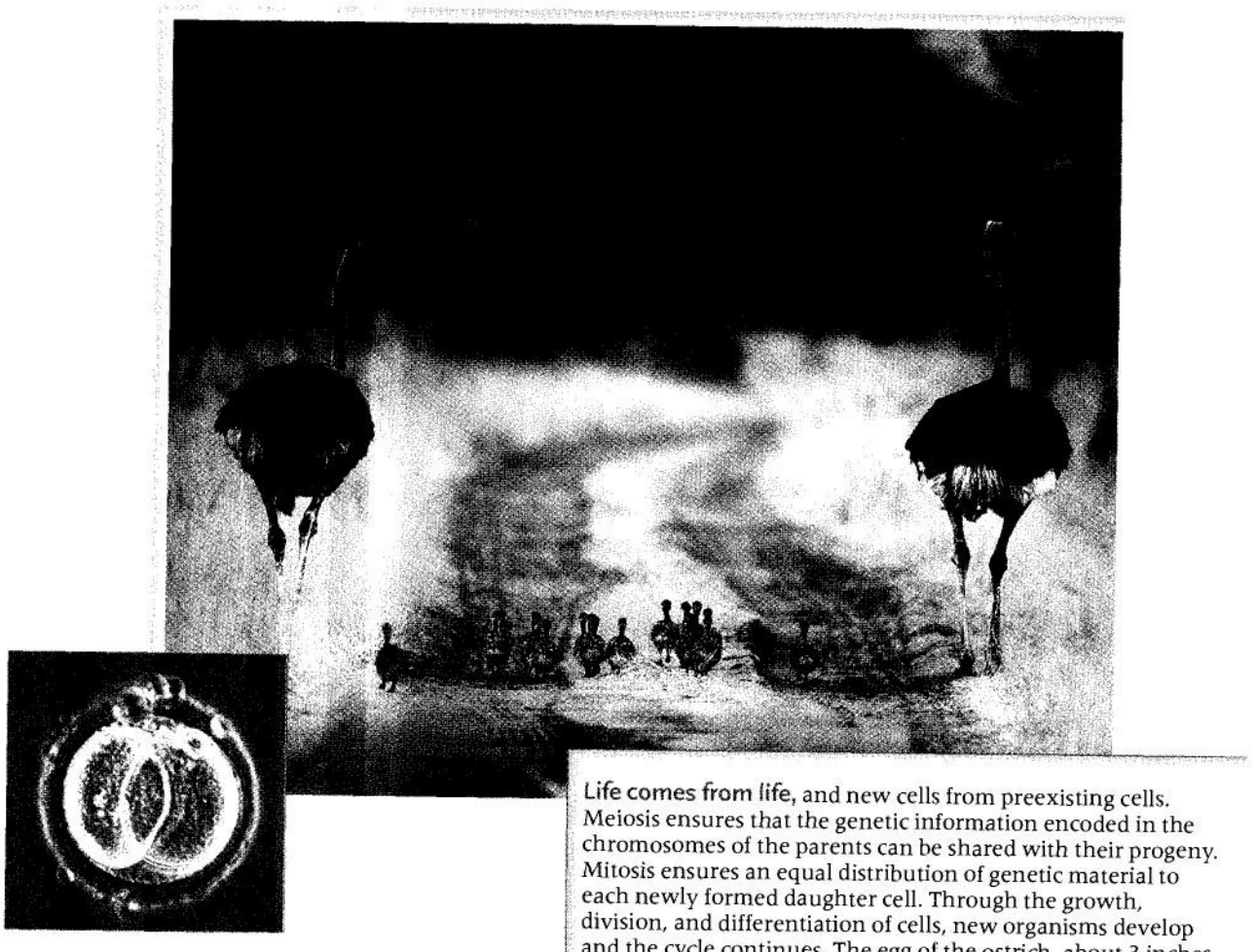
Sturtevant, A. H. 1965. *A Short History of Genetics.* New York: Harper & Row.

Sykes, B, ed. 1999. *The Human Inheritance.* New York: Oxford University Press.

CHAPTER

4

Genes and Chromosomes



Life comes from life, and new cells from preexisting cells. Meiosis ensures that the genetic information encoded in the chromosomes of the parents can be shared with their progeny. Mitosis ensures an equal distribution of genetic material to each newly formed daughter cell. Through the growth, division, and differentiation of cells, new organisms develop and the cycle continues. The egg of the ostrich, about 3 inches in diameter, is the largest known single cell. The 2-cell human embryo, shown left, is approximately the size of the period at the end of this sentence. [Ostriches: Art Wolfe/Tony Stone Images. Human embryo: Richard G. Rawlins/Custom Medical Stock Photo.]

CHAPTER OUTLINE

4.1 **The Stability of Chromosome Complements**

4.2 **Mitosis**

4.3 **Meiosis**
The First Meiotic Division: Reduction
The Second Meiotic Division: Equation

4.4 **Chromosomes and Heredity**
Chromosomal Determination of Sex
X-linked Inheritance
Pedigree Characteristics of Human X-linked Inheritance
Heterogametic Females
Nondisjunction as Proof of the Chromosome Theory of Heredity
Sex Determination in *Drosophila*

4.5 **Probability in the Prediction of Progeny Distributions**
Using the Binomial Distribution in Genetics

4.6 **Testing Goodness of Fit to a Genetic Hypothesis**
The Chi-square Method

4.7 **Are Mendel's Data Too Good to Be True?**

PRINCIPLES

- Chromosomes in eukaryotic cells are usually present in pairs.
- The chromosomes of each pair separate in meiosis, one going to each gamete.
- In meiosis, the chromosomes of different pairs undergo independent assortment because nonhomologous chromosomes move independently.
- In many animals, sex is determined by a special pair of chromosomes, the X and Y.
- The "criss-cross" pattern of inheritance of X-linked genes is determined by the fact that a male receives his X chromosome only from his mother and transmits it only to his daughters.
- Irregularities in the inheritance of an X-linked gene in *Drosophila* gave experimental proof of the chromosomal theory of heredity.
- The progeny of genetic crosses follow the binomial probability formula.
- The chi-square statistical test is used to determine how well the observed genetic data agree with the expectations derived from a hypothesis.

CONNECTIONS

Grasshopper, Grasshopper
E. Eleanor Carothers 1913
The Mendelian Ratio in Relation to Certain Orthopteran Chromosomes

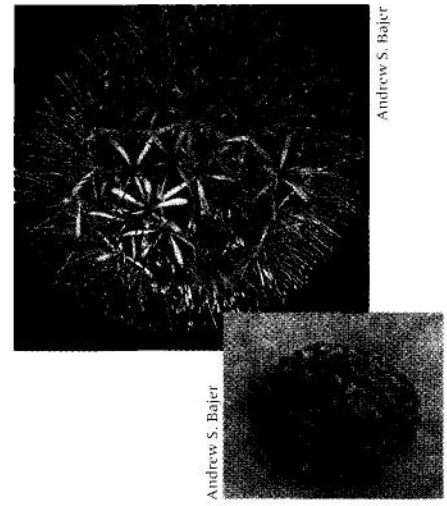
The White-Eyed Male
Thomas Hunt Morgan 1910
Sex Limited Inheritance in Drosophila

The Case against Mendel's Gardener
Ronald Aylmer Fisher 1936
Has Mendel's Work Been Rediscovered?

It came as no great revelation that genes are located in chromosomes. The parallel between their properties made this quite obvious. Genes come in pairs, chromosomes come in pairs. Alleles of a gene segregate, homologous chromosomes segregate. Unlinked genes undergo independent assortment, nonhomologous chromosomes undergo independent assortment. These parallels were first pointed out in 1903, and after that time there was little doubt that chromosomes are the cellular carriers of the genes. But parallels do not constitute scientific proof, nor does widespread agreement among scientists. In this chapter we shall examine some of the experimental evidence that was at the time—and still is—regarded as sufficient to prove the chromosome theory of heredity.

4.1 The Stability of Chromosome Complements

The cell nucleus was first discovered in 1831, but not until the late 1860s was it understood that nuclear division nearly always accompanies cell division. The importance of the nucleus in inheritance was reinforced by the nearly simultaneous discovery that the nuclei of two gametes fuse in the process of fertilization. The next major advance came a decade later with the discovery of **chromosomes,** easily visualized in the light microscope with the use of certain dyes. A few years later, chromosomes were found to segregate by an or-



Chromosomes of the African globe amaryllus (*Haemanthus*) as seen through a light microscope.

Table 4.1 Somatic (diploid) chromosome numbers of some plant and animal species

Organism	Chromosome number	Organism	Chromosome number
Field horsetail	216	Yeast (*Saccharomyces cerevisiae*)	32
Bracken fern	116	Fruit fly (*Drosophilia melanogaster*)	8
Giant sequoia	22	Nematode (*Caenorhabditis elegans*)	11 ♂, 12 ♀
Macaroni wheat	28	House fly	12
Bread wheat	42	Scorpion	4
Fava bean	12	Geometrid moth	224
Garden pea	14	Common toad	22
Wall cress (*Arabidopsis thaliana*)	10	Chicken	78
Corn (*Zea mays*)	20	Mouse	40
Lily	24	Gibbon	44
Snapdragon	16	Human being	46

derly process into the daughter cells formed by cell division as well as into the gametes formed by the division of reproductive cells. Three important regularities were observed about the **chromosome complement** (the complete set of chromosomes) of plants and animals.

1. The nucleus of each **somatic cell** (a cell of the body, in contrast to a **germ cell,** or gamete) contains a fixed number of chromosomes typical of the particular species. This number varies tremendously among species, and chromosome number bears little relation to the complexity of the organism (**Table 4.1**).
2. The chromosomes in the nuclei of somatic cells are usually present in pairs. For example, the 46 chromosomes of human beings consist of 23 pairs (**Figure 4.1**). Cells with nuclei that contain two similar sets of chromosomes are called **diploid.** A diploid individual carries two alleleic copies of each gene present in each pair of chromosomes. The chromosomes occur in pairs because one chromosome of each pair derives from the maternal parent and the other from the paternal parent of the organism.
3. The gametes that unite in fertilization to produce the diploid somatic cells have nuclei that contain only one set of chromosomes, consisting of one member of each pair. The gametic nuclei are **haploid.**

In multicellular organisms that develop from single cells, the presence of the diploid chromosome number in somatic cells and of the haploid chromosome number in germ cells indicates that there are *two* processes of nuclear division that differ in their outcome. One of these (mitosis) maintains the chromosome number; the other (meiosis) reduces the number by half. These two processes are examined in the following sections.

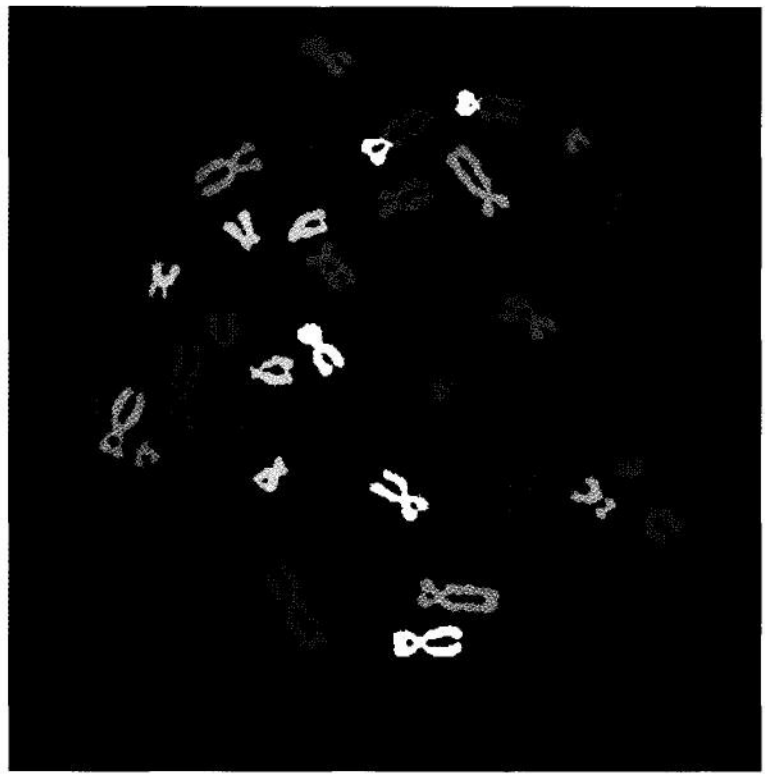


Figure 4.1 Chromosome complement of a human male. There are 46 chromosomes, present in 23 pairs. At the stage of the division cycle in which these chromosomes were observed, each chromosome consists of two identical halves lying side by side longitudinally. Except for the members of one chromosome pair (the pair that determines sex), the members of all the chromosome pairs are the same color because they contain DNA molecules that were labeled with the same mixture of fluorescent dyes. The colors differ from one pair to the next because the dye mixtures differ in color. In some cases, the long and the short arm have been labeled with a different color. [Courtesy of David C. Ward and Michael R. Speicher.]

4.2 Mitosis

Mitosis is a process of nuclear division that ensures that each of two daughter cells receives a diploid complement of chromosomes identical to the diploid complement of the parent cell. Mitosis is usually accompanied by **cytokinesis,** the process in which the cell itself divides to yield two daughter cells. The basic process of mitosis is remarkably uniform in all organisms:

1. Each chromosome is already present as a duplicated structure at the beginning of nuclear division. (The duplication of each chromosome coincides with the replication of the DNA molecule it contains.)
2. Each chromosome divides longitudinally into identical halves that separate from each other.
3. The separated chromosome halves move in opposite directions, and each becomes included in one of the two daughter nuclei that are formed.

In a cell not undergoing mitosis, the chromosomes are invisible with a light microscope. Each consists of an elongated thread too thin to be seen. This stage of the cell cycle is called **interphase.** In preparation for mitosis, the DNA in the chromosomes is replicated during a period of

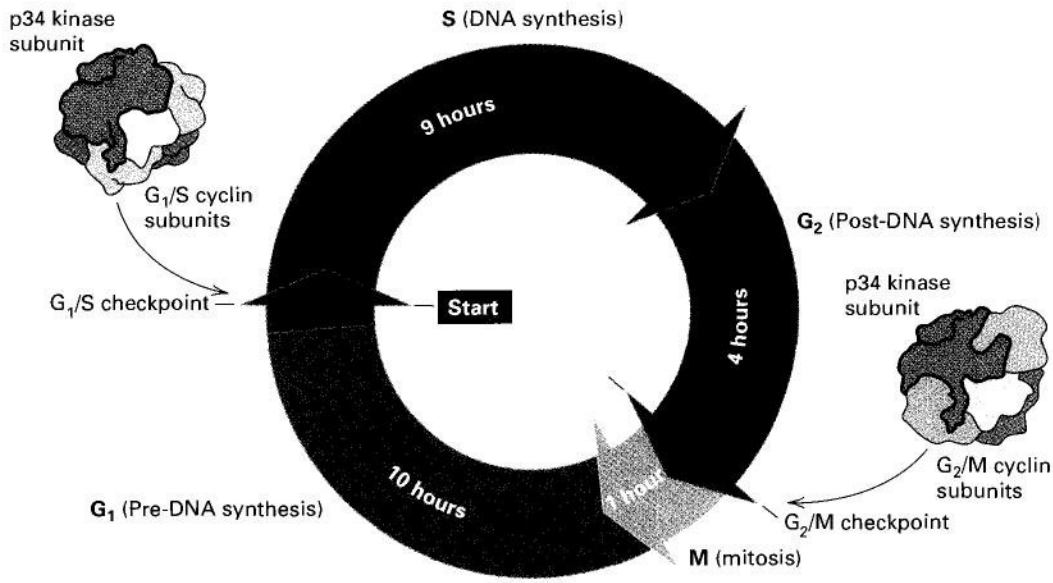


Figure 4.2 The cell cycle of a typical mammalian cell growing in tissue culture with a generation time of 24 hours.

interphase called **S** (Figure 4.2), which stands for *synthesis* of DNA. DNA replication is accompanied by chromosome duplication. Before and after S, there are periods, called $\mathbf{G_1}$ and $\mathbf{G_2}$, respectively, in which DNA replication does not take place. The **cell cycle** (the life cycle of a cell) is commonly described in terms of these three interphase periods followed by mitosis, **M.** The order of events is therefore $G_1 \rightarrow S \rightarrow G_2 \rightarrow M$, as shown in Figure 4.2. In this representation, the M period also includes the division of the cytoplasm (cytokinesis) into two approximately equal parts, each containing one daughter nucleus. The length of time required for a complete life cycle varies with cell type; it is 18–24 hours for the majority of cells in higher eukaryotes. The relative duration of the different periods in the cycle also varies considerably with cell type. Mitosis is usually the shortest period, requiring 0.5–2 hours.

The cell cycle itself is controlled by mechanisms that are essentially identical in all eukaryotes. The transitions from G_1 into S and from G_2 into M are called **checkpoints** because the transitions are delayed unless key processes have been completed (Figure 4.2). For example, at the G_1/S checkpoint, either sufficient time must have elapsed since the preceding mitosis (in some cell types) or (in other cell types) the cell must have attained sufficient size for DNA replication to be initiated. Similarly, the G_2/M checkpoint requires that DNA replication and repair of any DNA damage be completed for the M phase to commence. Both major control points are regulated in a similar manner and make use of a specialized protein kinase (called the p34 kinase subunit in Figure 4.2) that regulates the activity of target proteins by attaching phosphate groups (phosphorylation). The p34 kinase is one of numerous types of protein kinases that are used to regulate cellular processes. Mutant forms of many of these have been implicated in hereditary cancers, as we shall see in Chapter 15. To become activated, the p34 polypeptide subunit must combine with several other polypeptide chains known as **cyclins** because their abundance cycles in phase with the cell cycle. At the G_1/S control point, one set of cyclins combines with the p34 subunit to yield the active kinase that triggers DNA replication and other events of the S period. Similarly, at the G_2/M control point, a second set of cyclins combines with the p34 subunit to yield the active kinase that initiates condensation of the chromosomes, breakdown of the nuclear envelope, and reorganization of the cytoskeleton in preparation for cytokinesis.

Illustrated in Figure 4.3 are the essential features of chromosome behavior in mitosis. Mitosis is conventionally divided into four stages: **prophase, metaphase, anaphase,** and **telophase.** (If you have trouble remembering the order, you can jog your memory with "*p*eas *m*ake *a*wful *t*arts.") The stages have the following characteristics:

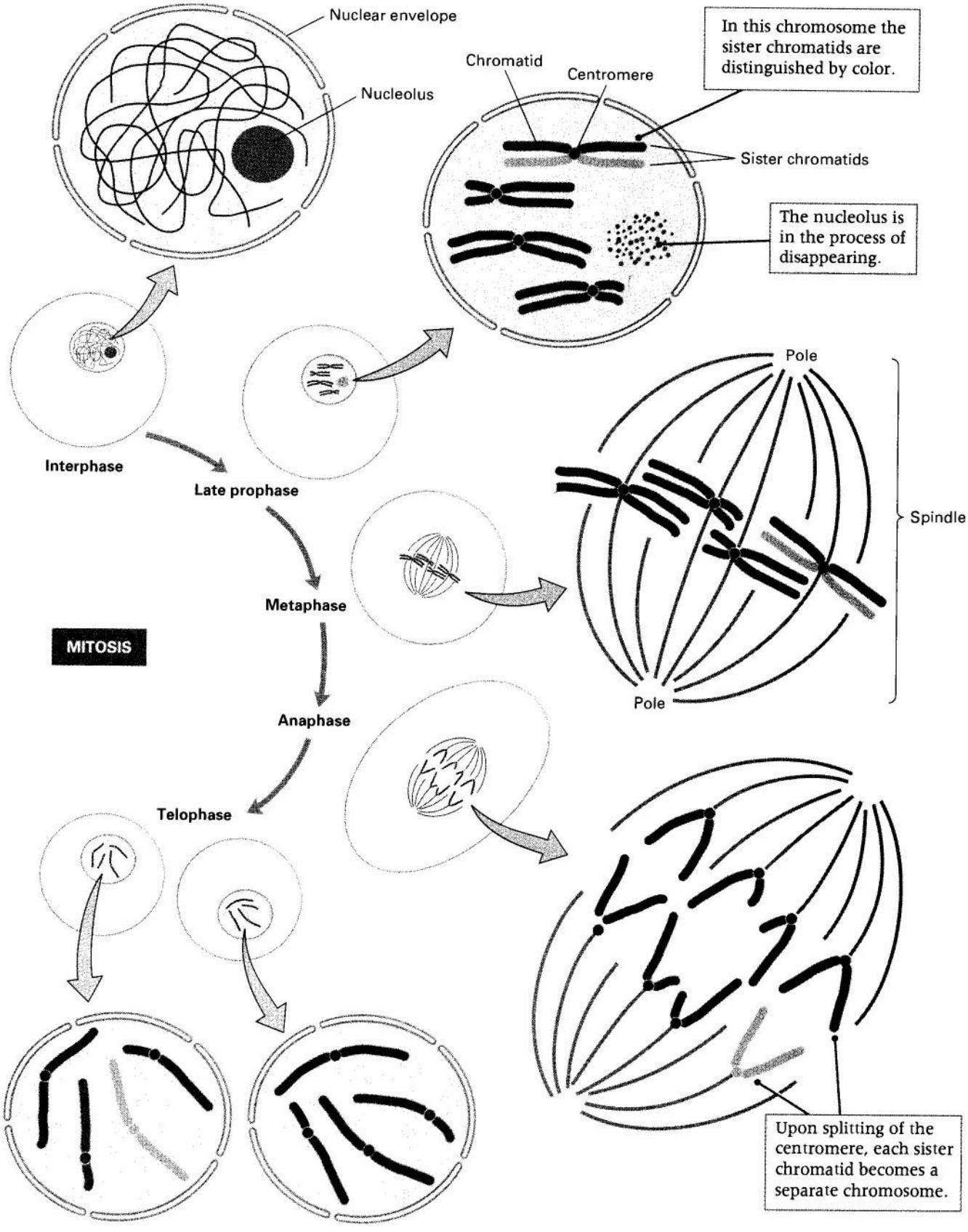
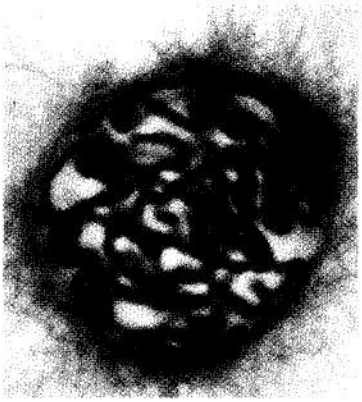
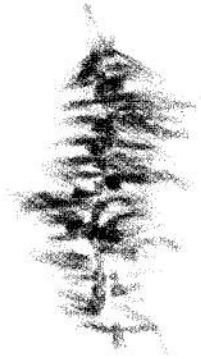


Figure 4.3 Chromosome behavior during mitosis in an organism with two pairs of chromosomes (red/rose versus green/blue). At each stage, the smaller inner diagram represents the entire cell, and the larger diagram is an exploded view showing the chromosomes at that stage.



Andrew S. Bajer

Prophase of *Haemanthus.*



Andrew S. Bajer

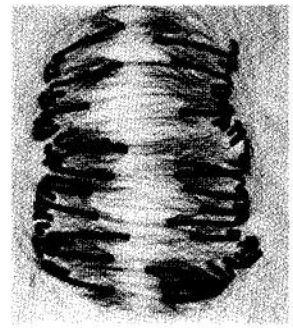
Metaphase of *Haemanthus.*

1. ***Prophase*** In interphase, the chromosomes have the form of extended filaments and cannot be seen with a light microscope as discrete bodies. Except for the presence of one or more conspicuous dark bodies (**nucleoli**), the nucleus has a diffuse, granular appearance. The beginning of prophase is marked by the condensation of chromosomes to form visibly distinct threads within the nucleus. Each chromosome is already longitudinally double, consisting of two closely associated subunits called **chromatids.** The longitudinally bipartite nature of each chromosome is readily seen later in prophase. Each pair of chromatids is the product of the duplication of one chromosome in the S period of interphase. The chromatids in a pair are held together at a specific region of the chromosome called the **centromere.** As prophase progresses, the chromosomes become shorter and thicker as a result of intricate coiling (Chapter 8). At the end of prophase, the nucleoli disappear and the nuclear envelope, a membrane surrounding the nucleus, abruptly disintegrates.

2. ***Metaphase*** At the beginning of metaphase, the **mitotic spindle** forms. The spindle is a bipolar structure consisting of fiber-like bundles of microtubules that extend through the cell between the poles of the spindle. The spindle fibers attach to each chromosome in the region of the centromere at a structure technically known as the **kinetochore.** After the chromosomes are attached to spindle fibers, they move toward the center of the cell until all the kinetochores lie on an imaginary plane equidistant from the spindle poles. This imaginary plane is called the **metaphase plate**. Aligned on the metaphase plate, the chromosomes reach their maximum contraction and are easiest to count and examine for differences in morphology.

Proper chromosome alignment is an important cell cycle control checkpoint at metaphase in both mitosis and meiosis. In a cell in which a chromosome is attached to only one pole of the spindle, the completion of metaphase is delayed. By grasping such a chromosome with a micromanipulation needle and pulling, one can mimic the tension that the chromosome would experience if it were attached on both sides; the mechanical tension allows the metaphase checkpoint to be passed, and the cell enters the next stage of division. The signal for chromosome alignment comes from the kinetochore, and the chemical nature of the signal seems to be the dephosphorylation of certain kinetochore-associated proteins. The role of the kinetochore is demonstrated by the finding that metaphase is not delayed by an unattached chromosome whose kinetochore has been destroyed by a focused laser beam. The role of dephosphorylation is demonstrated through the use of an antibody that reacts specifically with some kinetochore proteins only when they are phosphorylated. Unattached kinetochores combine strongly with the antibody, but attachment to the spindle weakens the reaction. In chromosomes that have been surgically detached from the spindle, the antibody reaction with the kinetochore reappears. Through the signaling mechanism, when all of the kinetochores are under tension and aligned on the metaphase plate, the metaphase checkpoint is passed and the cell continues the process of division.

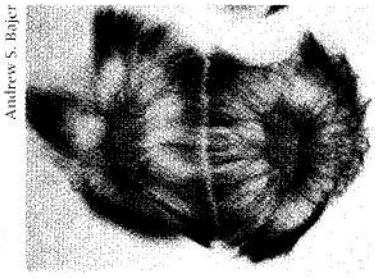
3. ***Anaphase*** In anaphase, the centromeres divide longitudinally, and the two **sister chromatids** of each chromosome move toward opposite poles of the spindle. Once the centromeres divide, each sister chromatid is regarded as a separate chromosome in its own right. Chromosome movement results in part from progressive shortening of the spindle fibers attached to the centromeres, which pulls the chromosomes in opposite directions toward the poles, and (usually) in part from a temporary elongation of the dividing cell in a direction paralleling the spindle. At the completion of anaphase, the chromosomes lie in two groups near opposite poles of the spindle. Each group contains the same number of chromosomes that was present in the original interphase nucleus.



Andrew S. Bajer

Anaphase of *Haemanthus.*

4. *Telophase* In telophase, a nuclear envelope forms around each compact group of chromosomes, nucleoli are formed, and the spindle disappears. The chromosomes undergo a reversal of condensation until they are no longer visible as discrete entities. The two daughter nuclei slowly assume a typical interphase appearance as the cytoplasm of the cell divides into two by means of a gradually deepening furrow around the periphery. (In plants, a new cell wall is synthesized between the daughter cells and separates them.)



Telophase of *Haemanthus*.

4.3 Meiosis

Meiosis is the mode of cell division that results in haploid daughter cells containing only one member of each pair of chromosomes. The process generates genetic diversity because each daughter cell contains a different set of alleles. Meiosis consists of two successive nuclear divisions. An overview of the chromosome behavior is outlined in **Figure 4.4**.

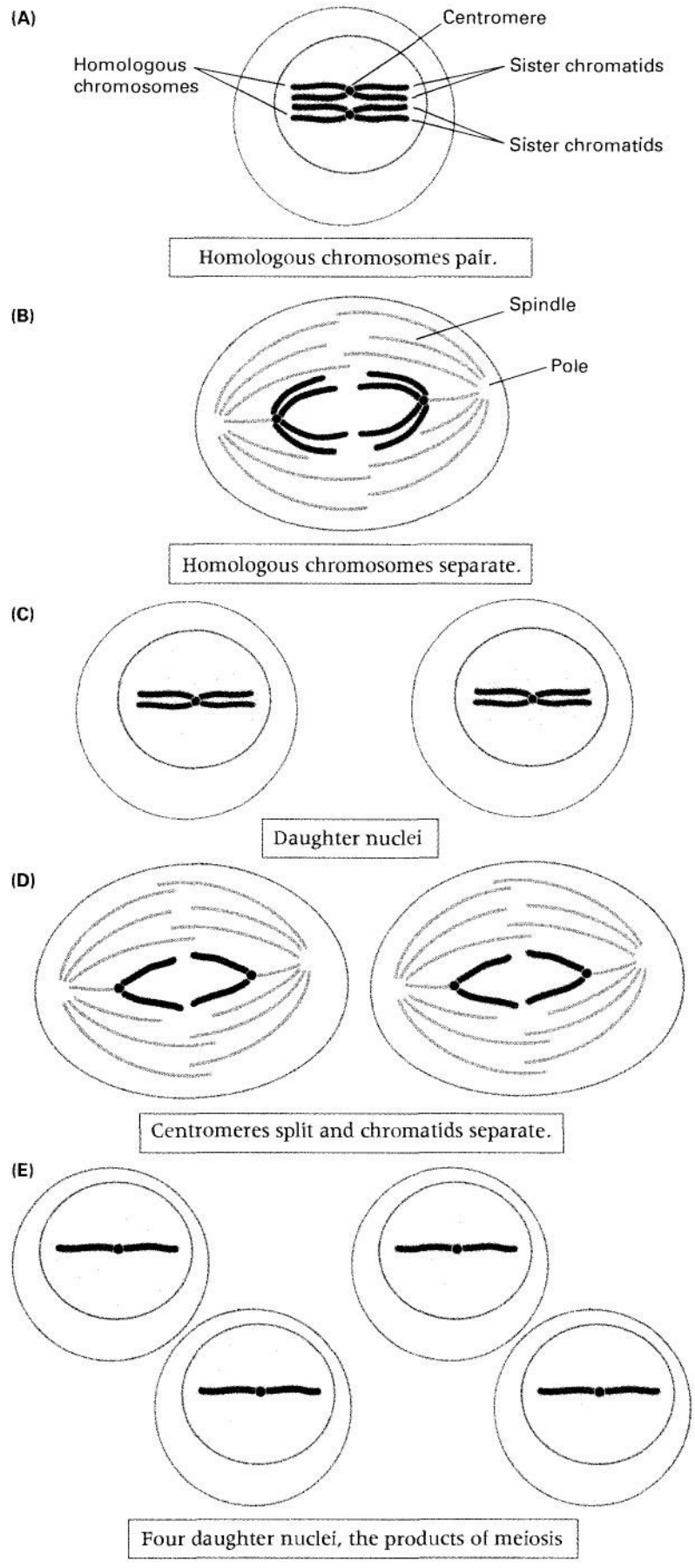
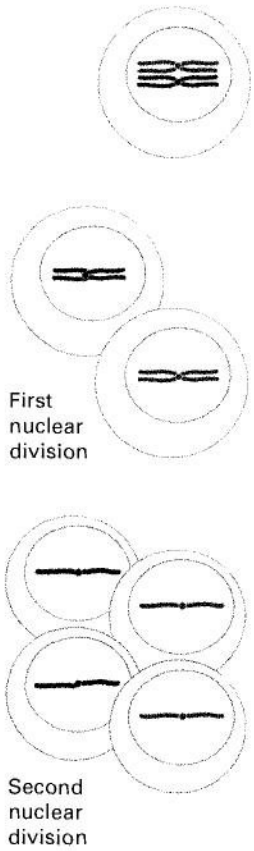


Figure 4.4 Overview of the behavior of a single pair of homologous chromosomes in meiosis. (A) The homologous chromosomes form a pair by coming together at the time when each chromosome already consists of two chromatids joined at a single centromere. (B) The members of each homologous pair separate. (C) At the end of the first meiotic division, each daughter nucleus carries one or the other of the homologous chromosomes. (D) In the second meiotic division, in each of the daughter nuclei formed in meiosis I, the sister chromatids separate. (E) The end result is four products of meiosis, each containing one of each pair of homologous chromosomes. For clarity, this diagram does not incorporate crossing-over, an interchange of chromosome segments that takes place at the stage depicted in part A.



1. Prior to the first nuclear division, the members of each pair of homologous chromosomes become closely associated along their length (Figure 4.4). Each member of the pair is already replicated and consists of two sister chromatids joined at the centromere. The pairing of the homologous chromosomes therefore produces a four-stranded structure.
2. In the first nuclear division, the homologous chromosomes are separated from each other, the members of each pair going to opposite poles of the spindle (Figure 4.4B). Each daughter chromosome consists of two chromatids attached to a common centromere (Figure 4.4C), so both of the two nuclei that are formed contain a haploid set of chromosomes. (Chromosomes are enumerated by counting the number of centromeres, not the number of chromatids.)
3. The second nuclear division loosely resembles a mitotic division, *but there is no DNA replication*. At metaphase, the chromosomes align on the metaphase plate; and at anaphase, the chromatids of each chromosome are separated into opposite daughter nuclei (Figure 4.4D). The net effect of the two divisions in meiosis is the creation of four haploid daughter nuclei, each containing the equivalent of a single sister chromatid from each pair of homologous chromosomes (Figure 4.4E).

Figure 4.4 does not show that the paired homologous chromosomes can exchange genes. The exchanges result in the formation of chromosomes that consist of segments from one homologous chromosome intermixed with segments from the other. In Figure 4.4, the exchanged chromosomes would be depicted as segments of alternating color. The exchange process is one of the critical features of meiosis, and it will be examined in the next section.

In animals, meiosis takes place in specific cells called *meiocytes*, a general term for the primary oocytes and spermatocytes in the gamete-forming tissues (**Figure 4.5**). The *oocytes* form egg cells, and the *spermatocytes* form sperm cells. Although the process of meiosis is similar in all sexually reproducing organisms, in the female of both animals and plants, only one of the four products develops into a functional cell while the other three disintegrate. In animals, the products of meiosis form either

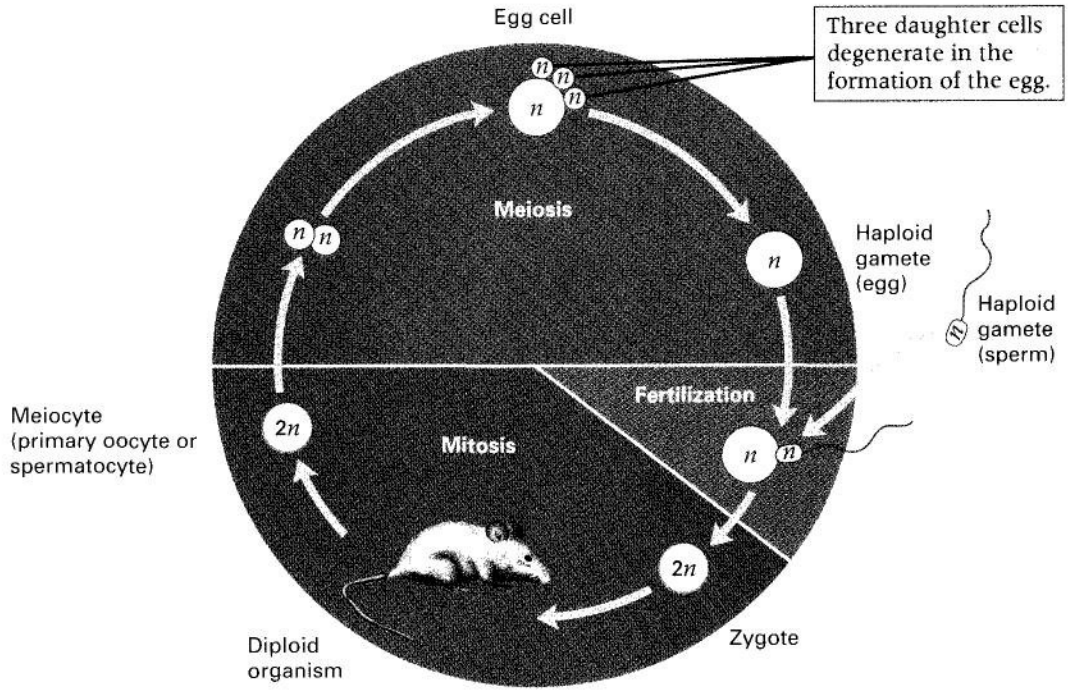


Figure 4.5 The life cycle of a typical animal. The number *n* is the number of chromosomes in the haploid chromosome complement. In males, the four products of meiosis develop into functional sperm; in females, only one of the four products develops into an egg.

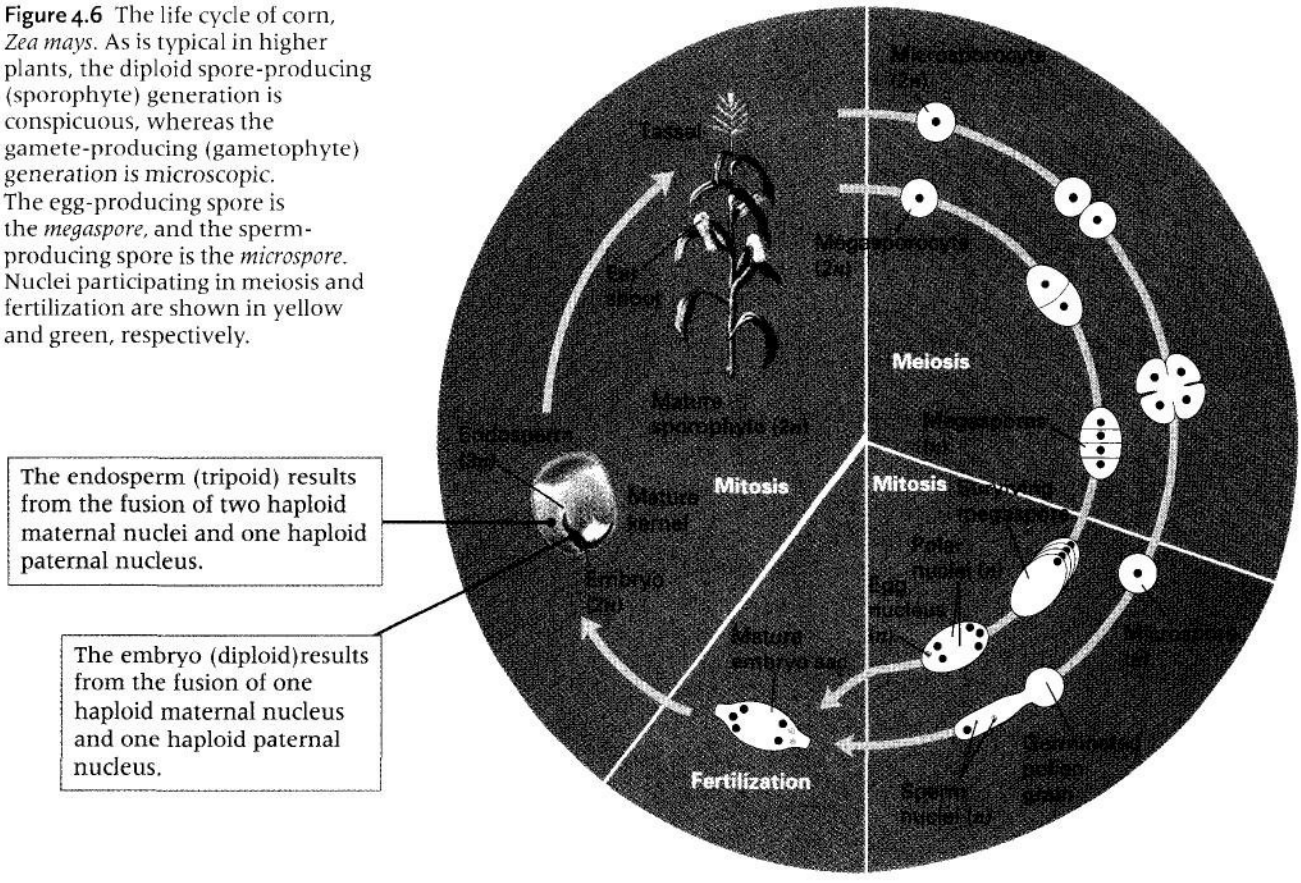


Figure 4.6 The life cycle of corn, *Zea mays*. As is typical in higher plants, the diploid spore-producing (sporophyte) generation is conspicuous, whereas the gamete-producing (gametophyte) generation is microscopic. The egg-producing spore is the *megaspore*, and the sperm-producing spore is the *microspore*. Nuclei participating in meiosis and fertilization are shown in yellow and green, respectively.

sperm or eggs. In plants, the situation is slightly more complicated:

1. The products of meiosis typically form **spores,** which undergo one or more mitotic divisions to produce a haploid **gametophyte** organism. The gametophyte produces gametes by mitotic division of a haploid nucleus (**Figure 4.6**).
2. Fusion of haploid gametes creates a diploid zygote that develops into the **sporophyte** plant, which undergoes meiosis to produce spores and so restarts the cycle.

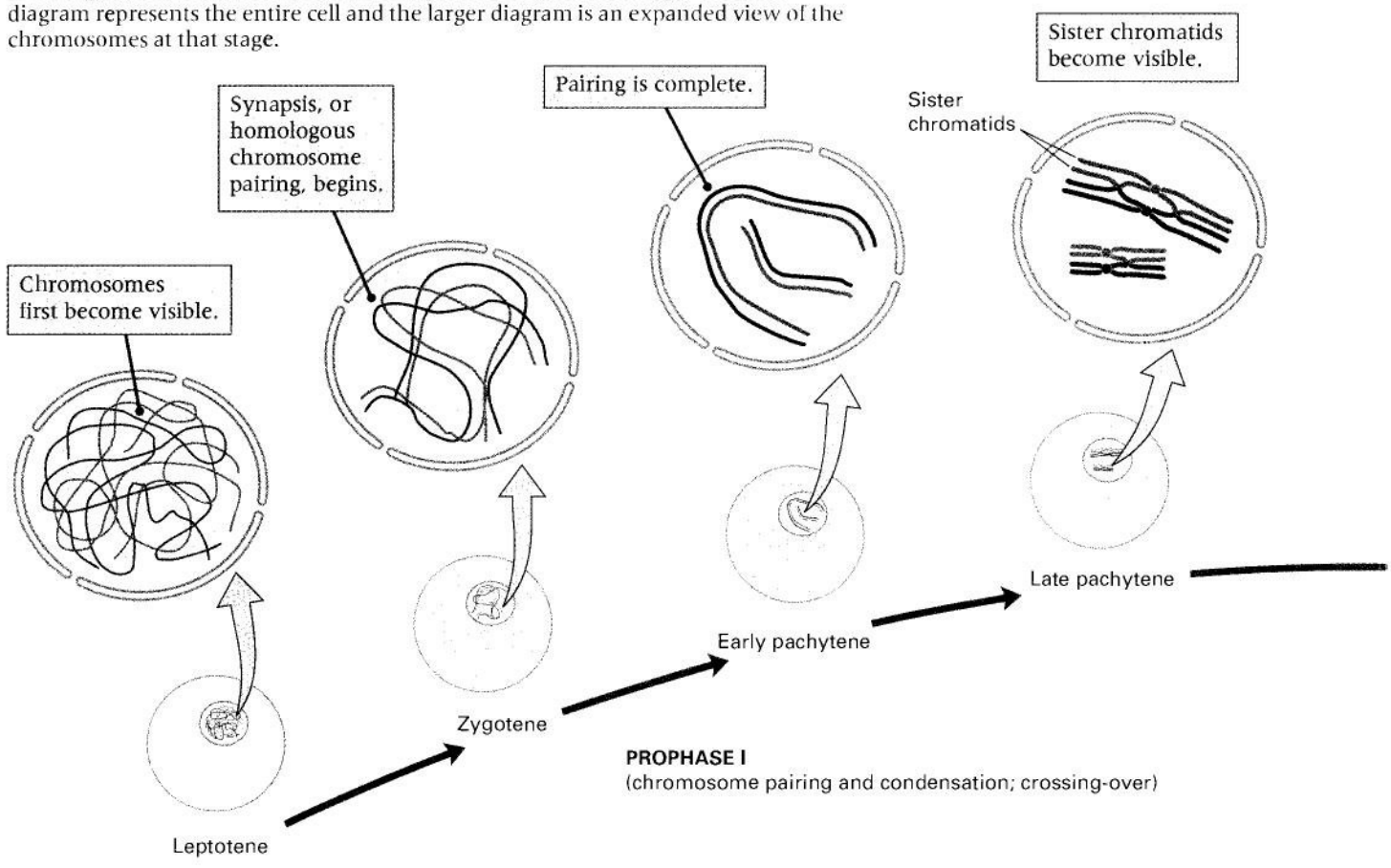
Meiosis is a more complex and considerably longer process than mitosis and usually requires days or even weeks. The entire process of meiosis is illustrated in its cellular context in **Figure 4.7**. The essence is that *meiosis consists of two divisions of the nucleus but only one duplication of the chromosomes.* The nuclear divisions—called the *first meiotic division* and the *second meiotic division*—can be separated into a sequence of stages similar to those used to describe mitosis. The distinctive events of this important process occur during the first division of the nucleus. These events are described in the following section.

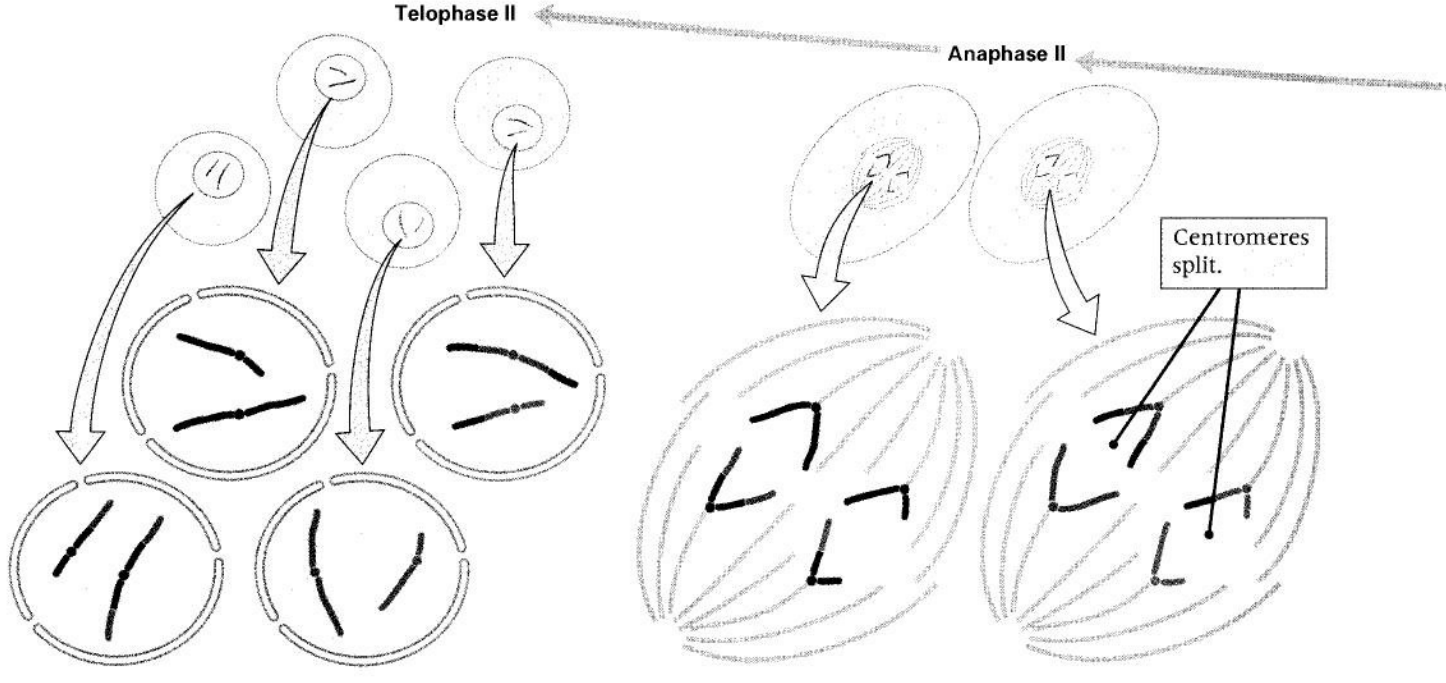
The First Meiotic Division: Reduction

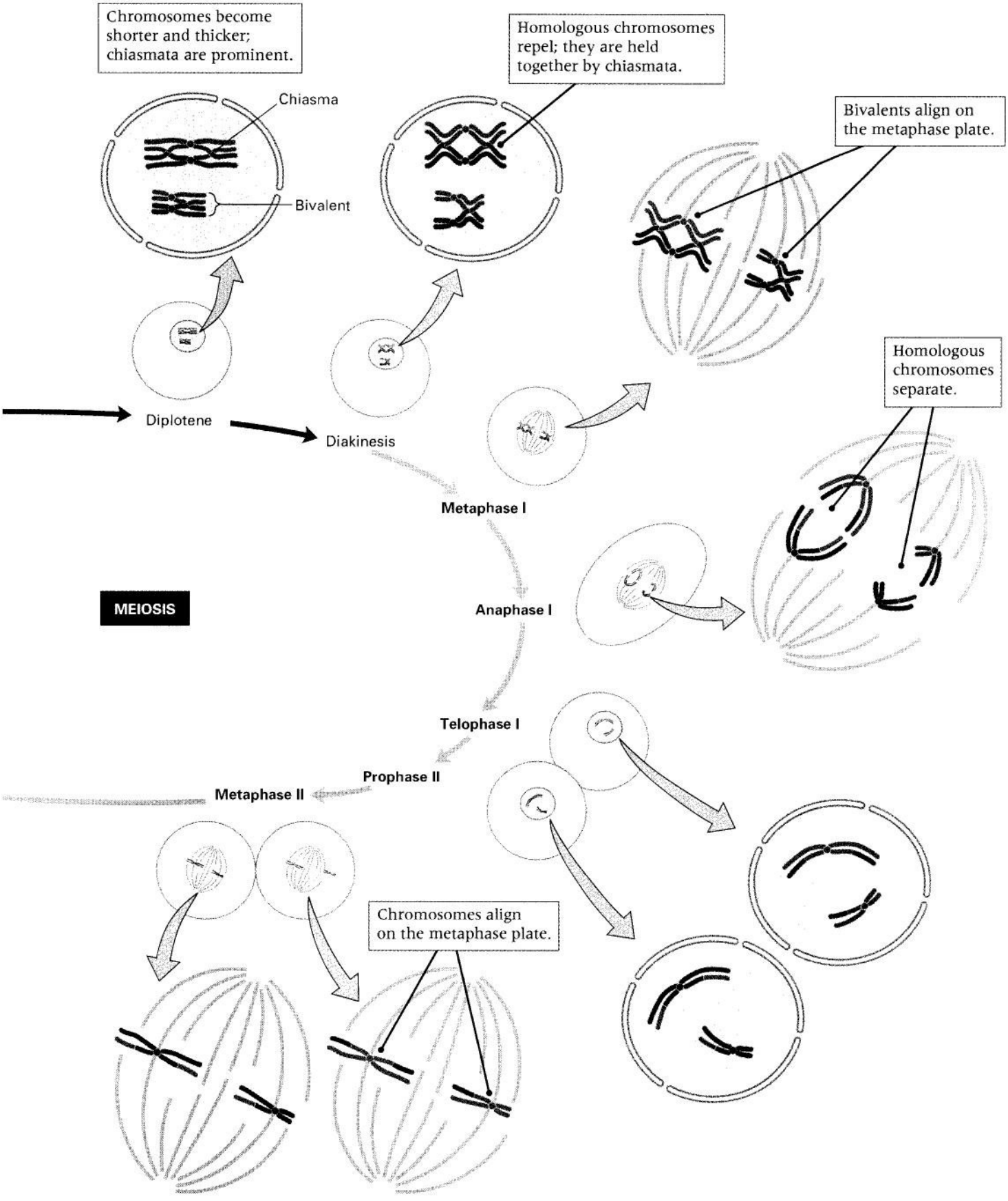
The first meiotic division (meiosis I) is sometimes called the **reductional division** because it divides the chromosome number in half. By analogy with mitosis, the first meiotic division can be split into four stages, which are called **prophase I, metaphase I, anaphase I,** and **telophase I.** These stages are generally more complex than their counterparts in mitosis. The stages and substages can be visualized with reference to **Figures 4.7 and 4.8.**

1. ***Prophase I*** This long stage lasts several days in most higher organisms. It is commonly divided into five substages: *leptotene, zygotene, pachytene, diplotene,* and *diakinesis.* These terms describe the appearance of the chromosomes at each substage.

Figure 4.7 Chromosome behavior during meiosis in an organism with two pairs of homologous chromosomes (red/rose and green/blue). At each stage, the small diagram represents the entire cell and the larger diagram is an expanded view of the chromosomes at that stage.






Chromosomes become shorter and thicker; chiasmata are prominent.
Chiasma
Bivalent
Homologous chromosomes repel; they are held together by chiasmata.
Bivalents align on the metaphase plate.
Homologous chromosomes separate.
Diplotene
Diakinesis
Metaphase I
MEIOSIS
Anaphase I
Telophase I
Prophase II
Metaphase II
Chromosomes align on the metaphase plate.

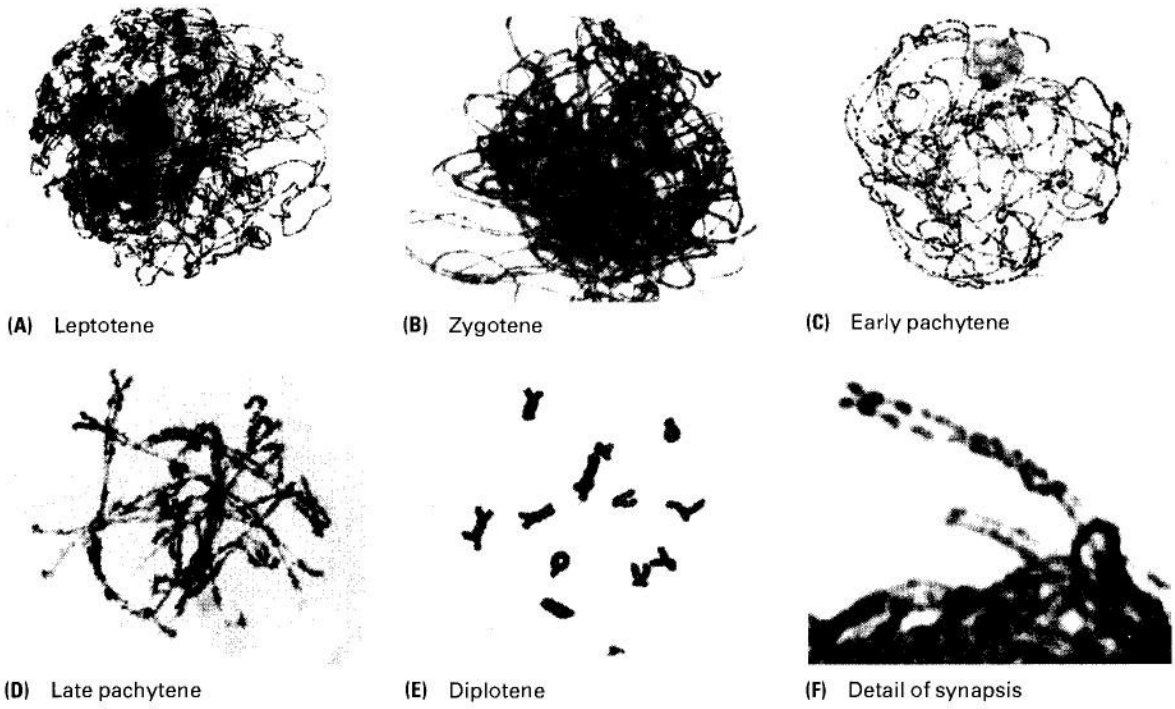


Figure 4.8 Substages of prophase I in microsporocytes of the lily (*Lilium longiflorum*). (A) Leptotene: condensation of the chromosomes is initiated and bead-like chromomeres are visible along the length of the chromosomes. (B) Zygotene: pairing (synapsis) of homologous chromosomes occurs (paired and unpaired regions can be seen particularly at the lower left in this photograph). (C) Early pachytene: synapsis is completed and crossing-over between homologous chromosomes occurs. (D) Late pachytene: continuation of the shortening and thickening of the chromosomes. (E) Diplotene: mutual repulsion of the paired homologous chromosomes, which remain held together at one or more cross points (chiasmata) along their length. Diakinesis follows (not shown): the chromosomes reach their maximum contraction. (F) Zygotene (at higher magnification in another cell) showing paired homologs and matching of chromomeres during synapsis. [Courtesy of Marta Walters (parts A, B, C, E, and F) and Herbert Stern (part D).]



Leptotene

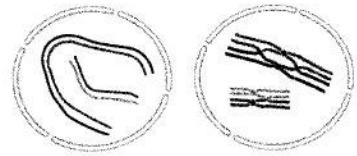
In **leptotene,** which literally means "thin thread," the chromosomes first become visible as long, thread-like structures. The pairs of sister chromatids can be distinguished by electron microscopy. In this initial phase of chromosome condensation, numerous dense granules appear at irregular intervals along their length. These localized contractions, called **chromomeres,** have a characteristic number, size, and position in a given chromosome (Figure 4.8A).



Zygotene

The **zygotene** period is marked by the lateral pairing, or **synapsis,** of homologous chromosomes, beginning at the chromosome tips. (The term *zygotene* means "paired threads.") As the pairing proceeds in zipper-like fashion along the length of the chromosomes, it results in a precise chromomere-by-chromomere association (Figure 4.8B and F). Each pair of synapsed homologous chromosomes is referred to as a **bivalent.**

Throughout **pachytene** (Figure 4.8C and D), which literally means "thick thread," the chromosomes continue to shorten and thicken (Figure 4.7). By late



Pachytene

pachytene, it can sometimes be seen that each bivalent (that is, each set of paired chromosomes) actually consists of a **tetrad** of four chromatids, but the two sister chromatids of each chromosome are usually juxtaposed very tightly. Genetic exchange by means of **crossing-over** takes place during pachytene, but the exchanges do not become apparent in the microscope until the transition to diplotene. In Figure 4.7, the sites of exchange are indicated by the points where chromatids of different colors cross over each other.

CoNnection

Grasshopper, Grasshopper

E. Eleanor Carothers 1913
University of Kansas,
Lawrence, Kansas
The Mendelian Ratio in Relation to Certain Orthopteran Chromosomes

As an undergraduate researcher, Carothers showed that nonhomologous chromosomes undergo independent assortment in meiosis. For this purpose she studied a grasshopper in which one pair of homologous chromosomes had members of unequal length. At the first anaphase of meiosis in males, she could determine by observation whether the longer or the shorter chromosome went in the same direction as the X chromosome. As detailed in this paper, she found 154 of the former and 146 of the latter, a result in very close agreement with the 1 : 1 ratio expected from independent assortment. There is no mention of the Y chromosome because in the grasshopper she studied, the females have the sex chromosome constitution XX, whereas the males have the sex chromosome constitution X. In the males she examined, therefore, the X chromosome did not have a pairing partner. The instrument referred to as a camera lucida was at that time in widespread use for studying chromosomes and other microscopic objects. It is an optical instrument containing a prism or an arrangement of mirrors that, when mounted on a microscope, reflects an image of the microscopic object onto a piece of paper where it can be traced.

The aim of this paper is to describe the behavior of an unequal bivalent in the primary spermatocytes of certain grasshoppers. The distribution of the chromosomes of this bivalent, in relation to the X chromosome, follows the laws of chance; and, therefore, affords direct cytological support of Mendel's laws. This distribution is easily traced on account of a very distinct difference in size of the homologous chromosomes. Thus another link is added to the already long chain of evidence that the chromosomes are distinct morphological individuals continuous from generation to generation, and, as such, are the bearers of the hereditary qualities. . . . This work is based chiefly on *Brachystola magna* [a short-horned grasshopper]. . . . The entire complex of chromosomes can be separated into two groups, one containing six small chromosomes and the other seventeen larger ones. [One of the larger ones is the X chromosome.] Examination shows that this group of six small chromosomes is composed of five of about equal size and one decidedly larger. [One of the small ones is the homolog of the decidedly larger one, making this pair of chromosomes unequal in size.] . . . In early metaphases the chromosomes appear as twelve separate individuals [the bivalents]. Side views show the X chromosome in its characteristic position near one pole. . . . Three hundred cells were drawn under the camera lucida to determine the distribution of the chromosomes in the asymmetrical bivalent in relation to the X chromosome. . . . In 228 cells the bivalent and the X chromosome were in the same section [the cells had been embedded in wax and thinly sliced]. In 107 cells the smaller chromosome was going to the same pole as the X chromosome, and in the remaining 121 the larger chromosome occupied this position. In the other 72 cells the X chromosome and the bivalent were in different sections, but great care was used to make sure that there was no mistake in identifying the cell or in labeling the drawings. The smaller chromosome is accompanying the X chromosome in 39 of the cells, and the larger in 33. As a net result, then, in the 300 cells drawn, the smaller chromosome would have gone to the same nucleus as the X chromosome 146 times, or in 48.7 percent of the cases; and the larger one, 154 times, or in 51.3 percent of the cases. . . . A consideration of the limited number of chromosomes and the large number of characters in any animal or plant will make it evident that each chromosome must control numerous different characters. . . . Since the rediscovery of Mendel's laws, increased knowledge has been constantly bringing into line facts that at first seemed utterly incompatible with them. There is no cytological explanation of any other form of inheritance. . . . It seems to me probable that all inheritance is, in reality, Mendelian.

Another link is added to the already long chain of evidence that the chromosomes are distinct morphological individuals continuous from generation to generation, and, as such, are the bearers of the hereditary qualities.

Source: Journal of Morphology 24: 487–511

At the onset of **diplotene,** the synapsed chromosomes begin to separate. Diplotene means "double thread," and the diplotene chromosomes are clearly double (Figure 4.8E and F). The homologous chromosomes remain held together at intervals by cross-connections resulting from crossing-over. Each cross-connection is called a **chiasma** (plural, chiasmata) and is formed by a breakage and rejoining between non-


Diplotene

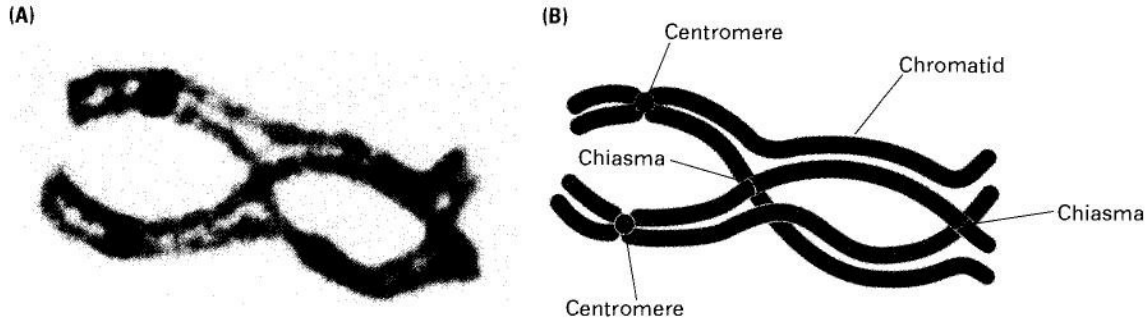


Figure 4.9 Light micrograph (A) and interpretive drawing (B) of a bivalent consisting of a pair of homologous chromosomes. This bivalent was photographed at late diplotene in a spermatocyte of the salamander *Oedipina poelzi*. It shows two chiasmata where the chromatids of the homologous chromosomes appear to exchange pairing partners. [From F. W. Stahl. 1964. *The Mechanics of Inheritance*. Prentice-Hall, Inc.; courtesy of James Kezer.]

sister chromatids. As shown in the chromosome and diagram in **Figure 4.9**, *a chiasma results from physical exchange between chromatids of homologous chromosomes*. In normal meiosis, each bivalent usually has at least one chiasma, and bivalents of long chromosomes often have three or more chiasmata.

The final period of prophase I is **diakinesis,** in which the homologous chromosomes seem to repel each other and the segments not connected by chiasmata move apart. (Diakinesis means "moving apart.") It is at this substage that the chromosomes attain their maximum condensation (Figure 4.8E). The homologous chromosomes in each bivalent remain connected by at least one chiasma, which persists until the first meiotic anaphase. Near the end of diakinesis, the formation of a spindle is initiated, and the nuclear envelope breaks down.

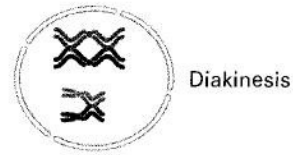
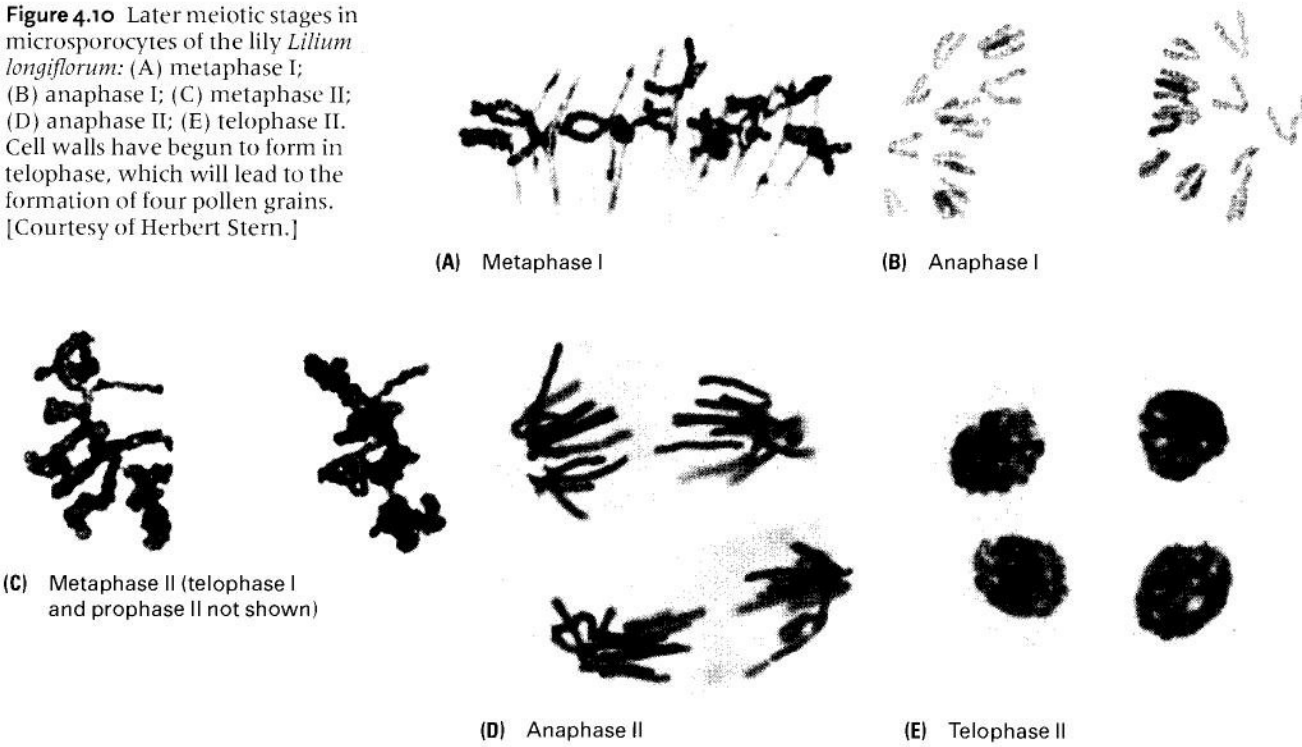
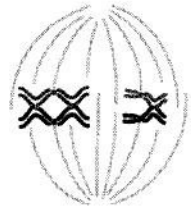


Figure 4.10 Later meiotic stages in microsporocytes of the lily *Lilium longiflorum:* (A) metaphase I; (B) anaphase I; (C) metaphase II; (D) anaphase II; (E) telophase II. Cell walls have begun to form in telophase, which will lead to the formation of four pollen grains. [Courtesy of Herbert Stern.]



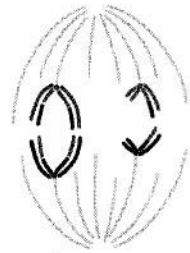
2. *Metaphase I* Each bivalent is maneuvered into a position straddling the metaphase plate with the centromeres of the homologous chromosomes oriented to opposite poles of the spindle (**Figure 4.10**A). The orientation of the centromeres determines which member of each bivalent will subsequently move to each pole, and whether the maternal or the paternal centromere is oriented toward a particular pole is completely a matter of chance. As shown in **Figure 4.11**, the bivalents formed from nonhomologous pairs of chromosomes can be oriented on the metaphase plate in either of two ways. If each of the nonhomologous chromosomes is heterozygous for a pair of alleles, then one type of alignment results in *A B* and *a b* gametes, whereas the other type results in *A b* and *a B* gametes (Figure 4.11). Because the metaphase alignment takes place at random, the two types of alignment—and therefore the four types of gametes—are equally frequent. The ratio of the four types of gametes is 1 : 1 : 1 : 1, which means that the *A*, *a* and *B*, *b* pairs of alleles undergo independent assortment. That is,

> Genes on different chromosomes undergo independent assortment because nonhomologous chromosomes align at random on the metaphase plate in meiosis I.



Metaphase I

3. *Anaphase I* In this stage, homologous chromosomes, each composed of two chromatids joined at an undivided centromere, separate from one another and move to opposite poles of the spindle (Figure 4.10B). Chromosome separation at



Anaphase I

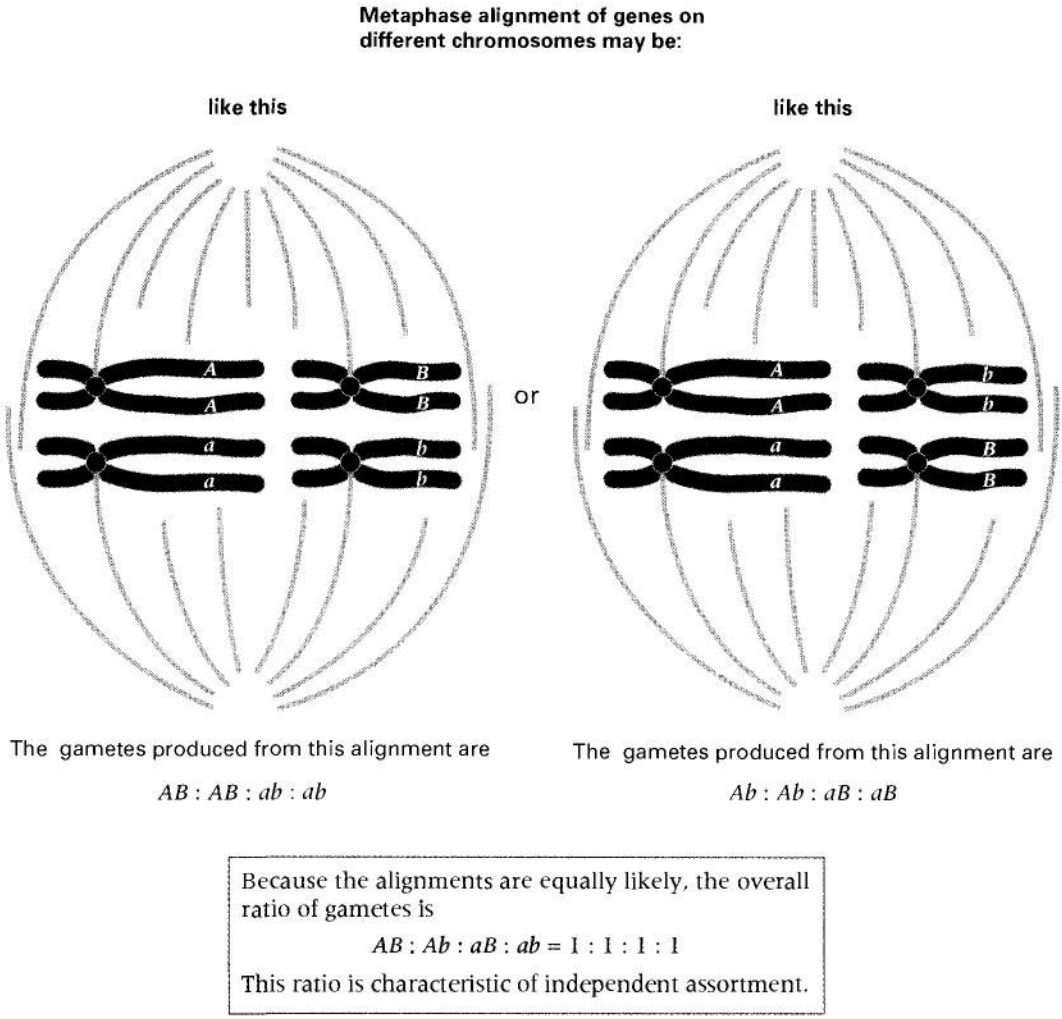


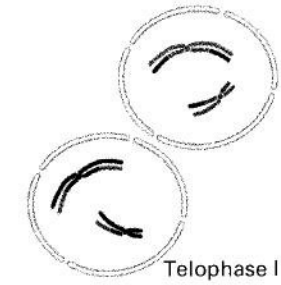
Figure 4.11 Independent assortment of genes on nonhomologous chromosomes results from random alignment of nonhomologous chromosomes at metaphase I.

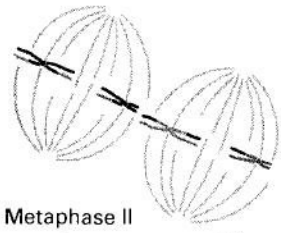
anaphase is the cellular basis of the segregation of alleles:

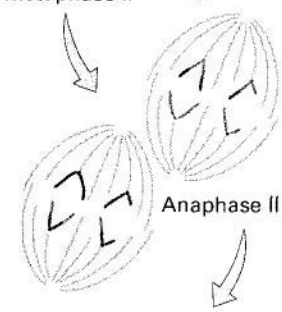
> The physical separation of homologous chromosomes in anaphase is the physical basis of Mendel's principle of segregation.

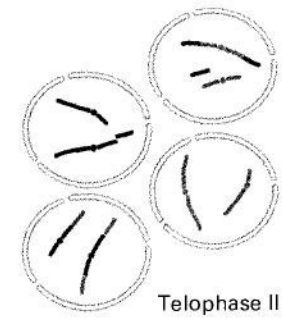
Note, however, that the centromeres of the sister chromatids are tightly stuck together and behave as a single unit. A specific protein, which in budding yeast is called the Rec8 protein, appears to be involved in cohesion of the sister centromeres. This protein appears in the centromeres and adjacent chromosome arms during S phase and persists throughout meiosis I. It disappears only at anaphase II, when sister-centromere cohesion is lost and the sister centromeres separate.

4. *Telophase I* At the completion of anaphase I, a haploid set of chromosomes consisting of one homolog from each bivalent is located near each pole of the spindle (Figure 4.6). In telophase, the spindle breaks down, and, depending on the species, either a nuclear envelope briefly forms around each group of chromosomes or the chromosomes enter the second meiotic division after only a limited uncoiling.









The Second Meiotic Division: Equation

The second meiotic division (meiosis II) is sometimes called the **equational division** because the chromosome number remains the same in each cell before and after the second division. In some species, the chromosomes pass directly from telophase I to **prophase II** without loss of condensation; in others, there is a brief pause between the two meiotic divisions and the chromosomes may "decondense" (uncoil) somewhat. *Chromosome replication never takes place between the two divisions;* the chromosomes present at the beginning of the second division are identical to those present at the end of the first.

After a short prophase (prophase II) and the formation of second-division spindles, the centromeres of the chromosomes in each nucleus become aligned on the central plane of the spindle at **metaphase II** (Figure 4.10C). In **anaphase II,** there is a breakdown of the tight physical coherence of the centromeres of the sister chromatids. As a result, the sister centromeres appear to split longitudinally, and the chromatids of each chromosome move to opposite poles of the spindle (Figure 4.10D). Once the centromeres split at anaphase II, each chromatid is considered to be a separate chromosome.

Telophase II (Figure 4.10E) is marked by a transition to the interphase condition of the chromosomes in the four haploid nuclei, accompanied by division of the cytoplasm. Thus the second meiotic division superficially resembles a mitotic division. However, there is an important difference:

> The chromatids of a chromosome are usually not genetically identical sisters along their entire length because of crossing-over associated with the formation of chiasmata during prophase of the first division.

4.4 Chromosomes and Heredity

In the early years of genetics, the strongest evidence that genes are located in chromosomes came from Mendel's principles of segregation and independent assortment, which were completely consistent with the behavior of chromosomes during meiosis. But the first rigorous experimental proof that genes are parts of chromosomes was obtained in experiments concerned with the pattern of transmission of the **sex chromosomes,** the chromosomes responsible for determination of the separate sexes in some plants and in nearly all animals. We will examine these results in this section.

Chromosomal Determination of Sex

The sex chromosomes are an exception to the rule that all chromosomes of diploid organisms are present in pairs of morphologically similar homologs. As early as 1891, microscopic analysis had shown that one of the chromosomes in males of some insect species does not have a homolog. This unpaired chromosome was called the **X chromosome,** and it was present in all somatic cells of the males but in only half the sperm cells. The biological significance

of these observations became clear when females of the same species were shown to have two X chromosomes.

In other species in which the females have two X chromosomes, the male has one X chromosome along with a morphologically unmatched chromosome. The unmatched chromosome is referred to as the **Y chromosome,** and it pairs with the X chromosome during meiosis in males, usually along only part of its length because of a limited region of homology. The difference in chromosomal constitution between males and females is a chromosomal mechanism for determining sex at the time of fertilization. Whereas every egg cell contains an X chromosome, half the sperm cells contain an X chromosome and the rest contain a Y chromosome. Fertilization of an X-bearing egg by an X-bearing sperm results in an XX zygote, which normally develops into a female; and fertilization by a Y-bearing sperm results in an XY zygote, which normally develops into a male (Figure 4.12). The result is a criss-cross pattern of inheritance of the X chromosome in which a male receives his X chromosome from his mother and transmits it only to his daughters.

The XX–XY type of chromosomal sex determination is found in mammals, including human beings, in many insects,

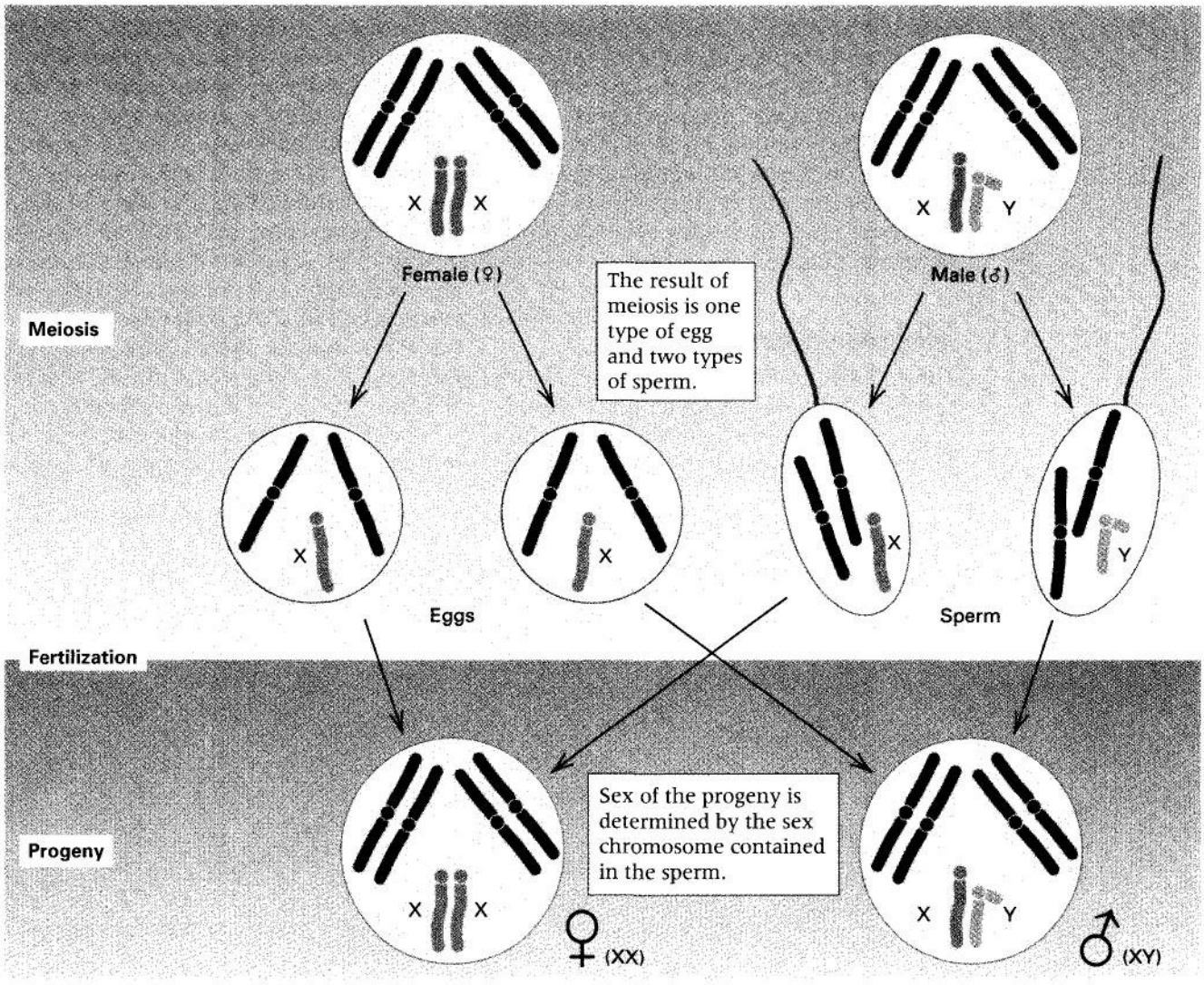


Figure 4.12 The chromosomal basis of sex determination in mammals, many insects, and other animals.

and in other animals, as well as in some flowering plants. The female is called the **homogametic** sex because only one type of gamete (X-bearing) is produced, and the male is called the **heterogametic** sex because two different types of gametes (X-bearing and Y-bearing) are produced. When the union of gametes in fertilization is random, a sex ratio at fertilization of 1 : 1 is expected because males produce equal numbers of X-bearing and Y-bearing sperm.

The X and Y chromosomes together constitute the sex chromosomes; this term distinguishes them from other pairs of chromosomes, which are called **autosomes.** Although the sex chromosomes control the developmental switch that determines the earliest stages of female or male development, the developmental process itself requires many genes scattered throughout the genome, including genes on the autosomes. The X chromosome also contains many genes with functions unrelated to sexual differentiation, as we will see in the next section. In most organisms, including human beings, the Y chromosome carries few genes other than those related to male determination.

X-linked Inheritance

The compelling evidence that genes are located in chromosomes came from the study of a *Drosophila* gene for white eyes, which proved to be present in the X chromosome. Recall that in Mendel's crosses, it did not matter which trait was present in the male parent and which in the female parent. Reciprocal crosses gave the same result. One of the earliest exceptions to this rule was found by Thomas Hunt Morgan in 1910 in an early study of a mutant fruit fly that had white eyes. The wildtype eye color is a brick-red combination of red and brown pigments (**Figure 4.13**). Although white eyes can result from certain combinations of autosomal genes that eliminate the pigments individually, the white-eye mutation that Morgan studied results in a metabolic block that knocks out both pigments simultaneously. (The gene codes for a type of transmembrane protein called an ABC transporter that is necessary to transport the eye pigment precursors into the pigment cells.)

Morgan's study started with a single male with white eyes that appeared in a wildtype laboratory population that had been maintained for many generations. In a mating of this male with wildtype females, all of the F_1 progeny of both sexes had red eyes, which showed that the allele for white eyes is recessive. In the F_2 progeny from the mating of F_1 males with females, Morgan observed 2459 red-eyed females, 1011 red-eyed males, and 782 white-eyed males. It was clear that the white-eyed phenotype was somehow connected with sex, because all of the white-eyed flies were males.

On the other hand, white eyes were not restricted to males. For example, when red-

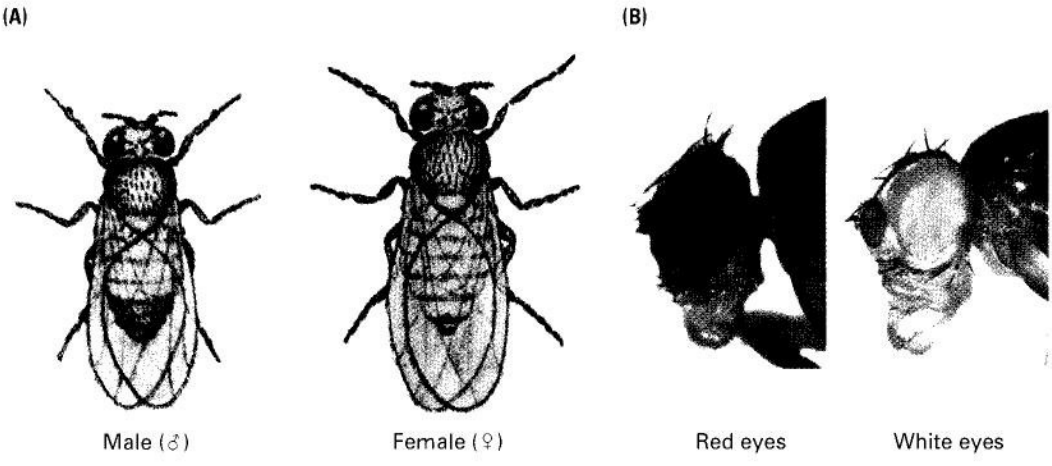


Figure 4.13 Drawings of a male and a female fruit fly, *Drosophila melanogaster.* The photographs show the eyes of a wildtype red-eyed male and a mutant white-eyed male. [Drawings courtesy of Carolina Biological Supply Company; photographs courtesy of E. R. Lozovsky.]

CoNnection

The White-Eyed Male

Thomas Hunt Morgan 1910
Columbia University,
New York, New York
Sex Limited Inheritance in Drosophila

Morgan's genetic analysis of the white-eye mutation marks the beginning of Drosophila *genetics. It is in the nature of science that as knowledge increases, the terms used to describe things change also. This paper affords an example, because the term* sex-limited inheritance *is used today to mean something completely different from Morgan's usage. What Morgan was referring to is now called X-linked inheritance or sex-linked inheritance. To avoid confusion, we have taken the liberty of substituting the modern equivalent wherever appropriate. Morgan was also unaware that* Drosophila *males had a Y chromosome. He thought that females were XX and males X, as in grasshoppers (see the Carothers paper). We have also supplied the missing Y chromosome. On the other hand, Morgan's gene symbols have been retained as in the original. He uses* R *for the wildtype allele for red eyes and* W *for the recessive allele for white eyes. This is a curious departure from the convention, already introduced by Mendel, that dominant and recessive alleles should be represented by the same symbol. Today we use* w *for the recessive allele and* w^+ *for the dominant allele.*

In a pedigree culture of *Drosophila* which had been running for nearly a year through a considerable number of generations, a male appeared with white eyes. The normal flies have brilliant red eyes. The white-eyed male, bred to his red-eyed sisters, produced 1,237 red-eyed offspring. . . . The F_1 hybrids, inbred, produced

2,459	red-eyed females
1,011	red-eyed males
782	white-eyed males

No white-eyed females appeared. The new character showed itself to be sex-linked in the sense that it was transmitted only to the grandsons. But that the character is not incompatible with femaleness is shown by the following experiment. The white-eyed male (mutant) was later crossed with some of his daughters (F_1), and produced

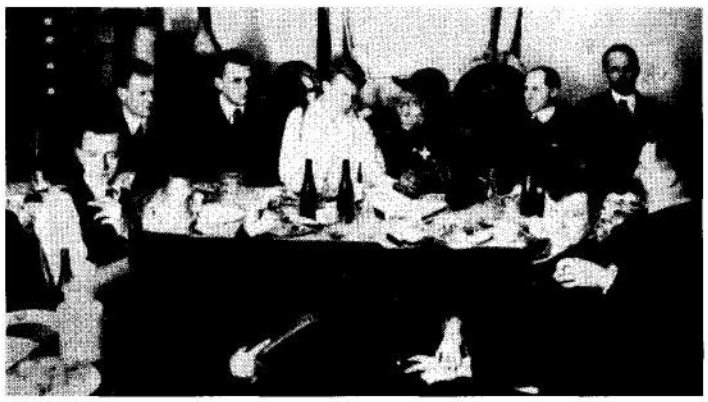
129	red-eyed females
132	red-eyed males
88	white-eyed females
86	white-eyed males

The results show that the new character, white eyes, can be carried over to the females by a suitable cross, and is in consequence in this sense not limited to one sex. It will be noted that the four classes of individuals occur in approximately equal numbers (25 percent). . . .

No white-eyed females appeared.

The results just described can be accounted for by the following hypothesis. Assume that all of the spermatozoa of the white-eyed male carry the "factor" for white eyes "W"; that half of the spermatozoa carry a sex factor "X," [and] the other half lack it, *i.e.*, the male is heterozygous for sex. [The male is actually XY.] Thus, the symbol for the male is "WXY", and for his two kinds of spermatozoa WX—Y. Assume that all of the eggs of the red-eyed female carry the red-eyed "factor" R; and that all of the eggs (after meiosis) carry one X each, the symbol for the red-eyed female will be therefore RRXX and that for her eggs will be RX. . . . The hypothesis just utilized to explain these results first obtained can be tested in several ways. [There follow four types of crosses, each yielding the expected result.] . . . In order to obtain these results it is necessary to assume that, when the two classes of spermatozoa are formed in the RXY male, R and X go together. . . . The fact is that this R and X are combined and have never existed apart.

Source: Science 32: 120–122



Morgan's group partying in the fly room at Columbia University in 1919, to celebrate A. H. Sturtevant's return from military service in World War I. T. H. Morgan is on the far right in the back row, with H. J. Miller beside him. Sturtevant is in the foreground, leaning back in his chair, and C. B. Bridges is in shirtsleeves sitting next to an apelike creature dubbed "Pithecanthropus," a dummy made up for the occasion. [Courtesy of the Archives, California Institute of Technology.]

eyed F_1 females from the cross of wildtype ♀ × white ♂ were backcrossed with their white-eyed fathers, the progeny consisted of both red-eyed and white-eyed females and red-eyed and white-eyed males in approximately equal numbers.

A key observation came from the mating of white-eyed females with wildtype males. All the female progeny had wildtype eyes, but all the male progeny had white eyes. This is the reciprocal of the original cross of wildtype ♀ × white ♂, which had given only wildtype females and wildtype males, so the reciprocal crosses gave different results.

Morgan realized that reciprocal crosses would yield different results if the allele for white eyes were present in the X chromosome. A gene on the X chromosome is said to be **X-linked.** The normal chromosome complement of *Drosophila melanogaster* is shown in **Figure 4.14.** Females have an XX chromosome complement, whereas males are XY. Morgan's hypothesis was that an X chromosome contains either a wildtype w^+ allele or a mutant w allele and that the Y chromosome does not contain a counterpart of the *white* gene. Using the *white* allele present in an X chromosome to represent the entire X chromosome, we can write the genotype of a white-eyed male as wY and that of a wildtype male as w^+Y. Because the w allele is recessive, white-eyed females are of genotype ww and wildtype females are either heterozygous w^+w or homozygous w^+w^+. The implications of this model for reciprocal crosses are shown in **Figure 4.15.** The mating wildtype ♀ × white ♂ is Cross A, and that of white ♀ × wildtype ♂ is Cross B.

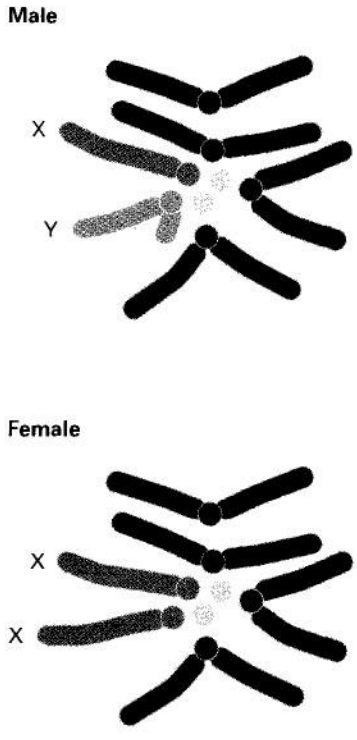


Figure 4.14 The diploid chromosome complements of a male and a female *Drosophila melanogaster.* The centromere of the X chromosome is nearly terminal, but that of the Y chromosome divides the chromosome into two unequal arms. The large autosomes (chromosomes 2 and 3, shown in blue and green) are not easily distinguishable in these types of cells. The tiny autosome (chromosome 4, shown in yellow) appears as a dot.

The X-linked mode of inheritance does account for the different phenotypic ratios observed in the F_1 and F_2 progeny from the crosses. The characteristics of X-linked inheritance can be summarized as follows:

1. Reciprocal crosses resulting in different phenotypic ratios in the sexes often indicate X-linked inheritance. In the case of white eyes in *Drosophila*, the cross of a red-eyed female with a white-eyed male yields all red-eyed progeny (Figure 4.15, Cross A), whereas the cross of a white-eyed female with a red-eyed male yields red-eyed female progeny and white-eyed male progeny (Figure 4.15, Cross B).
2. Heterozygous females transmit each X-linked allele to approximately half of their daughters and half of their sons; this is illustrated in the F_2 generation of Cross B in Figure 4.15.
3. Males that inherit an X-linked recessive allele exhibit the recessive trait because the Y chromosome does not contain a wildtype counterpart of the gene. Affected males transmit the recessive allele to all of their daughters but none of their sons; this principle is illustrated in the F_1 generation of Cross A in Figure 4.15. Any male that is not affected must carry the wildtype allele in his X chromosome.

The essence of X-linked inheritance is captured in the Punnett square in **Figure 4.16:** A male transmits his X chromosome only to his daughters, whereas a female transmits an X chromosome to offspring of both sexes.

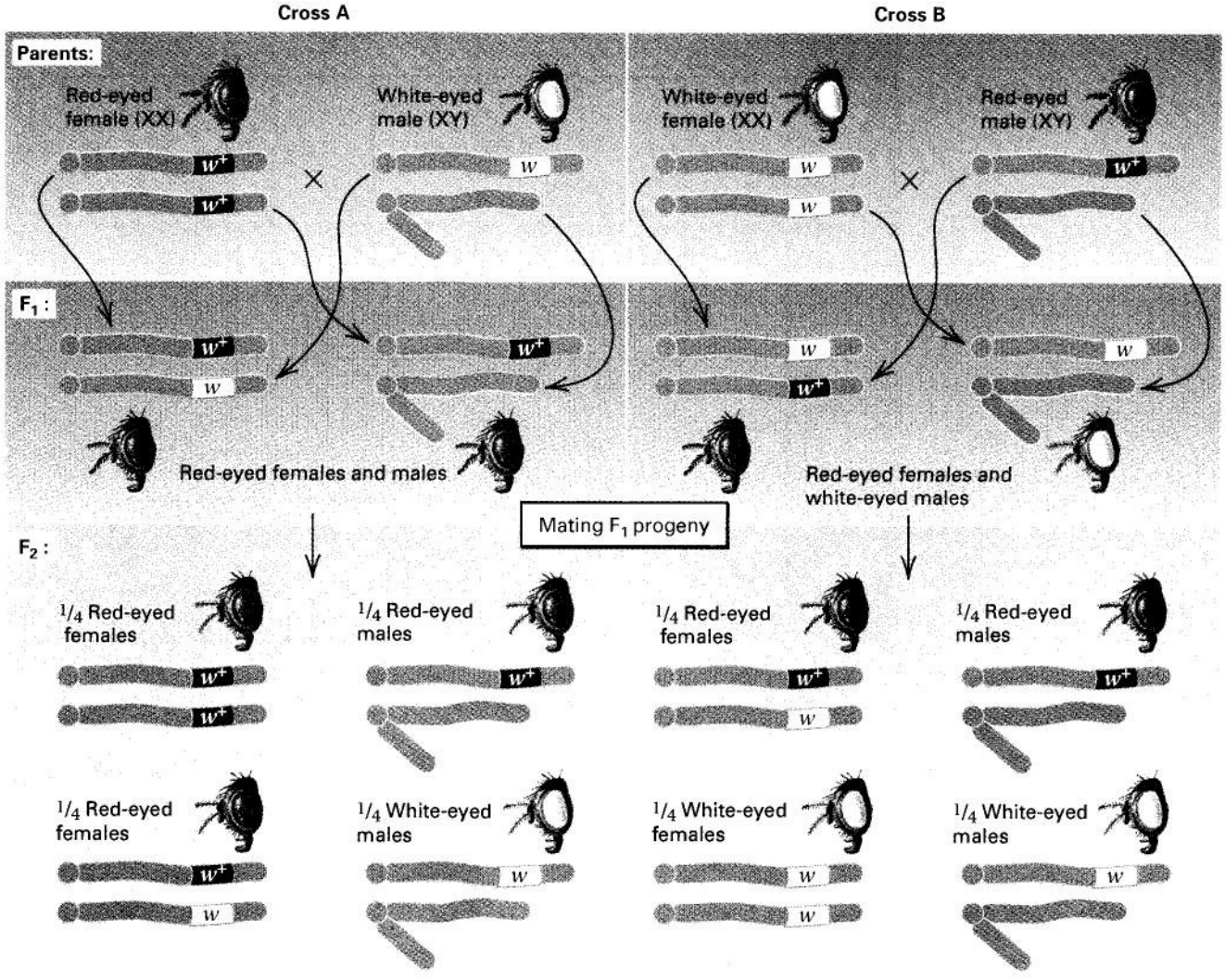


Figure 4.15 Chromosomal interpretation of the results obtained in F_1 and F_2 progenies in crosses of *Drosophila*. Cross A is a mating of a wildtype (red-eyed) female with a white-eyed male. Cross B is the reciprocal mating of a white-eyed female with a red-eyed male. In the X chromosome, the wildtype w^+ allele is shown in red, the mutant w allele in white. The Y chromosome does not carry either allele of the w gene.

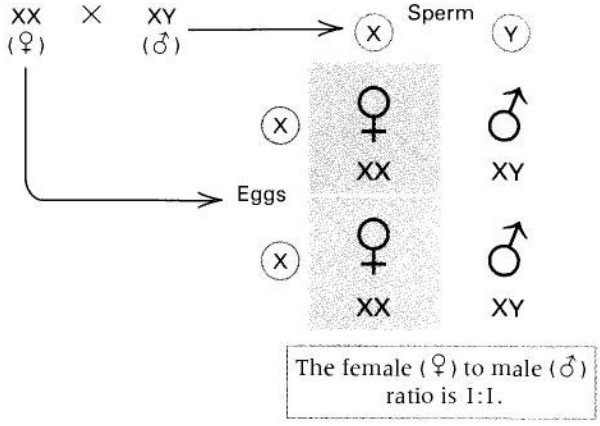


Figure 4.16 In chromosomal sex determination, each son gets his X chromosome from his mother and his Y chromosome from his father.

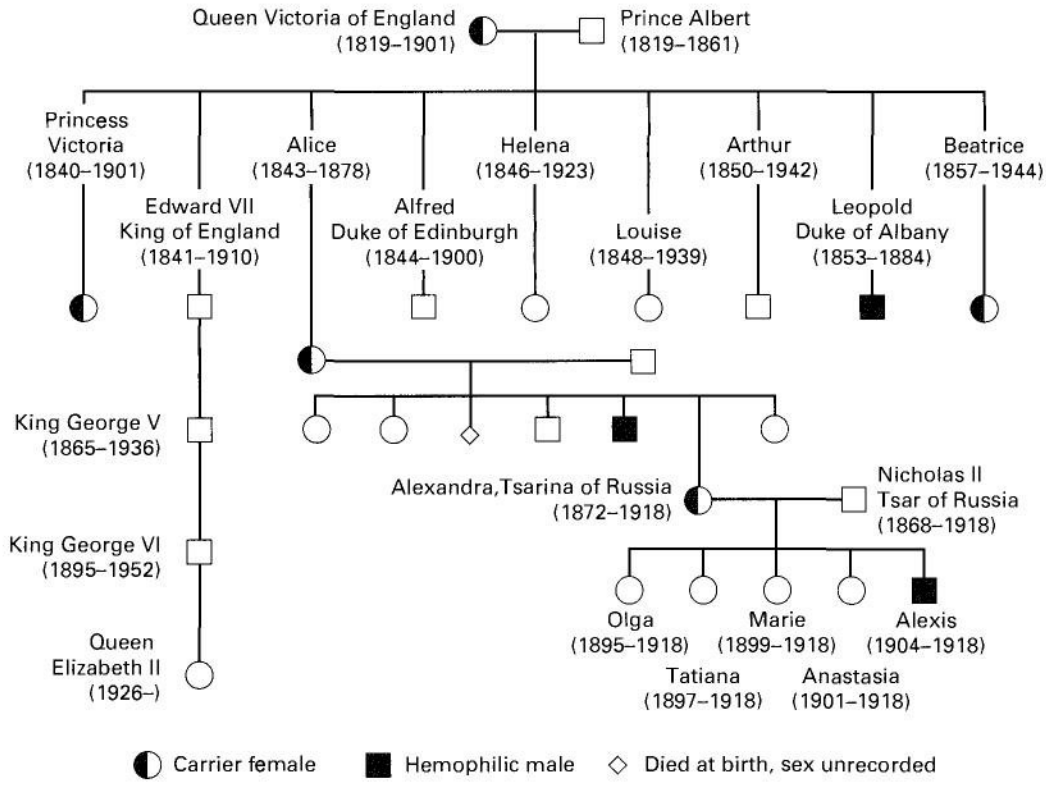


Figure 4.17 Genetic transmission of hemophilia A among the descendants of Queen Victoria of England, including her granddaughter, Tsarina Alexandra of Russia, and Alexandra's five children. The photograph is that of Tsar Nicholas II, Tsarina Alexandra, and the Tsarevich Alexis, who was afflicted with hemophilia. [*Source:* Culver Pictures.]



Pedigree Characteristics of Human X-linked Inheritance

An example of a human trait with an X-linked pattern of inheritance is *hemophilia A*, a severe disorder of blood clotting determined by a recessive allele. Affected persons lack a blood-clotting protein called factor VIII that is needed for normal clotting, and they suffer excessive, often life-threatening bleeding after injury. A famous pedigree of hemophilia starts with Queen Victoria of England (**Figure 4.17**). One of her sons, Leopold, was hemophilic, and two of her daughters were heterozygous carriers of the gene. Two of Victoria's granddaughters were also carriers, and by marriage they introduced the gene into the royal families of Russia and Spain. The heir to the Russian throne of the Romanoffs, Tsarevich Alexis, was afflicted with the condition. He inherited the gene from his mother, the Tsarina Alexandra, one of Victoria's granddaughters. The Tsar, the Tsarina, Alexis, and his four sisters were all executed by the Russian Bolsheviks in the 1918 revolution. Ironi-

cally, the present royal family of England is descended from a normal son of Victoria and is free of the disease.

X-linked inheritance in human pedigrees shows several characteristics that distinguish it from other modes of genetic transmission.

1. For any rare trait due to an X-linked recessive allele, the affected individuals are exclusively, or almost exclusively, male. There is an excess of males because females carrying the rare X-linked recessive are almost exclusively heterozygous and so do not express the mutant phenotype.
2. Affected males who reproduce have normal sons. This follows from the fact that a male transmits his X chromosome only to his daughters.
3. A woman whose father was affected has normal sons and affected sons in the ratio 1 : 1. This is true because any daughter of an affected male must be heterozygous for the recessive allele.

Heterogametic Females

In some organisms, the homogametic and heterogametic sexes are reversed; that is, the males are XX and the females are XY. This type of sex determination is found in birds, in some reptiles and fish, and in moths and butterflies. The reversal of XX and XY in the sexes results in an opposite pattern of nonreciprocal inheritance of X-linked genes. To distinguish sex determination in these organisms from the usual XX–XY mechanism, the sex chromosome constitution in the homogametic sex is designated WW and that in the heterogametic sex as WZ. Hence in organisms with heterogametic females, the chromosomal constitution of the females is designated WZ and that of the male WW.

A specific example of W-linked inheritance in chickens is shown in **Figure 4.18**. A few breeds of chickens have feathers with alternating transverse bands of light and dark color, resulting in a phenotype known as barred. In other breeds the feathers are uniformly colored and nonbarred. Reciprocal crosses between true-breeding barred and true-breeding nonbarred breeds give the results shown, which indicate that the gene that determines barring must be dominant and must be located in the W chromosome.

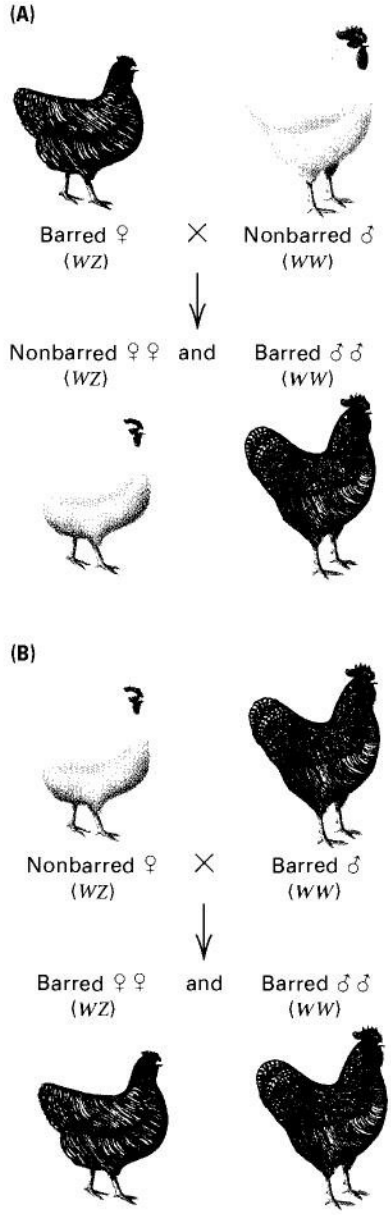


Figure 4.18 Barred feathers in chickens, a classic example showing that chromosomal sex determination in birds is the reverse of that in mammals. In birds, females are the heterogametic sex.

Nondisjunction as Proof of the Chromosome Theory of Heredity

The parallel between the inheritance of the *Drosophila white* mutation and the genetic transmission of the X chromosome supported the chromosome theory of heredity that genes are parts of chromosomes. Other

experiments with *Drosophila* provided the definitive proof.

One of Morgan's students, Calvin Bridges, discovered rare exceptions to the expected pattern of inheritance in crosses with several X-linked genes. For example, when white-eyed *Drosophila* females were mated with red-eyed males, most of the progeny consisted of the expected red-eyed females and white-eyed males. However, about 1 in every 2000 F_1 flies was an exception, either a white-eyed female or a red-eyed male. Bridges showed that these rare exceptional offspring resulted from occasional failure of the two X chromosomes in the mother to separate from each other during meiosis—a phenomenon called **nondisjunction.** The consequence of nondisjunction of the X chromosome is the formation of some eggs with two X chromosomes and others with none. Four classes of zygotes are expected from the fertilization of these abnormal eggs (**Figure 4.19**). Animals with no X chromosome are not detected because embryos that lack an X are not viable; likewise, most progeny with three X chromosomes die early in development. Microscopic examination of the chromosomes of the exceptional progeny from the cross white ♀ × wildtype ♂ showed that the exceptional white-eyed females had two X chromosomes *plus* a Y chromosome, and the exceptional red-eyed males had a single X but were *lacking* a Y. The latter, with a sex-chromosome constitution denoted XO, were sterile males.

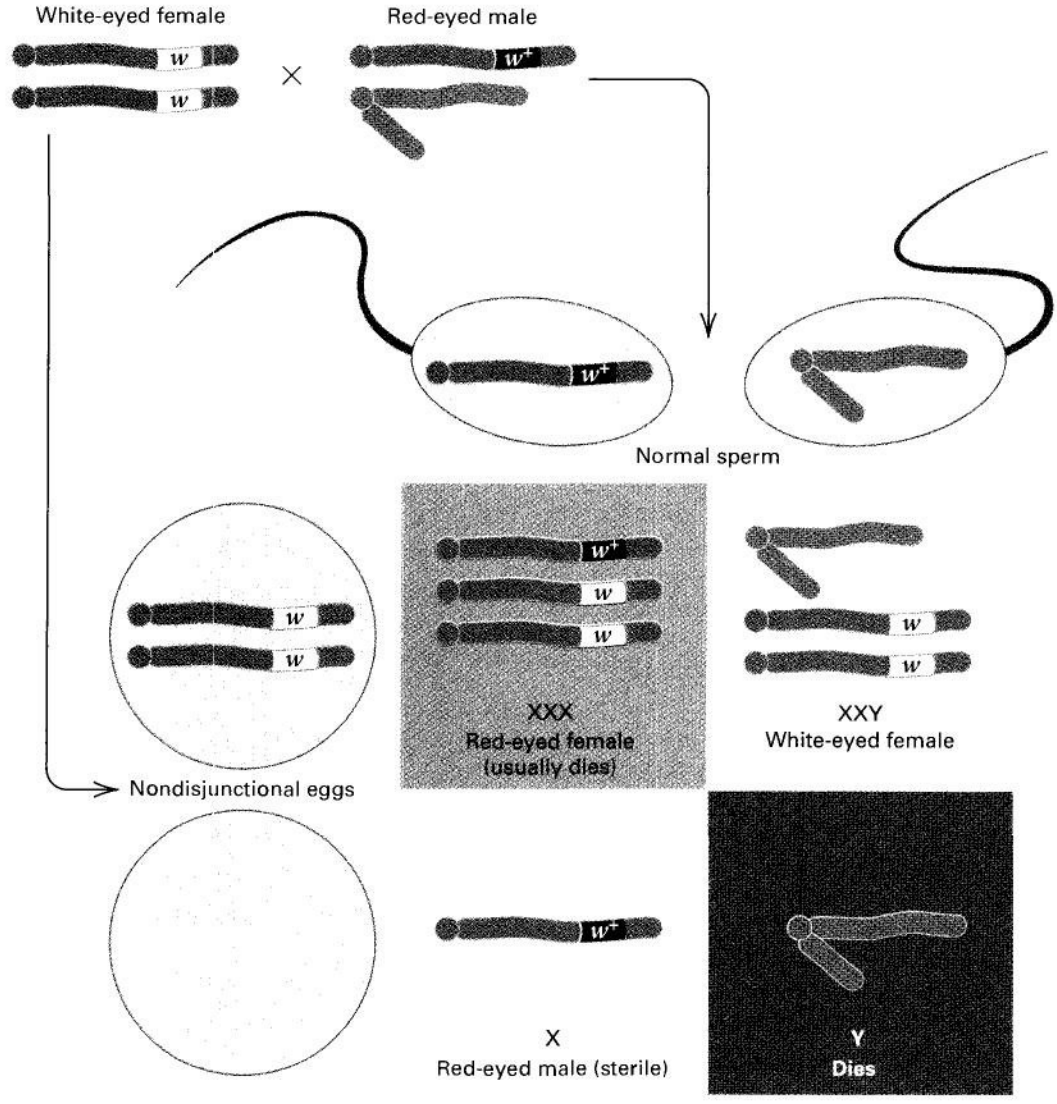


Figure 4.19 Results of nondisjunction of the X chromosomes in the first meiotic division in a female *Drosophila*.

These and related experiments demonstrated conclusively the validity of the chromosome theory of heredity.

Chromosome theory of heredity: Genes are physically located within chromosomes.

Bridges's evidence for the chromosome theory was that exceptional behavior on the part of chromosomes is precisely paralleled by exceptional inheritance of their genes. This proof of the chromosome theory ranks among the most important and elegant experiments in genetics.

Sex Determination in *Drosophila*

In the XX–XY mechanism of sex determination, the Y chromosome is associated with the male. In some organisms, including human beings, this association occurs because the presence of the Y chromosome triggers events in embryonic development that result in the male sexual characteristics. *Drosophila* is unusual among organisms with an XX–XY type of sex determination because the Y chromosome, although associated with maleness, is not male-determining. This is demonstrated by the finding, shown in **Figure 4.20**, that in *Drosophila*, XXY embryos develop into morphologically normal, fertile females, whereas XO embryos develop into morphologically normal, but sterile, males. (The O is written in the formula XO to emphasize that a sex chromosome is missing.) The sterility of XO males shows that the Y chromosome, though not necessary for male development, is essential for male fertility; in fact, the *Drosophila* Y chromosome contains six genes required for the formation of normal sperm.

The genetic determination of sex in *Drosophila* depends on the number of X chromosomes present in an individual fly compared with the number of sets of autosomes. In *Drosophila*, a haploid set of autosomes consists of one copy each of chromosomes 2, 3, and 4 (the autosomes). Normal diploid flies have two haploid sets of autosomes (a homologous pair each of chromosomes 2, 3, and 4) plus either two X chromosomes (in a female) or one X and one Y chromosome (in a male). If we use A to represent a complete haploid complement of autosomes, then

A = chromosome 2
+ chromosome 3
+ chromosome 4

In these terms, a normal male has the chromosomal complement XYAA, and the ratio of X chromosomes to sets of autosomes (the X/A ratio) equals 1X : 2A, or 1 : 2. Normal females have the chromosomal

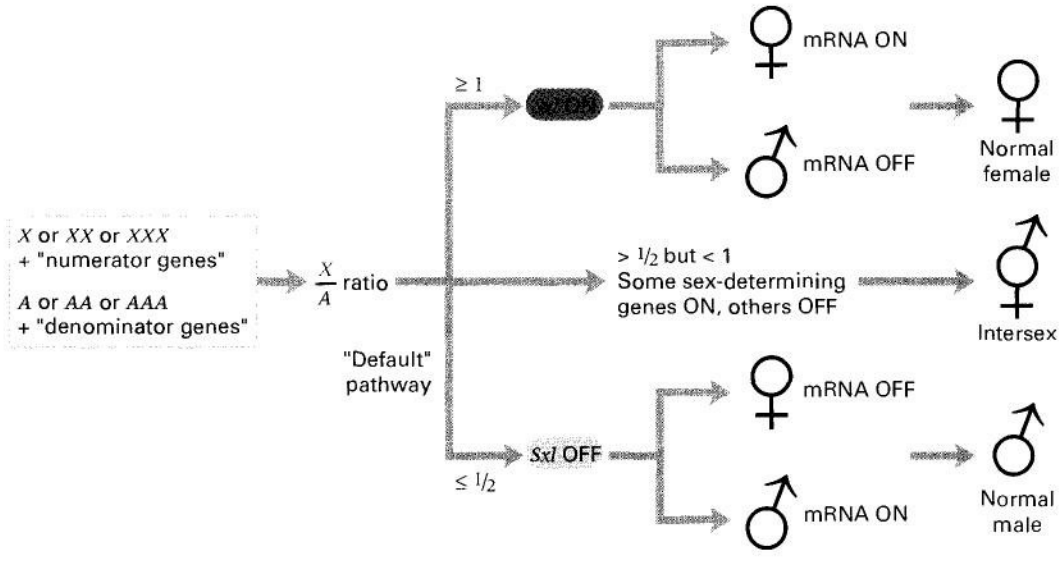


Figure 4.20 Early steps in the genetic control of sex determination in *Drosophila* through the activity of the Sex-lethal protein and ultimately through the "numerator" and "denominator" genes that signal the ratio of X chromosomes to sets of autosomes (the X/A ratio).

complement XXAA, so the female X/A ratio is 2X : 2A, or 1 : 1. Flies with X/A ratios smaller than 1 : 2 (for example, XAAA, which means one X chromosome and three sets of autosomes) are male; those with X/A ratios greater than 1 : 1 (for example, XXXAA, which implies three X chromosomes and two sets of autosomes) are female. Intermediate X/A ratios such as 2 : 3 (for example, XXAAA, with two X chromosomes and three sets of autosomes) develop as intersexes with some characteristics of each sex.

Sexual differentiation in *Drosophila* is controlled by a gene called *Sex-lethal (Sxl)*. The *Sxl* gene codes for two somewhat different proteins, depending on whether a male-specific coding region is included in the messenger RNA. Furthermore, the amount of Sxl protein present in the early embryo regulates the expression of the *Sxl* gene by a feedback mechanism. At low levels of Sxl protein, the male-specific form of the protein is made and shuts off further expression of the gene. At higher levels of the Sxl protein, the female-specific form of the protein is made and the gene continues to be expressed. In some unknown manner, the products of certain genes are sensitive to the X/A ratio and determine the amount of Sxl protein available to regulate the *Sxl* gene. The genes for sensing the number of X chromosomes are called *numerator genes* because they determine the "numerator" of the ratio X/A, and the genes for sensing the number of sets of autosomes are known as *denominator genes* because they determine the "denominator" in the ratio X/A. In normal males (X/A = 1 : 2), there is too little Sxl protein and the *Sxl* gene shuts down; in the absence of *Sxl* expression, sexual differentiation follows the male pathway, which is the "default" pathway. In normal females (X/A = 1 : 1), there is enough Sxl protein that the *Sxl* gene continues to be expressed. Continued expression of the *Sxl* gene initiates a cascade of genetic events, each gene in the cascade controlling one or more other genes downstream, and results in the expression of female-specific gene products and the repression of male-specific gene products. In intermediate situations, wherein the X/A ratio is between 1 : 2 and 1 : 1, some genes specific to each sex are expressed, and the resulting sexual phenotype is ambiguous—an intersex. The Sxl protein is an RNA-binding protein that determines the type of mRNA produced by some of the sex-determining genes. An outline of the genetic control of sex determination in *Drosophila* is shown in Figure 4.20.

4.5 Probability in the Prediction of Progeny Distributions

Genetic transmission includes a large component of chance. A particular gamete from an *Aa* organism might or might not include the *A* allele, depending on chance. A particular gamete from an *Aa Bb* organism might or might not include both the *A* and *B* alleles, depending on the chance orientation of the chromosomes on the metaphase I plate. Genetic ratios result not only from the chance assortment of genes into gametes but also from the chance combination of gametes into zygotes. Although exact predictions are not possible for any particular event, it is possible to determine the probability that a particular event will be realized, as we saw in Chapter 3. In this section, we consider some of the probability methods used in interpreting genetic data.

Using the Binomial Distribution in Genetics

The addition rule of probability deals with outcomes of a genetic cross that are mutually exclusive. Outcomes are "mutually exclusive" if they are incompatible in the sense that they cannot occur at the same time. For example, the possible sex distributions among three children consist of four mutually exclusive outcomes: 0, 1, 2, or 3 girls. These outcomes have probabilities 1/8, 3/8, 3/8, and 1/8, respectively. The addition rule states that the overall probability of any combination of mutually exclusive events is equal to the sum of the probabilities of the events taken separately. For example, the probability that a sibship of size three contains *at least one girl* includes the outcomes 1, 2, and 3 girls, so the overall probability of at least one girl equals 3/8 + 3/8 + 1/8 = 7/8.

The multiplication rule of probability deals with outcomes of a genetic cross that are independent. Any two possible out-

comes are independent if the knowledge that one outcome is actually realized provides no information about whether the other is realized also. For example, in a sequence of births, the sex of any particular child is not affected by the sex of any sibling born earlier and has no influence whatsoever on the distribution of sexes of any siblings born later. Each successive birth is independent of all the others. When possible outcomes are independent, the multiplication rule states that the probability of any combination of outcomes being realized equals the product of the probabilities of all the individual outcomes taken separately. For example, the probability that a sibship of three children will consist of three girls equals $1/2 \times 1/2 \times 1/2$, because the probability of each birth resulting in a girl is $1/2$, and the successive births are independent.

Probability calculations in genetics frequently use the addition and multiplication rules together. For example, the probability that all three children in a family will be of the same sex uses both the addition and the multiplication rules. The probability that all three will be girls is $(1/2)(1/2)(1/2) = 1/8$, and the probability that all three will be boys is also $1/8$. Because these outcomes are mutually exclusive (a sibship of size three cannot include three boys *and* three girls), the probability of either three girls or three boys is the sum of the two probabilities, or $1/8 + 1/8 = 1/4$. The other possible outcomes for sibships of size three are that two of the children will be girls and the other a boy, and that two will be boys and the other a girl. For each of these outcomes, three different orders of birth are possible—for example, GGB, GBG, and BGG—each having a probability of $1/2 \times 1/2 \times 1/2 = 1/8$. The probability of two girls and a boy, disregarding birth order, is the sum of the probabilities for the three possible orders, or $3/8$; likewise, the probability of two boys and a girl is also $3/8$. Therefore, the distribution of probabilities for the sex ratio in families with three children is

GGG	GGB	GBB	BBB
	GBG	BGB	
	BGG	BBG	

$$(1/2)^3 + 3(1/2)^2(1/2) + 3(1/2)(1/2)^2 + (1/2)^3 =$$
$$1/8 + 3/8 + 3/8 + 1/8 = 1$$



Several generations ago the average family size was much larger than it is today. Here is an example of a sibship with 13 children—4 girls and 9 boys—photographed in 1936. If the sex determination of each child is treated as an independent event, then the distribution of sex among sibships is governed by the binomial distribution. [Lucien Aigner/CORBIS]

The sex ratio information in this display can be obtained more directly by expanding the binomial expression $(p + q)^n$, in which p is the probability of the birth of a girl $(1/2)$, q is the probability of the birth of a boy $(1/2)$, and n is the number of children. In the present example,

$$(p + q)^3 = 1p^3 + 3p^2q + 3pq^2 + 1q^3$$

in which the red numerals are the possible number of birth orders for each sex distribution. Similarly, the binomial distribution of probabilities for the sex ratios in families of five children is

$$(p + q)^5 = 1p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + 1q^5$$

Each term tells us the probability of a particular combination. For example, the third term is the probability of three girls (p^3) and two boys (q^2) in a family that has five children:

$$10(1/2)^3(1/2)^2 = 10/32 = 5/16$$

There are $n + 1$ terms in a binomial expansion to the power n. The exponents of p decrease by one from n in the first term to

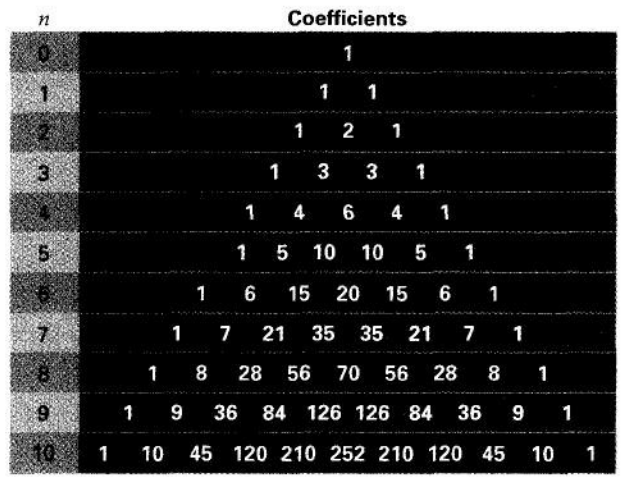


Figure 4.21 Pascal's triangle. The numbers in the nth row are the coefficients of the terms in the expansion of the polynomial $(p + q)^n$.

0 in the last term, and the exponents of q increase by one from 0 in the first term to n in the last term. The coefficients generated by successive values of n can be arranged in a regular triangle known as **Pascal's triangle,** shown for $n = 0$ to 10 in **Figure 4.21**. The horizontal rows of the triangle are symmetrical; each row begins and ends with a 1, and each other entry can be obtained as the sum of the two numbers on either side of it in the row above.

To generalize just a bit, if the probability of event A is p, that of event B is q, and the two events are independent and mutually exclusive, then the probability that A will be realized four times and B two times (in a specified order) is, by the multiplication rule, p^4q^2. Usually we are interested in a combination of events regardless of their order, such as "four A and two B." In this case, we multiply the probability that the combination 4A : 2B will be realized in any specified order by the number of possible orders. The number of different combinations of six events, four of type A and two of type B, is given by the coefficient of p^4q^2 in the expansion of $(p\ \mathrm{A} + q\ \mathrm{B})^6$. This coefficient can be found in the row for $n = 6$ in Pascal's triangle, as the fifth entry from the left. (It is the fifth because the successive entries are the coefficients of p^0q^6, p^1q^5, p^2q^4, p^3q^3, p^4q^2, p^5q^1, and p^6q^0.) Hence the overall probability of four realizations of A and two realizations of B in six trials is given by $15p^4q^2$.

The general rule for repeated trials of events with constant probabilities is as follows:

> If the probability of event A is p and the probability of the alternative event B is q, the probability that, in n trials, event A is realized s times and event B is realized t times is
>
> $$\frac{n!}{s!t!}p^sq^t \qquad (1)$$

in which $s + t = n$ and $p + q = 1$. The symbol $n!$ is read as "n factorial," and it stands for the product of all positive integers from 1 through n (that is, $n! = 1 \times 2 \times 3 \times \cdots \times n$). Values n factorial from $n = 0$ through $n = 15$ are given in **Table 4.2**. The value $0! = 1$ is defined arbitrarily to generalize its use in mathematical formulas. The magnitude of $n!$ increases very rapidly: 15! is more than a trillion.

Equation (1) applies even when either s or t equals 0 because $0! = 1$. (Remember also that any number raised to the zero power equals 1.) Any individual term in the expansion of the binomial $(p + q)^n$ is given by Equation (1) for the appropriate values of s and t. In Pascal's triangle, successive entries in the nth row are the values of $n!/(s!\,t!)$ for $s = 0, 1, 2, \ldots, n$.

It is worth taking a few minutes to justify the factorial part of the binomial expansion in Equation (1), which equals $n!/(s!t!)$. As we have noted, this ratio enumerates all possible ways in which s elements of one kind and t elements of another kind can be arranged in order, provided that the s elements and the t elements are not distinguished among themselves. A specific example might include s yellow peas and t green peas. Although the yellow peas and

Table 4.2 Factorials

n	n!	n	n!
0	1	8	40,320
1	1	9	362,880
2	2	10	3,628,800
3	6	11	39,916,800
4	24	12	479,001,600
5	120	13	6,227,020,800
6	720	14	87,178,291,200
7	5040	15	1,307,674,368,000

the green peas can be distinguished from each other because they have different colors, the yellow peas are not distinguishable from one another (because they are all yellow), nor are the green peas (because they are all green).

The reasoning behind the factorial formula begins with the observation that the total number of elements is $s + t = n$. Given n elements, each distinct from the next, the number of different ways in which they can be arranged is

$$n \times (n - 1) \times (n - 2) \times \cdots \times 3 \times 2 \times 1$$

Why? Because the first element can be chosen in n ways, and once it is chosen, the next can be chosen in $n - 1$ ways (because only $n - 1$ are left to choose from), and once the first two are chosen, the third can be chosen in $n - 2$ ways, and so on. Finally, once $n - 1$ elements have been chosen, there is only one way to choose the last element. The $s + t$ elements can be arranged in $n!$ ways, provided that the elements are all distinguished among themselves. However, applying again the argument we just used, each of the $n!$ particular arrangements must include $s!$ different arrangements of the s elements and $t!$ different arrangements of the t elements, or $s! \times t!$ altogether. Dividing $n!$ by $s! \times t!$ therefore yields the exact number of ways in which the s elements and the t elements can be arranged when the elements of each type are not distinguished among themselves.

Let us consider a specific application of Equation (1), in which we calculate the probability that a mating between two heterozygous parents will yield exactly the expected 3 : 1 ratio of the dominant and recessive traits among sibships of a particular size. The probability p of a child showing the dominant trait is 3/4, and the probability q of a child showing the recessive trait is 1/4. Suppose we wanted to know how often families with eight children contain exactly six children with the dominant phenotype and two with the recessive phenotype. This is the "expected" Mendelian ratio. In this case, $n = 8$, $s = 6$, $t = 2$, and the probability of this combination of events is

$$\frac{8!}{6!\,2!} p^6 q^2 = \frac{6! \times 7 \times 8}{6! \times 2!} (3/4)^6 (1/4)^2 = 0.31$$

That is, in only 31 percent of the families with eight children would the offspring exhibit the expected 3 : 1 phenotypic ratio; the other sibships would deviate in one direction or the other because of chance variation. The importance of this example is in demonstrating that although a 3 : 1 ratio is the "expected" outcome (and is also the single most likely outcome), the majority of the families (69 percent) actually have a distribution of offspring different from 3 : 1.

4.6 Testing Goodness of Fit to a Genetic Hypothesis

Geneticists often need to decide whether an observed ratio is in satisfactory agreement with a theoretical prediction. Mere inspection of the data is unsatisfactory because different investigators may disagree. Suppose, for example, that we crossed a plant having purple flowers with a plant having white flowers and, among the progeny, observed 14 plants with purple flowers and 6 with white flowers. Is this result close enough to be accepted as a 1 : 1 ratio? What if we observed 15 plants with purple flowers and 5 with white flowers? Is this result consistent with a 1 : 1 ratio? There is bound to be statistical variation in the observed results from one experiment to the next. Who is to say what results are consistent with a particular genetic hypothesis? In this section, we describe a test of whether observed results deviate too far from a theoretical expectation. The test is called a test for **goodness of fit,** where the word *fit* means how closely the observed results "fit," or agree with, the expected results.

The Chi-square Method

A conventional measure of goodness of fit is a value called **chi-square** (its symbol is χ^2), which is calculated from the number of progeny observed in each of various classes, compared with the number expected in each of the classes on the basis of some genetic hypothesis. For example, in a cross between plants with purple flowers and those with white flowers, we may be interested in testing the hypothesis that the parent with purple flowers is heterozygous for one pair of alleles determining flower color

and that the parent with white flowers is homozygous recessive. Suppose further that we examine 20 progeny plants from the mating and find that 14 are purple and 6 are white. The procedure for testing this genetic hypothesis (or any other genetic hypothesis) by means of the chi-square method is as follows:

1. ***State the genetic hypothesis in detail, specifying the genotypes and phenotypes of the parents and the possible progeny.*** In the example using flower color, the genetic hypothesis implies that the genotypes in the cross purple × white could be symbolized as *Pp* × *pp*. The possible progeny genotypes are either *Pp* or *pp*.
2. ***Use the rules of probability to make explicit predictions of the types and proportions of progeny that should be observed if the genetic hypothesis is true. Convert the proportions to numbers of progeny (percentages are not allowed in a χ^2 test).*** If the hypothesis about the flower-color cross is true, then we should expect the progeny genotypes *Pp* and *pp* to occur in a ratio of 1 : 1. Because the hypothesis is that *Pp* flowers are purple and *pp* flowers are white, we expect the phenotypes of the progeny to be purple or white in the ratio 1 : 1. Among 20 progeny, the expected numbers are 10 purple and 10 white.
3. ***For each class of progeny in turn, subtract the expected number from the observed number. Square this difference and divide the result by the expected number.*** In our example, the calculation for the purple progeny is $(14 - 10)^2/10 = 1.6$, and that for the white progeny is $(6 - 10)^2/10 = 1.6$.
4. ***Sum the result of the numbers calculated in step 3 for all classes of progeny. The summation is the value of χ^2 for these data.*** The sum for the purple and white classes of progeny is $1.6 + 1.6 = 3.2$, and this is the value of χ^2 for the experiment, calculated on the assumption that our genetic hypothesis is correct.

In symbols, the calculation of χ^2 can be represented by the expression

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

in which Σ means the summation over all the classes of progeny. Note that χ^2 is calculated using the observed and expected *numbers*, not the proportions, ratios, or percentages. Using something other than the actual numbers is the most common beginner's mistake in applying the χ^2 method.

The χ^2 value is reasonable as a measure of goodness of fit, because the closer the observed numbers are to the expected numbers, the smaller the value of χ^2. A value of $\chi^2 = 0$ means that the observed numbers fit the expected numbers perfectly.

As another example of the calculation of χ^2, suppose that the progeny of an $F_1 \times F_1$ cross includes two contrasting phenotypes observed in the numbers 99 and 45. In this case the genetic hypothesis might be that the trait is determined by a pair of alleles of a single gene, in which case the expected ratio of dominant : recessive phenotypes among the F_2 progeny is 3 : 1. Considering the data, the question is whether the observed ratio of 99 : 45 is in satisfactory agreement with the expected 3 : 1. Calculation of the value of χ^2 is illustrated in **Table 4.3**. The total number of progeny is $99 + 45 = 144$. The *expected* numbers in the two classes, on the basis of the genetic hypothesis that the true ratio is 3:1, are calculated as $(3/4) \times 144 = 108$ and $(1/4) \times 144 = 36$. Because there are two classes of data, there are two terms in the χ^2 calculation:

$$\chi^2 = \frac{(99 - 108)^2}{108} + \frac{(45 - 36)^2}{36}$$
$$= 0.75 + 2.25$$
$$= 3.00$$

Once the χ^2 value has been calculated, the next step is to interpret whether this value represents a good fit or a bad fit to the expected numbers. This assessment is done with the aid of the graphs in **Figure 4.22**. The x-axis gives the χ^2 values that reflect goodness of fit, and the y-axis gives the probability P that a worse fit (or one equally bad) would be obtained by chance, assuming that the genetic hypothesis is true. If the genetic hypothesis is true, then the observed numbers should be reasonably close to the expected numbers. Suppose that the observed χ^2 is so large that the probability of a fit as bad or worse is very small. Then the observed results do *not* fit the theoretical expectations. This means that the genetic hypothesis used to calculate the expected numbers of progeny must be rejected, because the observed numbers of progeny deviate too much from the expected numbers.

In practice, the critical values of P are conventionally chosen as 0.05 (the 5 percent level) and 0.01 (the 1 percent level). For P values ranging from 0.01 to 0.05, the probability that chance alone would lead to a fit as bad or worse is between 1 in 20 experiments and between 1 in 100, respectively. This is the middle region in Figure 4.22; if the P value falls in this range, the correctness of the genetic hypothesis is considered very doubtful. The result is said to be **statistically significant** at the 5 percent level. For P values smaller than 0.01, the probability that chance alone would lead to a fit as bad or worse is less than 1 in 100 experiments. This is the lower region in Figure 4.22; in this case, the result is said to be **highly significant** at the 1 percent level, and the genetic hypothesis is rejected outright. If the terminology of statistical

Table 4.3 Calculation of χ^2 for a monohybrid ratio

Phenotype (class)	Observed number	Expected number	Deviation from expected	$\frac{(\text{Deviation})^2}{\text{expected number}}$
Wildtype	99	108	−9	0.75
Mutant	45	36	19	2.25
Total	144	144		$\chi^2 = 3.00$

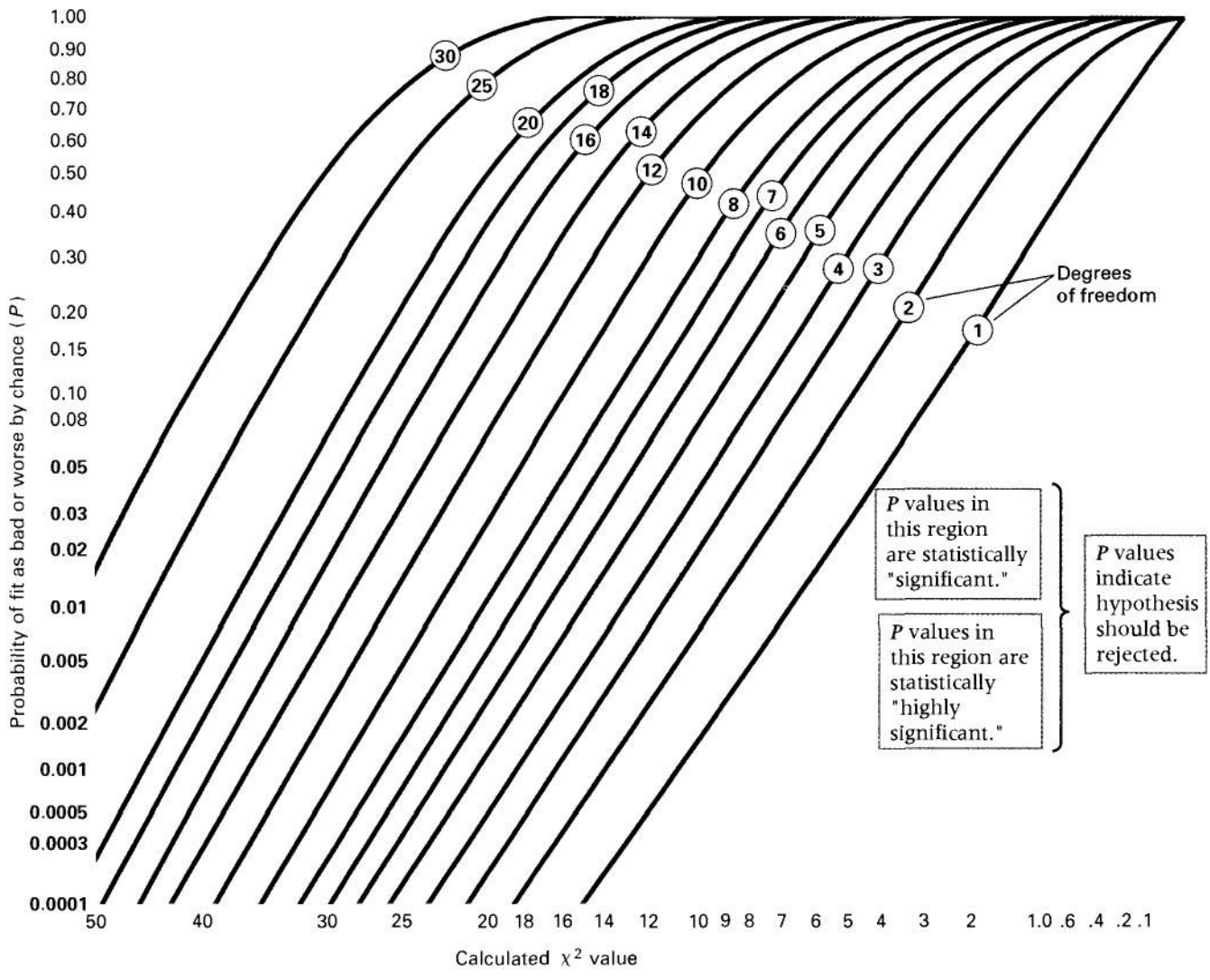


Figure 4.22 Graphs for interpreting goodness of fit to genetic predictions using the chi-square test. For any calculated value of χ^2 along the x-axis, the y-axis gives the probability P that chance alone would produce a fit as bad as or worse than that actually observed, when the genetic predictions are correct. Tests with P in the purple region (less than 5 percent) or in the green region (less than 1 percent) are regarded as statistically significant and normally require rejection of the genetic hypothesis that led to the prediction.

significance seems backward, it is because the term *significant* refers to the magnitude of the deviation between the observed and the expected numbers; in a result that is statistically significant, there is a large ("significant") difference between what is observed and what is expected.

To use Figure 4.22 to determine the P value corresponding to a calculated χ^2, we need the number of **degrees of freedom** of the particular χ^2 test. For the type of χ^2 test illustrated in Table 4.3, the number of degrees of freedom equals the number of classes of data minus 1. Table 4.3 contains two classes of data (wildtype and mutant), so the number of degrees of freedom is $2 - 1 = 1$. The reason for subtracting 1 is that, in calculating the expected numbers of progeny, we make sure that the total number of progeny is the same as that actually observed. For this reason, one of the classes of data is not really "free" to contain any number we might specify; because the expected number in one class must be adjusted to make the total come out correctly, one "degree of freedom" is lost. Analogous χ^2 tests with three classes of data have 2 degrees of freedom, and those with four classes of data have 3 degrees of freedom.

Once we have determined the appropriate number of degrees of freedom, we can interpret the χ^2 value in Table 4.3. Refer to Figure 4.22, and observe that each curve is labeled with its degrees of freedom. To determine the P value for the data in Table 4.3, in which the χ^2 value is 3.00, first find the location of $\chi^2 = 3$ (3.00) along the x-axis in Figure 4.22. Trace vertically on this line until you intersect the curve with 1 degree of freedom. Then trace horizontally to the left until you intersect the y-axis, and read the P value; in this case, $P = 0.08$. This means that chance alone would produce a χ^2 value as great as or greater than 3.00 in about 8 percent of experiments of the type in Table 4.3; and, because the P value is within the upper region, the goodness of fit to the hypothesis of a 3 : 1 ratio of wildtype: mutant is judged to be satisfactory.

As a second illustration of the χ^2 test, we will determine the goodness of fit of Mendel's round-versus-wrinkled data to the expected 3 : 1 ratio. Among the 7324 seeds that he observed, 5474 were round and 1850 were wrinkled. The expected numbers are $(3/4) \times 7324 = 5493$ round and $(1/4) \times 7324 = 1831$ wrinkled. The χ^2 value is calculated as

$$\chi^2 = \frac{(5474 - 5493)^2}{5493} + \frac{(1850 - 1831)^2}{1831}$$

$$= 0.26$$

The fact that the χ^2 is less than 1 already implies that the fit is very good. To find out how good, note that the number of degrees of freedom equals $2 - 1 = 1$ because there are two classes of data (round and wrinkled). From Figure 4.22, the P value for $\chi^2 = 0.26$ with 1 degree of freedom is approximately 0.65. This means that in about 65 percent of all experiments of this type, a fit as bad or worse would be expected simply because of chance. Only about 35 percent of all experiments would yield a better fit.

4.7 Are Mendel's Data Too Good to Be True?

Many of Mendel's experimental results are very close to the expected values. For the ratios listed in Table 3.1 in Chapter 3, the χ^2 values are 0.26 (round versus wrinkled seeds), 0.01 (yellow versus green seeds), 0.39 (purple versus white flowers), 0.06 (inflated versus constricted pods), 0.45 (green versus yellow pods), 0.35 (axial versus terminal flowers), and 0.61 (long versus short stems). (As an exercise in χ^2, you should confirm these calculations for yourself.) All of the χ^2 tests have P values of 0.45 or greater (Figure 4.22), which means that the reported results are in excellent agreement with the theoretical expectations.

The statistician Ronald Fisher pointed out in 1936 that Mendel's results are *suspiciously* close to the theoretical expectations. In a large number of experiments, some experiments can be expected to yield fits that appear doubtful simply because of chance variation from one experiment to the next. In Mendel's data, the doubtful values that are to be expected appear to be missing. **Figure 4.23** shows the observed deviations in Mendel's experiments compared

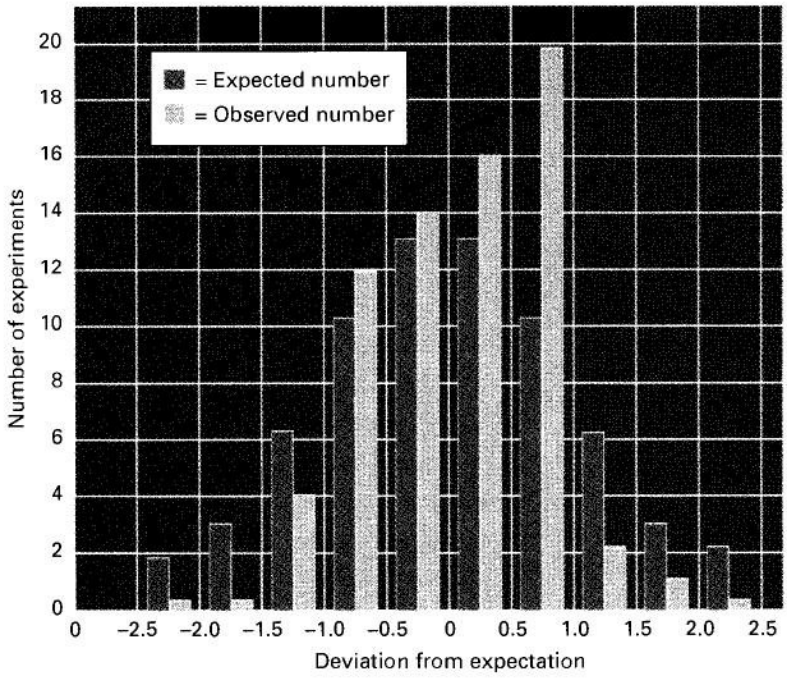


Figure 4.23 Distribution of deviations observed in 69 of Mendel's experiments (yellow bars) compared with expected values (orange bars). There is no suggestion that the data in the middle have been adjusted to improve the fit, but there are fewer experiments with large deviations than might be expected. Several experiments with large deviations may have been discarded or repeated.

with the deviations expected by chance. (The measure of deviation is the square root of the χ^2 value, assigned either a plus or a minus sign according to whether the dominant or the recessive phenotypic class was in excess of the expected number.) For each magnitude of deviation, the height of the bar on the right gives the number of experiments that Mendel observed with such a magnitude of deviation, and that of the bar on the left gives the number of experiments expected to deviate by this amount as a result of chance alone. There are clearly too few experiments with deviations smaller than -1 or larger than $+1$. This type of discrepancy could be explained if Mendel discarded or repeated a few experiments with large deviations that made him suspect that the results were not to be trusted.

Did Mendel cheat? Did he deliberately falsify his data to make them appear better? Mendel's paper reports extremely deviant ratios from individual plants, as well as experiments repeated when the first results were doubtful. These are not the kinds of things a dishonest person would admit. Only a small bias is necessary to explain the excessive goodness of fit in Figure 4.23. In a count of seeds or individual plants, only about 2 phenotypes per 1000 would need to be assigned to the wrong category to account for the bias in the 91 percent of the data generated by the testing of monohybrid ratios. The excessive fit could also be explained if three or four entire experiments were discarded or repeated because deviant results were attributed to pollen contamination or other accident. After careful reexamination of Mendel's data in 1966, the evolutionary geneticist Sewall Wright concluded,

> Mendel was the first to count segregants at all. It is rather too much to expect that he would be aware of the precautions now known to be necessary for completely objective data. . . . Checking of counts that one does not like, but not of others, can lead to systematic bias toward agreement. I doubt whether there are many geneticists even now whose data, if extensive, would stand up wholly satisfactorily under the χ^2 text. . . . Taking everything into account, I am confident that there was no deliberate effort at falsification.

Mendel's data are some of the most extensive and complete "raw data" ever published in genetics. Additional examinations of the data will surely be carried out as new statistical approaches are developed. However, the principal point to be emphasized is that up to the present time, no reputable statistician has alleged that Mendel knowingly and deliberately adjusted his data in favor of the theoretical expectation.

CoNnection

The Case Against Mendel's Gardener

Ronald Aylmer Fisher 1936
University College, London, England
Has Mendel's Work Been Rediscovered?

R. A. Fisher, one of the founders of modern statistics, was also interested in genetics. He gave Mendel's data a thorough going over and made an "abominable discovery." Fisher's unpleasant discovery was that some of Mendel's experiments yielded a better fit to the wrong expected values than they did to the right expected values. At issue are two series of experiments consisting of progeny tests in which F_2 plants with the dominant phenotype were self-fertilized and their progeny examined for segregation to ascertain whether each parent was heterozygous or homozygous. In the first series of experiments, Mendel explicitly states that he cultivated 10 seeds from each plant. What Mendel did not realize, apparently, is that inferring the genotype of the parent on the basis of the phenotypes of 10 progeny introduces a slight bias. The reason is shown in the accompanying illustration. Because a fraction $(3/4)^{10}$ of all progenies from a heterozygous parent will not exhibit segregation, purely as a result of chance, this proportion of Aa *parents gets misclassified as* AA. *The expected proportion of "apparent" AA plants is $(1/3) + (2/3)(3/4)^{10}$ and that of* Aa *plants is $(2/3)[1 - (3/4)^{10}]$, for a ratio of 0.37 : 0.63. In the first series of experiments, among 600 plants tested, Mendel reports a ratio of 0.335 : 0.665, which is in better agreement with the incorrect expectation of 0.33 : 0.67 than with 0.37 : 0.63. In the second series of experiments, among 473 progeny, Mendel reports a ratio of 0.32 : 0.68, which is again in better agreement with 0.33 : 0.67 than with 0.37 : 0.63. This is the "abominable discovery." The reported data differ highly significantly from the true expectation. How could this be? Fisher suggested that Mendel may have been deceived by an overzealous assistant. Mendel did have a gardener who tended the fruit orchards, a man described as untrustworthy and excessively fond of alcohol, and Mendel was also assisted in his pea experiments by two fellow monks. Another possibility, also suggested by Fisher, is that in the second series of experiments, Mendel cultivated more than 10 seeds from each plant. (Mendel does not specify how many seeds were tested from each plant in the second series.) If he cultivated 15 seeds per plant, rather than 10, then the data are no longer statistically significant and the insinuation of data tampering evaporates.*

The reconstruction [of Mendel's experiments] gives no doubt whatever that his report is to be taken entirely literally, and that his experiments were carried out in just the way and much in the order that they are recounted.

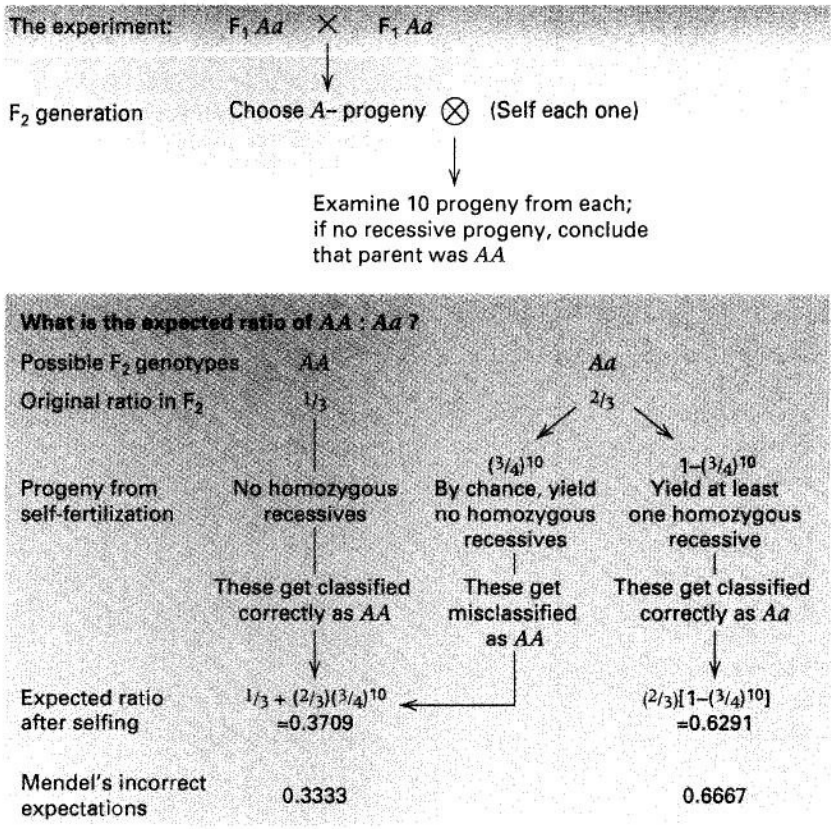
In connection with these tests of homozygosity by examining ten offspring formed by self-fertilization, it is disconcerting to find that the proportion of plants misclassified by this test is not inappreciable. Between 5 and 6 percent of the heterozygous plants will be classified as homozygous. . . . Now among 600 plants tested by Mendel 201 were classified as homozygous and 399 as heterozygous. . . . The deviation [from the true expected values of 222 and 378] is one to be taken seriously. . . . A deviation as fortunate as Mendel's is to be expected once in twenty-nine trials. . . . [In the second series of experiments], a total deviation of the magni-

Chapter Summary

The chromosomes in somatic cells of higher plants and animals are present in pairs of homologous chromosomes. Homologous chromosomes are usually identical in appearance, whereas nonhomologous chromosomes often show differences in size and structural detail that make them visibly distinct from each other. A cell whose nucleus contains two sets of homologous chromosomes is diploid. One set of chromosomes comes from the maternal parent and the other from the paternal parent. Gametes are haploid. A gamete contains only one set of chromosomes, consisting of one member of each pair of homologs.

Mitosis is the process of nuclear division that maintains the chromosome number when a somatic cell divides. Before mitosis, each chromosome replicates, forming a two-part structure consisting of two sister chromatids joined at the centromere (kinetochore). The chromosomes become visible at the onset of mitosis and, at metaphase, become aligned on the metaphase plate perpendicular to the spindle. At anaphase, the centromere of each chromosome divides, and the sister chromatids are pulled by spindle fibers to opposite poles of the cell. The separated sets of chromosomes present in telophase nuclei are genetically identical.

Meiosis is the type of nuclear division that takes place in germ cells, and it reduces the diploid number of chromosomes to the haploid number. The genetic material is replicated before the onset of meiosis, so each chromosome consists of two sister chromatids. The first meiotic division is the reductional division, which reduces the chromosome number by half. The homologous chromosomes first pair (synapsis) and then, at anaphase I, separate. The resulting products contain chromosomes that



tude observed, and in the right direction, is only to be expected once in 444 trials; there is therefore a serious discrepancy. . . . If we could suppose that larger progenies, say fifteen plants, were grown on this occasion, the greater part of the discrepancy would be removed. . . . Such an explanation, however, could not explain the discrepancy observed in the first group of experiments, in which the procedure is specified, without the occurrence of a coincidence of considerable improbability. . . . The reconstruction [of Mendel's experiments] gives no doubt whatever that his report is to be taken entirely literally, and that his experiments were carried out in just the way and much in the order that they are recounted. The detailed reconstruction of his programme on this assumption leads to no discrepancy whatsoever. A serious and almost inexplicable discrepancy has, however, appeared, in that in two series of results the numbers observed agree excellently with the two to one ratio, which Mendel himself expected, but differ significantly from what should have been expected had his theory been corrected to allow for the small size of his test progenies. . . . Although no explanation can be expected to be satisfactory, it remains a possibility among others that Mendel was deceived by some assistant who knew too well what was expected.

Source: Annals of Science 1: 115–137

consist of two chromatids attached to a common centromere. However, as a result of crossing-over, which takes place in prophase I, the chromatids may not be genetically identical along their entire length. In the second meiotic (equational) division, the centromeres divide and the homologous chromatids separate. The end result of meiosis is the formation of four genetically different haploid nuclei.

A distinctive feature of meiosis is the synapsis, or side-by-side pairing, of homologous chromosomes in the zygotene substage of prophase I. During the pachytene substage, the paired chromosomes become connected by chiasmata (the physical manifestations of crossing-over) and do not separate until anaphase I. This separation is called disjunction (unjoining), and failure of chromosomes to separate is called nondisjunction. Nondisjunction results in a gamete that contains either two copies or no copies of a particular chromosome. Meiosis is the physical basis of the segregation and independent assortment of genes. In *Drosophila*, an unexpected pattern of inheritance of the X-linked *white* gene was shown to be accompanied by nondisjunction of the X chromosome; these observations gave experimental proof of the chromosome theory of heredity.

Unlike other chromosome pairs, the X and Y sex chromosomes are visibly different and contain different genes. In mammals and in many insects and other animals, as well as in some flowering plants, the female contains two X chromosomes (XX) and hence is homogametic, and the male contains one X chromosome and one Y chromosome (XY) and hence is heterogametic. In birds, moths, butterflies, and some reptiles, the situa-

tion is the reverse: Females are the heterogametic sex (WZ) and males the homogametic sex (WW). The Y chromosome in many species contains only a few genes. In human beings and other mammals, the Y chromosome includes a male-determining factor. In *Drosophila*, sex is determined by a male-specific or female-specific pattern of gene expression that is regulated by the ratio of the number of X chromosomes to the number of sets of autosomes. In most organisms, the X chromosome contains many genes unrelated to sexual differentiation. These X-linked genes show a characteristic pattern of inheritance that is due to their location in the X chromosome.

The progeny of genetic crosses often conform to the theoretical predictions of the binomial probability formula. The degree to which the observed numbers of different genetic classes of progeny fit theoretically expected numbers is usually found with a chi-square (χ^2) test. On the basis of the criterion of the χ^2 test, Mendel's data fit the expectations somewhat more closely than chance would dictate. However, the bias in the data is relatively small and is likely to be due to recounting or to repeating certain experiments whose results were regarded as unsatisfactory.

Key Terms

anaphase
anaphase I
anaphase II
autosome
bivalent
cell cycle
checkpoints
centromere
chiasma
chi-square
chromatid
chromomere
chromosomes
chromosome complement
crossing-over
cyclins
cytokinesis
degrees of freedom
diakinesis
diploid
diplotene
equational division
G_1 period
G_2 period
gametophyte
germ cell
goodness of fit
haploid
heterogametic
highly significant
homogametic
interphase
kinetochore
leptotene
M period
meiosis
metaphase
metaphase plate
metaphase I
metaphase II
mitosis
mitotic spindle
nondisjunction
nucleoli
pachytene
Pascal's triangle
prophase
prophase I
prophase II
reductional division
S phase
sister chromatids
sex chromosome
somatic cell
spore
sporophyte
statistical significance
synapsis
telophase
telophase I
telophase II
tetrad
X chromosome
X-linked gene
Y chromosome
zygotene

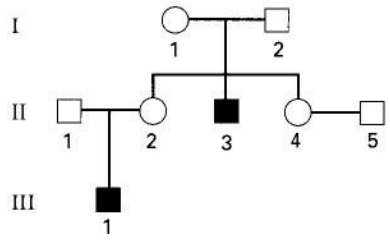
Review the Basics

- What is the genetic significance of the fact that gametes contain half the chromosome complement of somatic cells?
- The term *mitosis* derives from the Greek *mitos*, which means "thread." The term *meiosis* derives from the Greek *meioun*, which means "to make smaller." What feature, or features, of these types of nuclear division might have led to the choice of these terms?
- Explain the meaning of the terms *reductional division* and *equational division*. What is "reduced" or "kept equal"? To which nuclear divisions do the terms refer?
- Explain the following statement: "Independent alignment of nonhomologous chromosomes at metaphase I of meiosis is the physical basis of independent assortment of genes in different chromosomes."
- What are some of the important differences between the first meiotic division and the second meiotic division?
- Draw a diagram of a bivalent and label the following parts: centromere, sister chromatids, nonsister chromatids, homologous chromosomes, chiasma.
- What are the principal characteristics of human pedigrees in which a rare X-linked recessive allele is segregating?
- In what ways is the inheritance of a Y-linked gene different from that of an X-linked gene?
- How did nondisjunction "prove" the chromosome theory of heredity?

- Why is a statistical test necessary to determine whether an observed set of data yields an acceptable fit to the result expected from a particular genetic hypothesis? What statistical test is often used for this purpose?
- What are the conventional *P* values for "significant" and "highly significant" and what do these numbers mean?

Guide to Problem Solving

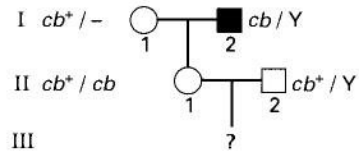
Problem 1 The accompanying diagrams show the appearance of a pair of homologous chromosomes in prophase I of meiosis. Arrange the diagrams in chronological order, and identify each stage as leptotene, zygotene, early pachytene, late pachytene, or diakinesis.



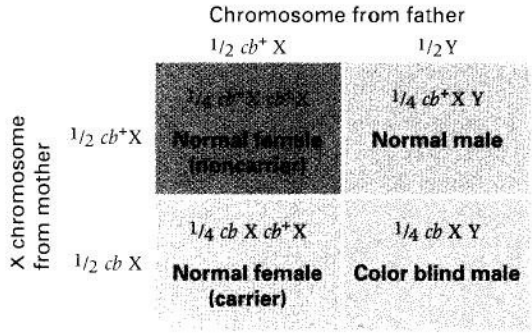
Answer First recall what distinguishes one stage of prophase I from the next. *Leptotene* literally means "thin thread," when each chromosome is in an extended, thread-like condition prior to synapsis; this stage corresponds to diagram B. *Zygotene* means "paired threads," and the pairing begins at the chromosome tips; this is configuration D. *Pachytene* means "thick thread"; it commences when pairing is completed and the homologous chromosomes still appear to be single, which corresponds to diagram E. By late pachytene each homologous chromosome clearly consists of two sister chromatids, and chiasmata are apparent, which is shown in diagram A. *Diakinesis* means "moving apart," and in this stage the synapsed homologous chromosomes begin to repel one another, being held together by the chiasmata, producing configuration C. Therefore, the order of the stages is B–D–E–A–C: leptotene (B), zygotene (D), early pachytene (E), late pachytene (A), and diakinesis (C).

Problem 2 Most color blindness in people is due to relatively common X-linked recessive alleles. A woman with normal color vision whose father was color blind marries a normal man. What types of color vision are expected in the offspring, and in what frequencies?

Answer In these kinds of problems it is helpful to draw a pedigree, showing the information given, and to identify the genotypes of the persons in the pedigree insofar as possible. In this case the pedigree is as indicated, the woman whose father was color blind being number II-1. Her father's genotype must be as shown, because he was color blind. We are told nothing about the mother's genotype, but because II-1 has normal color vision, the mother I-1 must have at least one nonmutant allele (designated cb^+). The normal male II-2 must have a nonmutant allele in his X chromosome, as shown.



The progeny in question are those in generation III, and their expected composition is shown in the Punnett square that follows. The expected offspring are 1/2 normal females, 1/4 normal males, and 1/4 color-blind males. Note that half of the female offspring are carriers of the recessive allele (heterozygous).



Problem 3 Ranch mink with the dark gray coat color known as aleutian are homozygous recessive, *aa*. Genotypes *AA* and *Aa* have the standard deep brown color. A mating of $Aa \times Aa$ produces eight pups.

(a) What is the probability that none of them has the aleutian coat color?

(b) What is the probability of a perfect 3 : 1 distribution of standard to aleutian?

(c) What is the probability of a 1 : 1 distribution in this particular litter?

Answer This kind of problem demonstrates the effect of chance variation in segregation ratios in small sibships. In the mating $Aa \times Aa$, the probability that any particular pup has the standard coat color is 3/4, and that of aleutian is 1/4. Therefore, in a litter of size 8, the probability of 0, 1, 2, 3, . . . pups with the standard coat color is given by successive terms in the binomial expansion of $[(3/4) + (1/4)]^8$. The specific probability of r standard pups and $8 - r$ aleutian pups is given by $8!/[r!(8 - r)!] \times (3/4)^r \times (1/4)^{(8 - r)}$.

(a) The probability that none is aleutian means that $r = 8$, so the probability is

$$(3/4)^8 = 6561/65{,}536 = 0.10$$

(b) A perfect 3 : 1 ratio in the litter means that $r = 6$, and the probability is

$$[8!/(6!\,2!)](3/4)^6(1/4)^2 = 28(729/4096)(1/16) = 0.31$$

(That is, a little less than a third of the litters would have the "expected" Mendelian ratio.)

(c) A 1 : 1 distribution in the litter means that $r = 4$, and the probability is

$$[8!/(4!\,4!)](3/4)^4(1/4)^4 = 70(81/256)(1/256) = 0.09$$

Problem 4 Certain varieties of maize are true-breeding either for colored aleurone (the outer layer of the seed) or for colorless aleurone. A cross of a colored variety with a colorless variety gave an F_1 with colored seeds. Among 1000 seeds produced by the cross $F_1 \times$ colored, all were colored; and, among 1000 seeds produced by the cross $F_1 \times$ colorless, 525 were colored, and 475 were colorless.

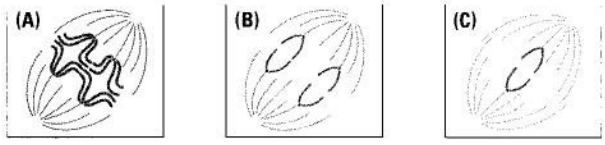
(a) What genetic hypothesis can explain these data?
(b) Using the criterion of a χ^2 test, evaluate whether the data are in satisfactory agreement with the model.

Answer **(a)** Mendelian segregation in a heterozygote is suggested by the 525 : 475 ratio, and dominance is implied by the crosses of colored × colorless and $F_1 \times$ colored. Therefore, a hypothesis that seems to fit the data is that there is a dominant allele for colored aleurone (say, *C*) and the true-breeding colored and colorless varieties are *CC* and *cc*, respectively. The F_1 has the genotype *Cc*. The cross $F_1 \times$ colored (*CC*) yields 1 *CC* : 1 *Cc* progeny (all colored), and the cross $F_1 \times$ colorless (*cc*) is expected to produce a 1 : 1 ratio of colored (*Cc*) to colorless (*cc*). **(b)** The expected numbers are 500 colored and 500 colorless, and therefore the χ^2 equals $(525 - 500)^2/500 + (475 - 500)^2/500 = 2.5$. Because there are two classes of data, there is 1 degree of freedom, and the *P* value is approximately 0.12 (see Figure 4.22). The *P* value is greater than 0.05, so the goodness of fit is regarded as satisfactory.

Analysis and Applications

4.1 At what stage in mitosis and meiosis are the chromosomes replicated? When do the chromosomes first become visible in the light microscope?

4.2 The diagrams shown here depict anaphase in cell division in a cell of a hypothetical organism with one pair of chromosomes. Identify the panels as being anaphase of mitosis, anaphase I of meiosis, or anaphase II of meiosis, stating on what basis you reached your conclusions.



4.3 If a cell contains 23 pairs of chromosomes immediately after completion of mitotic telophase, how many chromatids were present in metaphase?

4.4 The Greek roots of the terms *leptotene, zygotene, pachytene, diplotene,* and *diakinesis* literally mean *thin thread, paired thread, thick thread, doubled thread* and *moving apart,* respectively. How are these terms appropriate in describing the appearance and behavior of the chromosomes during prophase I? (Incidentally, the ending *-tene* denotes the adjective; the nouns are formed by dropping *-tene* and adding *-nema,* yielding the alternative terms *leptonema, zygonema, pachynema, diplonema,* and *diakinesis,* which are preferred by some authors.)

4.5 The first meiotic division is often called the *reductional* division and the second meiotic division the *equational* division. Which feature of the chromosome complement is reduced in the first meiotic division and which is kept equal in the second?

4.6 Maize is a diploid organism with 10 pairs of chromosomes. How many chromatids and chromosomes are present in the following stages of cell division:
(a) Metaphase of mitosis?
(b) Metaphase I of meiosis?
(c) Metaphase II of meiosis?

4.7 Garden peas have a somatic chromosome number of 14. If the centromeres of the 7 homologous pairs are designated as *Aa, Bb, Cc, Dd, Ee, Ff,* and *Gg:*
(a) How many different combinations of centromeres can be produced during meiosis?
(b) What is the probability that a gamete will contain only centromeres designated by capital letters?

4.8 Emmer wheat (*Triticum dicoccum*) has a somatic chromosome number of 28, and rye (*Secale cereale*) has a somatic chromosome number of 14. Hybrids produced by crossing these cereal grasses are highly sterile and have many characteristics intermediate between the parental species. How many chromosomes do the hybrids possess?

4.9 X-linked inheritance is occasionally called *crisscross* inheritance. In what sense does the X chromosome move back and forth between the sexes every generation? In what sense is the expression misleading?

4.10 The most common form of color blindness in human beings results from an X-linked recessive gene. A phenotypically normal couple have a normal daughter and a son who is color-blind. What is the probability that the daughter is heterozygous?

4.11 A chromosomally normal woman and a chromosomally normal man have a son whose sex-chromosome constitution is XYY. In which parent, and in which meiotic division, did the nondisjunction take place?

4.12 The mutation for Bar-shaped eyes in *Drosophila* has the following characteristics of inheritance:

(a) Bar males × wildtype females produce wildtype sons and Bar daughters,

(b) The Bar females from the mating in part **a**, when mated with wildtype males, yield a 1 : 1 ratio of Bar : wildtype sons and Bar : wildtype daughters.

What mode of inheritance do these characteristics suggest?

4.13 Vermilion eye color in *Drosophila* is determined by the recessive allele, *v*, of an X-linked gene, and the wildtype eye color determined by the v^+ allele is brick red. What genotype and phenotype ratios would be expected from the following crosses:

(a) vermilion male × wildtype female

(b) vermilion female × wildtype male

(c) daughter from mating in part **a** × wildtype male

(d) daughter from mating in part **a** × vermilion male

4.14 The autosomal recessive allele, *bw* for brown eyes in *Drosophila*, interacts with the X-linked recessive allele, *v*, for vermilion, to produce white eyes. What eye-color phenotypes, and in what proportions, would be expected from a cross of a white-eyed female (genotype *v v; bw bw*) with a brown-eyed male (genotype v^+ Y; *bw bw*)?

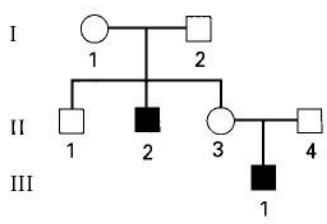
4.15 It is often advantageous to be able to determine the sex of newborn chickens from their plumage. How could this be done by using the W-linked dominant allele *S* for silver plumage and the recessive allele *s* for gold plumage? (Remember that, in chickens, the homogametic and heterogametic sexes are the reverse of those in mammals.)

4.16 A recessive mutation of an X-linked gene in human beings results in hemophilia, marked by a prolonged increase in the time needed for blood clotting. Suppose that phenotypically normal parents produce two normal daughters and a son affected with hemophilia.

(a) What is the probability that both of the daughters are heterozygous carriers?

(b) If one of the daughters mates with a normal man and produces a son, what is the probability that the son will be affected?

4.17 In the pedigree illustrated, the shaded symbols represent persons affected with an X-linked recessive form of mental retardation. What are the genotypes of all the persons in this pedigree?



4.18 For an autosomal gene with five alleles, there are fifteen possible genotypes (five homozygotes and ten heterozygotes). How many genotypes are possible with five alleles of an X-linked gene?

4.19 Attached-X chromosomes in *Drosophila* are formed from two X chromosomes attached to a common centromere. Females of genotype C(1)RM/Y, in which C(1)RM denotes the attached-X chromosomes, produce C(1)RM-bearing and Y-bearing gametes in equal proportions. What progeny are expected to result from the mating between a male carrying the X-linked allele, *w*, for white eyes and an attached-X female with wildtype eyes? How does this result differ from the typical pattern of X-linked inheritance? (Note: *Drosophila* zygotes containing three X chromosomes or no X chromosomes do not survive.)

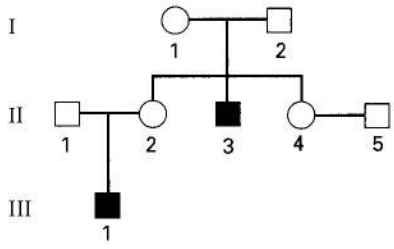
4.20 People who have the sex chromosome constitution XXY are phenotypically male. A woman heterozygous for an X-linked mutation for color blindness mates with a normal man and produces an XXY son, who is color-blind. What kind of nondisjunction can explain this result?

4.21 Mice with a single X chromosome and no Y chromosome (an XO sex-chromosome constitution) are fertile females. Assuming that at least one X chromosome is required for viability, what sex ratio is expected among surviving progeny from the mating XO ♀ × XY ♂?

4.22 In the accompanying pedigree, the shaded symbols represent persons affected with X-linked hemophilia, a blood-clotting disorder.

(a) If the woman identified as II-2 has two more children, what is the probability that neither will be affected?

(b) What is the probability that the first child of the mating II-4 × II-5 will be affected?



4.23 Assume a sex ratio at birth of 1 : 1 and consider two sibships, A and B, each with three children.

(a) What is the probability that A consists only of girls and B only of boys?

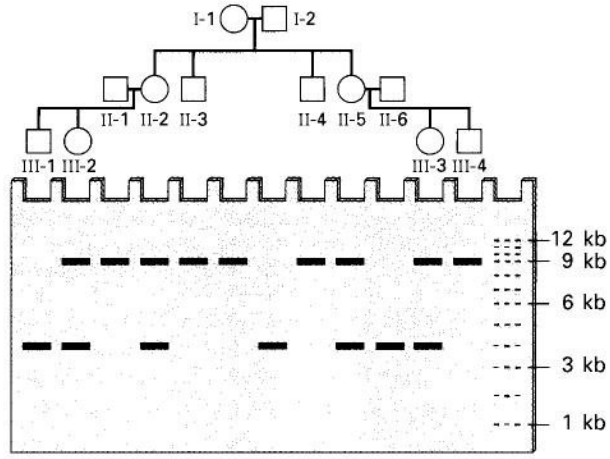
(b) What is the probability that one sibship consists only of girls and the other only of boys?

4.24 A litter of mice includes 8 pups. What is the probability that it contains an even number of males and an even number of females? Assume an equal likelihood of male and female pups, and for purposes of this problem regard 0 as an even number.

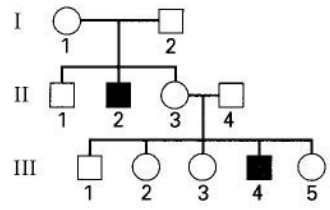
4.25 A litter of mice includes 7 pups. What is the probability that its sex ratio is 4 : 3 or 3 : 4? Assume an equal likelihood of male and female pups.

4.26 Shown at the right is a pedigree of X-linked muscular dystrophy (Duchenne muscular dystrophy). What is the probability that the woman III-3 is a carrier? What is the probability that a son of hers will be affected?

4.27 The accompanying pedigree and gel diagram shows the molecular phenotypes for an RFLP with two alleles. What mode of inheritance does the pedigree suggest? On the basis of this hypothesis, and using A_1 to represent the RFLP allele associated with the 4-kb band and A_2 to represent that associated with the 9-kb band, deduce the genotype of each individual in the pedigree.



4.28 A hybrid corn plant with green leaves was test-crossed with a plant having yellow-striped leaves. When 250 seedlings were grown, 140 of the seedlings had green leaves and 110 had yellow-striped leaves. Using the chi-squared method, test this result for agreement with the expected 1 : 1 ratio.



Problem 4.26

4.29 A cross was made to produce *D. melanogaster* flies heterozygous for two pairs of alleles: dp^+ and dp, which determine long versus short wings, and e^+ and e, which determine gray versus ebony body color. The following F_2 data were obtained:

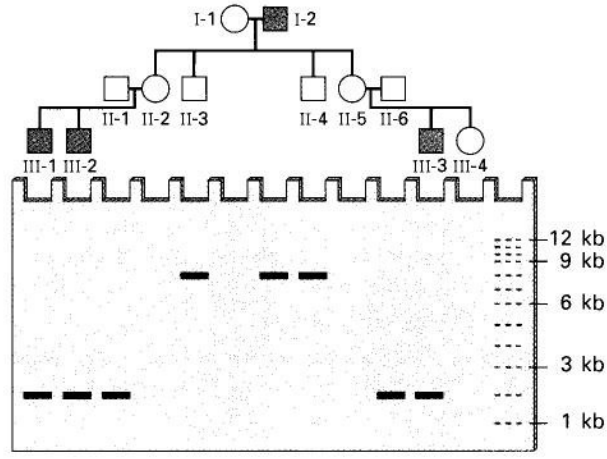
Long wing, gray body	462
Long wing, ebony body	167
Short wing, gray body	127
Short wing, ebony body	44

Test these data for agreement with the 9 : 3 : 3 : 1 ratio expected if the two pairs of alleles segregate independently.

4.30 Tall yellow-flowered crocus is mated with short white flowered crocus. Both varieties are true breeding. All the F_1 plants are backcrossed with the short white-flowered variety. This backcross yielded 234 tall yellow, 203 tall white; 175 short yellow, and 188 short white plants. Does the observed result fit the genetic hypothesis of 1 : 1 : 1 : 1 segregation as assessed by a χ^2 test?

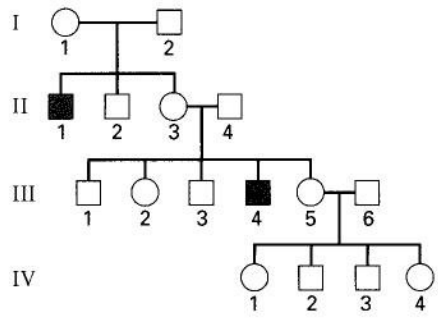
Challenge Problems

4.31 The pedigree and gel diagram shown here pertain to a morphological phenotype (shaded symbols in the pedigree) and a molecular phenotype (a two-allele RFLP). What mode of inheritance do these data suggest for each trait?



4.32 In mice, the dominant T allele results in a short tail, but the homozygous T/T genotype is lethal. A cross between two short-tailed mice produces a litter of five pups. What is the expected distribution of tailless to tailed mice? Why are the two most probable outcomes equally probable?

4.33 X-linked hemophilia is present in the pedigree illustrated here. Determine the probability that the woman III-5 is a carrier, taking into account the information that she has had two normal sons. What is the probability that her next son will be affected?



GeNETiCS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by-chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

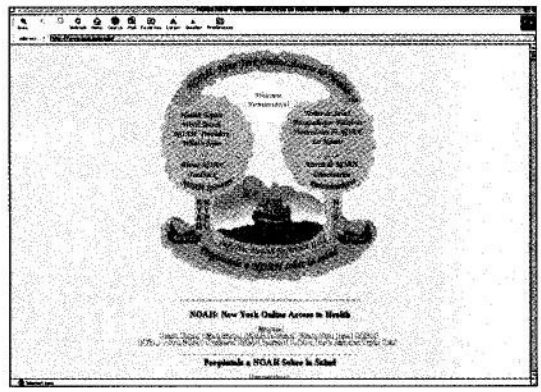
- A thorough command of chromosome behavior during cell division is an absolute prerequisite to understanding transmission genetics. The keyword **meiosis** links to a site that reviews key features of mitosis and meiosis and reinforces this information with animations.

- Duchenne muscular dystrophy (**DMD**) is a wasting disease of muscle that results from an X-linked mutation in one of the largest genes in the human genome, spanning more than 2.3 million nucleotide pairs. It affects about one in every 3500 newborn boys and has no cure. Early signs usually occur between the ages of 2 and 6 when affected boys are observed to fall frequently, to have difficulty getting up from a reclining position, and to walk with a waddling gait. Log onto this keyword site to learn more about Duchenne muscular dystrophy and other inherited diseases of muscle.

- Thomas Hunt **Morgan** was one of the most influential scientists of the twentieth century. His principal experimental work was carried out at Columbia University in the period 1910–1928, when he and his students (notably A. H. Sturtevant, C. B. Bridges, and H. J. Muller) pioneered the use of *Drosophila* as a model organism for studies in genetics. When this work began, it was generally assumed that the genetic information could not reside in the chromosomes, but Morgan's discovery of the *white-eye* mutation and his demonstration of X-linkage provided critical evidence for the chromosome theory. Consult this keyword site for an engaging early history of *Drosophila* genetics and *Drosophila* geneticists, enlivened by historical photographs.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 4 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 4.



Further Reading

Allshire, R. C. 1997. Centromeres, checkpoints and chromatid cohesion. *Current Opinion in Genetics & Development* 7: 264.

Bowen, I. D., S. M. Bowen, and A. H. Jones. 1998. *Mitosis and Apoptosis: Matters of Life and Death*. New York: Chapman & Hall.

Chandley, A. C. 1988. Meiosis in man. *Trends in Genetics* 4: 79.

Cohen, J. S., and M. E. Hogan. 1994. The new genetic medicine. *Scientific American*, December.

Garland, E. Allen. 1978. *Thomas Hunt Morgan: The Man and His Science*. Princeton, New Jersey: Princeton University Press.

McIntosh, J. R., and K. L. McDonald. 1989. The mitotic spindle. *Scientific American*, October.

McKusick, V. A. 1965. The royal hemophilia. *Scientific American*, August.

Miller, O. J. 1995. The fifties and the renaissance of human and mammalian genetics. *Genetics* 139: 484.

Page, A. W., and T. L. Orr-Weaver. 1997. Stopping and starting the meiotic cell cycle. *Current Opinion in Genetics & Development* 7: 23.

Sokal, R. R., and F. J. Rohlf. 1969. *Biometry*. New York: Freeman.

Sturtevant, A. H. 1965. *A Short History of Genetics*. Harper & Row.

Voeller, B. R., ed. 1968. *The Chromosome Theory of Inheritance: Classical Papers in Development and Heredity*. New York: Appleton-Century-Crofts.

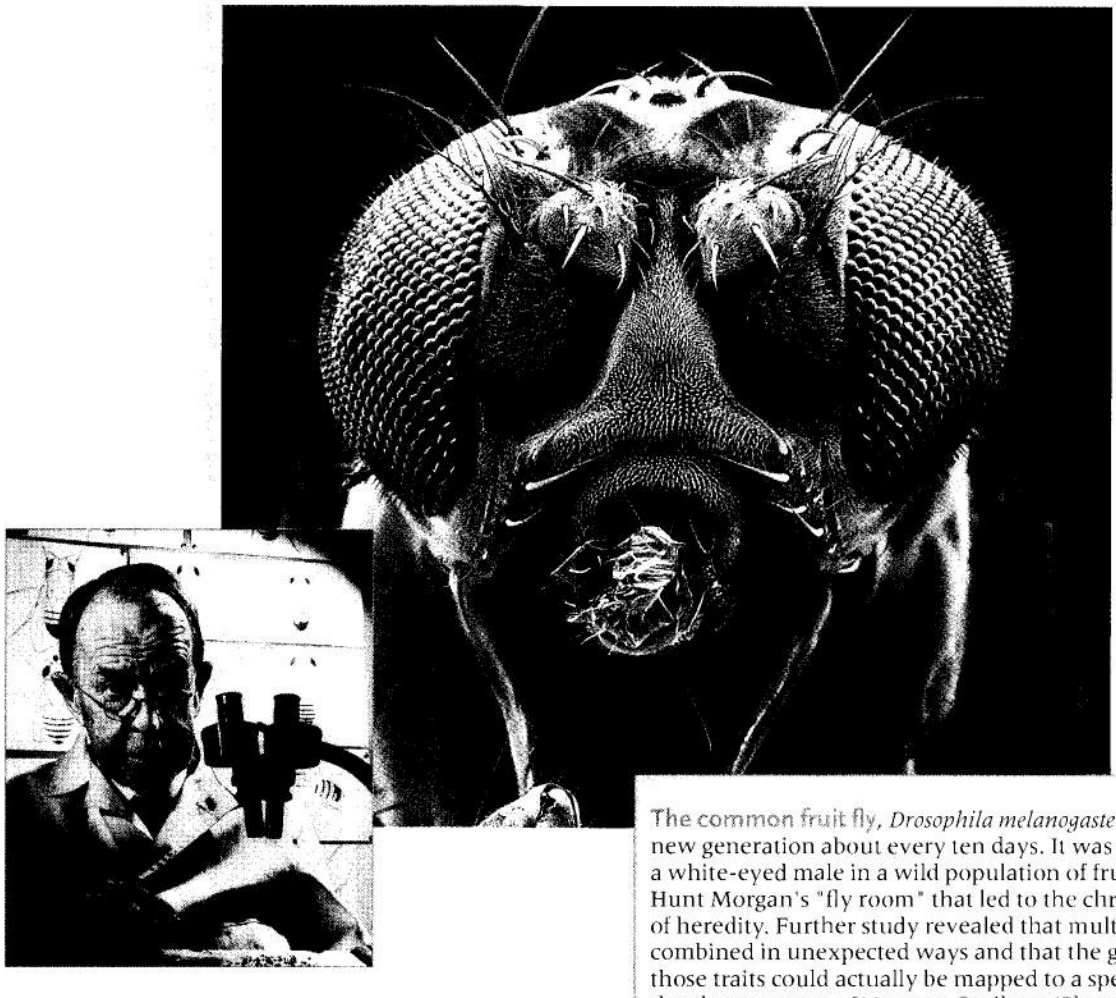
Wienberg, J., and R. Stanyon. 1997. Comparative painting of mammalian chromosomes. *Current Opinion in Genetics & Development* 7: 784–791.

Welsh, M. J., and A. E. Smith. 1995. Cystic fibrosis. *Scientific American*, December.

Zielenski, J., and L. C. Tsui. 1995. Cystic fibrosis: Genotypic and phenotypic variations. *Annual Review of Genetics* 29: 777.

CHAPTER 5

Genetic Linkage and Chromosome Mapping



The common fruit fly, *Drosophila melanogaster*, can produce a new generation about every ten days. It was the appearance of a white-eyed male in a wild population of fruit flies in Thomas Hunt Morgan's "fly room" that led to the chromosomal theory of heredity. Further study revealed that multiple traits combined in unexpected ways and that the genes controlling those traits could actually be mapped to a specific location in the chromosomes. [Morgan: Omikron/Photo Researchers, Inc.; fly: Oliver Meckes/Photo Researchers, Inc.]

CHAPTER OUTLINE

5.1 Linkage and Recombination of Genes in a Chromosome
- Coupling versus Repulsion of Syntenic Alleles
- The Chi-square Test for Linkage
- Each Pair of Linked Genes Has a Characteristic Frequency of Recombination
- Recombination in Females versus Males

5.2 Genetic Mapping
- Map Distance and Frequency of Recombination
- Crossing-over
- Crossing-over Takes Place at the Four-Strand Stage of Meiosis
- Multiple Crossovers

5.3 Genetic Mapping in a Three-Point Testcross
- Chromosome Interference in Double Crossovers
- Genetic Mapping Functions
- Genetic Map Distance and Physical Distance

5.4 Genetic Mapping in Human Pedigrees
- Maximum Likelihood and Lod Scores

5.5 Mapping by Tetrad Analysis
- Analysis of Unordered Tetrads
- Analysis of Ordered Tetrads

5.6 Recombination within Genes

5.7 Mitotic Recombination

PRINCIPLES

- Genes that are located sufficiently close together in a chromosome do not undergo independent assortment; they are said to be linked.
- The alleles of linked genes tend to be inherited as a group.
- Crossing-over between homologous chromosomes results in recombination that breaks up combinations of linked alleles.
- A genetic map depicts the relative positions of genes along a chromosome.
- The map distance between genes in a genetic map is related to the rate of recombination between the genes.
- Physical distance along a chromosome is often (but not always) correlated with map distance.
- Tetrads are sensitive indicators of linkage because each contains all the products of a single meiosis.
- Recombination can also take place between nucleotides within a gene.

CONNECTIONS

Genes All in a Row
Alfred H. Sturtevant 1913
The Linear Arrangement of Six Sex-Linked Factors in Drosophila, *as Shown by Their Mode of Association*

Human Gene Map
Jeffrey C. Murry and 26 other investigators 1994
A Comprehensive Human Linkage Map with Centimorgan Density

In meiosis, homologous chromosomes form pairs, and then the members of each pair separate from one another. Because homologous chromosomes behave as units when they undergo meiosis, one might expect that genes located in the same chromosome would not undergo the same pattern of independent assortment we see in nonhomologous chromosomes. Genes that are always transmitted together are said to show complete **linkage.** Pioneering studies of this issue were carried out by Thomas Hunt Morgan, who studied mutations located in the X chromosome of *Drosophila*. He did observe linkage, but what he observed was an incomplete form of linkage, evidenced by progeny who exhibited a combination of traits not exhibited in the parents.

Linkage between genes is usually incomplete because homologous chromosomes can undergo an exchange of segments when they are paired. So while the alleles present in each chromosome do tend to remain together in inheritance, some chromosomes can be transmitted that have new combinations of alleles. An exchange event (crossing-over) between homologous chromosomes results in the **recombination** of genes, yielding daughter chromosomes that carry combinations of alleles not present in the parental chromosomes. What we learn in this chapter is that the probability of crossing-over between any two genes serves as a measure of genetic distance between the genes. This makes it possible to construct a *genetic map*, a diagram of a chromosome showing the relative positions of the genes. The genetic mapping of linked genes is an important research tool in genetics because it enables a new gene to be assigned to a chromosome and even to a precise position relative to other genes in the same chromosome. Genetic mapping is often a first step in the identification and isolation of a new gene. Genetic mapping is essential in human genetics for the identification of genes associated with hereditary diseases, such as the genes whose presence predisposes women carriers to the development of breast cancer.

5.1 Linkage and Recombination of Genes in a Chromosome

As we saw in Chapter 4, a direct test of independent assortment is to carry out a testcross between an F_1 double heterozygote (*Aa Bb*) and the double recessive homozygote (*aa bb*). Shown in **Figure 5.1** are the expected gametes from the *Aa Bb* parent when the genes are on different chromo-

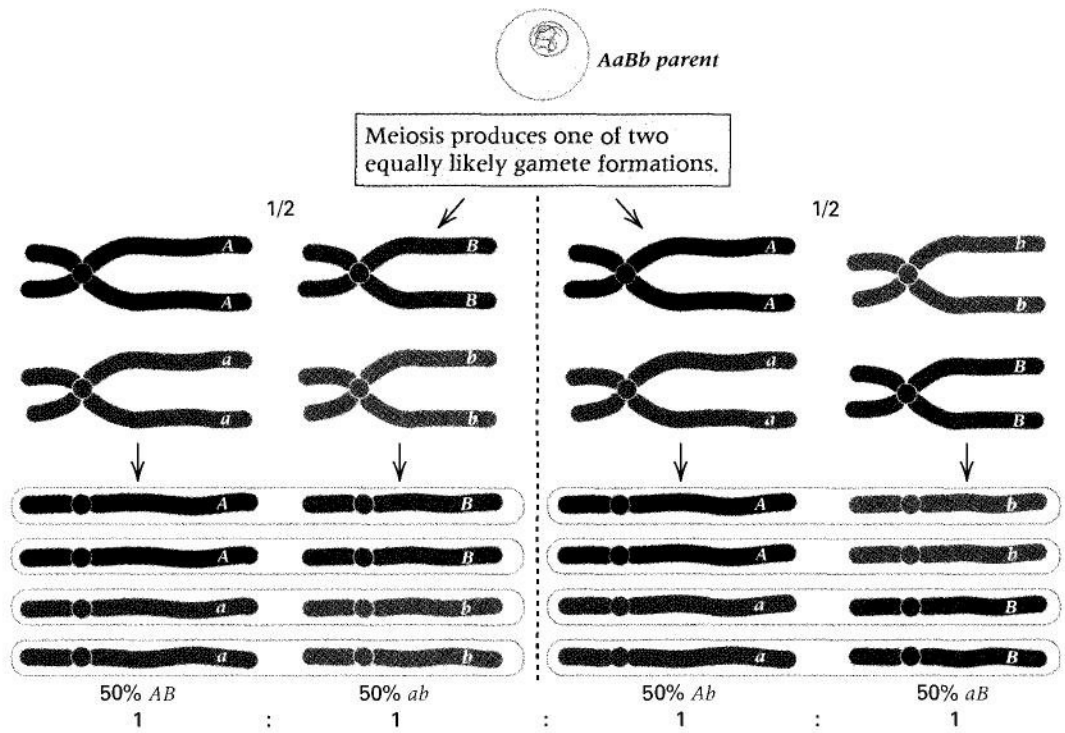


Figure 5.1 The physical basis of the independent assortment of genes in different chromosomes is that nonhomologous chromosomes orient independently at metaphase I of meiosis. In this example, the orientation on the left results in *A B* and *a b* gametes, whereas that on the right results in *A b* and *a B* gametes.

somes. Because the possible metaphase I alignments of homologous chromosomes are equally likely, the double heterozygote produces all four possible types of gametes in equal proportions. The alleles present in a gamete are typically written in a row with a space between each gene; in this case the four equally likely types of gametes are *A B*, *A b*, *a B*, and *a b*.

Independent assortment takes place in the *Aa Bb* genotype whether the *parents* were genotypically *AA BB* and *aa bb* or genotypically *AA bb* and *aa BB*. The four products of meiosis are still expected in equal proportions. Because of independent assortment, an expected 50 percent of the testcross progeny result from gametes carrying the same combination of alleles present in the parents of the double heterozygote (**parental types**), and 50 percent result from gametes carrying new combinations of the alleles (**recombinants**). For example, if the *Aa Bb* genotype came from the mating *AA BB* × *aa bb*, then the *A B* and *a b* gametes would be parental and the *A b* and *a B* gametes recombinant. On the other hand, if the *Aa Bb* genotype came from the mating *AA bb* × *aa BB*, then the *A b* and *a B* gametes would be parental and the *a b* and *A B* gametes recombinant.

In Morgan's early experiments with *Drosophila*, he examined the segregation of the X-linked alleles w versus w^+, which determine white versus normal red eyes, and the alleles m versus m^+, which determine miniature versus normal wing size. When genes are linked, it is important to keep track of which alleles are present together in the same chromosome. This is generally done by using a slash (also called a virgule) to separate the alleles present in homologous chromosomes. Morgan's initial cross was between females with white eyes and normal wings and males with red eyes and miniature wings. Using the slash, we would write this cross as

$$w\,m^+/w\,m^+ \;♀♀ \;\times\; w^+\,m/\mathrm{Y}\;♂♂$$

The resulting F_1 progeny consisted of wild-type females (genotype $w\,m^+/w^+\,m$) and white-eyed, nonminiature males (genotype $w\,m^+/\mathrm{Y}$). When these were crossed,

$$w\,m^+/w^+\,m\;♀♀ \;\times\; w\,m^+\,\mathrm{Y}\;♂♂$$

the female progeny were all nonminiature with a 1 : 1 ratio of red : white eyes (if you do not understand why, write out the Punnett square), and the male progeny were as follows:

white eye, normal wing ($w\,m^+/\mathrm{Y}$)	226	66.5 percent parental combinations
red eye, miniature wing ($w^+\,m/\mathrm{Y}$)	202	
red eye, normal wing ($w^+\,m^+/\mathrm{Y}$)	114	33.5 percent recombinant chromosomes
white eye, miniature wing ($w\,m/\mathrm{Y}$)	102	
	644	

Because each male receives his X chromosome from his mother, a male's phenotype reveals the genotype of the X chromosome that he inherited. The results of the experiment show a great departure from the 1 : 1 : 1 : 1 ratio of the four male phenotypes expected with independent assortment, yet the ratios of $w : w^+$ and of $m : m^+$ are both approximately 1 : 1. This type of deviation from independent assortment shows that the parental combination of the alleles for the w and m genes tend to remain together in inheritance, but they are not completely linked. In this experiment, the combinations of alleles in the parental chromosomes were present in 428/644 = 66.5 percent of the F_2 males, and recombinant (nonparental) chromosomes were present in 216/644 = 33.5 percent of the F_2 males. The 33.5 percent value for the recombinant chromosomes is called the **frequency of recombination;** it can also be written as 0.335. This frequency of recombination should be contrasted with the 50 percent recombination expected with independent assortment.

The recombinant X chromosomes w^+m^+ and $w\,m$ result from crossing-over in meiosis in F_1 females. In this example, the frequency of recombination between the linked w and m genes was 33.5 percent, but with other pairs of linked genes it ranges from near 0 to 50 percent. Even genes in the same chromosome, if they are sufficiently far apart, can undergo independent assortment, which means that their frequency of recombination equals 50 percent. This implies the following principle:

> Genes with recombination frequencies less than 50 percent are present in the same chromosome (linked). Two genes that undergo independent assortment, indicated by a

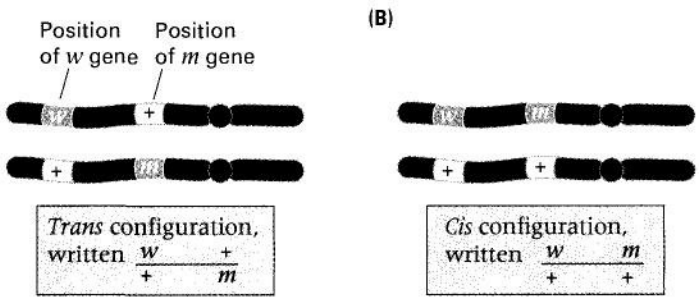


Figure 5.2 Possible configurations of the mutant alleles in a genotype heterozygous for two mutations. (A) The *trans*, or repulsion, configuration has the mutant alleles on opposite chromosomes. (B) The *cis*, or coupling, configuration has the mutant alleles on the same chromosome.

recombination frequency equal to 50 percent, either are in nonhomologous chromosomes or are located far apart in a single chromosome.

The possibility that two genes in the same chromosome may be unlinked creates a problem in terminology. We cannot say all genes in the same chromosome are "linked," because linkage means a recombination frequency of less than 50 percent, and genes that are sufficiently far apart are actually unlinked. The term that is used instead is *synteny:* Two genes are said to be **syntenic** if they are in the same chromosome, regardless of whether they show independent assortment or linkage.

Coupling Versus Repulsion of Syntenic Alleles

Geneticists use a "slash" notation for linked genes that has the general form $w^+ m / w m^+$. This notation is a simplification of the following more descriptive but cumbersome form:

$$\frac{w^+ \quad m}{w \quad m^+}$$

In this form of notation, the horizontal line replaces the slash in separating the two homologous chromosomes. For consistency, the linked genes are always written in the same linear order. This convention makes it possible to indicate the wildtype allele of a gene with a plus sign in the appropriate position. For example, the genotype $w m^+ / w^+ m$ could also be written without ambiguity as $w + / + m$. This system of gene notation is used for *Drosophila* and a number of other organisms.

A genotype that is heterozygous for each of two linked genes can have the alleles in either of two possible configurations, as shown in **Figure 5.2** for the *w* and *m* genes. In one configuration, called the ***trans,*** or **repulsion,** configuration, the mutant alleles are in opposite chromosomes, and the genotype is written as $w + / + m$ (Figure 5.2A). In the alternative configuration, called the ***cis,*** or **coupling,** configuration, the mutant alleles are present in the same chromosome, and the genotype is written as $w m / + +$.

Morgan's study of linkage between the *white* and *miniature* alleles began with the *trans* configuration. He also studied progeny from the *cis* configuration of the *w* and *m* alleles, which results from the cross of white, miniature females with red, nonminiature males. This cross is shown in the "+" notation as

$$\frac{w m}{w m} ♀♀ \times \frac{+ +}{\mathrm{Y}} ♂♂$$

In this case, the F_1 females were phenotypically wildtype double heterozygotes, and the males had white eyes and miniature wings. A cross of the F_1 progeny is written, again using the "+" notation, as

$$\frac{w m}{+ +} ♀♀ \times \frac{w m}{\mathrm{Y}} ♂♂$$

This cross produced the following progeny:

red eye, normal wing (++/*w m* ♀♀ and ++/Y ♂♂)	395	62.3 percent parental combinations
white eye, miniature wing (*w m*/*w m* ♀♀ and *w m*/Y ♂♂)	382	
white eye, normal wing (*w* +/*w m* ♀♀ and *w* +/Y ♂♂)	223	37.7 percent recombinant chromosomes
red eye, miniature wing (+ *m*/*w m* ♀♀ and + *m*/Y ♂♂)	247	
	1247	

This result demonstrates that the frequency of recombination with *w* and *m* in the *cis,* or coupling, configuration is the same as that observed with *w* and *m* in the *trans,* or repulsion, configuration: 37.7 percent versus 33.5 percent. The difference is within the range expected from sampling variation among experiments. However, in this case, the chromosomes constituting the

parental and recombinant classes of offspring are reversed. They are reversed because the original parents of the F_1 female were different: repulsion ($w +/+ m$) in one case and coupling ($w\ m/+\ +$) in the other. The repeated finding of equal recombination frequencies in experiments of this kind leads to the following conclusion:

> Recombination between linked genes takes place with the same frequency whether the alleles of the genes are in the *trans* (repulsion) configuration or in the *cis* (coupling) configuration; it is the same no matter how the alleles are arranged.

The Chi-square Test for Linkage

Any estimated value of r is subject to sampling variation from one experiment to the next. How can the statistical significance of an observed value of $r < 1/2$ be tested against the alternative hypothesis that $r = 0.5$? To take a specific example, how can we be sure that the value $r = 0.377$ observed in the test of $w\ m$ in the coupling configuration is not a statistical accident? Perhaps the true value is $r = 0.50$ and the genes, although syntenic, are nevertheless unlinked. Statistical significance for linkage can be tested on the basis of the observed numbers by means of the chi-square test examined in Chapter 4. As an example, we will use the $w\ m$ experiment yielding the estimate $r = 0.377$. The overall chi-square value in the testcross data is given in the top row of Table 5.1. With four classes of progeny there are $4 - 1 = 3$ degrees of freedom. But only one of these is relevant to the hypothesis of linkage. The others pertain to the segregation of w versus w^+ and of m versus m^+. This point is shown by the chi-square values in the bottom rows of the table. The test of 1 : 1 segregation of w versus w^+ entails pooling the m and m^+ data to obtain the observed ratio

$$w : w^+ = 382 + 223 : 395 + 247 = 605 : 642$$

To test for 1 : 1 of the m versus m^+ segregation, we pool the w and w^+ data

$$m : m^+ = 382 + 247 : 395 + 223 = 629 : 618$$

The hypothesis of no linkage ($r = 1/2$) predicts a 1 : 1 ratio of recombinant : nonrecombinant progeny, so in this case we pool the data as

$$\text{Recombinant} : \text{nonrecombinant} = 223 + 247 : 395 + 382 = 470 : 777$$

Table 5.1 Testing for linkage with chi-square

Chi-square test	Calculation		Degrees of freedom	*P* value
Overall chi square	$(395 - 311.75)^2/311.75 +$ $(382 - 311.75)^2/311.75 +$ $(223 - 311.75)^2/311.75 +$ $(247 - 311.75)^2/311.75 =$	76.775	3	$<< 0.001$
Segregation of w vs. w^+	$(605 - 623.5)^2/623.5 +$ $(642 - 623.5)^2/623.5 =$	1.098	1	0.295
Segregation of m vs. m^+	$(629 - 623.5)^2/623.5 +$ $(618 - 623.5)^2/623.5 =$	0.097	1	0.755
Linkage (recombinant vs. nonrecombinant)	$(470 - 623.5)^2/623.5 +$ $(777 - 623.5)^2/623.5 =$	75.580	1	$<< 0.001$
Sum of 1-d.f. chi-square values		76.775	3	

The three individual chi-square values, each with 1 degree of freedom (d.f.), are shown in Table 5.1. The overall chi-square with 3 d.f. has a P value that can be calculated by numerical methods as 1.5×10^{-16}; this leaves no doubt that the underlying ratio of the four classes of progeny differs from 1 : 1 : 1 : 1. To learn the source of the discrepancy, we examine the individual chi-square values that have 1 d. f. Neither the w versus w^+ nor the m versus m^+ segregation gives any evidence for departure from 1 : 1, the P values being 0.295 and 0.755, respectively. However, the chi-square value for recombinants versus nonrecombinants is highly significant, with a P value that can be calculated as 3.5×10^{-18}. From this test we can safely conclude that the cause of the discrepancy is linkage, and we have already estimated the frequency of recombination as $r = 470/1247 = 0.377$.

Note also in Table 5.1 that the 1-d.f. chi-square values sum to the overall chi-square of 76.775. The additivity means that the three hypotheses being tested are independent in the sense that the truth or falsehood of any of them has no bearing on the validity of the others. In practice, we are not often interested in such a detailed analysis of segregation as is outlined in Table 5.1. Usually we want to test only for linkage, and

this test is based on the chi-square value for recombinants versus nonrecombinants:

> The chi-square test for linkage is a test for an equal number (1 : 1 ratio) of recombinant versus nonrecombinant progeny, which has 1 degree of freedom.

Testing the individual 1 : 1 segregations is therefore unnecessary.

Each Pair of Linked Genes Has a Characteristic Frequency of Recombination

The recessive allele *y* of another X-linked gene in *Drosophila* results in yellow body color instead of the usual gray color determined by the y^+ allele. In one experiment, white-eyed females were mated with males having yellow bodies, and then the wildtype F_1 females were testcrossed with yellow-bodied, white-eyed males:

$$+\,w/+\,w\ ♀♀ \times y\,+/\mathrm{Y}\ ♂♂$$
$$\downarrow$$
$$+\,w/y\,+\ ♀♀ \times y\,w/\mathrm{Y}\ ♂♂$$

The progeny were

Wildtype body, white eye (maternal gamete, + *w*)	4292	98.6 percent parental combinations
Yellow body, red eye (maternal gamete, *y* +)	4605	
Wildtype body, red eye (maternal gamete, + +)	86	1.4 percent are recombinant chromosomes
Yellow body, white eye (maternal gamete, *y w*)	44	
	9027	

In a second experiment, yellow-bodied, white-eyed females were crossed with wildtype males, and then the F_1 wildtype females and F_1 yellow-bodied, white-eyed males were intercrossed:

$$y\ w/y\ w\ ♀♀ \times +\,+/\mathrm{Y}\ ♂♂$$
$$\downarrow$$
$$+\,+/y\ w\ ♀♀ \times y\ w/\mathrm{Y}\ ♂♂$$

In this case, 98.6 percent of the F_2 progeny had parental chromosomes and 1.3 percent had recombinant chromosomes. The parental and recombinant types were reversed in the reciprocal crosses, but the recombination frequency was virtually the same. Females with the *trans* genotype *y* +/+ *w* produced about 1.4 percent recombinant progeny, carrying either of the recombinant chromosomes *y w* or + +; similarly, females with the *cis* genotype *y w*/+ + produced about 1.4 percent recombinant progeny, carrying either of the recombinant chromosomes *y* + or + *w*. However, the recombination frequency between the genes for yellow body and white eyes was much lower than that between the genes for white eyes and miniature wings (1.4 percent versus about 35 percent). These and other experiments reinforce the following conclusions:

- The recombination frequency is a characteristic of a particular pair of genes.
- Recombination frequencies are the same in *cis* (coupling) and *trans* (repulsion) heterozygotes.

Recombination in Females versus Males

Drosophila is unusual in that recombination does not take place in males. Although it is not known how (or why) crossing-over is prevented in males, the result of the absence of recombination in *Drosophila* males is that all syntenic genes (those located in the same chromosome) show complete linkage in the male. For example, the genes *cn* (cinnabar eyes) and *bw* (brown eyes) are both in chromosome 2 but are so far apart that, in females, there is 50 percent recombination. Thus the cross

$$\frac{cn\ bw}{+\ +}\ ♀♀ \times \frac{cn\ bw}{cn\ bw}\ ♂♂$$

yields progeny of genotype + +/*cn bw* and *cn bw*/*cn bw* (the nonrecombinant types) as well as *cn* +/*cn bw* and + *bw*/*cn bw* (the recombinant types) in the proportions 1 : 1 : 1 : 1. However, because there is no crossing-over in males, the reciprocal cross

$$\frac{cn\ bw}{cn\ bw}\ ♀♀ \times \frac{cn\ bw}{+\ +}\ ♂♂$$

yields progeny only of the nonrecombinant genotypes ++/*cn bw* and *cn bw*/*cn bw* in equal proportions. The absence of recombination in *Drosophila* males is a convenience often made use of in experimental design. As shown in the case of *cn* and *bw*, all the alleles present in any chromosome in a male must be transmitted as a group, without being recombined with alleles present in the homologous chromosome.

The complete absence of crossing-over in *Drosophila* males is atypical, but it is not unusual for frequencies of recombination to differ between the sexes. For example, the human genome also exhibits more recombination in females than in males, but the difference is not nearly so extreme as that in *Drosophila*. On the average, across the entire human genome, the recombination frequency between genetic markers in males is approximately 60 percent of the value observed between the same genetic markers in females.

5.2 Genetic Mapping

The linkage of the genes in a chromosome can be represented in the form of a **genetic map,** which shows the linear order of the genes along the chromosome with the distances between adjacent genes proportional to the frequency of recombination between them. A genetic map is also called a **linkage map** or a **chromosome map.** The concept of genetic mapping was first developed by Morgan's student Alfred H. Sturtevant in 1913. The early geneticists understood that recombination between genes takes place by an exchange of segments between homologous chromosomes in the process now called crossing-over. Each crossover is manifested physically as a chiasma, or cross-shaped configuration, between homologous chromosomes; chiasmata are observed in prophase I of meiosis (Chapter 4). Each chiasma results from the breaking and rejoining of chromatids during synapsis, with the result that there is an exchange of corresponding segments between them. The theory of crossing-over is that each chiasma results in a new association of genetic markers. This process is illustrated in **Figure 5.3.** When there is no crossing-over (Figure 5.3A), the alleles present in each homologous chromosome remain in the same combination. When crossing-over does take place (Figure 5.3B), the outermost (*distal*) alleles in two of the chromatids are interchanged (recombined).

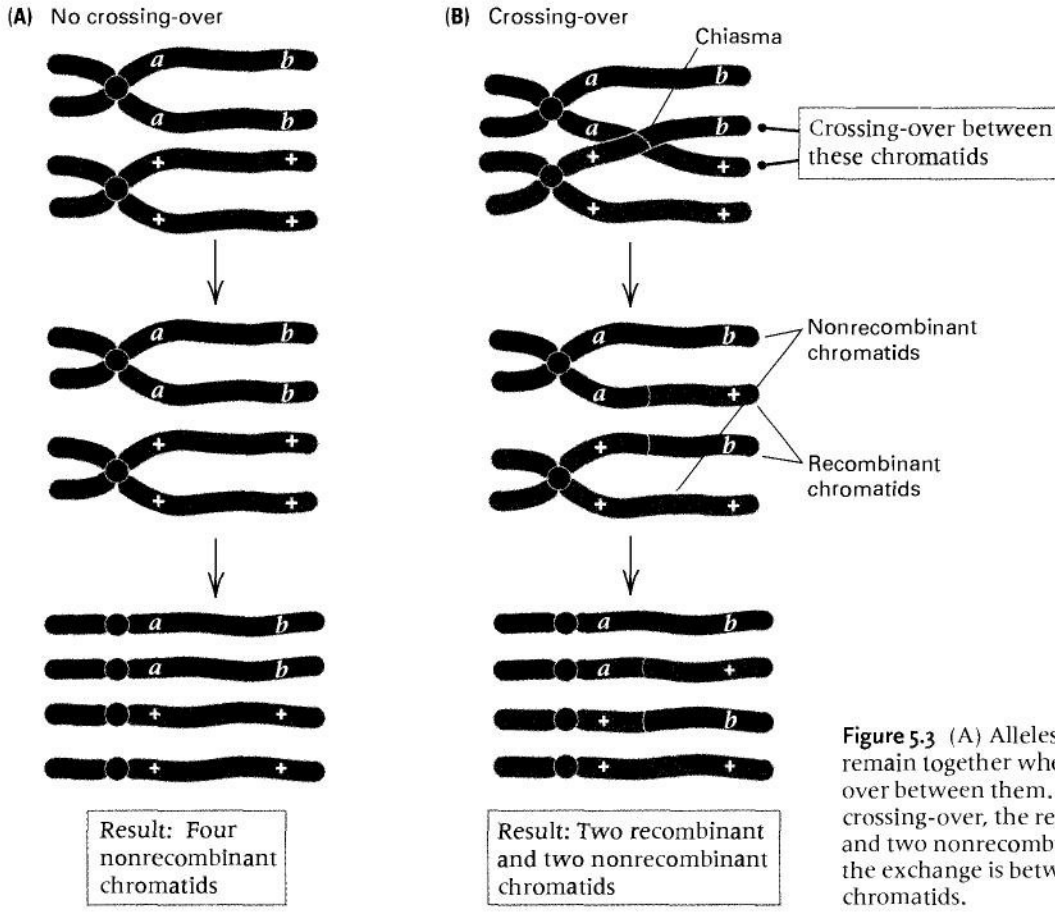


Figure 5.3 (A) Alleles in the same chromosome remain together when there is no crossing-over between them. (B) When there is crossing-over, the result is two recombinant and two nonrecombinant products, because the exchange is between only two of the four chromatids.

Map Distance and Frequency of Recombination

The unit of distance in a genetic map is called a **map unit.** Over short intervals, 1 map unit equals 1 percent recombination. For example, two genes that recombine with a frequency of 3.5 percent are said to be located 3.5 map units apart. One map unit is also called a **centimorgan** (abbreviated cM) in honor of T. H. Morgan. A distance of 3.5 map units therefore equals 3.5 centimorgans and indicates 3.5 percent recombination between the genes. For ease of reference, we list the four completely equivalent ways in which a genetic distance between two genes may be represented.

- As a *frequency of recombination* (in the foregoing example, 0.035)
- As a *percent recombination* (here 3.5 percent)
- As a *map distance* in map units (in this case, 3.5 map units)
- As a map distance in *centimorgans* (here 3.5 centimorgans, abbreviated 3.5 cM)

Physically, 1 map unit can be defined as the length of the chromosome in which, on the average, 1 crossover is formed in every 50 cells undergoing meiosis. This principle is illustrated in **Figure 5.4.** If 1 meiotic cell in 50 has a crossover, the frequency of *crossing-over* equals 1/50, or 2 percent. Yet the frequency of *recombination* between the genes is 1 percent. The correspondence of 1 percent recombination with 2 percent crossing-over is a little confusing until one realizes that each crossover results in 2 recombinant chromatids and 2 nonrecombinant chromatids (Figure 5.4). A frequency of crossing-over of 2 percent means that, among the 200 chromosomes that result from meiosis in 50 cells, exactly 2 chromosomes (those involved in the exchange) are recombinant for genetic markers flanking the crossover. To put the matter in another way, 2 percent crossing-over corresponds to 1 percent recombination because only half the chromatids in each cell with an exchange are actually recombinant.

In situations in which there are genetic markers along the chromosome, such as the *A, a* and *B, b* pairs of alleles in Figure 5.4, recombination between the marker genes takes place only when crossing-over occurs *between* the genes. **Figure 5.5** illustrates a case in which a crossover takes place between the gene *A* and the centromere, rather than between the genes *A* and *B*. The crossover does result in the physical exchange of segments between the innermost chromatids. But because the exchange is located outside the region between *A* and *B*,

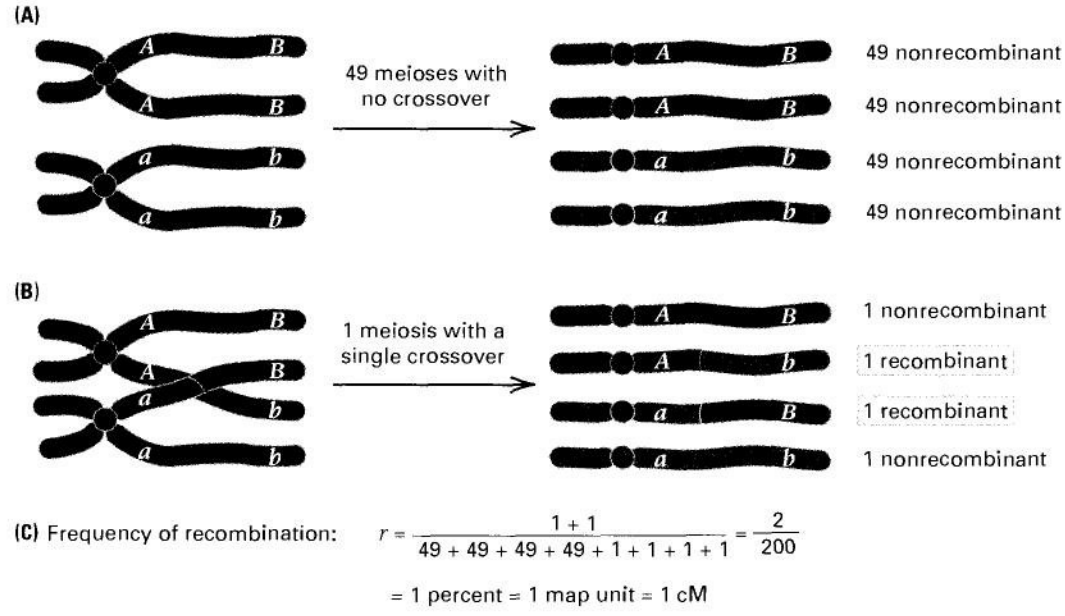


Figure 5.4 Chromosomal configurations in 50 meiotic cells, in which 1 cell has a crossover between two genes. (A) The 49 cells without a crossover result in 98 *A B* and 98 *a b* chromosomes; these are all nonrecombinant. (B) The cell with a crossover yields chromosomes that are *A B, A b, a B,* and *a b,* of which the middle two types are recombinant chromosomes. (C) The recombination frequency equals 2/200, or 1 percent (also called 1 map unit or 1 cM). Therefore, 1 percent recombination means that 1 meiotic cell in 50 has a crossover in the region between the genes.

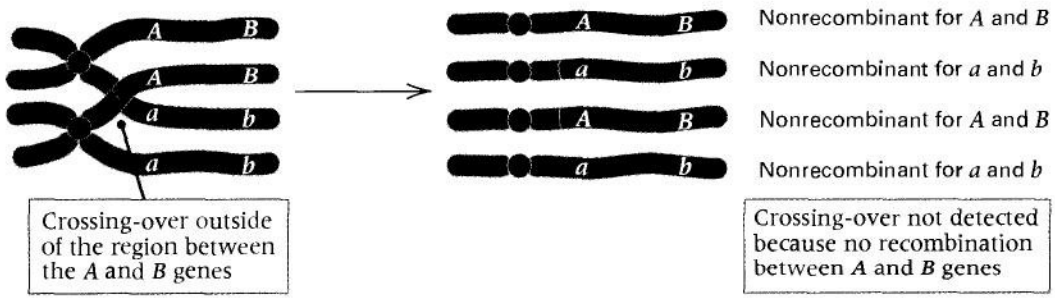


Figure 5.5 Crossing-over outside of the region between two genes is not detectable through recombination. Although a segment of chromosome is exchanged, the genetic markers stay in the nonrecombinant configurations, in this case *A B* and *a b*.

all of the resulting gametes must carry either the *A B* or the *a b* allele combination. These are nonrecombinant chromosomes. The presence of the crossover is undetected because it is not in the region between the genetic markers.

In some cases, the region between genetic markers is large enough that two (or even more) crossovers can be formed in a single meiotic cell. One possible configuration for two crossovers is shown in **Figure 5.6**. In this example, both crossovers are between the same pair of chromatids. The result is that there is a physical exchange of a segment of chromosome between the marker genes, but the double crossover remains undetected because the markers themselves are not recombined. The absence of recombination results from the fact that insofar as recombination between *A* and *B* is concerned, the second crossover reverses the effect of the first. The resulting chromosomes are either *A B* or *a b*, both of which are nonrecombinant.

Given that double crossing-over in a region between two genes can remain undetected because it does not result in recombinant chromosomes, there is an important distinction between the distance between two genes as measured by the *recombination frequency* and that distance as measured in *map units*. Map units measure how much crossing-over takes place between the genes. For any two genes, the map distance between them is defined as follows:

$$\text{Map distance (cM)} = (1/2) \times \text{average number of crossovers in the region per meiotic cell} \times 100 \quad (1)$$

The recombination frequency, on the other hand, reflects how much recombination is actually *observed* in a particular experiment. Double crossovers that do not yield recombinant gametes, such as the one in Figure 5.6, do contribute to the map distance but do not contribute to the recombination frequency. The distinction is important only when the region in question is large enough that double crossing-over can occur. If the region between the genes is so short that no more than one crossover can be formed in

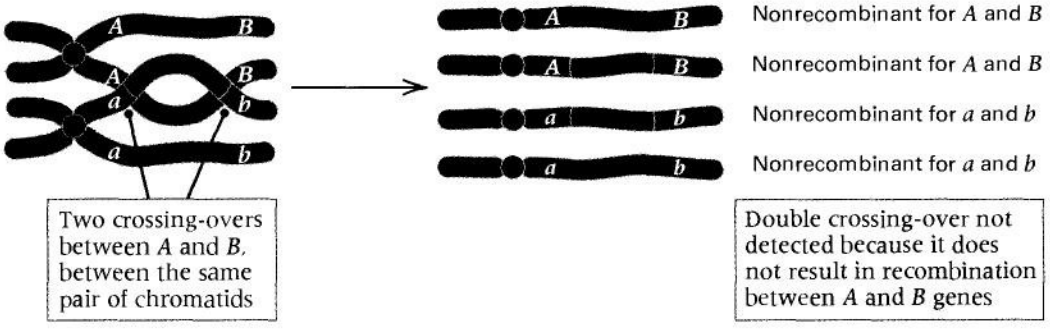


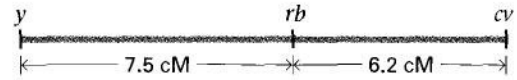
Figure 5.6 If two crossovers take place between marker genes, and both involve the same pair of chromatids, then neither crossover is detected because all of the resulting chromosomes are nonrecombinant *A B* or *a b*.

the region in any one meiosis, then map units and recombination frequencies are the same (because there are no multiple crossovers that can undo each other). This is the basis for saying that 1 map unit equals 1 percent recombination. Over an interval so short as to yield 1 percent observed recombination, multiple crossovers are usually precluded, so the map distance equals the recombination frequency in this case.

For a series of linked genes, if the chromosomal region between each adjacent pair is sufficiently short that multiple crossovers are not formed in the region, the recombination frequencies (and hence the map distances) between the genes are additive. This important feature of recombination, and also the logic used in genetic mapping, is illustrated in **Figure 5.7**. The genes are all in the X chromosome of *Drosophila*: *y* (yellow body), *rb* (ruby eye color), and *cv* (shortened wing crossvein). The recombination frequency between genes *y* and *rb* is 7.5 percent, and that between *rb* and *cv* is 6.2 percent. The genetic map might be any one of three possibilities, depending on whether *y*, *cv*, or *rb* is in the middle. Map A, which has *y* in the middle, can be excluded because it implies that the recombination frequency between *rb* and *cv* should be greater than that between *rb* and *y*, which contradicts the observed data.

Maps B and C are both consistent with the recombination frequencies. They differ in their predictions regarding the recombination frequency between *y* and *cv*. In map B the predicted distance is 1.3 map units, whereas in map C the predicted distance is 13.3 map units. In reality, the observed recombination frequency between *y* and *cv* is 13.3 percent. Map C is therefore correct.

There are actually two genetic maps corresponding to map C. They differ only in whether *y* is placed at the left or at the right. One map is



The other map is



These two ways of depicting the genetic map are completely equivalent.

By this type of reasoning, all the known genes in a chromosome can be assigned a position in a genetic map of the chromosome by considering successive subsets of three markers. Each set of syntenic genes constitutes a **linkage group**. The number of linkage groups is the same as the haploid number of chromosomes of the species. For example, cultivated corn *(Zea mays)* has ten pairs of chromosomes and ten linkage groups. A partial genetic map of chromosome 10 is shown in **Figure 5.8**, along with the phenotypes shown by some of the mutants. The ears of corn in Figure 5.8C and

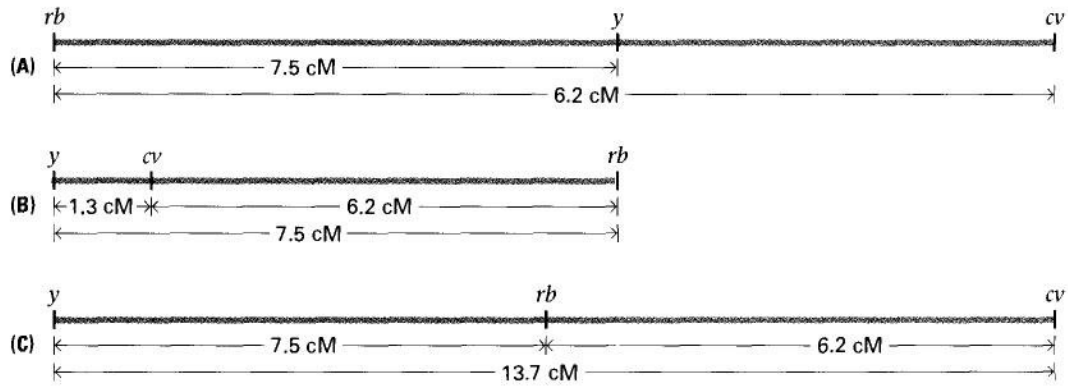


Figure 5.7 In *Drosophila*, the genes *y* (yellow body) and *rb* (ruby eyes) have a recombination frequency of 7.5 percent, and *rb* and *cv* (shortened wing crossvein) have a recombination frequency of 6.2 percent. There are three possible genetic maps, depending on whether *y* is in the middle (A), *cv* is in the middle (B), or *rb* is in the middle (C). Map A can be excluded because it implies that *rb* and *y* are closer than *rb* and *cv*, whereas the observed recombination frequency between *rb* and *y* is larger than that between *rb* and *cv*. Maps B and C are compatible with the data given.

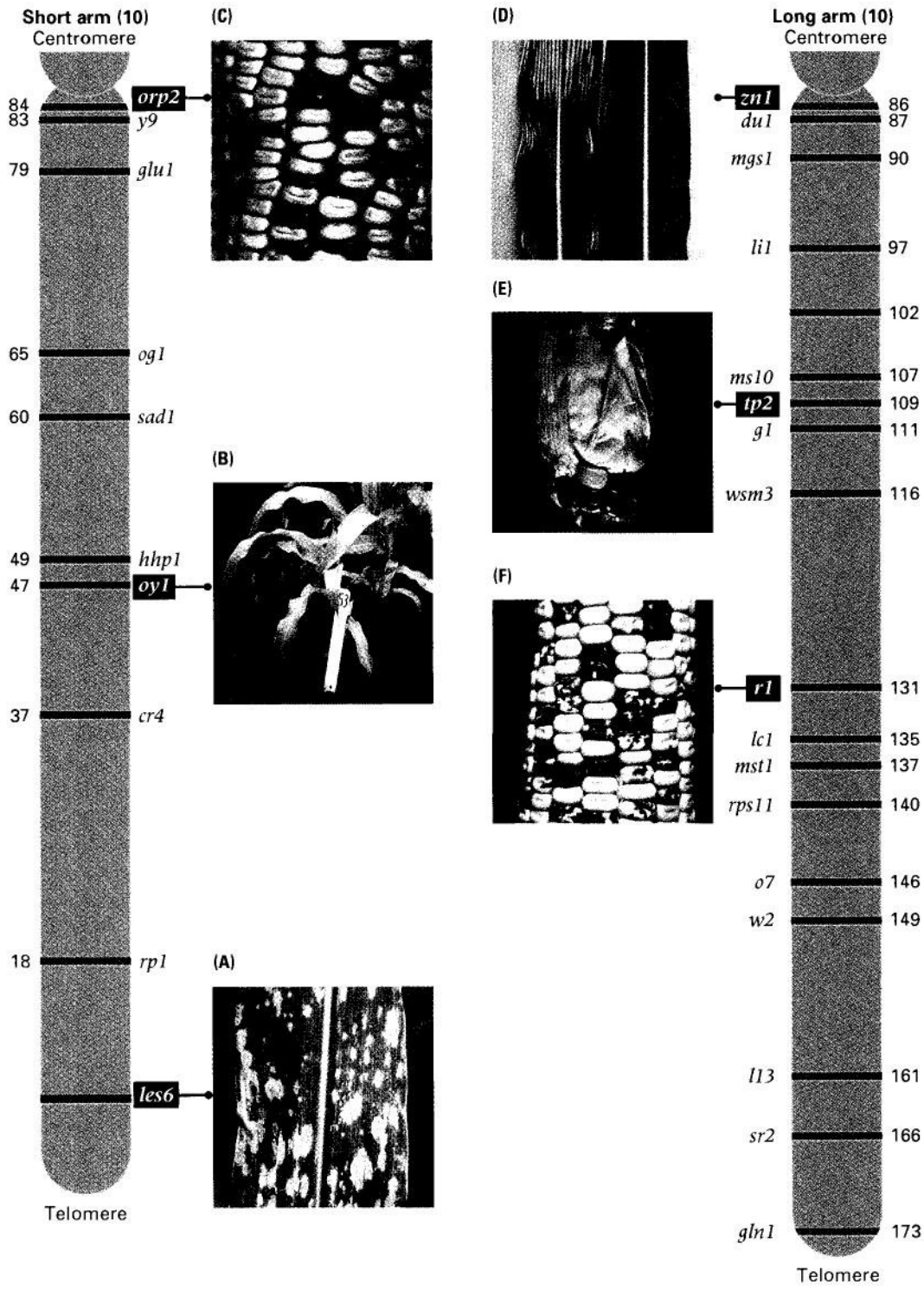
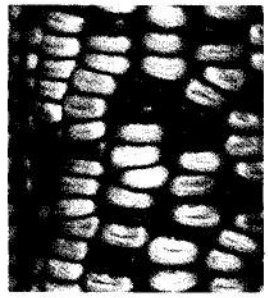


Figure 5.8 Genetic map of chromosome 10 of corn, *Zea mays*. The map distance to each gene is given in standard map units (centimorgans) relative to a position 0 for the telomere of the short arm (lower left). (A) Mutations in the gene *lesion-6 (les6)* result in many small to medium-sized, irregularly spaced, discolored spots on the leaf blade and sheath. The photograph shows the phenotype of a heterozygote for *Les6*, a dominant allele. (B) Mutations in the gene *oil yellow-1 (oy1)* result in a yellow-green plant. The plant in the foreground is heterozygous for the dominant allele *Oy1*; behind is a normal plant. (C) The *orp2* allele is a recessive expressed as orange pericarp, a maternal tissue that surrounds the kernels. The ear shows the segregation of *orp2* in a cross between two heterozygous genotypes, yielding a 3 : 1 ratio of yellow : orange seeds. (D) The gene *zn1* is *zebra necrotic-1*, in which dying tissue appears in longitudinal leaf bands. The leaf on the left is homozygous *zn1*, that on the the right is wildtype. (E) Mutations in the gene *teopod-2 (tp2)* result in many small, partially podded ears and a simple tassle. An ear from a plant heterozygous for the dominant allele *Tp2* is shown. (F) The mutation *R1-mb* is an allele of the *r1* gene—resulting in red or purple color in the aleurone layer of the seed. Note the marbled color in kernels of an ear segregating for *R1-mb*. [Photographs courtesy of M. G. Neuffer; genetic map courtesy of E. H. Coe.]

5.8F demonstrate the result of Mendelian segregation. Figure 5.8C shows a 3 : 1 segregation of yellow : orange kernels produced by the recessive *orange pericarp*-2



(*orp*-2) allele in a cross between two heterozygous genotypes, and Figure 5.8F shows a 1 : 1 segregation of marbled : white kernels produced by the dominant allele *R1-mb* in a cross between a heterozygous genotype and a homozygous normal.



Crossing-over

The orderly arrangement of genes represented by a genetic map is consistent with the conclusion that each gene occupies a well-defined site (*locus*) in the chromosome. In a heterozygote, the alleles occupy corresponding locations in the homologous chromosomes. Crossing-over is brought about by a physical exchange of segments resulting in a new association of genes in the same chromosome. This process has the following features:

1. The exchange of segments between parental chromatids takes place in the first meiotic prophase, *after the chromosomes have duplicated.* At this stage the four chromatids (strands) of a pair of homologous chromosomes are closely synapsed. Crossing-over is a physical exchange between chromatids in a pair of homologous chromosomes.
2. The exchange process consists of the breaking and rejoining of the two chromatids, resulting in the *reciprocal* exchange of equal and corresponding segments between them (see Figure 5.3).
3. The sites of crossing-over are more or less random along the length of a chromosome pair. Hence the probability of crossing-over between two genes increases as the physical distance between the genes along the chromosome becomes larger. This principle is the basis of genetic mapping.

A classic demonstration that recombination is associated with a physical exchange of segments between homologous chromosomes made use of two structurally altered X chromosomes of *Drosophila* that permitted parental and recombinant chromosomes to be recognized in a microscope. This experiment, performed in 1936 by Curt Stern, is outlined in **Figure 5.9**. One of the abnormal X chromosomes was missing a segment that had become attached to chromosome 4; this X chromosome could be identified by its missing terminal segment. The second aberrant X chromosome could be identified because it had a small second arm consisting of a piece of the Y chromosome. The mutant alleles *car* (a recessive allele resulting in carnation eye color instead of wildtype red) and *B* (a dominant allele resulting in bar-shaped eyes instead of round) were present in the first abnormal X chromosome, and the wildtype alleles of these genes were in the second abnormal X. Females with the two structurally and genetically marked X chromosomes were mated with males having a normal X that carried the recessive alleles of the genes (Figure 5.9). Among the progeny from this cross, flies with parental or recombinant chromosomes could be recognized by their eye color and shape. In the genetically recombinant progeny from the cross, the X chromosome had the morphology that would be expected if recombination of the genes were accompanied by an exchange that recombined the chromosome markers. In particular, the progeny with wildtype, bar-shaped eyes had an X chromosome with a missing terminal segment and the attached Y arm; and the progeny with carnation-colored, round eyes had a structurally normal X chromosome with no missing terminus and no Y arm. As

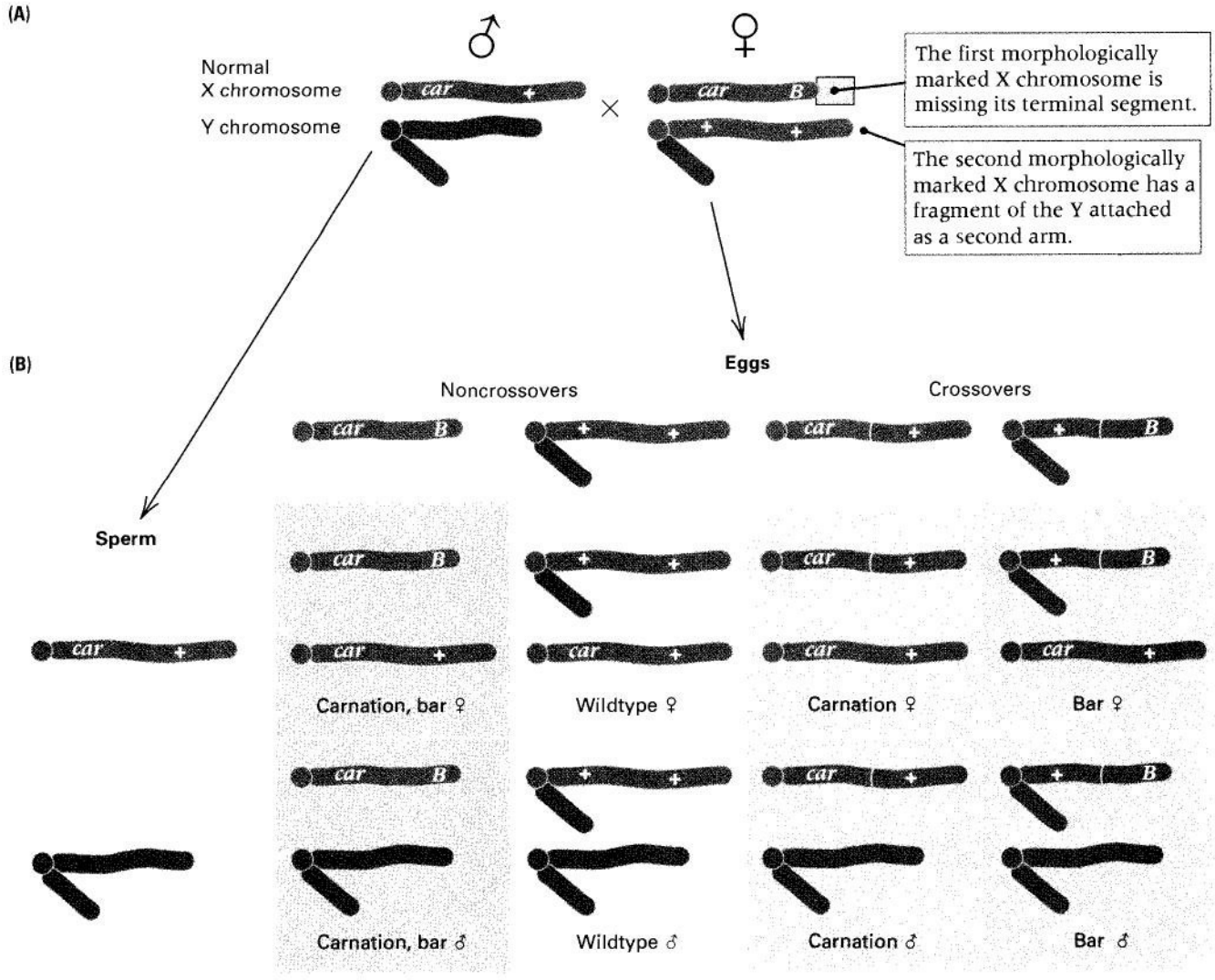


Figure 5.9 (A) Diagram of a cross in which the two X chromosomes in a *Drosophila* female are morphologically distinguishable from each other and from a normal X chromosome. One X chromosome has a missing terminal segment, and the other has a second arm consisting of a fragment of the Y chromosome. (B) Result of the cross. The carnation offspring contain a structurally normal X chromosome, and the bar offspring contain an X chromosome with both morphological markers. The result demonstrates that genetic recombination between marker genes is associated with physical exchange between homologous chromosomes. Segregation of the missing terminal segment of the X chromosome, which is attached to chromosome 4, is not shown.

expected, the nonrecombinant progeny were found to have an X chromosome morphologically identical with one of the X chromosomes in the mother.

Crossing-over Takes Place at the Four-Strand Stage of Meiosis

So far we have asserted, without citing experimental evidence, that crossing-over takes place in meiosis after the chromosomes have duplicated, at the stage when each bivalent has four chromatid strands. One experimental proof that crossing-over takes place after the chromosomes have duplicated came from a study of laboratory stocks of *D. melanogaster* in which the two X chromosomes in a female are joined to a common centromere to form an aberrant chromosome called an **attached-X,** or **compound-X,** chromosome. The normal X chromosome in *Drosophila* has a centromere almost at the end of the chromosome.

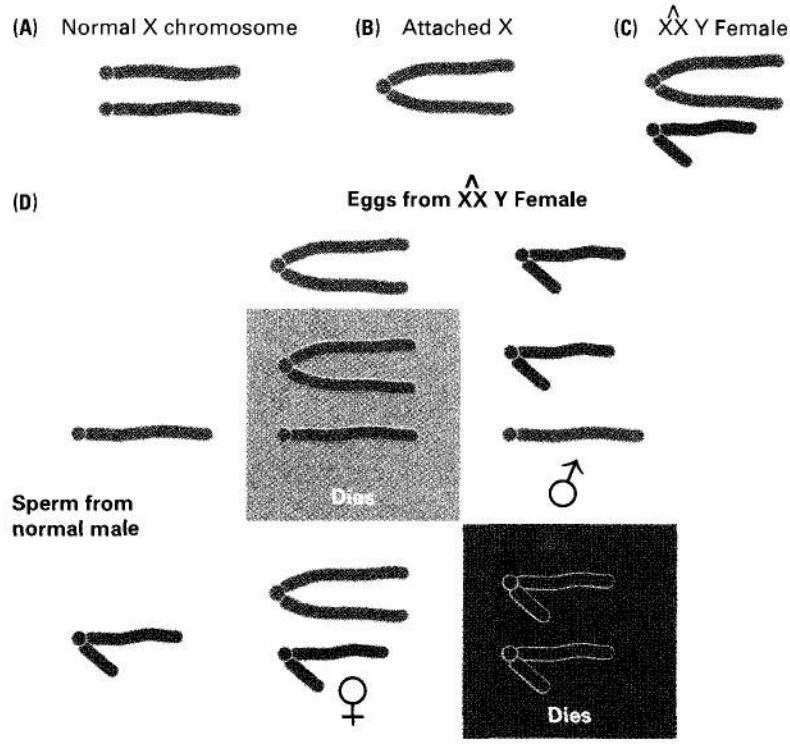


Figure 5.10 Attached-X (compound-X) chromosomes in *Drosophila*. (A) A structurally normal X chromosome in a female. (B) An attached-X chromosome, with the long arms of two normal X chromosomes attached to a common centromere. (C) Typical attached-X females also contain a Y chromosome. (D) Outcome of a cross between an attached-X female and a normal male. Genotypes with either three X chromosomes or no X chromosomes are lethal. Note that in this case, a male fly receives its X chromosome from its father and its Y chromosome from its mother—the opposite of the usual situation in *Drosophila*.

The attachment of two of these chromosomes to a single centromere results in a chromosome with two equal arms, each consisting of a virtually complete X. Females with a compound-X chromosome usually contain a Y chromosome as well, and they produce two classes of viable offspring: females who have the maternal compound-X chromosome along with a paternal Y chromosome, and males with the maternal Y chromosome along with a paternal X chromosome (**Figure 5.10**). Attached-X chromosomes are frequently used to study X-linked genes in *Drosophila*, because when an attached-X female is mated with a male carrying any X-linked mutation, the sons receive the mutant X chromosome and the daughters receive the attached-X chromosome. In matings with attached-X females, therefore, the inheritance of an X-linked gene in the male passes from father to son to grandson, and so forth, which is the opposite of usual X-linked inheritance.

In an attached-X chromosome in which one X carries a recessive allele and the other carries the wildtype nonmutant allele, crossing-over between the X-chromosome arms can yield attached-X products in which the recessive allele is present in both arms of the attached-X chromosome (**Figure 5.11**). Hence, attached-X females that are heterozygous can produce some female progeny that are homozygous for the recessive allele. The frequency with which homozygosity is observed increases with increasing map distance of the gene from the centromere. From the diagrams in Figure 5.11, it is clear that homozygosity can result only if the crossover between the gene and the centromere takes place after the chromosome has duplicated. The finding of homozygous attached-X female progeny therefore implies that crossing-over takes place at the *four-strand* stage of meiosis. If this were not the case, and crossing-over happened before duplication of the chromosome (at the *two-strand* stage), it would

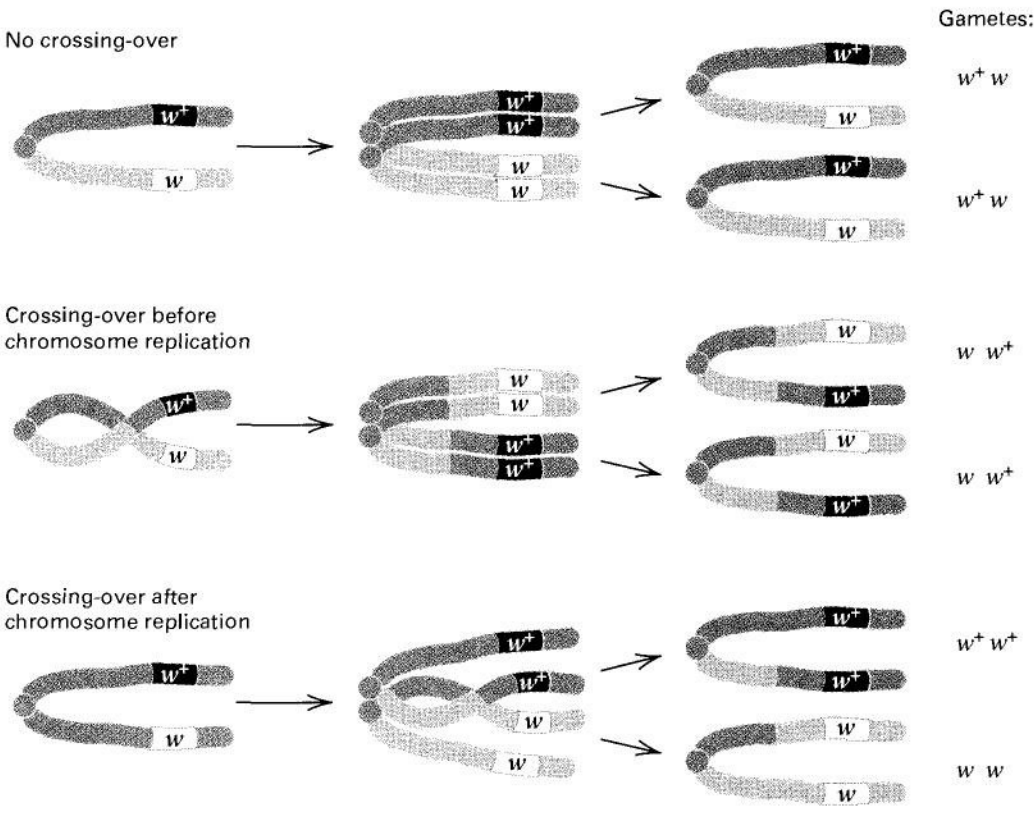


Figure 5.11 Crossing-over must take place at the four-strand stage in meiosis to produce a homozygous attached-X chromosome from one that is heterozygous for an allele. To yield homozygosity, the exchange must take place between the centromere and the gene.

result only in a swap of the alleles between the chromosome arms and would never yield the homozygous products that are actually observed.

Multiple Crossovers

When two genes are located far apart along a chromosome, more than one crossover can occur between them in a single meiosis, and the probability of multiple crossovers increases with the distance between the genes. Multiple crossing-over complicates the interpretation of recombination data. The problem is that some of the multiple exchanges between genes do not result in recombination and hence are not detected. As we saw in Figure 5.6, the effect of one crossover can be canceled by another crossover farther along the way. If two exchanges between the same two chromatids take place between the genes *A* and *B*, then their net effect will be that all chromosomes are nonrecombinant, either *A B* or *a b*. Two of the products of this meiosis have an interchange of their middle segments, but the chromosomes are not recombinant for the genetic markers, so they are genetically indistinguishable from noncrossover chromosomes. The possibility of such canceling means that the true map distance, defined in Equation (1), is *underestimated* by setting it equal to the observed frequency of recombination.

In higher organisms, double crossing-over is effectively precluded in chromosome segments that are sufficiently short. Therefore, by using recombination data for closely linked markers to build up genetic linkage maps, we can avoid multiple crossovers that cancel each other's effects. For genes that are distant along a chromosome, the true map distance is estimated by summing the recombination frequencies across shorter subintervals within which multiple crossovers do not occur.

CoNnection

Genes All in a Row

Alfred H. Sturtevant 1913
Columbia University,
New York, New York
The Linear Arrangement of Six Sex-Linked Factors in Drosophila, *As Shown by Their Mode of Association*

Genetic mapping remains the cornerstone of genetic analysis. It is the principal technique used in modern human genetics to identify the chromosomal location of mutant genes associated with inherited diseases, as we saw with Huntington disease in Chapter 2. The genetic markers used in human genetics are homologous DNA fragments that differ in length from one person to the next, but the basic principles of genetic mapping are the same as those originally enunciated by Sturtevant. In this excerpt, we have substituted the symbols presently in use for the genes, y *(yellow body),* w *(white eyes),* v *(vermilion eyes),* m *(miniature wings), and* r *(rudimentary wings). (The sixth gene mentioned is another mutant allele of white, now called white-eosin.) In this paper, Sturtevant uses the term* crossing-over *instead of* recombination *and* crossovers *instead of* recombinant chromosomes. *We have retained his original terms but, in a few cases, have put the modern equivalent in brackets.*

Morgan, by crossing white eyed, long winged flies to those with red eyes and rudimentary wings (the new sex-linked character), obtained, in F_2, white eyed rudimentary winged flies. This could happen only if "crossing-over" [recombination] is possible; which means, on the assumption that both of these factors are in the X chromosome, that an interchange of materials between homologous chromosomes occurs (in the female only, since the male has only one X chromosome). A point not noticed at this time came out later in connection with other sex-linked factors in *Drosophila*. It became evident that some of the sex-linked factors are associated, i. e., that crossing-over does not occur freely between some factors, as shown by the fact that the combinations present in the F_1 flies are much more frequent in the F_2 than are new combinations of the same characters. This means, on the chromosome view, that the chromosomes, or at least certain segments of them, are much more likely to remain intact during meiosis than they are to interchange materials. . . . It would seem, if this hypothesis be correct, that the proportion of "crossovers" [recombinant chromosomes] could be used as an index of the distance between any two factors. Then by determining the distances (in the above sense) between A and B and between B and C, one should be able to predict AC. . . . Just how far our theory stands the test is shown by the data below, giving observed percent of crossovers [recombinant chromosomes] and the distances calculated [from the summation of shorter intervals].

These results form a new argument in favor of the chromosome view of inheritance, since they strongly indicate that the factors investigated are arranged in a linear series.

It will be noticed at once that the longer distances, *y—r* and *w—r*, give smaller per cent of crossovers, than the calculation calls for. This is a point which was to be expected and is probably due to the occurrence of two breaks in the same chromosome, or "double crossing-over." But in the case of the shorter distances the correspondence with expectation is perhaps as close as was to be expected with the small numbers that are available. . . . It has been found possible to arrange six sex-linked factors in *Drosophila* in a linear series, using the number of crossovers per 100 cases [the frequency of recombination] as an index of the distance between any two factors. A source of error in predicting the strength of association between untried factors is found in double crossing-over. The occurrence of this phenomenon is demonstrated. . . . These results form a new argument in favor of the chromosome view of inheritance, since they strongly indicate that the factors investigated are arranged in a linear series.

Factors	Calculated distance	Observed percentage of crossovers
y—v	30.7	32.2
y—m	33.7	35.5
y—r	57.6	37.6
w—m	32.7	33.7
w—r	56.6	45.2

Source: Journal of Experimental Zoology 14: 43–59

The *minimum* recombination frequency between two genes is 0. The recombination frequency also has a maximum:

> No matter how far apart two genes may be, the maximum frequency of recombination between any two genes is 50 percent.

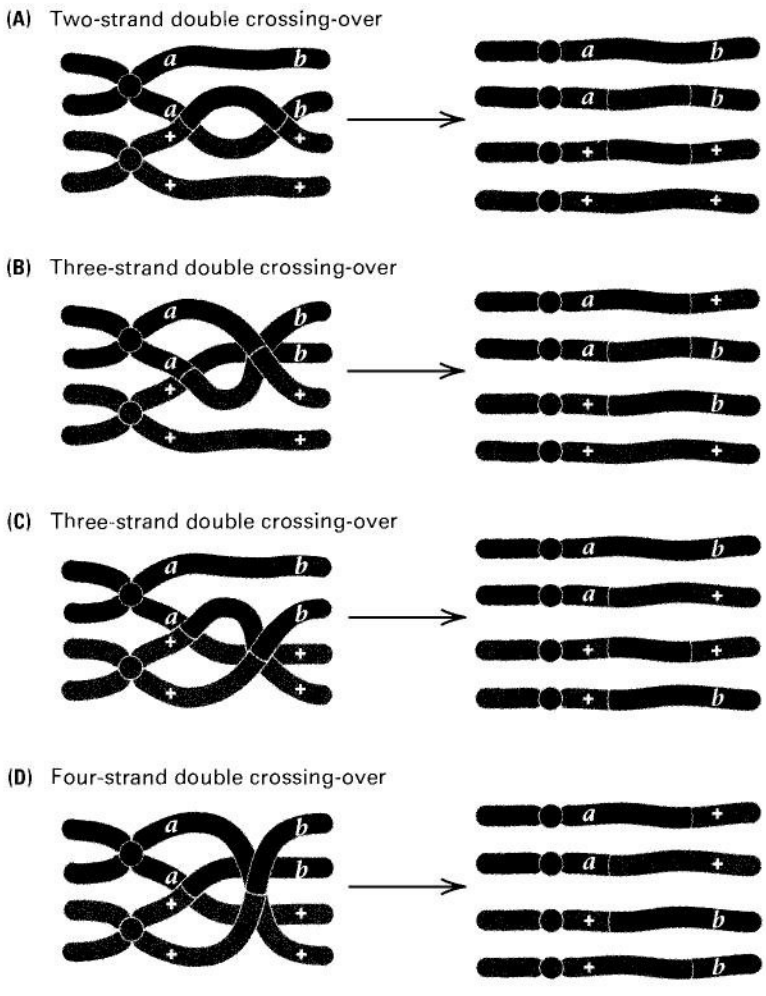
Fifty percent recombination is the same value that would be observed if the genes were on nonhomologous chromosomes and assorted independently. The maximum frequency of recombination is observed when the genes are so far apart in the chro-

mosome that at least one crossover is nearly always formed between them. Figure 5.3B shows that a single exchange in every meiosis would result in half of the products having parental combinations and the other half having recombinant combinations of the genes. Two exchanges between two genes have the same effect, as shown in **Figure 5.12**. Figure 5.12A shows a two-strand double crossover, in which the same chromatids participate in both exchanges; no recombination of the marker genes is detectable. When the two exchanges have one chromatid in common (three-strand double crossover, Figure 5.12B and C), the result is indistinguishable from that of a single exchange: Two products with parental combinations and two with recombinant combinations are produced. Note that there are two types of three-strand doubles, depending on which three chromatids participate. The final possibility is that the second exchange involves the chromatids that did not participate in the first exchange (four-strand double crossover, Figure 5.12D), in which case all four products are recombinant.

In most organisms, when double crossovers are formed, the chromatids that take part in the two exchange events are selected at random. In this case, the expected proportions of the three types of double exchanges are 1/4 two-strand doubles, 1/2 three-strand doubles, and 1/4 four-strand doubles. This means that, on the average,

$$(1/4)(0) + (1/2)(2) + (1/4)(4) = 2$$

recombinant chromatids will be found among the 4 chromatids produced from meioses with two exchanges between a pair of genes. This is the same proportion obtained with a single exchange between the



Chromosomes recovered	
Parental	Recombinant
4	0
2	2
2	2
0	4
8	8

Figure 5.12 The result of two crossovers in the interval between two genes is indistinguishable from independent assortment of the genes, provided that the chromatids participate at random in the exchanges. (A) A two-strand double crossover. (B, C) The two types of three-strand double crossovers. (D) A four-strand double crossover.

genes. Moreover, a maximum of 50 percent recombination is obtained for any number of exchanges.

In the discussion of Figure 5.12, we emphasized that in most organisms, the chromatids taking part in double-exchange events are selected at random. Then the maximum frequency of recombination is 50 percent. When there is a *nonrandom* choice of chromatids in successive crossovers, the phenomenon is called **chromatid interference.** It can be seen in Figure 5.12 that, relative to a random choice of chromatids, an excess of four-strand double crossing-over (*positive chromatid interference*) results in a maximum frequency of recombination greater than 50 percent; likewise, an excess of two-strand double crossing-over (*negative chromatid interference*) results in a maximum frequency of recombination less than 50 percent. Therefore, the finding that the maximum frequency of recombination between two genes in the same chromosome is not 50 percent can be regarded as evidence for chromatid interference. Positive chromatid interference has not yet been observed in any organism; negative chromatid interference has been reported in some fungi.

Double crossing-over is detectable in recombination experiments that employ **three-point crosses,** which include three pairs of alleles. If a third pair of alleles, c^+ and c, is located between the two with which we have been concerned (the outermost genetic markers), then double exchanges in the region can be detected when the crossovers flank the *c* gene (**Figure 5.13**). The two crossovers, which in this example take place between the same pair of chromatids, would result in a reciprocal exchange of the c^+ and c alleles between the chromatids. A three-point cross is an efficient way to obtain recombination data; it is also a simple method for determining the order of the three genes, as we shall see in the next section.

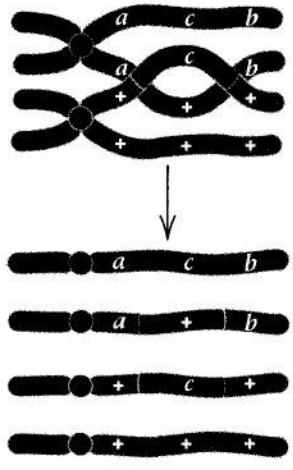


Figure 5.13 A two-strand double crossover that spans the middle pair of alleles in a triple heterozygote results in a reciprocal exchange of the middle pair of alleles between the two participating chromatids.

5.3 Genetic Mapping in a Three-Point Testcross

The data in **Table 5.2** result from a testcross in corn with three linked genes. We will use these to illustrate the analysis of a three-point cross. The recessive alleles are *lz* (for lazy or prostrate growth habit), *gl* (for glossy leaf), and *su* (for sugary endosperm), and the multiply heterozygous parent in the cross had the genotype

$$\frac{Lz\ Gl\ Su}{lz\ gl\ su}$$

Therefore, the two classes of progeny that inherit noncrossover (parental-type) gametes are the normal plants and plants with the lazy-glossy-sugary phenotype. These classes are far larger than any of the crossover classes. If the combination of dominant and recessive alleles in the chromosomes of the heterozygous parent were unknown, then we could deduce from their relative frequencies among the progeny that the noncrossover gametes were *Lz Gl Su* and *lz gl su*. This point is important enough to state as a general principle:

> In any genetic cross involving linked genes, no matter how complex, the two most frequent types of gametes with respect to any pair of genes are *nonrecombinant;* these provide the linkage phase (*cis* versus *trans*) of the alleles of the genes in the multiply heterozygous parent.

In mapping experiments, the order of genes along the chromosome is usually not known in advance. In Table 5.2, the order

in which the three genes are shown is entirely arbitrary. There is, however, an easy way to determine the correct order from three-point data. Simply identify the genotypes of the double-crossover gametes produced by the heterozygous parent and compare them with the nonrecombinant gametes. Because the probability of two simultaneous exchanges is considerably smaller than that of either single exchange, the double-crossover gametes will be the least frequent types. Table 5.2 shows that the classes composed of four plants with the sugary phenotype and two plants with the lazy-glossy phenotype (products of the *Lz Gl su* and *lz gl Su* gametes, respectively) are the least frequent and therefore constitute the double-crossover progeny.

The effect of double crossing-over, as Figure 5.13 shows, is to interchange the members of the *middle* pair of alleles between the chromosomes.

This means that if the parental chromosomes are

Lz Gl Su and *lz gl su*

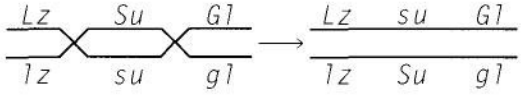
and the double-crossover chromosomes are

Lz Gl su and *lz gl Su*

then *Su* and *su* are interchanged by the double crossing-over and must be the middle pair of alleles. Therefore, the genotype of the heterozygous parent in the cross should be written as

$$\frac{Lz \quad Su \quad Gl}{lz \quad su \quad gl}$$

which is now diagrammed correctly with respect to both the order of the genes and the configuration of alleles in the homologous chromosomes. A two-strand double crossover between chromatids of these parental types is diagrammed below, and the products can be seen to correspond to the two types of gametes identified in the data as the double crossovers.



From this diagram, it can also be seen that the reciprocal products of a single crossover between *lz* and *su* would be *Lz su gl* and *lz Su Gl* and that the products of a single exchange between *su* and *gl* would be *Lz Su gl* and *lz su Gl*.

Table 5.2 Progeny from a three-point testcross in corn

Phenotype of testcross progeny	Genotype of gamete from hybrid parent	Number
Normal (wildtype)	*Lz Gl Su*	286
Lazy	*lz Gl Su*	33
Glossy	*Lz gl Su*	59
Sugary	*Lz Gl su*	4
Lazy, glossy	*lz gl Su*	2
Lazy, sugary	*lz Gl su*	44
Glossy, sugary	*Lz gl su*	40
Lazy, glossy, sugary	*lz gl su*	272

We can now summarize the data in a more informative way, writing the genes in correct order and identifying the numbers of the different chromosome types produced by the heterozygous parent that are present in the progeny.

Lz	*Su*	*Gl*	286	Parental types
lz	*su*	*gl*	272	
Lz	*su*	*gl*	40	Single crossovers between *lz* and *su*
lz	*Su*	*Gl*	33	
Lz	*Su*	*gl*	59	Single crossovers between *su* and *gl*
lz	*su*	*Gl*	44	
Lz	*su*	*Gl*	4	Double-crossover types
lz	*Su*	*gl*	2	
			740	

Note that each class of single recombinants includes two reciprocal products and that these are found in approximately equal frequencies (40 versus 33, and 59 versus 44). This observation illustrates an important principle:

The two reciprocal products that result from any crossover, or any combination of crossovers, are expected to appear in approximately equal frequencies among the progeny.

In calculating the frequency of recombination from the data, remember that the double-recombinant chromosomes result from *two* exchanges, one in each of the

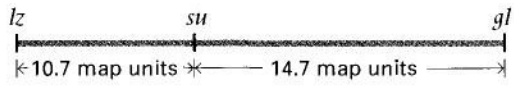
chromosome regions defined by the three genes. Therefore, chromosomes that are recombinant between *lz* and *su* are represented by the following chromosome types:

Lz	*su*	*gl*	40
lz	*Su*	*Gl*	33
Lz	*su*	*Gl*	4
lz	*Su*	*gl*	2
			79

Because 79/740, or 10.7 percent, of the chromosomes recovered in the progeny are recombinant between the *lz* and *su* genes, the map distance between these genes is estimated as 10.7 map units, or 10.7 centimorgans. Similarly, the chromosomes that are recombinant between *su* and *gl* are represented by

Lz	*Su*	*gl*	59
lz	*su*	*Gl*	44
Lz	*su*	*Gl*	4
lz	*Su*	*gl*	2
			109

The recombination frequency between the *su* and *gl* genes is 109/740, or 14.7 percent, so the map distance between them is estimated as 14.7 map units, or 14.7 centimorgans. The genetic map of the chromosome segment in which the three genes are located is therefore



One error that many students make in analyzing three-point crosses is to forget to include the double recombinants when calculating the recombination frequency between adjacent genes. You can keep from falling into this trap by remembering that the double-recombinant chromosomes have single recombination in *both* regions.

Chromosome Interference in Double Crossovers

The ability to detect double crossovers makes it possible to determine whether exchanges in two different regions are formed independently. Using the information from the example with corn, we know from the recombination frequencies that the probability of recombination is 0.107 between *lz* and *su* and 0.147 between *su* and *gl*. If crossing-over is independent in the two regions (which means that the formation of one exchange does not alter the probability of the second exchange), then the probability of an exchange in both regions is the product of these separate probabilities, or $0.107 \times 0.147 = 0.0157$ (1.57 percent). This implies that in a sample of 740 gametes, the expected number of double crossovers would be $740 \times 0.0157 = 11.6$, whereas the number actually observed is only 6. Such deficiencies in the observed number of double crossovers are common and identify a phenomenon called **chromosome interference,** in which crossing-over in one region of a chromosome reduces the probability of a second crossover in a nearby region. Because chromosome interference is nearly universal, whereas chromatid interference is virtually unknown, the term *interference*, when used without qualification, nearly always refers to chromosome interference.

The **coefficient of coincidence** is the observed number of double-recombinant chromosomes divided by the expected number. Its value provides a quantitative measure of the degree of interference, defined as

$$i = \text{interference} = 1 - \text{coefficient of coincidence}$$

From the data in our example, the coefficient of coincidence is $6/11.6 = 0.51$, which means that the observed number of double crossovers was only 51 percent of the number we would expect to observe if crossing-over in the two regions were independent. The value of the interference depends on the distance between the genetic markers and on the species. In some species, the interference increases as the distance between the two outside markers becomes smaller, until a point is reached at which double crossing-over is eliminated; that is, no double crossovers are found, and the coefficient of coincidence equals 0 (or, to say the same thing, the interference equals 1). In *Drosophila*, this distance is about 10 map units. In yeast, by contrast, interference is incomplete even over short

distances: For markers separated by 3 map units, the interference is in the range 0.3 to 0.6; and for those separated by 7 map units, it is in the range 0.1 to 0.3. In most organisms, when the total distance between the genetic markers is greater than about 30 map units, interference essentially disappears and the coefficient of coincidence approaches 1.

Genetic Mapping Functions

The effect of interference on the relationship between genetic map distance and the frequency of recombination is illustrated in Figure 5.14. Each curve is an example of a **mapping function,** which is the mathematical relation between the genetic distance across an interval in map units (centimorgans) and the observed frequency of recombination across the interval. In other words, a mapping function converts a *map distance* between genetic markers into a *recombination frequency* between the markers. As we have seen, when the map distance between the markers is small, the recombination frequency equals the map distance. This principle is reflected in the curves in Figure 5.14 in the region in which the map distance is smaller than about 10 cM. At less than this distance, all of the curves are nearly straight lines, which means that map distance and recombination frequency are equal: 1 map unit equals 1 percent recombination, and 10 map units equal 10 percent recombination.

For distances greater than 10 map units, the curves differ. The upper curve is based on the assumption of complete interference i, so that $i = 1$. With this mapping function, the linear relation holds all the way to a map distance of 50 cM. Then the curve terminates abruptly at a map distance of 50 map units. It terminates because complete interference means that a chromosome can have no more than one crossover along its length; because one crossover corresponds to 50 map units, this is the maximum length of the chromosome.

The other two curves in Figure 5.14 are for different patterns of interference along the chromosome. The bottom curve is usually called *Haldane's mapping function* after its inventor. It assumes no interference ($i = 0$), and the mathematical form of the function is $r = (1/2)(1 - e^{-d/50})$, where d is the map distance in centimorgans. Any mapping function for which i is between 0 and 1 must lie in the interval between the top and bottom curves. The particular example shown is *Kosambi's mapping function,* in which the interference is assumed to

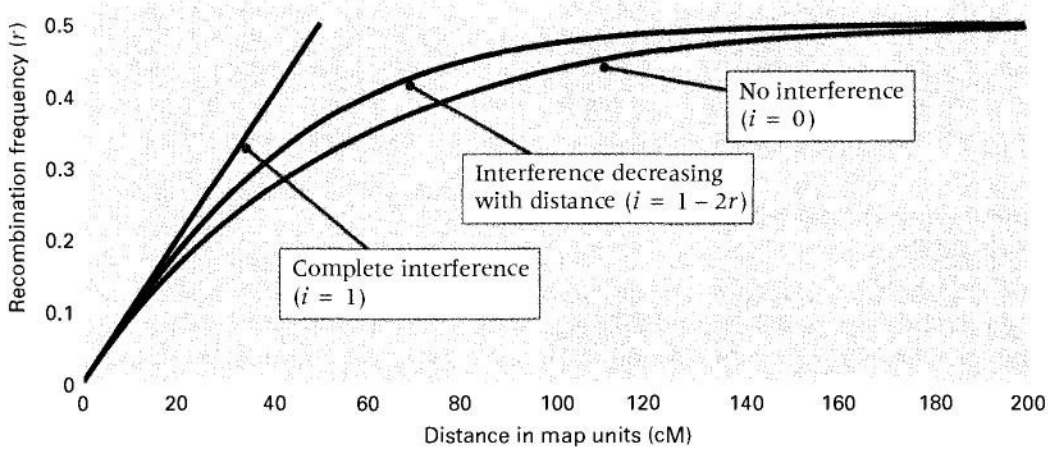


Figure 5.14 A mapping function is the relation between genetic map distance across an interval and the observed frequency of recombination across the interval. Map distance is defined as one-half the average number of crossovers converted into a percentage. The three mapping functions correspond to different assumptions about interference, i. In the first curve, $i = 1$ (complete interference); in this case the curve terminates abruptly because there can be at most one crossover per chromosome. In the third curve, $i = 0$ (no interference). The mapping function in the middle is based on the assumption that i decreases as a linear function of distance.



Fruit flies are grown in sterilized bottles, partly filled with a growth medium. The larvae feed on the nutrient and then pupate on the filter paper. A new generation appears every 10 days. [Dr. Jeremy Burgess/Science Photo Library/Photo Researchers, Inc.]

decrease as a linear function of distance according to $i = 1 - 2r$. Although simple in its underlying assumptions, the formula for Kosambi's function is not simple.

Haldane developed his mapping function in 1919, Kosambi his in 1943. Between these years and long afterward, geneticists had little interest in different mapping functions other than as curiosities, because there were few sets of data large enough to distinguish one reasonable function from the next. In recent years, with the explosion in the number of molecular markers available in virtually all organisms, and with the resurgence of interest in genetic mapping because of its role in identifying the position of mutations as precisely as possible prior to cloning (isolating the DNA), mapping functions have again become moderately fashionable. Checked against large data sets, none of the simple mapping functions in Figure 5.14 fits perfectly, but alternatives that fit better are much more complex even than Kosambi's mapping function.

Most mapping functions are almost linear near the origin, as are those in Figure 5.14. This near linearity implies that for map distances smaller than about 10–15 cM, whatever the pattern of chromosome interference, there are so few double recombinants that the recombination frequency in percent essentially equals the map distance. Hence the map distance between two widely separated genetic markers can be estimated with some confidence by summing the map distances across smaller segments between the markers, provided that each of the smaller segments is less than about 15 map units in length.

Genetic Map Distance and Physical Distance

Generally speaking, the greater the physical separation between genes along a chromosome, the greater the map distance between them. Physical distance and genetic map distance are usually correlated because a greater distance between genetic markers affords a greater chance for crossing-over to take place.

On the other hand, the general correlation between physical distance and genetic map distance is by no means absolute. We have already noted that the frequency of recombination between genes may differ in males and females. An unequal frequency of recombination means that the sexes can have different map distances in their genetic maps, although the physical chromosomes of the two sexes are the same and the genes must have the same linear order. An extreme example is that of *Drosophila* males, in which no crossing-over occurs. In *Drosophila* males, the map distance between any pair of genes located in the same chromosome is 0; nevertheless, genes in different chromosomes do undergo independent assortment.

The general correlation between physical distance and genetic map distance can break down even in a single chromosome. For example, crossing-over is much less frequent in some regions of the chromosome than in other regions. The term **heterochromatin** refers to certain regions of the chromosome that have a dense, compact structure in interphase; these regions take up many of the standard dyes used to make chromosomes visible. The rest of the chromatin, which becomes visible only after chromosome condensation in mitosis or meiosis, is called **euchromatin.** In most organisms, the major heterochromatic regions are adjacent to the centromere; smaller blocks are present at the ends of the chromosome arms (the telomeres) and interspersed with the euchromatin. In general, crossing-over is much less frequent in regions of heterochromatin than in regions of euchromatin.

Because there is less crossing-over in heterochromatin, a given length of heterochromatin will appear much shorter in the genetic map than an equal length of euchromatin. In heterochromatic regions, therefore, the genetic map gives a distorted picture of the physical map. An example of such distortion appears in **Figure 5.15**, which compares the physical map and the genetic map of chromosome 2 in *Drosophila*. The physical map is depicted as the chromosome appears in metaphase of mitosis. Two genes near the tips and two near the euchromatin–heterochromatin junction are indicated in the genetic map. The map distances across the euchromatic arms are 54.5 and 49.5 map units, respectively, for a total euchromatic map distance of 104.0 map

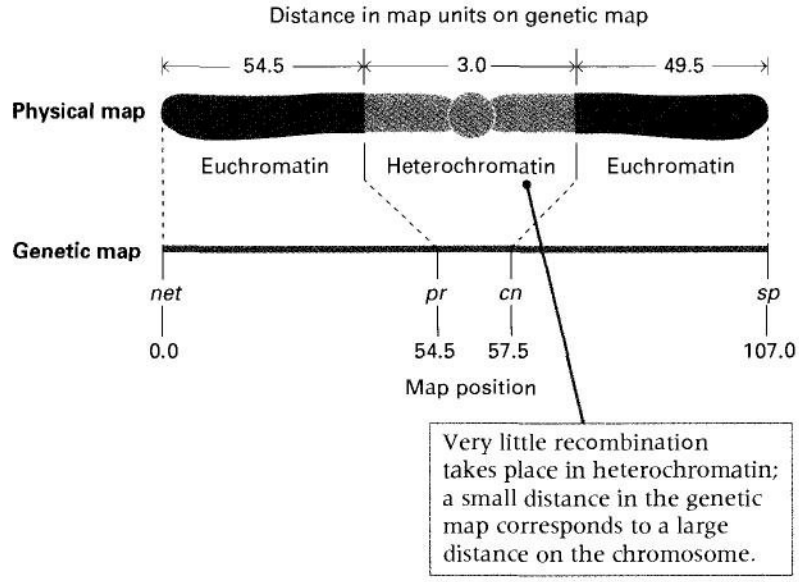


Figure 5.15 Chromosome 2 in *Drosophila* as it appears in metaphase of mitosis (physical map, top) and in the genetic map (bottom). Heterochromatin and euchromatin are in contrasting colors. The genes are *net* (net wing veins), *pr* (purple eye color), *cn* (cinnabar eye color), and *sp* (speck of wing pigment). The total map length is 54.5 + 49.5 + 3.0 = 107.0 map units. The heterochromatin accounts for 3.0/107.0 = 2.8 percent of the total map length but constitutes approximately 25 percent of the physical length of the metaphase chromosome.

units. However, the heterochromatin, which constitutes approximately 25 percent of the entire chromosome, has a genetic length in map units of only 3.0 percent. The distorted length of the heterochromatin in the genetic map results from the reduced frequency of crossing-over in the heterochromatin. In spite of the distortion of the genetic map across the heterochromatin, in the regions of euchromatin there is a good correlation between the physical distance between genes and their distance in map units in the genetic map.

5.4 Genetic Mapping in Human Pedigrees

In Chapter 2 we emphasized the utility of DNA polymorphisms as genetic markers in the human genome for mapping genetic risk factors ("disease genes"). Precise genetic mapping is often the first step in cloning a genetic risk factor. The present human genetic map is based on studies of a standard set of large pedigrees established by the Centre d'Etude du Polymorphisme Humain (CEPH) in Paris. The map includes more than 12,000 genetic markers, primarily STRPs, each heterozygous in an average of 65 percent of people tested. Because there is more recombination in females than in males, the female and male genetic maps differ in length. The female map is about 4400 cM, the male map about 2700 cM. Averaged over both sexes, the length of the human genetic map for all 23 pairs of chromosomes is about 3500 cM. Because the total DNA content per haploid set of chromosomes is 3154 million base pairs, there is, very roughly, 1 cM per million base pairs in the human genome.

The methods used for mapping in human genetics are somewhat different from those for experimental organisms, because

- Matings cannot be arranged for experimental convenience.
- Coupling (*A B*/*a b*) versus repulsion (*A b*/*a B*) heterozygotes cannot usually be distinguished.
- Progeny numbers are small.

In the next section we examine a method commonly used for linkage analysis in human genetics, which allows data from multiple pedigrees to be pooled.

CoNnection
Human Gene Map

Jeffrey C. Murry and 26 other investigators 1994
University of Iowa and 9 other research institutions
A Comprehensive Human Linkage Map with Centimorgan Density

Assessed by the distance between genetic markers, the human genetic map is one of the most dense genetic maps of any organism. At present, the human map includes more than 5000 genetic markers. The assembly of the human genetic map represents a major achievement in international science. It was pieced together by hundreds of investigators working in many countries throughout the world. This paper is one example of a progress report illustrating several important features. First, the human genetic map was made possible by researchers studying a single, very extensive set of families assembled by CEPH (Centre d'Étude du Polymorphisme Humain, Paris) so that their data could be collated. Second, a very large proportion of the markers in the human map are detected by means of the polymerase chain reaction (PCR) to amplify short tandem repeats (STRPs) that are abundant and highly polymorphic in the human genome. Third, the data are so extensive that they are not presented in the paper itself but rather in a large wall chart and (more conveniently) at a site on the Internet. Fourth, the authors point out the utility of the genetic map in determining the chromosomal locations of mutations that cause disease, using the methods of this chapter, as a first step in identifying the genes themselves. Finally, as is clear from the closing passages of the excerpt, the authors are well aware of the implications of modern human genetics for ethics, law, and social policy.

It should be emphasized that the new opportunities and challenges for biomedical research provided by these marker-dense maps also create an urgency to face the parallel challenges in the areas of ethics, law, and social policy.

For the first time, humans have been presented with the capability of understanding their own genetic makeup and how it contributes to morbidity of the individual and the species. Rapid scientific advances have made this possible, and developments in molecular biology, genetics, and computing, coupled with a cooperative and interactive biomedical community, have accelerated the progress of investigation into human inherited disorders. A primary engine driving these advances

Maximum Likelihood and Lod Scores

Figure 5.16 depicts four pedigrees segregating for a polymorphic DNA marker and a trait determined by a dominant allele D. For simplicity, we assume complete penetrance of the dominant allele. The symbols M and m represent alleles of the molecular marker, and for the sake of illustration, we assume that the linkage phase of the double-heterozygous parents is known to be $M\,D/m\,d$ in each pedigree. For linkage studies, each offspring is classified as to whether it is or is not recombinant. In this example, there is only 1 recombinant progeny, indicated by the arrow in pedigree B, and 9 nonrecombinant progeny. Using the same analysis as in Section 5.1 and other examples in this chapter, we would estimate the frequency of recombination as $r = 1/(1 + 9) = 0.10$. The problem is that the numbers are so small that a chi-square test for statistical significance is inappropriate.

How can we decide whether the pedigrees in Figure 5.16 yield significant evidence for linkage? The first step is to deduce the probability of the offspring from each mating, using the binomial distribution discussed in Chapter 4, in terms of the frequency of recombinant or nonrecombinant r. In this application, we classify each progeny only with respect to whether it is recombinant or nonrecombinant and ignore sex, affected status, and other traits. For example, the sibship in pedigree B consists of 1 recombinant and 3 nonrecombinants, and so the binomial probability of this sibship is $(4!/1!3!)r^1(1-r)^3 = 4r(1-r)^3$. This and the other probablities are shown immediately below each pedigree. The next step is to calculate, for each pedigree, the probability of its distribution of recombinant and nonrecombinant progeny under the assumption that $r = 0.5$ (independent assortment), which assumes that the genes are either in different chromosomes or widely separated in the same

has been the development and use of human gene maps that allow the rapid positional assignment of an inherited disorder as a starting point for gene identification and characterization. . . . The genetic maps were based on genotypes generated from DNA samples obtained from the CEPH reference pedigree set. Since this material is publicly available, individual research groups can add their own genetic markers in the future. . . . Although only markers genotyped on CEPH reference families are included in the maps, the list is extensive. The CEPH database contains blood group markers and protein polymorphisms as well as a variety of DNA-based markers including RFLPs and short tandem repeat polymorphisms (STRPs). There has been an emphasis on PCR-based markers. . . . The 3617 STRP markers alone provide about one marker every 1×10^6 base pairs (bp) throughout the genome. Once an initial localization [of a new gene] is identified by means of a set of genome-wide genetic maps, subsequent steps can be undertaken to increasingly narrow the region to be searched and eventually to select a specific interval on which studies of physical reagents, such as cloned DNA fragments, can be carried out. . . . Once the physical reagents are in hand, gene and mutation searches are possible with a number of strategies, . . . but considerable problems and difficulties still remain. . . . Public access to databases is an especially important feature of the Human Genome Project. They are easy to use and facilitate rapid communication of new findings and can be updated efficiently. We have made primary data available electronically on the World Wide Web at http://gdbwww.gdb.org/. . . . Finally, it should be emphasized that the new opportunities and challenges for biomedical research provided by these marker-dense maps also create an urgency to face the parallel challenges in the areas of ethics, law, and social policy. Our ability to distinguish individuals for forensic purposes, identify genetic predispositions for rare and common inherited disorders, and to characterize, if present, the underlying nature of the genetic components of normal trait variability such as height, intelligence, sexual preference, or personality type, has never been greater. Although technically feasible, whether these maps should be used for these ends should be resolved after open dialogue to review those implications and devise policy to deal with the as yet unpredictable outcomes.

Source: Science 265: 2049–2054

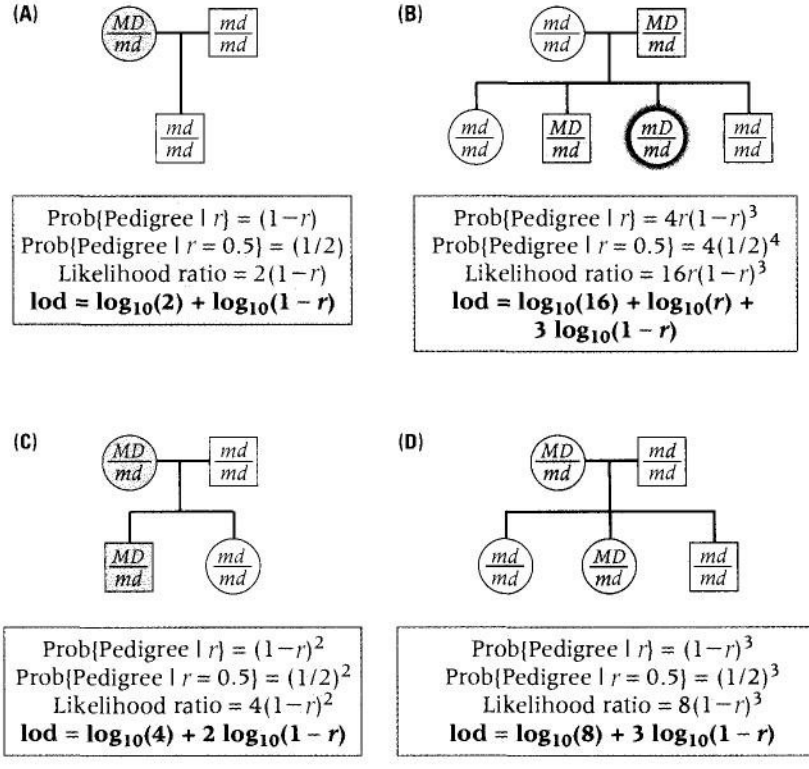


Figure 5.16 Pedigrees showing genetic linkage of a molecular marker M to a disease gene D. For each pedigree in turn, each progeny is classified as either recombinant or nonrecombinant, and the probability of the particular offspring distribution is calculated for recombination frequency r using the binomial distribution. This probability is divided by that for the special case $r = 1/2$ (no linkage) to obtain the likelihood ratio. (Note that the binomial coefficients cancel.) For any value of r, the likelihood ratio is the relative probability of the pedigree assuming linkage to that assuming no linkage. The logarithm (base 10) of the likelihood ratio constitutes the lod score for the pedigree.

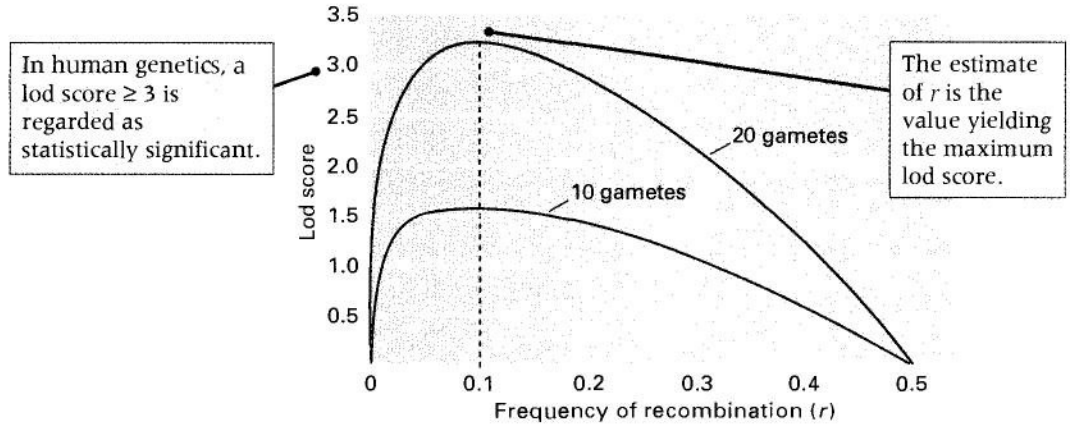


Figure 5.17 Overall lod score for the pedigrees in Figure 5.16 showing maximum lod at $r = 0.10$. The lod based on 10 gametes is not statistically significant; that based on 20 gametes is statistically significant (lod ≥ 3).

chromosome. These probabilities are given in the second line below each pedigree. For each pedigree, the ratio of the probability of the pedigree given an arbitrary value of r, symbolized Prob{Pedigree $\mid r$}, to that with $r = 0.5$, symbolized Prob{Pedigree $\mid r = 0.5$}, is called the **likelihood ratio;** this is given in the third line. (Note that, in calculating the likelihood ratio, the binomial coefficients in the numerator and denominator cancel.) From the likelihood ratios we obtain the **lod score** for each pedigree, which is the logarithm (base 10) of the likelihood ratio. These are given in the bottom line under each pedigree. (The term *lod* stands for *log-odds,* because the likelihood ratio is sometimes called the *odds.*)

Because the pedigrees are independent, the probabilities multiply across pedigrees, which means that the likelihood ratios also multiply. Hence the lod scores are additive across pedigrees, because $\log_{10}(xy) = \log_{10}(x) + \log_{10}(y)$. For the pedigrees in Figure 5.16, therefore, we can add the lod scores to obtain the overall lod:

$$\text{Lod} = \log_{10}(2 \times 16 \times 4 \times 8) + \log_{10}(r) + 9\log_{10}(1 - r)$$

The frequency of recombination is estimated as the value of r that maximizes the lod score for all the pedigrees. The curve of lod against r is plotted in **Figure 5.17**. In this example the maximum lod is at $r = 0.10$. This estimate makes intuitive sense because 1 progeny, among the total of 10, carries a recombinant chromosome. To assess the strength of the evidence for linkage, the convention in human genetics is to regard a lod score greater than 3 as being statistically significant. The rationale is that for two genes chosen at random, the odds against linkage in the human genome are about 50 : 1 (modern genetic maps put it closer to 66 : 1), and conventional statistical significance at the 5 percent level corresponds to odds of 20 : 1. Hence for pedigree data to override the *a priori* odds against linkage at a 5 percent level requires odds of $(50 \times 20) : (1 \times 1)$, or 1000 : 1; this corresponds to a lod score of 3 because $\log_{10}(1000) = 3$. On the other side, a lod score < -2 is generally regarded as significant evidence *against* linkage. Any value $-2 <$ lod < 3 is considered uninformative in regard to linkage, and additional data are required. In the example we are considering, the maximum lod at $r = 0.10$ is lod $= 1.60$, so the evidence for linkage is not regarded as statistically significant. There are too few total progeny.

Although 10 progeny are not sufficient to make a strong case for linkage in this example, 20 progeny would be sufficient. Suppose that the size of each sibship in Figure 5.16 were doubled and that 2 of the total of 20 progeny carried a recombinant chromosome. In this case the overall lod score would be

$$\text{Lod} = \log_{10}(2^{20}) + 2\log_{10}(r) + 18\log_{10}(1 - r)$$

This curve is plotted in Figure 5.17 also. The maximum is again at $r = 0.10$, but in this case the maximum lod equals 3.20. Because the maximum lod is greater than 3, the evidence for linkage is regarded as statistically significant.

The analysis in Figure 5.17 illustrates the underlying principles of linkage analy-

sis in human pedigrees. The CEPH human linkage database currently includes about 3.5 million lod scores for pairs of syntenic markers in the human genome. These allow the frequency of recombination between any pair of genetic markers to be estimated, along with a confidence interval around each estimate. In practice, the genetic mapping of disease genes presents many complications not dealt with here. For example, if the linkage phase is not known, the probability for each pedigree is calculated as a weighted average, assuming an equal likelihood of coupling and repulsion. There are also problems with incomplete penetrance. For example, the major genetic risk factor for human breast cancer *BRCA1* has a penetrance of 37 percent by age 40, 66 percent by age 55, and 85 percent by age 80, so the ages of women in breast cancer pedigrees must be taken into account. There are also different genetic as well as environmental causes of breast cancer, yielding risks for women who do not carry *BRCA1* of 0.4 percent by age 40, 3 percent by age 55, and 8 percent by age 80. Sorting out complexities such as these requires considerably more data than for simple Mendelian inheritance. For example, for a gene that doubles the risk of a complex disease, localizing a gene to within 1 cM ($\approx 10^6$ bp) through comparisons of marker genotypes in pairs of affected siblings requires a median of 700 sib pairs.

5.5 Mapping by Tetrad Analysis

Some species of fungi are especially favorable for genetic analysis. They are haploid, so recessive mutations are expressed, and they produce very large numbers of progeny. Most important, each meiotic tetrad is contained in a sac-like structure called an **ascus** and can be recovered as an intact group. Each product of meiosis is included in a separate reproductive cell called an **ascospore,** and all of the ascospores formed from one meiotic cell remain together in the ascus (**Figure 5.18**). The advantage of using these organisms to study recombination is the potential for analyzing all of the products from each meiotic division. Certain aberrant asci showing segregation ratios other than 1 : 1 provide critical information about the molecular

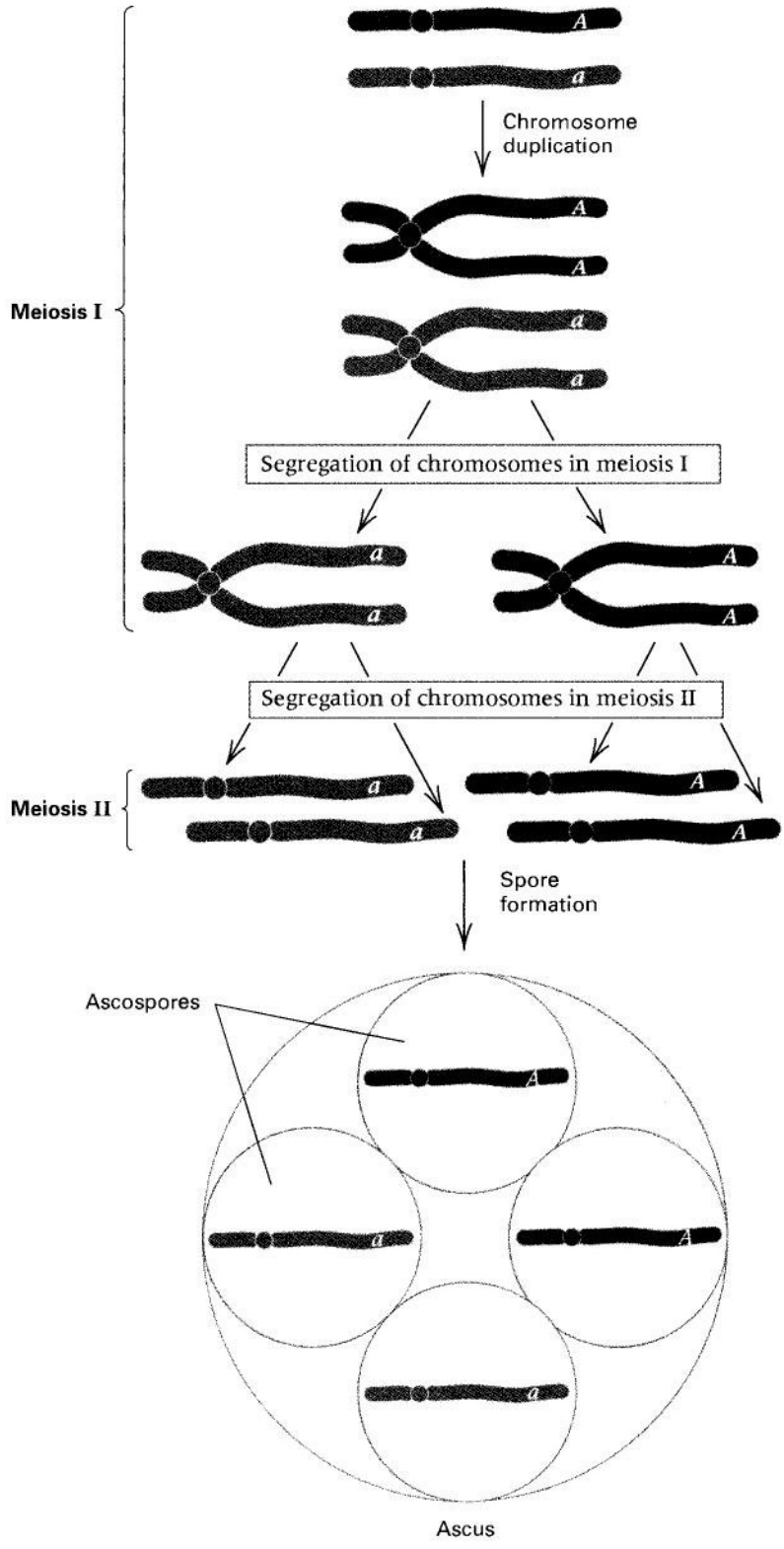


Figure 5.18 Formation of an ascus containing all of the four products of a single meiosis. Each product of meiosis forms a reproductive cell called an ascospore; these cells are held together in the ascus. Segregation of one chromosome pair is shown.

mechanism of recombination. This will be discussed in Chapter 6. In this section we show how the analysis of meiotic products in asci (*tetrad analysis*) can be used to study genetic linkage.

In these organisms, the life cycle is relatively short. In yeast, the diploid is induced to undergo meiosis by growth in medium that is deficient in its nutritional composition. The resulting haploid meiotic products, which form the ascospores, germinate to yield the vegetative stage of the organism (**Figure 5.19**). In some species, each of the four products of meiosis subsequently undergoes a mitotic division, with the result that each member of the tetrad yields a *pair* of genetically identical ascospores. In most of the organisms, the meiotic products, or their derivatives, are not arranged in any particular order in the ascus. However, bread molds of the genus *Neurospora* and related species have the useful characteristic that the meiotic products are arranged in a definite order directly related to the planes of the meiotic divisions. We will examine the ordered system after first looking at unordered tetrads.

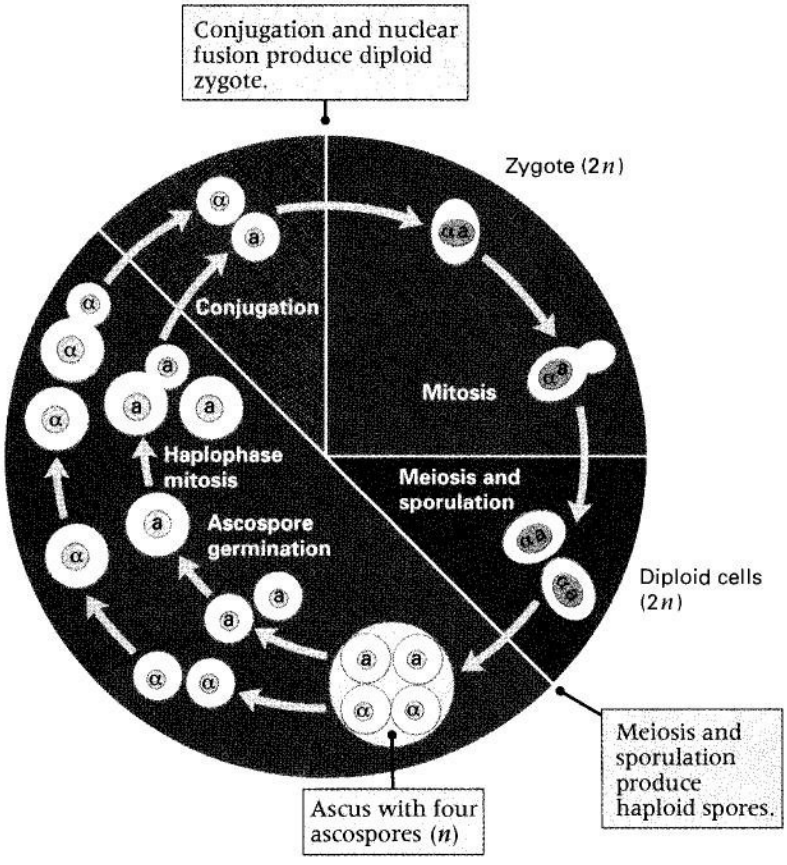


Figure 5.19 Life cycle of the yeast *Saccharomyces cerevisiae*. Mating type is determined by the alleles **a** and α. Both haploid and diploid cells normally multiply by mitosis (budding). Depletion of nutrients in the growth medium induces meiosis and sporulation of cells in the diploid state. Diploid nuclei are orange; haploid nuclei are yellow.

Analysis of Unordered Tetrads

In the tetrads, when two pairs of alleles are segregating, three patterns of segregation are possible. For example, in the cross

$$A\,B \times a\,b$$

the three types of tetrads are

AB AB ab ab	Referred to as a **parental-ditype (PD) tetrad.** Only two genotypes are represented, and the alleles have the same combinations found in the parents.
Ab Ab aB aB	Referred to as a **nonparental-ditype (NPD) tetrad.** Only two genotypes are represented, but the alleles have nonparental combinations.
AB Ab aB ab	Referred to as a **tetratype (TT) tetrad.** All four possible genotypes are present.

It is because of the following principle that tetrad analysis is an effective way to determine whether two genes are linked.

> When genes are *unlinked*, the parental-ditype tetrads and nonparental-ditype tetrads are expected in equal frequencies (PD = NPD).

The reason for the equality PD = NPD for unlinked genes is shown in **Figure 5.20** (part A) for two pairs of alleles located in different chromosomes. In the absence of crossing-over between either gene and its centromere, the two chromosomal configurations are equally likely at metaphase I, so PD = NPD. When there is crossing-over between either gene and its centromere (Figure 5.20B), a tetratype tetrad results, but this does not change the fact that PD = NPD.

In contrast, when genes are linked, parental ditypes are far more frequent than nonparental ditypes. To see why, assume that the genes are linked and consider the events required for the production of the

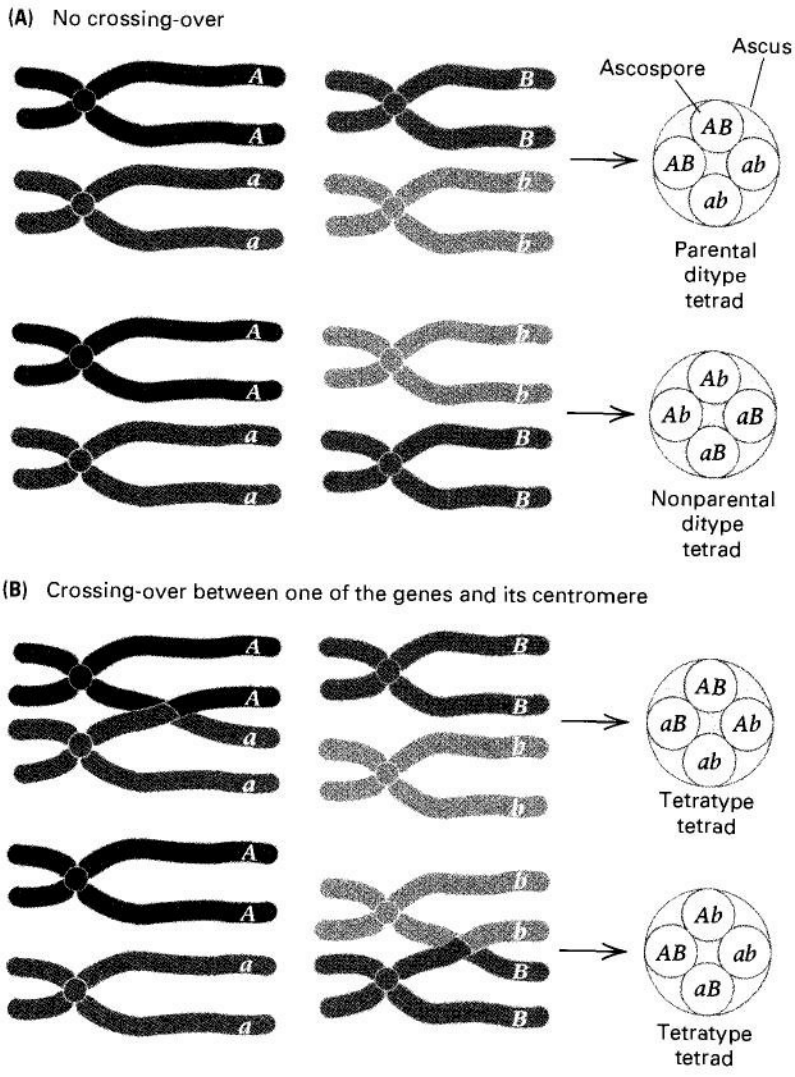


Figure 5.20 Types of unordered asci produced with two genes in different chromosomes. (A) In the absence of crossing-over, random arrangement of chromosome pairs at metaphase I results in two different combinations of chromatids, one yielding PD tetrads and the other NPD tetrads. (B) When crossing-over takes place between one gene and its centromere, the two chromosome arrangements yield TT tetrads. If both genes are closely linked to their centromeres (so that crossing-over between each gene and its centromere is rare), few TT tetrads are produced.

three types of tetrads. **Figure 5.21** shows that when no crossing-over takes place between the genes, a PD tetrad is formed. Single crossing-over between the genes results in a TT tetrad. The formation of a two-strand, three-strand, or four-strand double crossover results in a PD, TT, or NPD tetrad, respectively. With linked genes, meiotic cells with no crossovers will always outnumber those with four-strand double crossovers. Therefore,

> Linkage is indicated when nonparental-ditype tetrads appear with a much lower frequency than parental-ditype tetrads (NPD << PD).

The relative frequencies of the different types of tetrads can be used to determine the map distance between two linked genes, assuming that the genes are sufficiently close that triple crossovers and higher levels of crossing-over can be neglected. To determine the map distance, we first use the number of NPD tetrads to estimate the frequency of double crossovers. The four types of double crossovers in Figure 5.21 are expected in equal frequency. Thus the number of tetrads resulting from double crossovers (parts C, D, E, and F of Figure 5.21) should equal four times the number of NPD tetrads

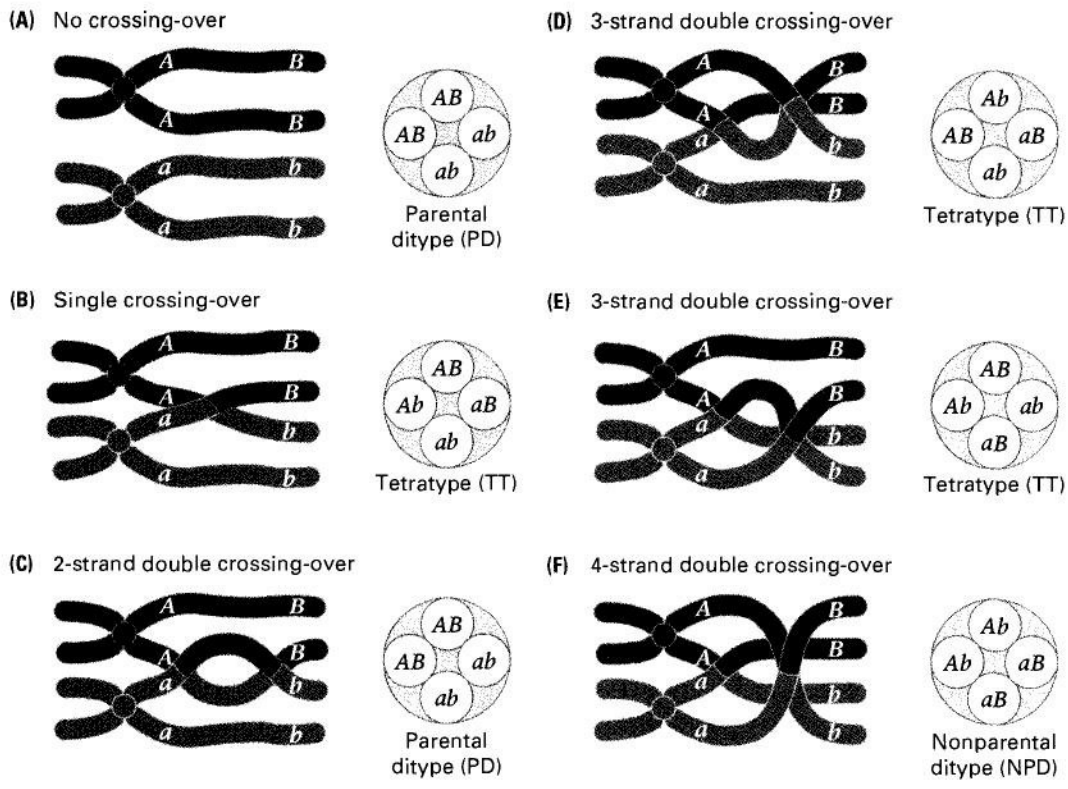


Figure 5.21 Types of tetrads produced with two linked genes. In the absence of crossing-over (A), a PD tetrad is produced. With a single crossover between the genes (B), a TT tetrad is produced. Among the four possible types of double crossovers between the genes (C, D, E, and F), only the four-strand double crossover in part F yields an NPD tetrad.

(part F), or 4 × [NPD], where the square brackets denote the observed number of NPD tetrads. At the same time, the number of TT tetrads resulting from three-strand double crossovers (parts D and E) should equal twice the number of NPD tetrads, or 2 × [NPD] (part F). Subtracting 2 × [NPD] from the total number of tetratypes yields the number of tetratype tetrads that originate from single crossovers, or [TT] − 2[NPD]. By definition (Equation 1), the map distance between two genes equals half the frequency of single-crossover tetrads ([TT] − 2[NPD]) plus the frequency of double-crossover tetrads (4[NPD]) times 100. Thus

$$\text{Map distance} = \frac{(1/2)([\text{TT}] - 2[\text{NPD}]) + 4[\text{NPD}]}{\text{Total number of tetrads}} \times 100$$

$$= \frac{(1/2)[\text{TT}] + 3[\text{NPD}]}{\text{Total number of tetrads}} \times 100 \qquad (2)$$

The 1/2 in these expressions corrects for the fact that only half of the meiotic products in a TT tetrad produced by a single crossover are recombinant. In the special case where the linked genes are close enough that [NPD] = 0 (no double crossovers), the map distance equals the percentage of TT tetrads divided by 2.

A systematic format for analyzing linkage relationships for allelic pairs in unordered tetrads is presented in the form of a logical flow diagram in **Figure 5.22.** All of the equations are based on the assumption that there is no chromatid interference. The right-hand side summarizes the equations for linked genes that we derived on the basis of Figure 5.21. The left-hand side pertains to unlinked genes, and it shows how the frequency of tetratypes can be used to make inferences about the distance of each gene from its own centromere.

As an example of the calculations for linked genes, consider a two-factor cross that yields 112 PD, 4 NPD, and 24 TT tetrads. The fact that NPD << PD (that is, 4 << 112) indicates that the two genes are linked and leads us down the right branch of the tree. Because both TT and NPD tetrads are recovered, we take the first left fork, and because there are NPD tetrads, we also take the next left fork. The appropriate equation is therefore Equation (2). Substitution of the values yields a map distance of

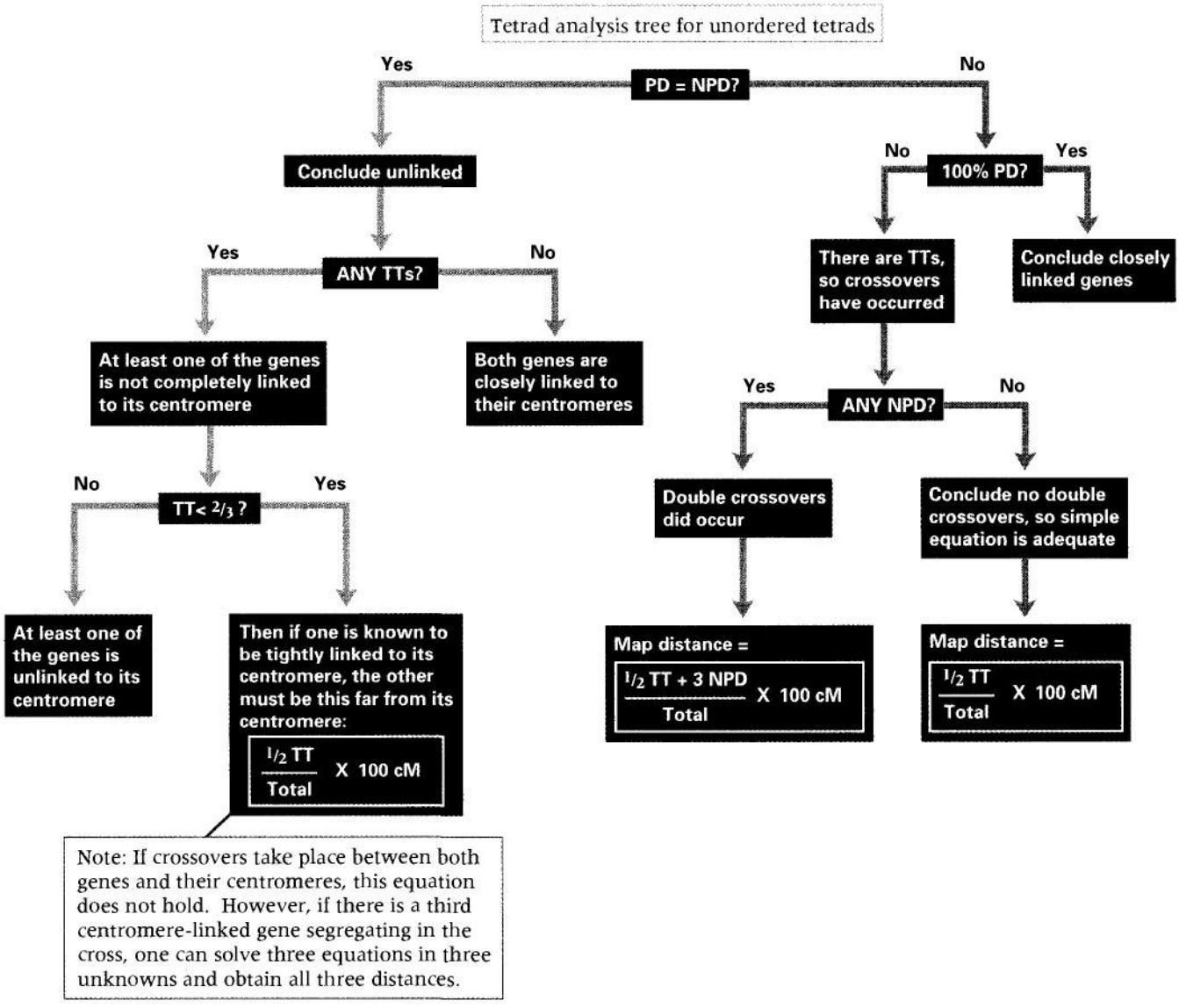


Figure 5.22 Tree diagram for analyzing linkage relations from unordered tetrads.

Map distance

$$= \frac{[(1/2) \times 24] + (3 \times 4)}{112 + 4 + 24} \times 100$$

$$= 17.1 \text{ cM}$$

Note that this mapping procedure differs from that presented earlier in the chapter in that recombination frequencies are not calculated directly from the number of recombinant and nonrecombinant chromatids but, rather, from the inferred types of crossovers.

As Figure 5.22 indicates, tetrad analysis yields a great deal of information about the linkage relationship between genetic markers and the distance of unlinked markers to their respective centromeres. However, it is not necessary to carry out a full tetrad analysis for estimating linkage. The alternative is to examine spores chosen at random after allowing the tetrads to break open and disseminate their spores. This procedure is called **random-spore analysis,** and the linkage relationships are determined exactly as described earlier for *Drosophila* and *Zea mays*. In particular, the frequency of recombination equals the number of spores that are recombinant for the genetic markers divided by the total number of spores.

Analysis of Ordered Tetrads

In *Neurospora crassa*, a species used extensively in genetic investigations, the products of meiosis are contained in an *ordered* array of ascospores (**Figure 5.23**). A zygote nucleus contained in a sac-like ascus undergoes meiosis almost immediately after it is formed. The four nuclei produced by meiosis are in a linear, ordered sequence in the ascus, and each of them undergoes a mitotic division to form two genetically identical and adjacent ascospores. Each mature ascus contains eight ascospores arranged in four pairs, each pair derived from one of the products of meiosis. The ascospores can be removed one by one from an ascus and each germinated in a culture tube to determine its genotypes.

Ordered asci also can be classified as PD, NPD, or TT with respect to two pairs of alleles; hence the tree diagram in Figure 5.22 can be used to analyze the linkage data. In addition, the ordered arrangement of mei-

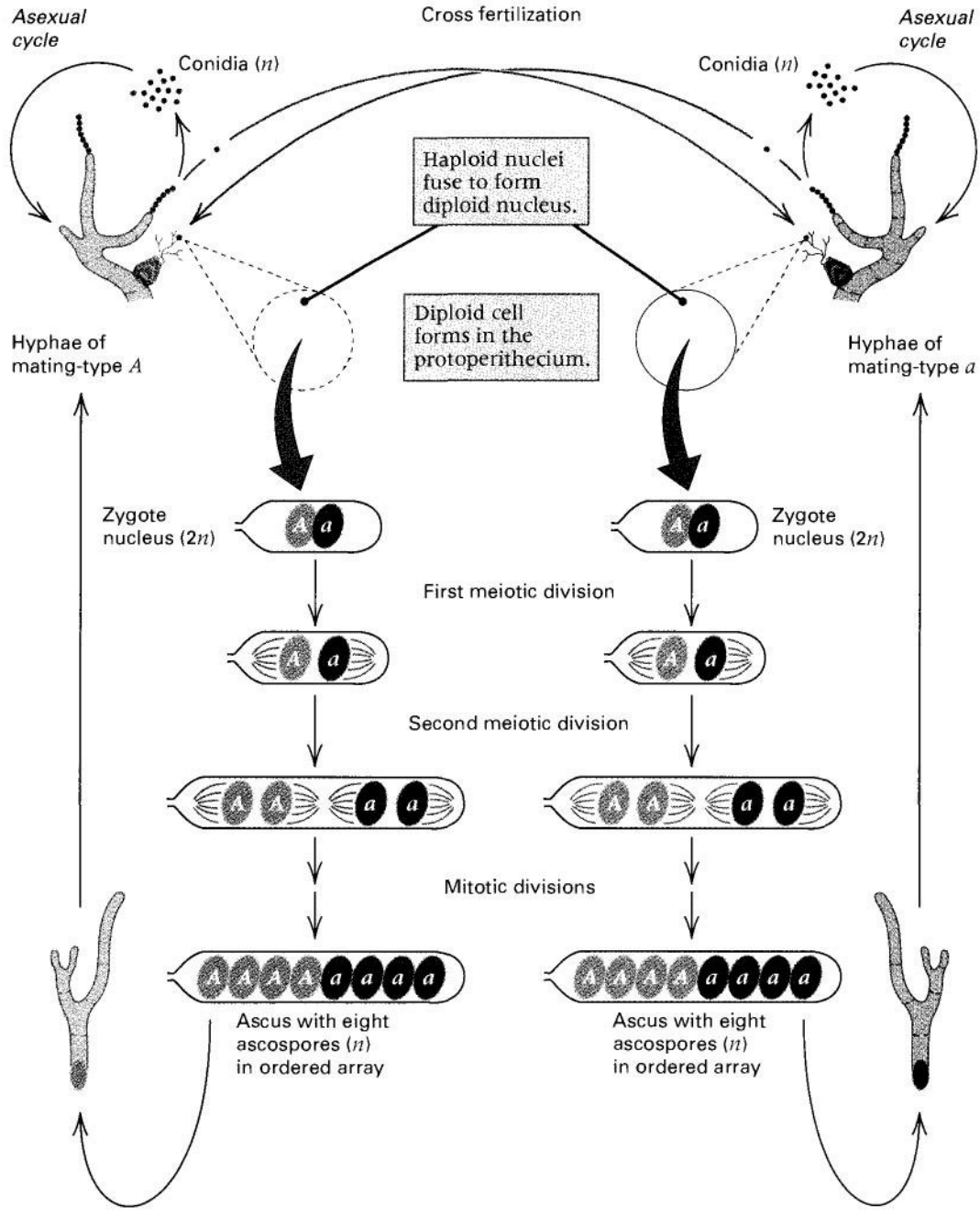


Figure 5.23
The life cycle of *Neurospora crassa*. The vegetative body consists of partly segmented filaments called hyphae. Conidia are asexual spores that function in the fertilization of organisms of the opposite mating type. A protoperithecium develops into a structure in which numerous cells undergo meiosis.

otic products makes it possible to determine the recombination frequency between any particular gene and its centromere. The logic of the mapping technique is based on the feature of meiosis shown in **Figure 5.24**:

Homologous centromeres of parental chromosomes separate at the first meiotic division; the centromeres of sister chromatids separate at the second meiotic division.

Thus, in the absence of crossing-over between a gene and its centromere, the alleles of the gene (for example, *A* and *a*) must separate in the first meiotic division; this separation is called **first-division**

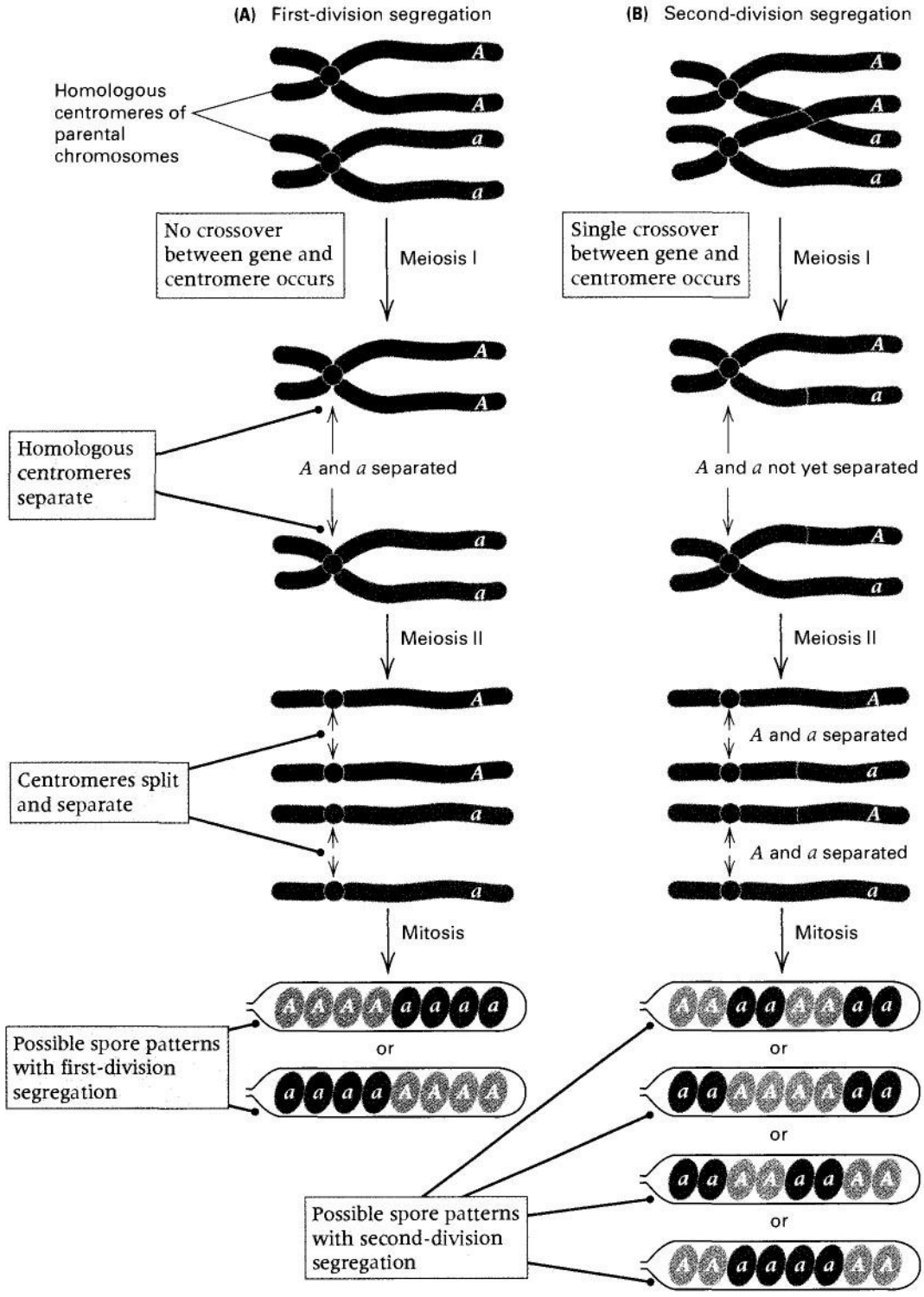


Figure 5.24 First- and second-division segregation in *Neurospora*. (A) First-division segregation patterns are found in the ascus when crossing-over between the gene and centromere does not take place. The alleles separate (segregate) in meiosis I. Two spore patterns are possible, depending on the orientation of the pair of chromosomes on the first-division spindle. (B) Second-division segregation patterns are found in the ascus when crossing-over between the gene and the centromere delays separation of *A* from *a* until meiosis II. Four patterns of spores are possible, depending on the orientation of the pair of chromosomes on the first-division spindle and that of the chromatids of each chromosome on the second-division spindle.

segregation. If, instead, a crossover is formed between the gene and its centromere, the *A* and *a* alleles do not become separated until the second meiotic division; this separation is called **second-division segregation.** The distinction between first-division and second-division segregation is shown in Figure 5.24. As shown in part A, first-division segregation is indicated by either of two possible arrangements of the spores: *A A a a* and *a a A A*. However, as shown in part B, second-division segregation is indicated by any of four patterns: *A a A a*, *a A a A*, *A a a A*, and *a A A a*.

The percentage of asci with second-division segregation can be used to map a gene with respect to its centromere. For example, let us assume that 30 percent of a sample of asci from a cross exhibit second-division segregation for the *A* and *a* alleles. This means that 30 percent of the cells undergoing meiosis had a crossover between the *A* gene and its centromere. Furthermore, in each cell in which crossing-over takes place, two of the chromatids are recombinant and two are nonrecombinant. In other words, a frequency of crossing-over of 30 percent corresponds to a recombination frequency of 15 percent. By convention, map distance refers to the frequency of recombinant meiotic products rather than to the frequency of cells with crossovers. Therefore, the map distance between a gene and its centromere is given by

$$\frac{(1/2)(\text{Asci with second-division segregation patterns})}{\text{Total number of asci}} \times 100 \qquad (3)$$

This equation is valid as long as the gene is close enough to the centromere for multiple crossovers to be neglected. Reliable linkage values are best determined for genes that are near the centromere. The location of more distant genes is accomplished by mapping them relative to other genes nearer the centromere.

If a gene is far from its centromere, crossing-over between the gene and its centromere is so frequent that the *A* and *a* alleles become randomized with respect to the four chromatids. The result is that the six possible spore arrangements shown in Figure 5.24 are all equally likely. This equal likelihood reflects the patterns that can result from choosing randomly among 2 *A* and 2 *a* spore pairs in the ascus, as shown by the branching diagram in **Figure 5.25.**

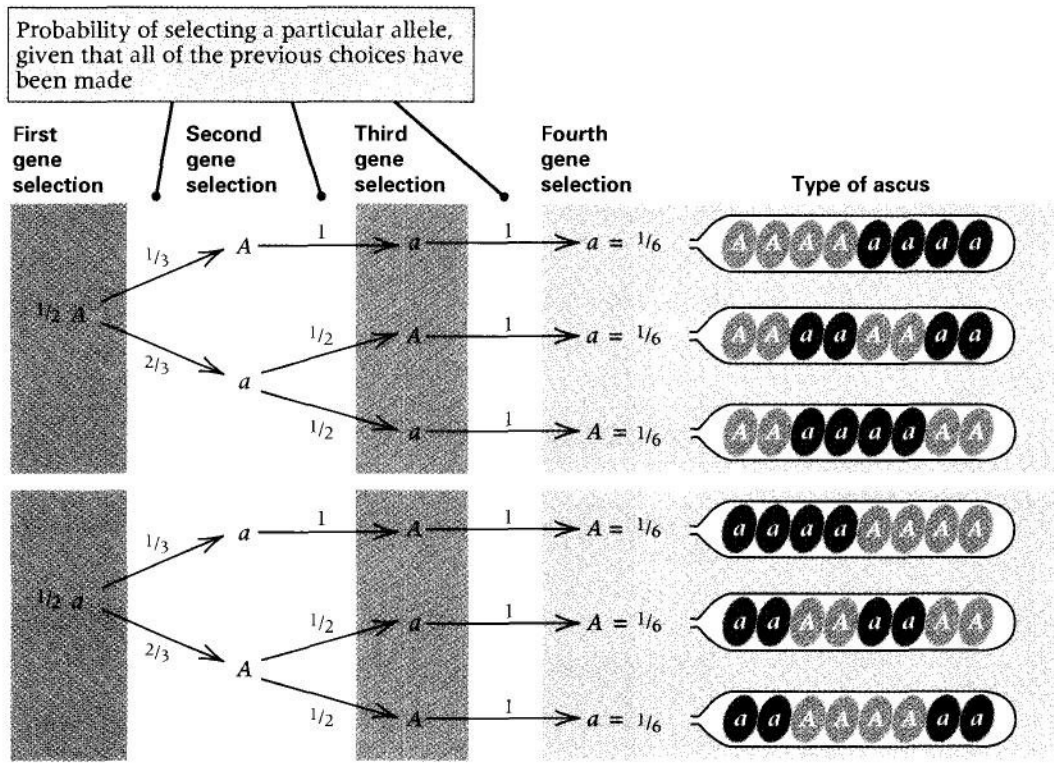


Figure 5.25 Diagram showing the result of free recombination between an allelic pair, *A* and *a*, and the centromere. The frequency of second-division segregation equals 2/3. This is the maximum frequency of second-division segregation, provided there is no chromatid interference.

Therefore, in the absence of chromatid interference,

> The maximum frequency of second-division segregation asci is 2/3.

5.6 Recombination within Genes

Genetic analyses and microscopic observations made before 1940 led to the "beads-on-a-string" model of the chromosome, wherein each gene was a "bead" joined in some manner to the next in the "string." The gene was regarded as the smallest unit of inheritance and the smallest unit that could undergo mutation. Recombination was assumed to occur only between genes, not within them. This concept began to change when geneticists realized that mutant alleles of a gene might be the consequence of alterations at different mutable sites within the gene. This now seems obvious, because we know that a gene is a sequence of nucleotides, but at the time, the best evidence that a gene was composed of subunits was the occurrence of recombination within the gene. The initial observation came in 1940 from investigations by C. P. Oliver of a gene in the X chromosome of *D. melanogaster* known as *lozenge* (*lz*). Mutant alleles of *lozenge* are recessive, and their phenotypic effects include a disturbed arrangement of the facets of the compound eye and a reduction in the eye pigments. Numerous *lz* alleles with distinguishable phenotypes are known.

Females heterozygous for two different *lz* alleles, one in each homologous chromosome, have lozenge eyes. When such heterozygous females were crossed with males that had one of the *lz* mutant alleles in the X chromosome, and large numbers of progeny were examined, flies with normal eyes were occasionally found. For example, one cross carried out was

$$X\ \frac{+\quad lz^{BS}+\quad +}{ct\quad +lz^{g}\quad v}\ X \times \frac{+\quad +lz^{g}\quad v}{\qquad\qquad\qquad}\ X\ Y$$

in which lz^{BS} and lz^{g} are mutant alleles of *lozenge*, and *ct* (cut wing) and *v* (vermilion eye color) are genetic markers that map 7.7 units to the left and 5.3 units to the right, respectively, of the *lozenge* locus. From this cross, 134 males and females with wildtype eyes were found among more than 16,000 progeny, for a frequency of 8×10^{-3}. These exceptional progeny might have resulted from a reverse mutation of one or of the other of the mutant *lozenge* alleles to lz^{+}, but the observed frequency, though small, was much greater than the known frequencies of reverse mutations. The genetic constitution of the maternally derived X chromosomes in the normal-eyed offspring was also inconsistent with such an explanation: The male offspring had cut wings and the females were *ct*/+ heterozygotes. That is, all of the rare nonlozenge progeny had an X chromosome with the constitution

$$\underline{ct\quad ++\quad +}$$

which could be accounted for by recombination between the two *lozenge* alleles. Proof of this conclusion came from the detection of the reciprocal recombinant chromosome

$$\underline{+\quad lz^{BS}lz^{g}\quad v}$$

in five male progeny that had a lozenge phenotype distinctly different from the phenotype resulting from the presence of either a lz^{BS} or lz^{g} allele alone. The exceptional males also had vermilion eyes, as expected from reciprocal recombination. The observation of intragenic recombination indicated that genes have a fine structure, now known to be a sequence of nucleotides, and that the multiplicity of allelic forms might be due to mutations at different sites (nucleotides) in the gene.

5.7 Mitotic Recombination

Genetic exchange can also take place in mitosis, although at a frequency about 1000-fold lower than in meiosis and probably by a somewhat different mechanism than meiotic recombination. The first evidence for mitotic recombination was obtained by Curt Stern in 1936 in experiments with *Drosophila*, but the phenomenon has been studied most carefully in fungi such as yeast and *Aspergillus*, in which the frequencies are higher than in most other organisms.

Genetic maps can be constructed from mitotic recombination frequencies. In some organisms in which a sexual cycle is unknown, mitotic recombination is the only method of obtaining linkage data. In organisms in which both meiotic recombination and mitotic recombination are found, mitotic recombination is always at a much lower frequency. In mitotic recombination, the relative map distances between particular genes sometimes correspond to those based on meiotic recombination frequencies, but for unknown reasons, the distances are often markedly different. The discrepancies may reflect different mechanisms of exchange and perhaps a nonrandom distribution of potential sites of exchange.

The genetic implications of mitotic recombination are illustrated in **Figure 5.26.** Each of the homologous chromosomes has two genetic markers distal to the site of the breakage and reunion. At the following anaphase, each centromere splits and the daughter cells receive one centromere of each color, along with the chromatid attached to it. The disjunction can happen in two ways, as shown on the right. If the two nonexchanged and the two exchanged chromatids go together (top), the result of the exchange is not detectable genetically. However, if each nonexchanged chromatid goes with one of the exchanged chromatids, then the result is that the region of the chromosome distal to the site of the exchange becomes homozygous. With an appropriate configuration of genetic markers, the exchange is detectable as a **twin spot** in which cells of one homozygous genotype (in this example, a pair of *a B* spots) are adjacent to cells of a different homozygous genotype (in this example, a pair of *A b* spots). In the original experiments with *Drosophila*, Stern used the X-linked markers $A = y^+$ and $a = y$ (*y* is *yellow* body color) along with $B = sn^+$ and $b = sn$ (*sn* is *singed* bristles). In females, the result of mitotic recombination was observed as a twin spot in which a patch of cuticle with a yellow color and normal bristles was adjacent

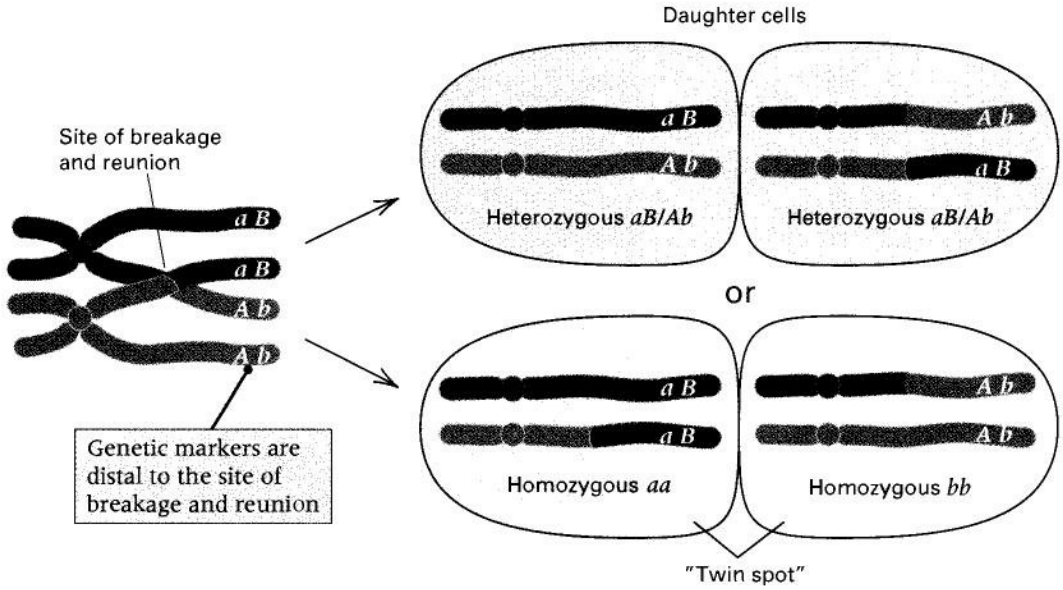


Figure 5.26 Mitotic recombination. Homologous chromosomes in prophase of mitosis are shown in light and dark blue. At anaphase, each centromere splits, and each daughter cell receives one centromere of each color and its attached chromatid. When a rare reciprocal exchange takes place between nonsister chromatids of homologous chromosomes, the daughter cells can be either of two types. If the nonexchanged chromatids and the exchanged chromatids go together (top right), both cells are genetically like the parent—in this example, *A b*/*a B*. If one nonexchanged and one exchanged chromatid go to each cell together (bottom right), the result is that alleles distal to the point of exchange become homozygous. In this example, one daughter cell is *a B*/*a B* and the other *A b*/*A b*. Adjacent cell lineages are therefore homozygous for either *aa* or *bb* and can be detected phenotypically as a "twin spot"of *aa* somatic tissue adjacent to *bb* somatic tissue.

to a patch of cuticle with normal color and singed bristles.

The rate of mitotic recombination can be increased substantially through x-ray treatment and the use of certain mutations. Mitotic recombination is a useful tool in genetics because it results in the production of **somatic mosaics,** organisms that contain two or more genetically different types of tissue. For example, the mosaic *Drosophila* females discussed in the previous paragraph consist of predominantly wildtype tissue but include some patches of y^+ *sn*/y^+ *sn* and some patches of *y* sn^+/*y* sn^+. Each twin spot of mutant tissue derives from repeated division of a single cell that became homozygous through a mitotic exchange. In modern developmental genetics, the movement and differentiation of particular cell lineages are tracked by the study of patches of genetically marked tissue resulting from mitotic recombination.

Chapter Summary

Nonallelic genes located in the same chromosome tend to remain together in meiosis rather than to undergo independent assortment. This phenomenon is called linkage. The indication of linkage is a significant deviation from the 1 : 1 : 1 : 1 ratio of phenotypes in the progeny of a cross of the form *Aa Bb* × *aa bb*. When alleles of two linked genes segregate, more than 50 percent of the gametes produced have parental combinations of the segregating alleles, and fewer than 50 percent have nonparental (recombinant) combinations of the alleles. The recombination of linked genes results from crossing-over, a process in which nonsister chromatids of the homologous chromosomes exchange corresponding segments in the first meiotic prophase.

The frequency of recombination between different genes can be used to determine the relative order and locations of the genes in chromosomes. This type of analysis is called genetic mapping. The distance between adjacent genes in such a map (a genetic or linkage map) is defined to be proportional to the frequency of recombination between them; the unit of map distance (the map unit or centimorgan) is defined as 1 percent recombination. One map unit corresponds to a physical length of the chromosome in which a crossover event takes place, on the average, once in every 50 meiotic divisions. For short distances, map units are additive. (For example, for three genes with order *a b c*, if the map distances *a* to *b* and *b* to *c* are 2 and 3 map units, respectively, then the map distance *a* to *c* is 2 + 3 = 5 map units.) The recombination frequency underestimates actual genetic distance if the region between the genes being considered is too great. This discrepancy results from multiple crossover events, which yield either no recombinants or the same number produced by a single event. For example, two crossovers in the region between two genes may yield no recombinants, and three crossover events may yield recombinants of the same type as that from a single crossover.

When many genes are mapped in a particular species, they form linkage groups equal in number to the haploid chromosome number of the species. The maximum frequency of recombination between any two genes in a cross is 50 percent; this happens when the genes are in nonhomologous chromosomes and assort independently or when the genes are sufficiently far apart in the same chromosome that at least one crossover is formed between them in every meiosis. The map distance between two genes may be considerably greater than 50 centimorgans, because the map distance is equal to half of the average number of crossovers per chromosome times 100. A mapping function is the mathematical relation between the genetic map distance across an interval and the observed percent recombination in the interval.

In many organisms, including human beings, model experimental organisms, and agricultural animals and plants, the genetic map includes hundreds or thousands of genetic markers distributed more or less uniformly throughout the euchromatin. Some of the most useful genetic markers are DNA polymorphisms (changes in nucleotide sequence present in wildtype organisms that are not associated with any phenotypic abnormalities) because of their high level of heterozygosity in natural populations. In human genetics, evidence for linkage is assessed by determining the frequency of recombination that maximizes the lod score and therefore maximizes the ratio of the probability of the observed pedigrees with linkage to that of the observed pedigrees with no linkage.

The four haploid products of each of a series of meiotic divisions can be used to analyze linkage and recombination in some species of fungi and unicellular algae. The method is called tetrad analysis. In *Neurospora* and related fungi, the meiotic tetrads are contained in a tubular sac, or ascus, in a linear order, which makes it possible to determine whether a pair of alleles segregated in the first or the second meiotic division. With such asci, it is possible to use the centromere as a genetic marker; in fact, the centromere serves as a reference point to which all genes in the same chromosome can be mapped. Linkage analysis in unordered tetrads is based on the frequencies of parental-ditype (PD), nonparental-ditype (NPD), and tetratype (TT) tetrads. The observation of NPD << PD is a sensitive indicator of linkage.

Key Terms

ascospore
ascus
attached-X chromosome
centimorgan
chromatid interference
chromosome interference
chromosome map
cis configuration
coefficient of coincidence
compound-X chromosome
coupling configuration
euchromatin
first-division segregation
frequency of recombination
genetic map
heterochromatin
likelihood ratio
linkage
linkage group
linkage map
lod score
map unit
mapping function
nonparental-ditype (NPD) tetrad
parental type
parental-ditype (PD) tetrad
random-spore analysis
recombinant
recombination
repulsion configuration
second-division segregation
somatic mosaics
syntenic genes
tetratype (TT) tetrad
three-point cross
trans configuration
twin spot

Review the Basics

- In genetic analysis, why is it important to know the position of a gene along a chromosome?
- How is the recombination frequency estimated? What does it mean to say that two genes have a recombination frequency of 10 percent? What is the maximum frequency of recombination between two genes? Is there a maximum map distance between genes?
- Why is the frequency of recombination over a long interval of a chromosome always smaller than the map distance over the same interval?
- For a region between two genes in which no more than one crossover can take place in each cell undergoing meiosis, the frequency of recombination is equal to one-half the frequency of meiotic cells in which a chiasma occurs in the region. Explain why the factor of one-half occurs when we compare recombination frequency to chiasma frequency.
- What is a double crossover? How many different kinds of double crossovers are possible? Diagram each kind of double crossover in a region between two heterozygous genes in a bivalent, along with the resulting chromatids.
- What is interference? In a region of high interference, would you observe more or fewer double crossovers? Explain.
- What is special about the inheritance of an attached-X chromosome, compared with the inheritance of unattached X chromosomes?
- Explain how in fungi, the occurrence of tetratype tetrads for two linked genes implies that crossing-over must take place at the four-strand stage of meiosis.
- Explain how the analysis of fungal tetrads for two linked genes allows double crossovers to be detected in a two-point cross (a cross in which only two genes are segregating).
- Explain how the recovery of ordered tetrads allows any gene to be mapped relative to the centromere of the chromosome on which the gene is located.

Guide to Problem Solving

Problem 1 The mutations *cinnabar* (*cn*, bright red eyes) and *vestigial* (*vg*, malformed wings) are linked in the second chromosome of *Drosophila*. Among 1000 progeny of the cross *cn vg* / + + females × *cn vg* / *cn vg* males, the following genotypes of progeny were observed. From these data, estimate the frequency of recombination between the *cn* and *vg* genes.

cn vg / *cn vg*	455
cn + / *cn vg*	35
+ *vg* / *cn vg*	45
+ + / *cn vg*	465

Answer In linkage problems of this type, you must first classify the progeny in terms of whether they are recombinant or nonrecombinant. Because the parental genotype is given as *cn vg* / + +, the nonrecombinant progeny are *cn vg* / *cn vg* (455) and + + / *cn vg* (465), and the recombinant progeny are *cn* + / *cn vg* (35) and + *vg* / *cn vg* (45). If the parental genotypes had not been given, it would still have been possible to classify the progeny using the principle that when there is linkage, the recombinant progeny are always less frequent than the nonrecombinant progeny. Because the frequency of recombination is determined by the frequency of progeny carrying recombinant

chromosomes, the frequency of recombination between *cn* and *vg* is estimated as $r = (35 + 45)/1000 = 0.08$. The frequency of recombination can also be expressed as a percentage. For recombination frequencies smaller than about 10 percent, the percent recombination equals the number of map units between the genes; in this case, $r =$ 8 percent recombination, or 8 map units.

Problem 2 In addition to *cn* and *vg*, the genes *curved* (*c*, curved wings) and *plexus* (*px*, extra wing veins) are linked in the second chromosome of *Drosophila*. In a cross of *cn c px* / + + + females × *cn c px* / *cn c px* males, the following progeny were counted:

cn c px / *cn c px*	296
cn c + / *cn c px*	63
cn + + / *cn c px*	119
cn + *px* / *cn c px*	10
+ *c px* / *cn c px*	86
+ *c* + / *cn c px*	15
+ + + / *cn c px*	329
+ + *px* / *cn c px*	82
Total	1000

(a) What is the frequency of recombination between *cn* and *c*?

(b) What is the frequency of recombination between *c* and *px*?

(c) What is the frequency of recombination between *cn* and *px*?

(d) Why is the frequency of recombination between *cn* and *px* smaller than the sum of that between *cn* and *c* and that between *c* and *px*?

(e) What is the coefficient of coincidence across this region? What is the value of the interference?

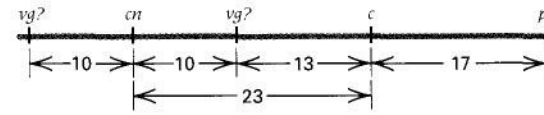
(f) Draw a genetic map of the region, showing the locations of *cn*, *c*, and *px* and the map distances between the genes.

(g) From the data in Problem 1, where in this map would you put the gene *vg*?

Answer Do not try to hurry through linkage problems! You will be rewarded by taking time to organize the information in the optimal manner. First, group the progeny types into reciprocal pairs—*cn c px* with + + +, *cn c* + with + + *px*, and so forth—and make a new list organized as shown here. (Ignore the *cn c px* chromosome from the father because it contributes no information about recombination.)

cn c px	296	625
+ + +	329	
cn c +	63	145
+ + *px*	82	
cn + +	119	205
+ *c px*	86	
cn + *px*	10	25
+ *c* +	15	
Total		1000

In this tabulation, braces have been used to group the pairs of reciprocal products in order to keep the groups separate. The number next to each brace is the total number of chromosomes in the group. The most numerous group of reciprocal chromosomes (in this case, *cn c px* and + + +) consists of the nonrecombinants, and the least numerous group of reciprocal chromosomes (in this case, *cn* + *px* and + *c* +) consists of the double recombinants. Rearrange the order of the groups, if necessary, so that the nonrecombinants are at the top of the list and the double recombinants are at the bottom. (In the present example, rearrangement is not necessary.) At this point, also make sure that the order of the genes is correct as given, by comparing the genotypes of the double recombinants with those of the nonrecombinants. If the gene order is correct, then it will require two recombination events (one in each interval) to derive the double-recombinant chromosomes from the nonrecombinants. If this is not the case, rearrange the order of the genes. (The "odd man out" in comparing the double recombinants with the nonrecombinants is always the gene in the middle.) In this particular example, the gene order is correct as given. Finally, with this preliminary bookkeeping done, we can proceed to tackle the questions. **(a)** The frequency of recombination between *cn* and *c* is given by the totals of all classes of progeny showing recombination in the *cn*–*c* interval, in this case $(205 + 25)/1000 = 0.23$. **(b)** The frequency of recombination between *c* and *px* equals $(145 + 25)/1000 = 0.17$. **(c)** The frequency of recombination between *cn* and *px* equals $(145 + 205)/1000 = 0.35$. (Note that the double recombinants are not included in this total, because the double recombinants are not recombined for *cn* and *px*; their allele combinations for *cn* and *px* are the same as in the nonrecombinants.) **(d)** The frequency of recombination between *cn* and *px* (0.35) is smaller than the sum of that between *cn* and *c* and that between *c* and *px* ($0.23 + 0.17 = 0.40$) because of double recombination. **(e)** The coefficient of coincidence equals the observed number of double recombinants divided by the expected number. The observed number is 25 and the expected number is $0.23 \times 0.17 \times 1000 = 39.1$; the coefficient of coincidence therefore equals $25/39.1 = 0.64$. The interference equals $1 -$ *coefficient of coincidence*, so the interference equals $1 - 0.64 = 0.36$. **(f)** The genetic map is shown in the accompanying diagram. The distances are in map units (centimorgans). However, the map distances of 23 and 17 map units are based on the 23 percent and 17 percent recombination observed between *cn* and *c* and between *c* and *px*, respectively; the actual distances in map units are probably a little greater than these estimates because of a small amount of double recombination within each of the intervals. **(g)** The position of *vg* is located 10 map units from *cn* because there is 10 percent recombination observed between *vg* and *cn*. However, there is no way of knowing from the information given whether *vg* is to the left of *cn* or the right, so both possible positions for *vg* are indicated. (In fact, *vg* is located between *cn* and *c*.)



Problem 3 In the yeast *Saccharomyces cerevisiae*, the gene *MET14* codes for an enzyme used in synthesis of the amino acid methionine and is very close to the centromere of chromosome XI. For simplicity, denote *MET14* with the symbol *a*, and suppose that *a* is used in mapping a new mutation—say, *b*—whose location in the genome is unknown. Diploids are made by mating *a B* × *A b*, in which the uppercase symbols denote the wildtype alleles, and asci are examined. In this mating, what kinds of asci correspond to parental ditype (PD), nonparental ditype (NPD) and tetratype (TT)? What relative proportions of PD, NPD and TT asci would be expected in each of the following cases:

(a) *b* closely linked with *a*?

(b) *b* on a different chromosome from the one *a* is on but close to the centromere?

(c) *b* on a different chromosome from the one *a* is on and far from the centromere?

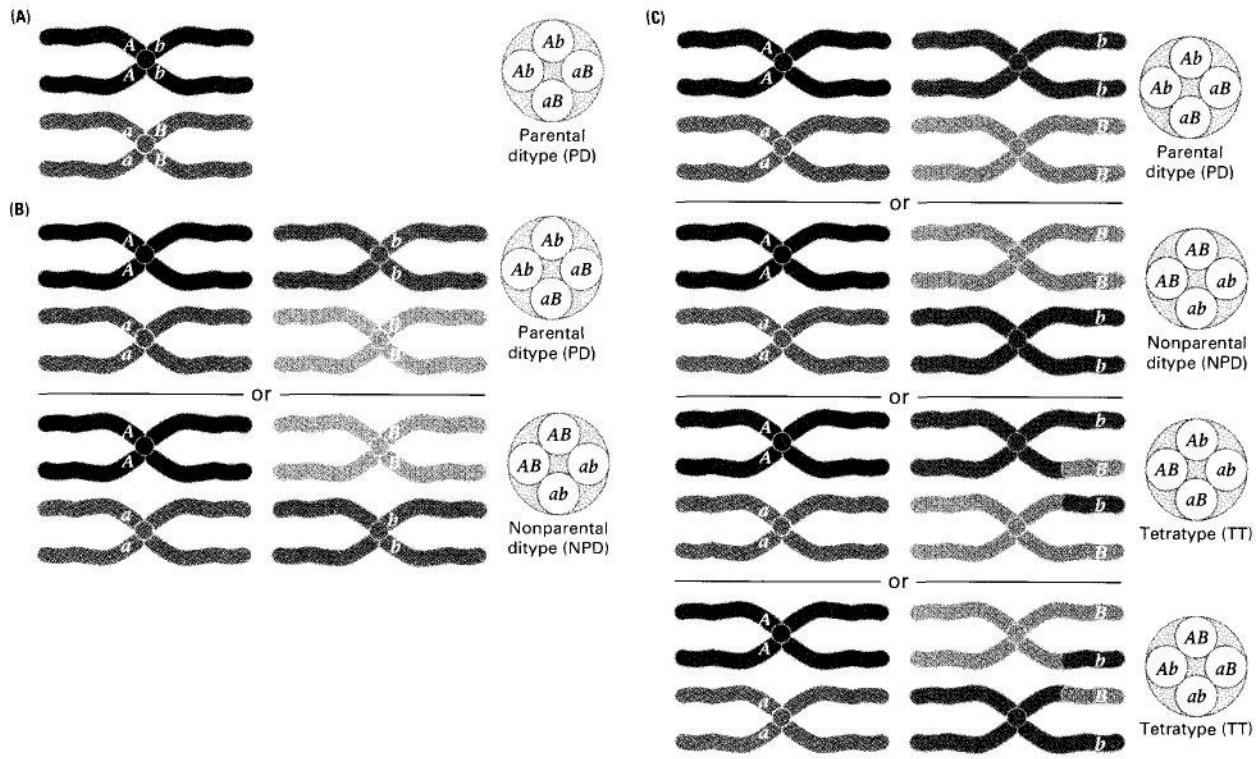
Answer The genotype of the diploid is the *trans*, or repulsion, heterozygote *a B*/*A b*, so the parental ditype (PD) tetrads contain two *a B* spores and two *A b* spores; the nonparental ditype (NPD) tetrads contain two *a b* spores and two *A B* spores; and the tetratype (TT) tetrads contain one spore each of *a B*, *A b*, *a b*, and *A B*. In this problem, the specific questions illustrate the principle that *any marker tightly linked to its centromere in a fungus with unordered tetrads serves to identify first-division and second-division segregation of every other gene in the organism.* This principle makes possible the mapping of other genes with respect to their centromeres. Reasoning through problems of tetrad analysis is best approached by considering the possible chromosome configurations during the meiotic divisions. The various possibilities are illustrated in the figure below, in which parts A through C correspond to questions (a) through (c).

(a) Part A shows the genes closely linked to each other (and therefore to the same centromere). They are shown in different arms, but they could just as well be in the same arm. With no crossing-over between either gene and the centromere, all tetrads are PD.

(b) In part B, the genes are closely linked to different centromeres. With no crossing-over between either gene and its centromere, half the tetrads are PD and half NPD.

(c) In part C, the *b* gene is not centromere-linked. Again PD = NPD (indicating that *A* and *B* are in different chromosomes), but crossing-over between the gene and the centromere yields TT tetrads. *The frequency of TT tetrads increases with the recombination frequency between the gene and the centromere, to a maximum of 2/3 when the gene and centromere undergo free recombination.*

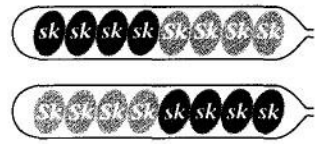
Problem 4 Genes called spore killers are relatively frequent in natural populations of *Neurospora crassa*. In asci produced by a heterozygous spore-killer genotype, all spores that do not carry the spore-killer allele are killed, and they are unable to be germinated. For example, in a cross of *Sk* × *sk*, where *Sk* represents the spore-killer allele, all spores carrying *sk* are killed. For one spore-killer allele, a cross of *Sk* × *sk* produced 125 asci. All of the asci had four dead ascospores and four live ascospores. In 95 of the asci, the four dead spores were all adjacent at either



Problem 3 Answer

the top end of the ascus or the bottom end. What is the map distance between the *Sk* allele and its centromere?

Answer Centromere mapping in *Neurospora* is based on the relative frequencies of first-division and second-division segregation. One feature that distinguishes these two types of asci is that in first-division segregation, the four spores carrying a particular allele are at one end of the ascus or the other. In the *Sk* × *sk* cross, the first-division asci would have one of the configurations shown in the diagram. Because the *sk*-bearing asci die, in these asci (and only in these) the dead ascospores are all either at the top or the bottom of the ascus. The observation is that 95/125 asci are of this type, which implies that 95/125 = 76 percent of the asci show first-division segregation. Because the distance between a gene and its centromere is estimated as 1/2 times the frequency of *second*-division segregation, this spore-killer gene is estimated to be at a distance of (1 − 0.76)/2 = 12 map units from its centromere.



Analysis and Applications

5.1 What is the principal difference between the genotype *A B / a b* and the genotype *A b / a B*? What is the role of the slash (/, technically called a virgule) in such an expression? Which genotype is the coupling (*cis*) genotype and which is the repulsion (*trans*) genotype?

5.2 What gametes, and in what ratios, are produced by male and female *Drosophila* of genotype *A B a b* when the genes are present in the same chromosome and the frequency of recombination between them is 5 percent?

5.3 If a large number of genes were mapped, how many linkage groups would be found in

(a) a haploid organism with 17 chromosomes per somatic cell?

(b) a bacterial cell with a single, circular DNA molecule?

(c) the rat, with 42 chromosomes per diploid somatic cell?

5.4 How is it possible for two genes located in the same chromosome to undergo independent assortment?

5.5 If the recombination frequency between genes *A* and *B* is 6.2 percent, what is the distance between the genes in map units in the linkage map?

5.6 Consider an organism heterozygous for two genes located in the same chromosome, *A B / a b*. If a single crossing-over occurs between the genes in every cell undergoing meiosis, and no multiple crossovers occur between the genes, what is the recombination frequency between the genes? What effect would multiple crossing-over have on the recombination frequency?

5.7 Normal plant height in wheat requires the presence of either or both of two dominant alleles, *A* and *B*. Plants homozygous for both recessive alleles (*a b/a b*) are dwarfed but otherwise normal. The two genes are in the same chromosome and recombine with a frequency of 16 percent. From the cross *A b/a B* × *A B/a b*, what is the expected frequency of dwarfed plants among the progeny?

5.8 The dominant allele for Huntington disease is segregating in the pedigree shown below. Also shown is the pattern of DNA bands observed for a restriction fragment polymorphism. The bands are designated by the numbers 1 and 2. Use *HD* and *hd* for the Huntington alleles, and A_1 and A_2 for the RFLP alleles that yield bands 1 and 2, respectively.

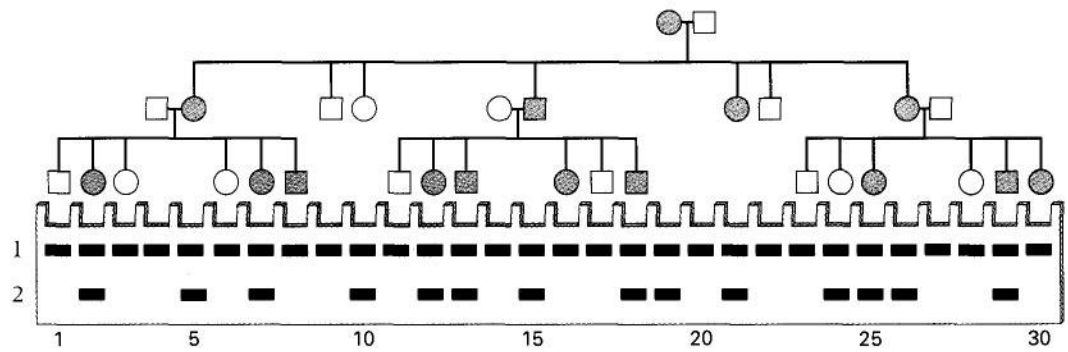
(a) Deduce the genotype of each person in the pedigree with respect to the Huntington gene and to the RFLP.

(b) Suggest a reason why no persons are observed who have only the smaller DNA fragment (fragment 2, the lower band).

(c) Indicate whether there is evidence of linkage between the Huntington gene and the RFLP.

(d) If so, list each person who carries a recombinant chromosome according to the numbers across the bottom of the gel.

(e) Estimate the frequency of recombination between the Huntington gene and the RFLP.

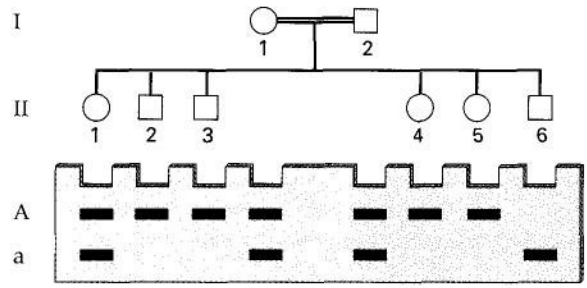


Problem 5.8

5.9 In the accompanying human pedigree, the parents I-1 and I-2 are both heterozygous for a recessive allele for cystic fibrosis (*cf*; the wildtype allele is *CF*) and heterozygous for a restriction fragment length polymorphism (RFLP) with phenotypes shown in the diagram of the electrophoresis gel. DNA fragments *A* and *a* are the result of a restriction-enzyme cleavage site that differs between alleles *A* and *a*. The RFLP is linked to the *CF* gene. I-1 and I-2 are first cousins, and a molecular analysis of their parents and siblings indicates that both I-1 and I-2 have the genotype *CF A / cf a*. The frequency of recombination in females is 16 percent, and in males is 10 percent. None of the children II-1 through II-5 is affected with cystic fibrosis.

(a) What is the probability that the child II-6 is affected?

(b) What is the probability that the child II-1 is a carrier?



5.10 In the domestic cat, the genes for yellow coat color (actually, a sort of orange) and for black coat color are alleles of an X-linked gene. Heterozygous females show patches of yellow and patches of black fur, a pattern called *calico*. The pedigree illustrated below shows the distribution of coat color in a number of litters of kittens, along with the patterns of DNA bands observed for two restriction fragment polymorphisms. Bands 1 and 2 result from DNA fragments produced by alleles A_1 and A_2; bands 3 and 4 result from DNA fragments produced by alleles B_3 and B_4 of a different gene. Examine the pedigree and assign complete genotypes to as many animals as possible.

(a) Is there any evidence that the coat color gene and the A_1, A_2 allelic pair are linked?

(b) Is there any evidence that the coat color gene and the B_3, B_4 allelic pair are linked?

(c) If there is evidence of linkage, identify which animals carry a recombinant chromosome.

(d) Estimate the frequency of recombination between the linked genes.

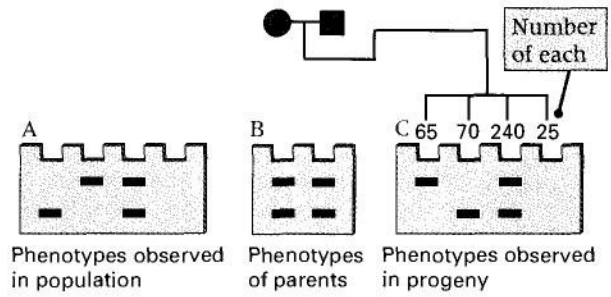
5.11 A human geneticist discovers the molecular variation in DNA sequence illustrated in the accompanying diagrams of electrophoresis gels. In the human population as a whole, she finds any of four phenotypes, shown in panel A. She believes that this may be a simple genetic polymorphism with three alleles, like the ABO blood groups. There are two alleles that yield DNA fragments of different sizes, fast (*F*) or slow (*S*) migration, and a "null" allele (*O*) in which the DNA fragment is deleted. The genotypes in panel A would therefore be, from left to right, *FF* or *FO*, *SS* or *SO*, *FS*, and *OO*. In the population as a whole, the putative *OO* genotype is extremely common, and the *FS* genotype is quite rare. To investigate this hypothesis further, the geneticist studies offspring of matings between parents who have the putative *FS* genotype (panel B). The types of progeny, and their numbers, are shown in panel C.

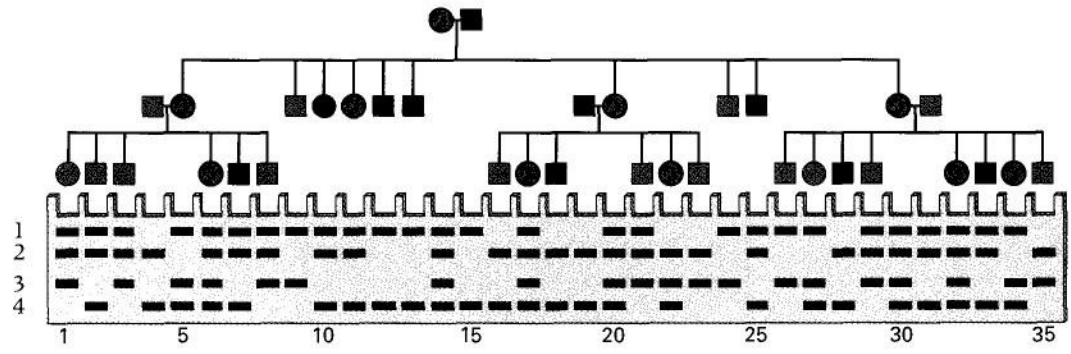
(a) What result would be expected from the three-allele hypothesis?

(b) Are the observed data consistent with this result? Why or why not?

(c) Suggest a genetic hypothesis that can explain the data in panel C.

(d) Are the data consistent with your hypothesis?





Problem 5.10

5.12 In the yellow-fever mosquito, *Aedes aegypti*, a dominant gene *DDT* for DDT resistance (DDT is dichlorodiphenyltrichloroethane, an insecticide) and a dominant gene *Dl* for Dieldrin resistance (Dieldrin is another long-lasting insecticide) are known to be in the same chromosome. A cross was carried out between a DDT-resistant strain and a Dieldrin-resistant strain, and female progeny resistant to both insecticides were testcrossed with wild-type males. The progeny were 99 resistant to both insecticides, 88 resistant to DDT only, 89 resistant to Dieldrin only, and 106 sensitive to both insecticides.

(a) Are *DDT* and *Dl* alleles of the same gene? How can you tell?

(b) Are *DDT* and *Dl* linked?

(c) What can you say about the positions of *DDT* and *Dl* along the chromosome?

5.13 In the yellow-fever mosquito, *Aedes aegypti*, sex is determined by a single pair of alleles, *M* and *m*. Heterozygous *Mm* mosquitos are male, homozygous *mm* are female. The recessive gene *bz* for bronze body color (the dominant phenotype is black) is linked to the sex-determining gene and recombines with it at a frequency of 6 percent. In the cross of a heterozygous bz^+ *M*/*bz m* male with a bronze female, what genotypes and phenotypes of progeny are expected?

5.14 Two genes in chromosome 7 of corn are identified by the recessive alleles *gl* (glossy), determining glossy leaves, and *ra* (ramosa), determining branching of ears. When a plant heterozygous for each of these alleles was crossed with a homozygous recessive plant, the progeny consisted of the following genotypes with the numbers of each indicated:

Gl ra/*gl ra*	88	*gl Ra*/*gl ra*	103
Gl Ra/*gl ra*	6	*gl ra*/*gl ra*	3

Calculate the frequency of recombination between these genes.

5.15 The recessive alleles *b* and *cn* of two genes in chromosome 2 of *Drosophila* determine black body color and cinnabar eye color, respectively. If the genes are 8 map units apart, what are the expected genotypes and phenotypes and their relative frequencies among the progeny of the cross ++/*b cn* × ++/*b cn*? (Remember that crossing-over does not occur in *Drosophila* males.)

5.16 In a testcross of an individual heterozygous for each of three linked genes, the most frequent classes of progeny were *A B c*/*a b c* and *a b C*/*a b c*, and the least frequent classes were *A B C*/*a b c* and *a b c*/*a b c*. What was the genotype of the triple heterozygote parent, and what is the order of the genes?

5.17 Construct a map of a chromosome from the following recombination frequencies between individual pairs of genes: *r*–*c*, 10; *c*–*p*, 12; *p*–*r*, 3; *s*–*c*, 16; *s*–*r*, 8. You will discover that the distances are not strictly additive. Why aren't they?

5.18 Yellow versus gray body color in *D. melanogaster* is determined by the alleles *y* and y^+, vermilion versus wild-type eyes by the alleles *v* and v^+, and singed versus straight bristles by the alleles *sn* and sn^+. When females heterozygous for each of these X-linked genes were testcrossed with yellow, vermilion, singed males, the following classes and numbers of progeny were obtained:

yellow, vermilion, singed	53
yellow, vermilion	108
yellow, singed	331
yellow	5
vermilion, singed	3
vermilion	342
singed	95
wildtype	63

(a) What is the order of the three genes? Construct a linkage map with the genes in their correct order, and indicate the map distances between the genes.

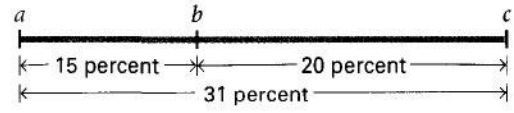
(b) How does the frequency of double crossovers observed in this experiment compare with the frequency expected if crossing-over occurs independently in the two chromosome regions? Determine the coefficient of coincidence and the interference.

5.19 In corn, the alleles *C* and *c* result in colored versus colorless seeds, *Wx* and *wx* in nonwaxy versus waxy endosperm, and *Sh* and *sh* in plump versus shrunken endosperm. When plants grown from seeds heterozygous for each of these pairs of alleles were testcrossed with plants from colorless, waxy, shrunken seeds, the progeny seeds were as follows:

colorless, nonwaxy, shrunken	84
colorless, nonwaxy, plump	974
colorless, waxy, shrunken	20
colorless, waxy, plump	2349
colored, waxy, shrunken	951
colored, waxy, plump	99
colored, nonwaxy, shrunken	2216
colored, nonwaxy, plump	15
Total	6708

Determine the order of the three genes, and construct a linkage map showing the genetic distances between adjacent genes.

5.20 The accompanying diagram summarizes the recombination frequencies observed in a large experiment to study three linked genes.



What was the observed frequency of double crossing-over in this experiment? Calculate the interference.

5.21 The recessive mutations *b* (black body color), *st* (scarlet eye color), and *hk* (hooked bristles) identify three autosomal genes in *D. melanogaster*. The following progeny were obtained from a testcross of females heterozygous for all three genes.

black, scarlet	243
black	241
black, hooked	15
black, hooked, scarlet	10
hooked	235
hooked, scarlet	226
scarlet	12
wildtype	18

What conclusions are possible concerning the linkage relations of these three genes? Calculate any appropriate map distances.

5.22 In the nematode *Caenorhabditis elegans*, the mutations *dpy*-21 (dumpy) and *unc*-34 (uncoordinated) identify linked genes that affect body conformation and coordination of movement. The frequency of recombination between the genes is 24 percent. If the heterozygote *dpy*-21 +/+ *unc*-34 undergoes self-fertilization (the normal mode of reproduction in this organism), what fraction of the progeny is expected to be both dumpy and uncoordinated?

5.23 Dark eye color in rats requires the presence of a dominant allele of each of two genes, *R* and *P*. Animals homozygous for either or both of the recessive alleles have light-colored eyes. In one experiment, homozygous dark-eyed rats were crossed with doubly recessive light-eyed rats, and the resulting F_1 animals were testcrossed with rats of the homozygous light-eyed strain. The progeny from the testcross consisted of 628 dark-eyed and 889 light-eyed rats. In another experiment, *R p*/*R p* animals were crossed with *r P*/*r P* animals, and the F_1 progeny were crossed with animals from an *r p*/*r p* strain. The progeny consisted of 86 dark-eyed and 771 light-eyed rats. Determine whether the genes are linked, and if they are, estimate the frequency of recombination between the genes from each of the experiments.

5.24 A *Drosophila* geneticist exposes flies to a mutagenic chemical and obtains nine mutations in the X chromosome that are lethal when homozygous. The mutations are tested in pairs in complementation tests, with the results shown in the accompanying table. A+ indicates complementation (that is, flies carrying both mutations survive), and a − indicates noncomplementation (flies carrying both mutations die). How many genes (complementation groups) are represented by the mutations, and which mutations belong to each complementation group?

	1	2	3	4	5	6	7	8	9
1	−	+	−	+	+	+	−	+	+
2		−	+	+	+	+	+	+	+
3			−	+	+	+	−	+	+
4				−	+	−	+	+	+
5					−	+	+	+	−
6						−	+	+	+
7							−	+	+
8								−	+
9									−

5.25 The following classes and frequencies of ordered tetrads were obtained from the cross $a^+ b^+ \times a\ b$ in *Neurospora*. (Only one member of each pair of spores is shown.)

Spore pair				Number
1–2	3–4	5–6	7–8	of asci
a^+b^+	a^+b^+	$a\ b$	$a\ b$	1766
a^+b^+	$a\ b$	a^+b^+	$a\ b$	220
a^+b^+	$a\ b^+$	a^+b	$a\ b$	14

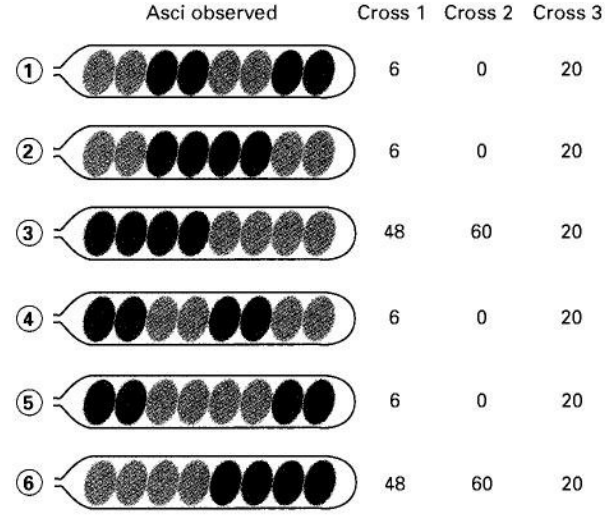
What is the order of the genes in relation to the centromere?

5.26 The following spore arrangements were obtained in the indicated frequencies from ordered tetrads in a cross between a *Neurospora* strain *com val* (*c v*), which exhibits a compact growth form and is unable to synthesize the amino acid valine, and a wildtype strain, ++. (Only one member of each pair of spores is shown.)

Spore pair	Ascus composition				
1–2	*c v*	*c* +	*c v*	+ *v*	*c v*
3–4	*c v*	*c* +	*c* +	*c* +	+ *v*
5–6	++	+ *v*	+ *v*	*c v*	*c* +
7–8	++	+ *v*	++	++	++
Number:	34	36	20	1	9

What can you conclude about the linkage and location of the genes with respect to each other and to the centromere?

5.27 In *Neurospora* several genes are needed for production of black pigment in spores. Recessive alleles at any one of these can result in tan spores. Six possible arrangements of spore colors are possible in crosses segregating for one of these allelic pairs, as shown in the accompanying diagram. Three different allelic pairs were analyzed in crosses 1, 2, and 3, respectively. The results are tabulated in the illustration.



For each cross, indicate how the different arrangements arise and what you can deduce about the location of the

segregating gene. In crosses in which it is relevant, calculate a map distance.

5.28 In baker's yeast (*Saccharomyces cerevisiae*), which produces unordered asci, the following cross was made:

leu2 *trp1* + × + + *met14*

The resulting diploid was induced to undergo meiosis, and the asci were dissected and spore clones analyzed. The following tetrad types were found, with the observed number of each shown below. For convenience, *leu2*, *trp1*, and *met14* are abbreviated *l*, *t*, and *m*, respectively.

l t m	l t +	+ t +	+ t m	+ t m	+ t +
l t m	l t +	+ t +	+ t m	l t m	l t +
+ + +	+ + m	l + m	l + +	+ + +	+ + m
+ + +	+ + m	l + m	l + +	l + +	l + m
230	235	215	220	54	46

Analyze these data as fully as you can with respect to the linkage of each gene with respect to every other gene and with respect to its centromere. Where a genetic linkage is indicated, calculate the map distance.

5.29 Genes called "spore-killers" are relatively frequent in natural populations of *Neurospora crassa*. In asci produced by a heterozygous spore-killer genotype, all spore not carrying the spore-killer allele are killed and are unable to be germinated. For example, a cross of *Sk-1* × *sk-1*, where *Sk-1* represents the spore-killer allele, all spores carrying *sk-1* are killed. Suppose *Sk-1* and *Sk-2* are spore-killer alleles of different genes. The cross *Sk-1 sk-2* × *sk-1 Sk-2* is carried out, and the resulting asci are examined. Because there are two different spore-killer systems, the only spores that survive are those containing both *Sk-1* and *Sk-2*. Among 180 asci examined from the cross, 76 asci had eight dead spores, 32 asci had four dead spores, and 72 asci had six dead spores.

(a) Classify each type of ascus as parental ditype (PD), nonparental ditype (NPD), or tetratype (TT).

(b) Determine whether *Sk-1* and *Sk-2* are linked genes.

5.30 In baker's yeast (*Saccharomyces cerevisiae*), which produces unordered asci, the following cross was made:

rad6 *trp5* *leu1* + × + + + *met14*

The resulting diploid was induced to undergo meiosis, the asci were dissected, and the spore clones analyzed. The following tetrad types were found in the numbers shown below each tetrad. For convenience, *rad6*, *trp5*, *leu1*, and *met14* are abbreviated as *r*, *t*, *l*, and *m*, respectively.

r t l +	r t l m	r t l +	r t l m	r t l +
r t l +	r t l m	+ t l +	+ t l m	+ + l +
+ + + m	+ + + +	r + + m	r + + +	r t + m
+ + + m	+ + + +	+ + + m	+ + + +	+ + + m
188	206	105	92	140

r t l m	r t l +	r t l m	r + l +	+ t l m
+ + l m	+ + + +	+ + l m	+ t l +	+ + l m
r t + +	r t l m	+ t + +	r t + m	r + + +
+ + + +	+ + + m	r + + +	+ + + m	r t + +
154	109	3	2	1

Use what you learned about *met14* in Problem 5.28 to analyze these data as fully as you can for linkage of the genes to one another and of each to its respective centromere. When genetic linkage is indicated, calculate the map distance.

Challenge Problems

5.31 The linkage relationships of four genes were studied in the flowering plant *Arabidopsis thaliana*. The recessive allele *cnx* confers resistance to chlorate, *ga3* results in gibberellin-responsive dwarfism, *sex1* causes overproduction of starch, and *th2* creates a requirement for thiamine. A plant homozygous for all four genetic markers was crossed with wildtype, and the F_1 progeny were testcrossed. The gametes from the heterozygous parent were found to be as follows:

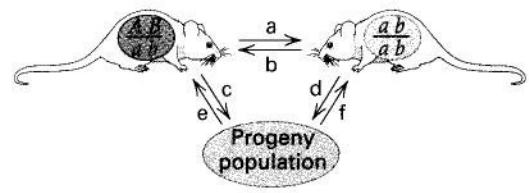
cnx	*ga3*	*sex1*	*th2*	185	*cnx*	*ga3*	*sex1*	+	29
cnx	*ga3*	+	*th2*	205	*cnx*	*ga3*	+	+	28
cnx	+	*sex1*	*th2*	2	*cnx*	+	*sex1*	+	26
cnx	+	+	*th2*	3	*cnx*	+	+	+	19
+	*ga3*	*sex1*	*th2*	24	+	*ga3*	*sex1*	+	1
+	*ga3*	+	*th2*	21	+	*ga3*	+	+	4
+	+	*sex1*	*th2*	26	+	+	*sex1*	+	198
+	+	+	*th2*	31	+	+	+	+	198

(a) Which of the genes, if any, are linked?

(b) For any genes that are linked, construct a genetic map.

(c) What is the value of the chromosome interference across any region of linked genes?

5.32 Tissue antigens, such as those that dictate the success or failure of skin grafts, are determined in a codominant fashion. A graft will be rejected if, and only if, the donor tissue contains an antigen not present in the recipient. The accompanying diagram shows all possible skin grafts between two strains of mice and their progeny. Each of the allele pairs *A*, *a* and *B*, *b* determines a different tissue antigen, and the genes are linked with a frequency of recombination of 20 percent. For each of the arrows, what is the probability of acceptance of a graft in which the donor is an animal chosen at random from the population shown at the base of the arrow and the recipient is an animal chosen at random from the population indicated by the arrowhead?



GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes,* choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

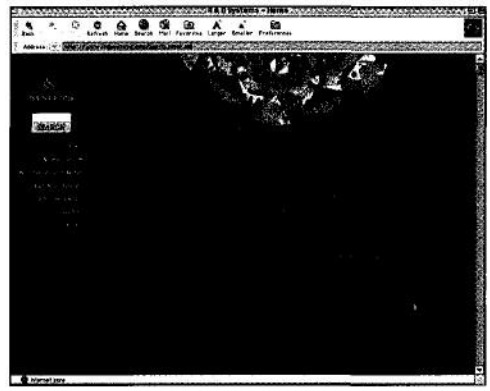
- The linkage map of the human chromosomes currently includes more than 5000 genetic markers. Detailed linkage maps of the **human chromosomes** are available in graphical format at this keyword site. The site also includes tables of the physical length of each chromosome in nucleotide pairs, the genetic length in centimorgans in males and females, and the number of informative meioses that are required to yield statistically significant lod scores for any specified recombination fraction.

- The database **SGD** includes information on the molecular biology and genetics of *Saccharomyces cerevisiae,* commonly known as budding yeast or baker's yeast. Genetic linkage analysis remains an important tool in this organism, even though its DNA sequence is completely known. Genetic research is often driven by the discovery of a new mutation that exhibits an arresting phenotype, and the first step in the analysis is to map the mutation genetically to learn its position and ultimately its identity. The keyword **tetrads** will access extensive listings of the numbers of parental ditype, nonparental ditype, and tetratype tetrads observed in two-point crosses.

- One of the earliest cases of genetic linkage in a human autosome, reported in 1965, was that between the ABO blood-group gene and a gene that, when mutated, causes nail–patella syndrome (**NPS**). The frequency of recombination is about 10% but is higher in females than in males. At this keyword site you can learn more about NPS, as well more recent linkage data and the lod scores on which the inference of linkage is based.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 5 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **Pic Site** for Chapter 5.



5.33 The Couch Potato has a genetic mapping function given by the formula

$$r = 50[1 - (0.2)^{d/100}]$$

where r is the frequency of recombination (in percent) and d is the map distance (in centimorgans).

(a) What frequency of recombination corresponds to 10 map units? to 20 map units? to 50 map units? to 100 map units?

(b) What map distance corresponds to 1 percent recombination? to 5 percent recombination? to 25 percent recombination? to 40 percent recombination? What happens when $r = 50$ percent?

(c) Draw a graph of frequency of recombination against map distance using this mapping function.

Further Reading

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314.

Carlson, E. A. 1987. *The Gene: A Critical History.* 2d ed. Philadelphia: Saunders.

Creighton, H. S., and B. McClintock. 1931. A correlation of cytological and genetical crossing over in *Zea mays*. *Proceedings of the National Academy of Sciences, USA* 17: 492.

Fincham, J. R. S., P. R. Day, and A. Radford. 1979. *Fungal Genetics.* Oxford, England: Blackwell.

Green, M. M. 1996. The "Genesis of the White-Eyed Mutant" in *Drosophila melanogaster*: A reappraisal. *Genetics* 142: 329.

Kohler, R. E. 1994. *Lords of the Fly.* Chicago, IL: University of Chicago Press.

Levine, L. 1971. *Papers on Genetics.* St. Louis, MO: Mosby.

Lewis, E. B. 1995. Remembering Sturtevant. *Genetics* 141: 1227.

Lucchesi, J. C. 1999. On the origin of sex chromosomes. *Bioessays* 21: 188.

Morton, N. E. 1995. LODs past and present. *Genetics* 140: 7.

Risch, N. J. 2000. Searching for genetic determinants in the new millenium. *Nature* 405: 847.

Stewart, G. D., T. J. Hassold, and D. M. Kurnit. 1988. Trisomy 21: Molecular and cytogenetic studies of nondisjunction. *Advances in Human Genetics* 17: 99.

Sturtevant, A. H. 1965. *A History of Genetics.* New York: Harper & Row.

Sturtevant, A. H., and G. W. Beadle. 1962. *An Introduction to Genetics.* New York: Dover.

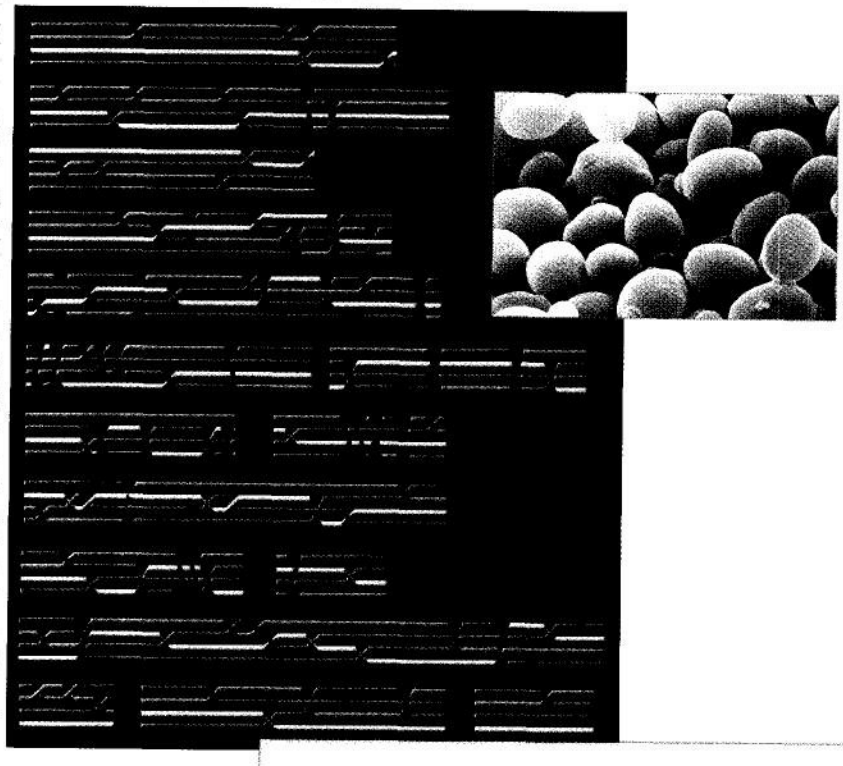
Voeller, B. R., ed. 1968. *The Chromosome Theory of Inheritance: Classical Papers in Development and Heredity.* New York: Appleton-Century-Crofts.

Wayne, R. K., and E. A. Ostrander. 1999. Origin, genetic diversity, and genome structure of the domestic dog. *Bioessays* 21: 247.

White, R., and J.-M. Lalouel. 1988. Chromosome mapping with DNA markers. *Scientific American,* February.

CHAPTER 6

Molecular Biology of DNA Replication and Recombination



Positions of all crossovers in a single meiotic cell of budding yeast. The two outer strands and the two inner strands represent the pairs of sister chromatids, hence what at first appear to be sister-strand exchanges are actually exchanges between nonsister chromatids. Locations of crossovers detected by microarray hybridization methods are described in Chapter 13. [Diagram: Reprinted with permission from E. Winzeler, D. Richards, and R. Davis, *Science* 281: 5380. Copyright 1998 American Association for the Advancement of Science. Yeast: Andrew Syred/Science Photo Library/Photo Researchers, Inc.]

CHAPTER OUTLINE

6.1 Problems of Initiation, Elongation, and Incorporation Error

6.2 Semiconservative Replication of Double-Stranded DNA
The Meselson–Stahl Experiment
Semiconservative Replication of DNA in Chromosomes
Theta Replication of Circular DNA Molecules
Rolling-Circle Replication
Multiple Origins and Bidirectional Replication in Eukaryotes

6.3 Unwinding, Stabilization, and Stress Relief

6.4 Initiation by a Primosome Complex

6.5 Chain Elongation and Proofreading

6.6 Discontinuous Replication of the Lagging Strand
Fragments in the Replication Fork
The Joining of Precursor Fragments

6.7 Terminator Sequencing of DNA
The Sequencing Procedure
Shotgun Sequencing
Clinical Use of Dideoxynucleoside Analogs

6.8 Molecular Mechanisms of Recombination
Gene Conversion and Mismatch Repair
The Holliday Model of Recombination
Single-Strand Break and Repair Model
Double-Strand Break and Repair Model

PRINCIPLES

- DNA replication is semiconservative; the parental strands remain intact and serve as templates for daughter strand synthesis.
- New polynucleotide strands are initiated by a primosome containing an RNA polymerase that synthesizes a short RNA primer complementary to a region of the template strand.
- Each RNA primer is elongated by DNA polymerase, which adds successive deoxyribonucleotides to the 3' end of the growing chain. The leading strand, whose 3' end faces the replication fork, is synthesized continuously; the lagging strand is synthesized in relatively short precursor fragments (Okazaki fragments).
- The major DNA polymerase (Pol III in *E. coli*) has a 3'-to-5' exonuclease function that serves for proofreading, in which the last-added nucleotide is removed if it contains an incorrect base.
- Dideoxy sequencing is a chain-termination method of DNA sequencing in which the nucleotide sequence of a growing strand is deduced from the lengths of successive fragments whose elongation was terminated by the incorporation of a dideoxyribonucleotide lacking a 3'-OH group.
- In the Holliday model of homologous recombination, two nicked duplexes come together to form heteroduplex regions, which can be substrates for mismatch repair and gene conversion, and a Holliday junction that can be resolved with recombination of outside markers.

CONNECTIONS

Replication by Halves
Matthew Meselson and Franklin W. Stahl 1958
The Replication of DNA in Escherichia coli

Happy Holliday
Robin Holliday 1964
A Mechanism for Gene Conversion in Fungi

Duplex DNA consists of two complementary strands of deoxyribonucleotide polymers usually paired and wound in a right-handed helix. (The details of this structure are examined in Chapter 2.) Immediately upon its discovery, the structure of DNA suggested how it replicates. Watson and Crick suggested that the strands could come apart and each separated strand serve as a template for the synthesis of a new partner strand complementary in nucleotide sequence. This process would create two daughter duplexes, both identical in nucleotide sequence to the original duplex. Although this hypothesis proved wrong in nearly all of its details, the fundamental concepts were accurate:

- In DNA replication, each parental strand remains intact as an uninterrupted unit.
- Each parental strand serves as a template for the synthesis of a complementary strand.
- The end result of DNA replication is the formation of two daughter duplexes, both identical in nucleotide sequence to the parental duplex.

The main problems with the Watson–Crick model are that new DNA strands cannot be initiated, but only elongated, and that elongation of a DNA strand can occur in only one direction. In coping with these and other molecular constraints, cells have evolved a relatively complex apparatus for DNA replication that involves a number of "molecular machines" whose structures and functions are discussed in this chapter. The synthesis of DNA is also an important component of genetic recombination, molecular models of which we also examine.

6.1 Problems of Initiation, Elongation, and Incorporation Error

A DNA molecule or a region of DNA that replicates as an individual unit is known as a **replicon.** The circular DNA molecule of *E. coli,* 4.6 Mb in length, is a single replicon in which replication originates at a specific region of 245 base pairs called the **origin of replication.** Most other bacterial chromosomes are also single replicons, as are the small, usually circular molecules known as *plasmids* that are often found in bacterial cells. Nearly all eukaryotic chromosomes consist of many replicons and therefore contain many origins of replication. Some human chromosomes contain thousands of origins of replication. In budding yeast, the sequences at which DNA replication can initiate are known as *autonomously replicating sequences* (*ARS*). It is not known whether an origin of replication is a specific DNA sequence or a region in the chromosome, how origins are recognized by the replication machinery, or how their numbers and activities are coordinated so that each chromosomal region is replicated once, and only once, in the S phase of each cell cycle.

The primary function of DNA replication is to duplicate the base sequence of the parent molecule. The specificity of base pairing—adenine with thymine (replaced by uracil in RNA) and guanine with cytosine—provides the mechanism used by all genetic replication systems. Each parental strand remains as an intact molecule and serves as a template for the synthesis of a newly replicated strand, or *daughter strand,* which is complementary in nucleotide sequence. For example, wherever an adenine nucleotide is present in the parental strand, a thymine nucleotide will be added to the growing end of the daughter strand. There are three principal molecular constraints on this process.

- ***Initiation*** No mechanism is known by which a new DNA strand can be initiated. All known mechanisms initiate DNA synthesis with a short RNA molecule known as a **primer,** which is elongated by adding successive deoxyribonucleotides to the 3' end. At a later stage in strand synthesis, the primer RNA is degraded and replaced with an equivalent stretch of DNA.
- ***Elongation*** Nucleotide monomers can be added only to the 3' end of the growing daughter strand. This immediately presents a problem, which is evident in **Figure 6.1.** Because the strands in duplex DNA are antiparallel (one is oriented in the 5'-to-3' direction, the other in the 3'-to-5' direction), strand elongation can take place continuously only in the direction in which the DNA duplex is unwinding. The elongating strand with its 3' end facing in the direction of

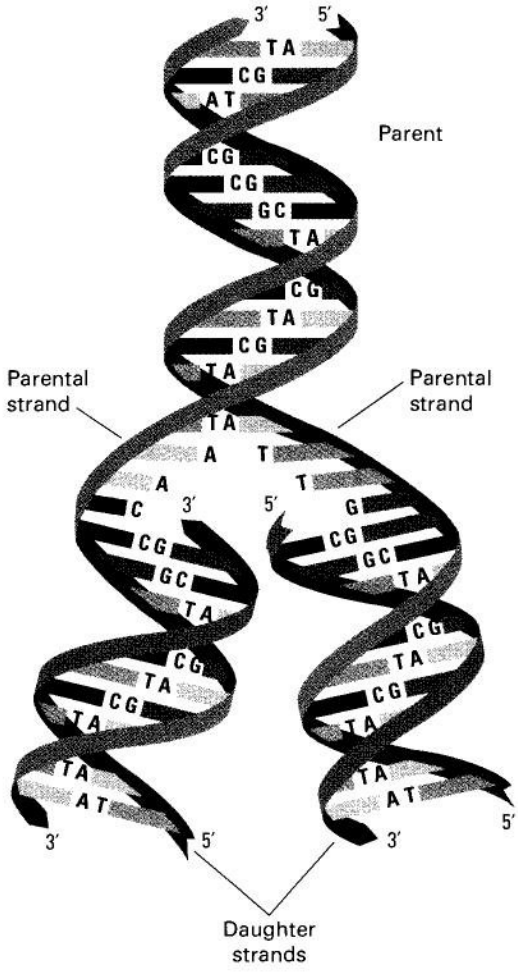


Figure 6.1 DNA replication as originally envisaged by Watson and Crick.

unwinding (up, in Figure 6.1) can be elongated as the helix opens. This continuously replicated strand is called the **leading strand.** The other strand, which is called the **lagging strand,** has its 5′ end facing in the direction of unwinding, and the 5′ end cannot be elongated. The lagging strand must be synthesized in some other way.

- ***Incorporation error*** The molecular specificity of base pairing is not perfect. On average, an incorrect (mismatched) nucleotide would be incorporated at the rate of about 10^{-5} per template nucleotide per round of replication. In human DNA, this would result in 60,000 new base-substitution mutations per cell cycle, which would cause an insupportable load of genetic disease and disability. As we shall see, the vast majority of the incorporation errors are corrected immediately after they occur by an enzymatic function called **proofreading.**

6.2 Semiconservative Replication of Double-Stranded DNA

The mechanism in which each parental DNA strand remains intact and serves as a template for a daughter strand is known as **semiconservative replication.** The experimental evidence for semiconservative replication is examined next.

The Meselson–Stahl Experiment

In theory, DNA could be replicated by a number of mechanisms other than the semiconservative mode, but the reality of semiconservative replication was demonstrated experimentally by Matthew Meselson and Franklin Stahl in 1958. The experiment made use of a newly developed high-speed centrifuge (an **ultracentrifuge**) that could spin a solution so fast that molecules differing only slightly in density could be separated. In their experiment, the heavy ^{15}N isotope of nitrogen was used to alter the density of DNA and allow the parental and daughter DNA molecules to be separated. DNA isolated from the bacterium *E. coli* grown in a medium containing ^{15}N as the only available source of nitrogen is denser than DNA from bacteria grown in media with the normal ^{14}N isotope. These DNA molecules can be separated in an ultracentrifuge, because they have about the same density as a very concentrated solution of cesium chloride (CsCl).

When a CsCl solution containing DNA is centrifuged at high speed, the Cs^+ ions gradually sediment toward the bottom of the centrifuge tube. This movement is counteracted by diffusion (the random movement of molecules), which prevents complete sedimentation. At equilibrium, a linear gradient of CsCl concentration and therefore density is present, increasing in cesium concentration and density from the top of the centrifuge tube to the bottom. DNA in the tube also moves upward or downward to a position in the gradient at which the local density of the solution is equal to its own density. At equilibrium, a mixture of ^{14}N-containing ("light") and ^{15}N-containing ("heavy") *E. coli* DNA will separate into two

distinct zones in a density gradient, even though they differ only slightly in density. DNA from *E. coli* containing ^{14}N in the purine and pyrimidine rings has a density of 1.708 g/cm^3, whereas DNA with ^{15}N in the purine and pyrimidine rings has a density of 1.722 g/cm^3. These molecules can be separated because a solution of 6.6 molar CsCl has a density of 1.700 g/cm^3. When spun in a centrifuge, the CsCl solution forms a gradient of density that brackets the densities of the light and heavy DNA molecules. It is for this reason that the separation technique is called **equilibrium density-gradient centrifugation.**

Imagine an experiment using bacteria grown for many generations in a ^{15}N-containing medium so that all parental DNA strands are "heavy." At the beginning of the experiment, the cells are transferred to a ^{14}N-containing medium so that newly synthesized DNA strands are "light." Duplex DNA is isolated from samples of cells taken from the culture at intervals, and equilibrium density-gradient centrifugation is carried out to determine the density of the molecules. With semiconservative replication, the expected result of the experiment is as shown in **Figure 6.2**. After one round of replication, each duplex consists of one heavy and one light strand, so all daughter molecules are of intermediate density. After two rounds of replication, the duplexes containing an original parental strand are again intermediate in density, but now there are an equal number of duplexes consisting of two light strands, so two bands differing in density

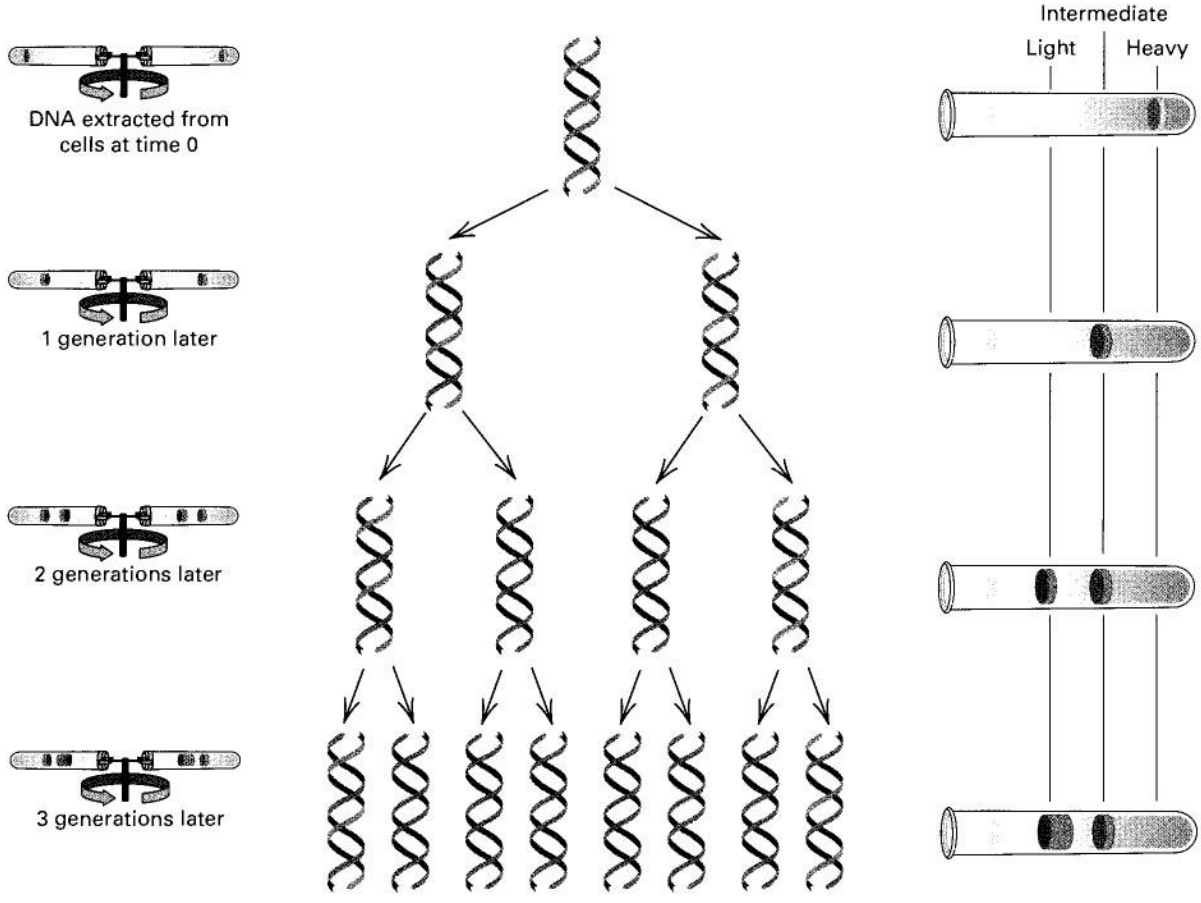


Figure 6.2 Expected density differences in parental and daughter DNA strands after 0, 1, 2, and 3 rounds of replication, on the hypothesis of semiconservative replication. DNA strands labeled with ^{15}N are shown in red, those labeled with ^{14}N in blue. The diagrams at the right show the positions at which the molecules equilibrate in a CsCl density gradient.

result from centrifugation. After a third round of replication, DNA duplexes of light and intermediate density are again expected, but in this generation their expected ratio of abundances is 3 : 1, as shown by the ribbon diagrams.

The actual result of the Meselson–Stahl experiment is shown in part A of **Figure 6.3.** Each photograph or absorption trace is that of the CsCl solution in a centrifuge tube oriented horizontally as in Figure 6.2, with the bottom (more dense) at the right and the top (less dense) at the left. Each photograph in Figure 6.3A shows the photographic image taken in ultraviolet light of wavelength 260 nm (nanometers), which is absorbed by DNA. The positions of the DNA molecules in the density gradient are therefore indicated by the dark bands that absorb the light. The traces show the absorption of the ultraviolet light along the tube; the peaks correspond to the positions of the bands in the photographs, but the height and width of each peak make it possible to quantify the DNA in each band.

At the start of the experiment (time 0), all of the DNA was heavy (^{15}N). After the transfer to ^{14}N, a band of lighter density began to appear, and it gradually became more prominent as the cells replicated their DNA and divided (Figure 6.3A). After one generation of growth (one round of replication of the DNA molecules and a doubling of the number of cells), all of the DNA had a "hybrid" density exactly intermediate between the densities of ^{15}N-DNA and ^{14}N-DNA. The finding of molecules with a hybrid density indicates that the replicated molecules contain equal amounts of the two nitrogen isotopes. After a second generation of replication in the ^{14}N medium (1.9 generations in the original experiment), half of the DNA had the density of DNA with ^{14}N in both strands ("light" DNA) and the other half had the hybrid density. After three generations, the ratio of light to hybrid DNA was approximately 3 : 1, and after four generations (4.1 in the original experiments), it was approximately 7 : 1. This distribution of ^{15}N atoms is precisely the result predicted from semiconservative replication as illustrated in Figure 6.2. Similar experiments with replicating DNA from numerous viruses and bacteria later confirmed the semiconservative mode of replication in these groups.

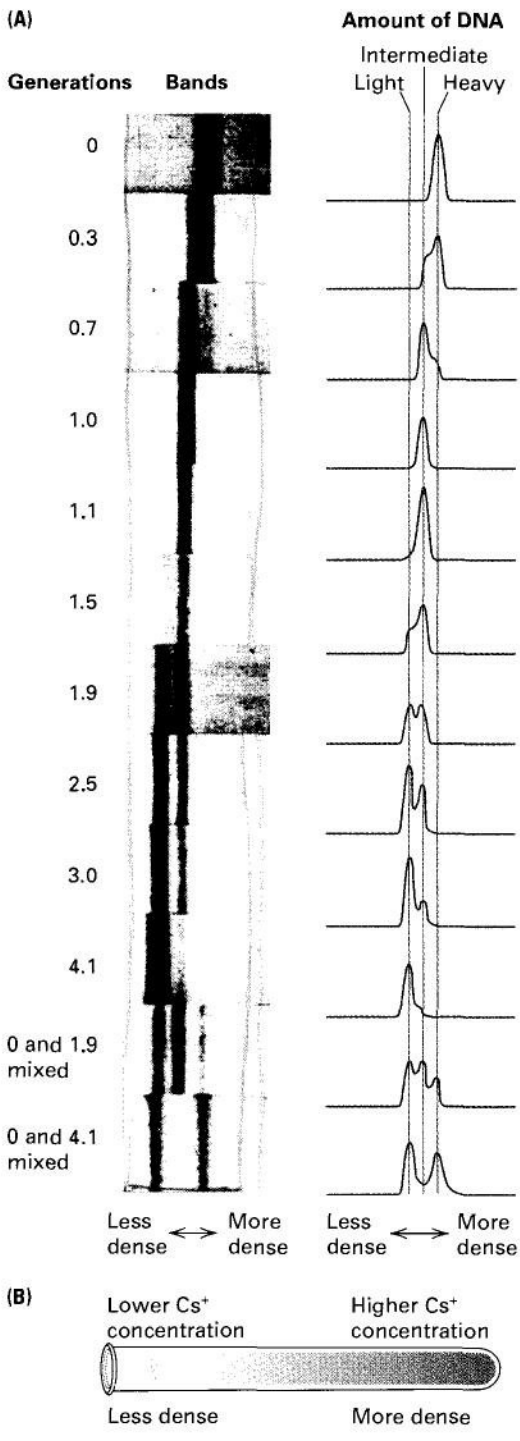


Figure 6.3 The Meselson–Stahl experiment on DNA replication. (**A**) Photographs of the centrifuge tubes taken with ultraviolet light are shown at the left. The top of the tube is at the left, the bottom at the right, as shown in part B. The smooth curves in part A show quantitatively the amount of absorption of the ultraviolet light across the tube. [Photograph courtesy of Matthew Meselson. From M. Meselson and F. Stahl, *Proc. Natl. Acad. Sci. USA* 1958. 44:671.]

CoNnection

Replication by Halves

Matthew Meselson and Franklin W. Stahl 1958
California Institute of Technology, Pasadena, California
The Replication of DNA in Escherichia coli

Replication of DNA by separation of its strands, followed by the synthesis of a new partner strand for each separated parental strand, is so fundamental a biological mechanism that it is easy to forget that it had to be proved experimentally somewhere along the way. Although semiconservative replication is a simple and logical way to proceed, biological systems might have found a different way to replicate their DNA. The Meselson–Stahl experiment that demonstrated semiconservative replication made use of a novel technique, density-gradient centrifugation, which was invented by the authors specifically for the purpose of separating DNA molecules differing slightly in density. The use of cesium chloride was mandated, because in concentrated solution, it has a density similar to that of DNA. The method of cesium chloride density centrifugation is still important in modern molecular biology; it is used in the isolation of DNA molecules.

Studies of bacterial transformation and bacteriophage infection strongly indicate that deoxyribonucleic acid (DNA) can carry and transmit hereditary information and can direct its own replication. Hypotheses for the mechanism of DNA replication differ in the predictions they make concerning the distribution among progeny molecules of atoms derived from parental molecules. . . . We anticipated that a label which imparts to the DNA molecule an increased density might permit an analysis of this distribution by sedimentation techniques. To this end, a method was developed for the detection of small density differences among macromolecules. . . . A small amount of DNA in a concentrated solution of cesium chloride is centrifuged until equilibrium is closely approached. The opposing processes of sedimentation and diffusion have then produced a stable concentration gradient of the cesium chloride, with a continuous increase in density along the direction of centrifugal force. The macromolecules of DNA present in this density gradient are driven by the centrifugal field into the region where the solution density is equal to their own buoyant density. . . . Bacteria uniformly labeled with N^{15} were abruptly changed into N^{14} medium. . . . Samples were withdrawn from the culture immediately and afterward at intervals for several generations. . . . Until one generation time has elapsed, half-labeled molecules accumulate. One generation after the switch to N^{14}, these half-labeled or "hybrid" molecules alone are observed. Subsequently, only half-labeled DNA and completely unlabeled DNA are found. When two generation times have elapsed, half-labeled and unlabeled DNA are present in equal amounts. These results permit the following conclusions to be drawn: 1. *The nitrogen of a DNA molecule is divided equally between two subunits which remain intact through many generations.* . . . 2. *Following replication, each daughter molecule has received one parental subunit.* . . . 3. *The replicative act results in a molecular doubling.* . . . A molecular structure for DNA has been proposed by Watson and Crick that . . . suggested to them a definite and structurally plausible hypothesis for the duplication of the DNA molecule. . . . The results of the present experiment are in exact accord with the expectation of the Watson–Crick model for DNA duplication. . . . The results presented here direct our attention to [other important problems]. What are the molecular structures of the subunits of DNA which are passed on intact to each daughter molecule? What is the relationship of these subunits to each other in a DNA molecule? What is the mechanism of the synthesis and dissociation of the subunits in vivo?

The results of the present experiment are in exact accord with the expectation of the Watson–Crick model for DNA replication.

Source: Proceedings of the National Academy of Sciences of the USA 44: 671–682

Semiconservative Replication of DNA in Chromosomes

DNA replication in the chromosomes of eukaryotic cells is also semiconservative. Direct evidence comes from labeling studies that are similar in principle to those of Meselson and Stahl but use a label that can be detected in chromosomes in the light microscope. An example with human chromosomes is shown in **Figure 6.4**. In this case the label is *bromodeoxyuridine* (*BUdR*), a nucleoside analog of thymidine that contains the base 5-bromouracil, which base-pairs with adenine. When DNA replication takes place in the presence of BUdR, it is efficiently incorporated into the newly synthesized strands in place of thymidine.

The interpretation of the chromosomes in Figure 6.4 is outlined in **Figure 6.5**. Part A

shows a chromosome that has undergone one round of replication in the presence of BUdR; each DNA molecule is half-labeled. Part B shows a chromosome after a second round of BUdR replication; in this case the DNA duplex in one chromatid is fully labeled, and that in the other is half-labeled. When these chromosomes are stained with a fluorescent dye (Hoechst 33258 is normally used for this purpose), the chromatid containing the half-labeled duplex fluoresces very brightly (light gray in Figure 6.4), whereas in the chromatid containing the fully labeled DNA, the fluorescence is "quenched" or suppressed (dark gray in Figure 6.4). Note that most of the chromosomes in Figure 6.4 resemble those in Figure 6.5C. This observation demonstrates that replication of the DNA in human chromosomes is semiconservative, because the half-labeling corresponds to that in the Meselson–Stahl experiment after one round of replication (Figure 6.2).

Figure 6.4 also includes a number of chromosomes with a labeling pattern like that in Figure 6.5D. These result from semi-

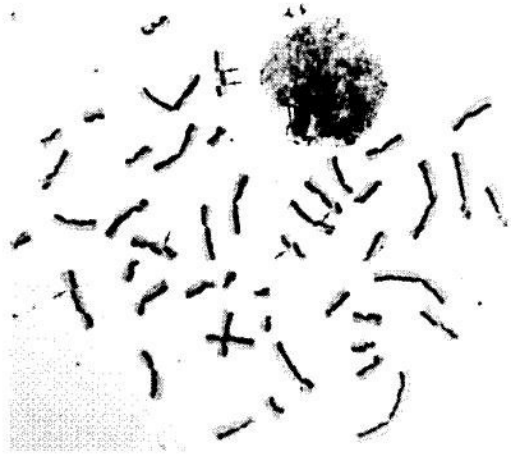


Figure 6.4 BUdR-labeled human chromosomes. [Courtesy of Ann P. Craig-Holmes.]

conservative replication followed by **sister-chromatid exchange.** In this process, the sister chromatids undergo a breakage and reunion, corresponding at the DNA level to breakage and rejoining of strands that have the same polarity in both chromatids. Because sister chromatids are identical in DNA sequence, sister-chromatid exchange normally has no genetic consequence. Molecular and cytological evidence for

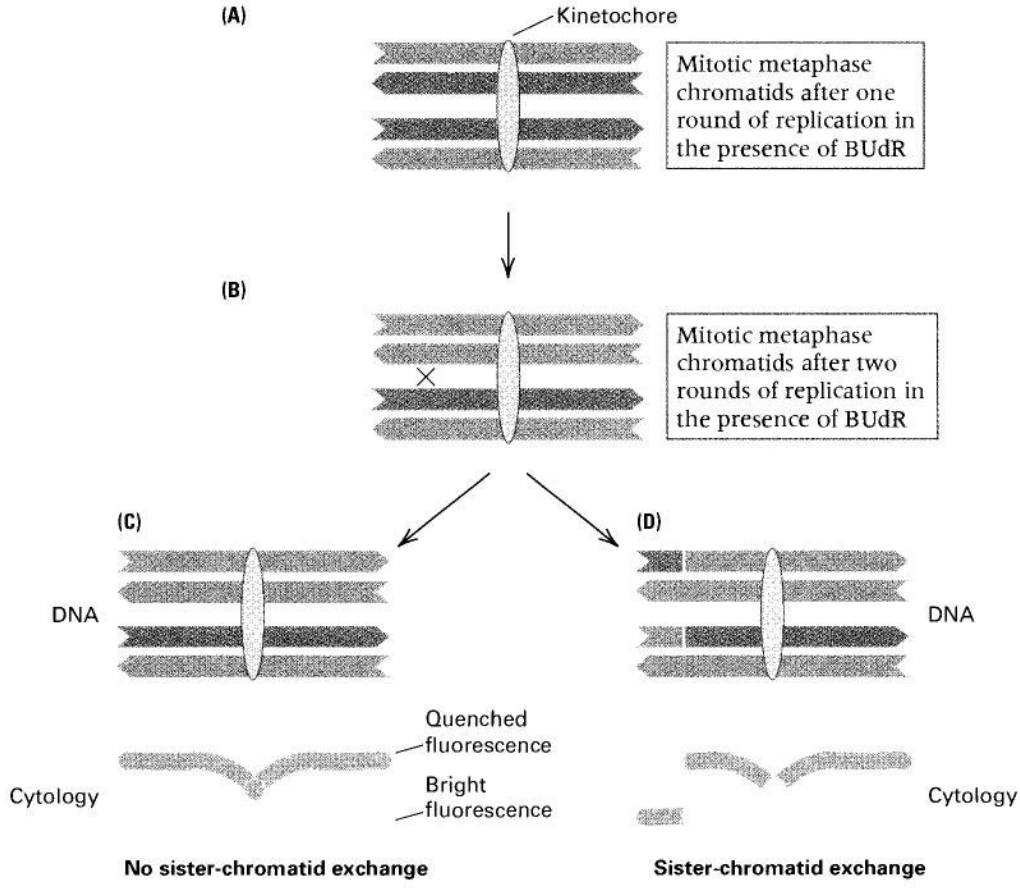


Figure 6.5 (A) One round of chromosome replication in the presence of BUdR results in half-labeled DNA duplexes. (B) Two rounds result in one fully labeled duplex and one half-labeled duplex. (C) Full labeling quenches (dulls) the fluorescence caused by Hoechst 33258 and certain other dyes, so that the half-labeled chromatid is fluorescent but the fully labeled chromatid is not. This finding confirms semiconsevative replication of chromosomal DNA. (D) A sister-chromatid exchange results in an abrupt switch in fluorescence from one chromatid to its partner.

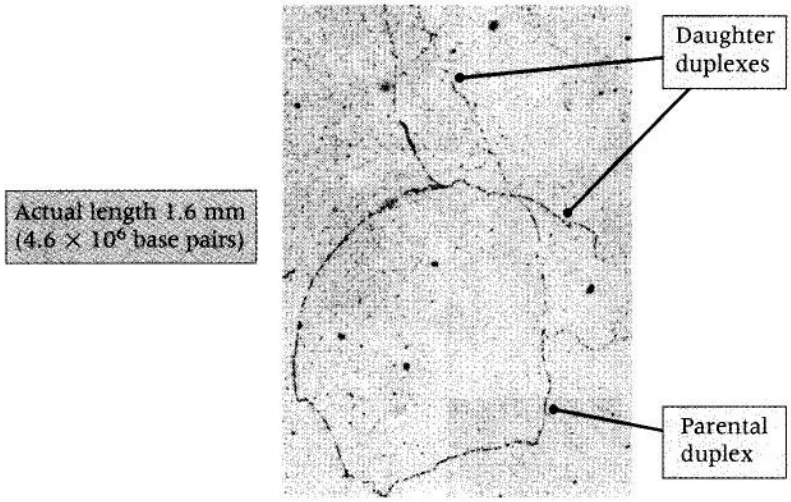


Figure 6.6 Autoradiogram of the intact replicating chromosome of an *E. coli* cell that has grown in a medium containing ^{3}H-thymine for slightly less than two generations. The continuous lines of dark grains were produced by electrons emitted by decaying ^{3}H atoms in the DNA molecule. The image is observed with a light microscope. [From J. Cairns, *Cold Spring Harbor Symp. Quant. Biol.* 1963. 28:44].

sister-chromatid exchange are illustrated in Figures 6.4 and 6.5D. With respect to the human cell in Figure 6.4, the culture had been exposed to x rays, which break chromatids. Hence this cell shows more sister-chromatid exchanges than would normally be found.

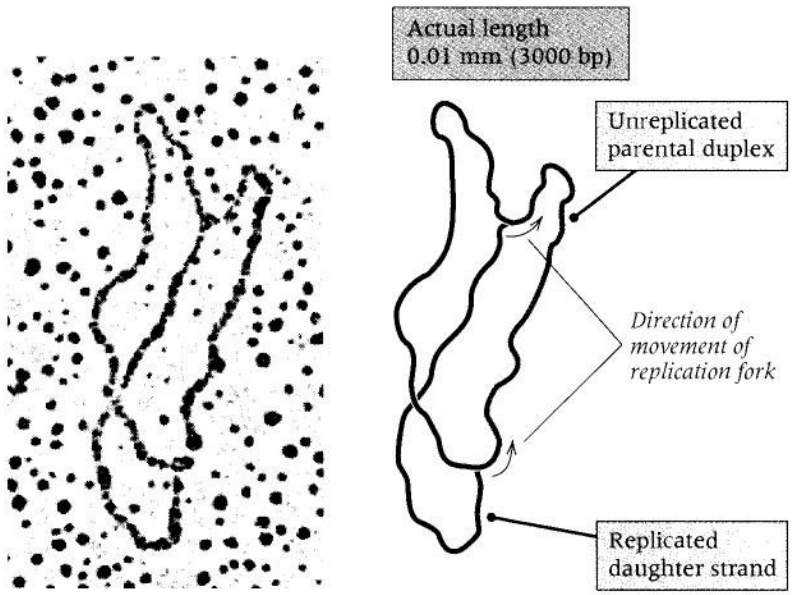


Figure 6.7 Electron micrograph of a small, circular DNA molecule replicating by the θ mode. The parental and daughter segments are shown in the drawing. [Electron micrograph courtesy of Donald Helinski.]

Theta Replication of Circular DNA Molecules

In the Meselson–Stahl experiment, the *E. coli* DNA was extensively sheared and fragmented during isolation, so the original form of the molecule was unknown. Genetic mapping experiments had suggested that the bacterial chromosome is circular (Chapter 10), but not until later was this demonstrated directly through the isolation of unbroken molecules and their examination by electron microscopy.

The first physical evidence that *E. coli* DNA replicates as a circle came from an autoradiographic experiment. Cells were grown in a medium containing radioactive thymine (^{3}H-thymine) so that all DNA synthesized would be radioactive. The DNA was isolated without fragmentation and placed on photographic film. Each radioactive decay caused a tiny black spot to appear in the film, and after several months there were enough spots so that the DNA could be visualized with a microscope; the pattern of black spots on the film traced the contour of the molecule. One of the now-famous autoradiograms from this experiment is shown in **Figure 6.6**. The actual length of the *E. coli* chromosome is 1.6 mm (4.6 million base pairs). On a smaller scale, **Figure 6.7** shows a replicating circular molecule of a plasmid found inside certain bacterial cells. The total length of the molecule is only 0.01 mm (3000 base pairs). The replicating circle is schematically like the Greek letter θ (theta), so this mode of replication is usually called **θ replication.**

In a replicating DNA molecule, the region in which parental strands are separating and new strands are being synthesized is called a **replication fork.** A circular DNA molecule replicating in the θ mode can have either one or two replication forks, as shown in **Figure 6.8**. When there is only one replication fork, it proceeds around the circle in a single direction from the origin of replication and comes back again to the origin. When there are two replication forks, they proceed in opposite directions from the origin and come together on the other side of the circle at a region known as the **terminus of replication.** In both circular and linear molecules, with only a few exceptions,

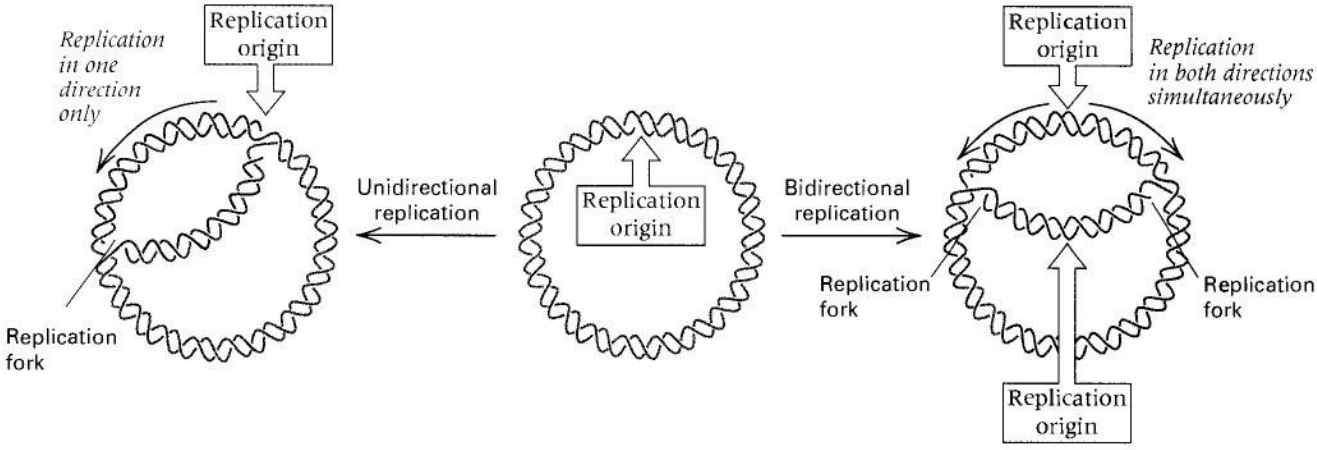


Figure 6.8 The distinction between unidirectional and bidirectional DNA replication. In unidirectional replication, there is only one replication fork; bidirectional replication requires two replication forks. The curved arrows indicate the direction of movement of the forks. Most DNA replicates bidirectionally.

when replication is initiated at an origin of replication, two replication forks are created that move outward from the origin. In other words, DNA nearly always undergoes **bidirectional replication.**

Rolling-Circle Replication

Some circular DNA molecules, including those of a number of bacterial and eukaryotic viruses, replicate by a process that does not include a θ-shaped intermediate. This replication mode is called **rolling-circle replication.** In this process, replication starts with a cut at a specific sugar–phosphate bond in a double-stranded circle (**Figure 6.9**). This cut produces two chemically distinct ends: a 3' end (at which the nucleotide has a free 3'-OH group) and a 5' end (at which the nucleotide has a free 5'-P group). The DNA is synthesized by the addition of deoxyribonucleotides to the 3' end with simultaneous displacement of the 5' end from the circle. As replication proceeds around the circle, the 5' end "rolls out" as a tail of increasing length.

In most cases, as the single-stranded tail is extended, a complementary chain is synthesized with the single strand as a template; this results in a double-stranded DNA

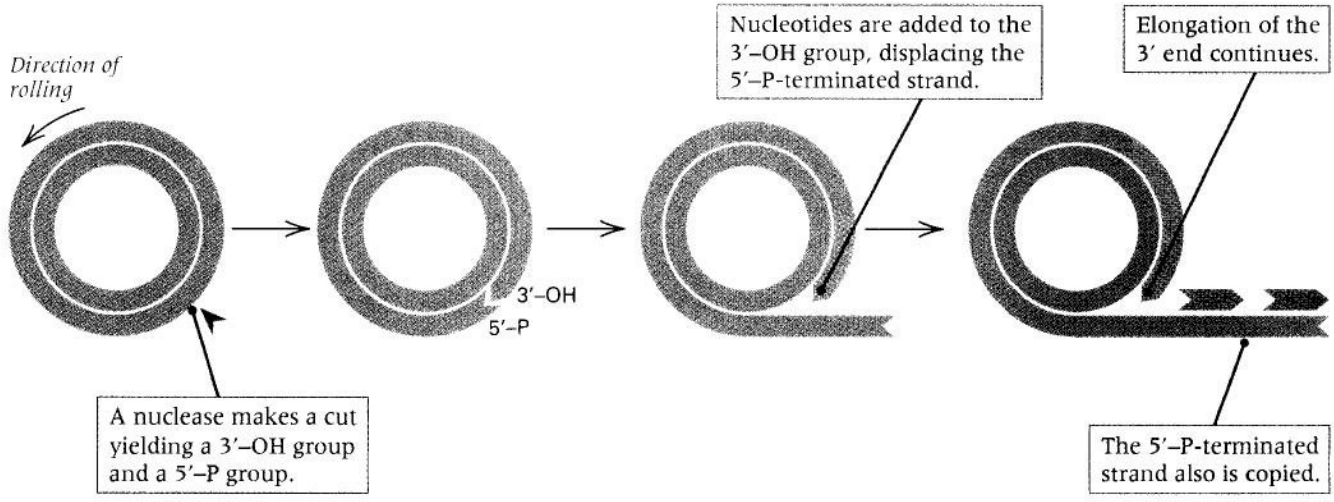


Figure 6.9 Rolling-circle replication. Newly synthesized DNA is shown in red. The displaced strand is replicated in short fragments.

tail. Because the displaced strand is covalently linked to the newly synthesized DNA in the circle, replication does not terminate, and extension proceeds without interruption, forming a tail that may be many times longer than the circumference of the circle. The long tail therefore consists of tandem linear repeats of the DNA sequence within the circle, and each complete replication around the circle adds another tandem repeat to the tail (**Figure 6.10**). Rolling-circle replication is a common feature in the late stages of replication of double-stranded DNA phages that have circular intermediates. An important example of rolling-circle replication will also be seen in Chapter 10, where matings between donor and recipient *E. coli* cells are described.

Multiple Origins and Bidirectional Replication in Eukaryotes

The DNA duplex in a eukaryotic chromosome is linear and replicates bidirectionally. Replication is initiated at many sites in the DNA. The structures that result from the numerous origins of replication are seen in electron micrographs as multiple loops along a DNA molecule (**Figure 6.11**). The loops are called *replication bubbles*. Each replication bubble originates at an origin of replication from which bidirectional replication takes place, and as the replication forks proceed outward from the origin, the bubble grows larger (**Figure 6.12**). The replication forks of adjacent bubbles fuse where they meet.

Multiple initiation is a means of reducing the total replication time of a large molecule. In eukaryotic cells, movement of each replication fork proceeds at a rate of approximately 10–100 nucleotide pairs per second. For example, in *D. melanogaster*, the rate of replication is about 50 nucleotide pairs per second at 25°C. Because the DNA molecule in the largest chromosome in *Drosophila* contains about 7×10^7 nucleotide pairs, replication from a single bidirectional origin of replication would take about 8 days. Developing *Drosophila* embryos actually use about 8500 replication origins per chromosome, which reduces the replication time to a few minutes. In a typical eukaryotic cell, origins are spaced about 40,000 nucleotide pairs apart, which allows each chromosome to replicate in 15–30 minutes. Because chromosomes do not all replicate simultaneously, complete replication of all chromosomes in eukaryotes usually takes 5–10 hours.

DNA replication is much faster in prokaryotes than in eukaryotes. In *E. coli*, for example, replication of a DNA molecule takes place at a rate of approximately 1500 nucleotide pairs per second; similar rates are found in the replication of all phage DNA molecules that have been studied. Complete replication of the *E. coli* DNA molecule from its single bidirectional origin of replication requires approximately 30 minutes because the molecule contains about 4.6×10^6 nucleotide pairs. Yet under optimal culture conditions, an *E. coli* cell can divide every 20 minutes. How is this possible? It is possible because a new round of replication is started at the origin every 20 minutes, even though the previous round of replication has not yet been completed. In a rapidly dividing culture of *E. coli* cells, the DNA molecule passed to each daughter cell in division is already in the process of replicating in preparation for the next division.

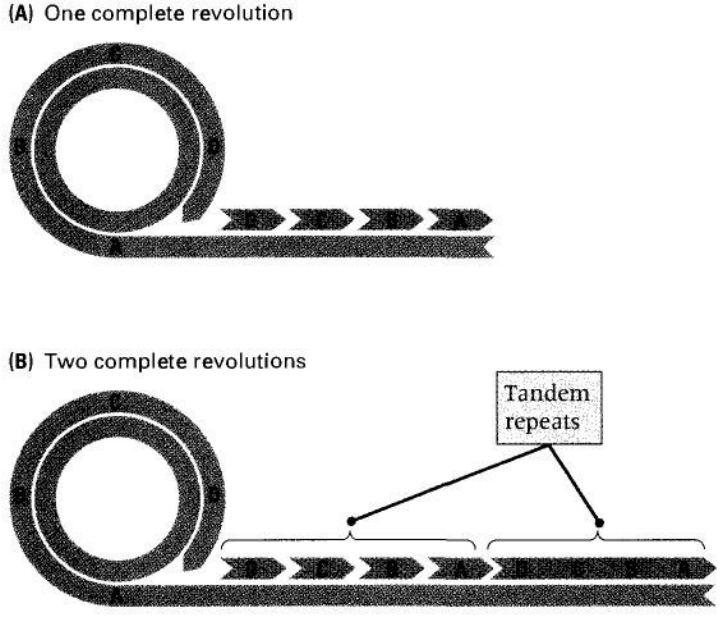


Figure 6.10 Continued rolling-circle replication results in a linear extension that contains tandem repeats of sequences found in the parental circular molecule.

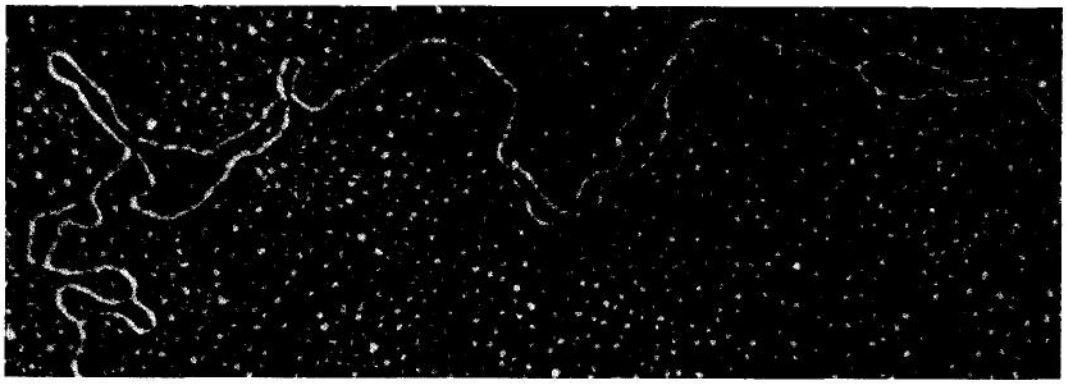


Figure 6.11 Replicating DNA of *Drosophila melanogaster.* The electron micrograph shows a 30-kb region containing five replication loops. [Electron micrograph courtesy of David Hogness.]

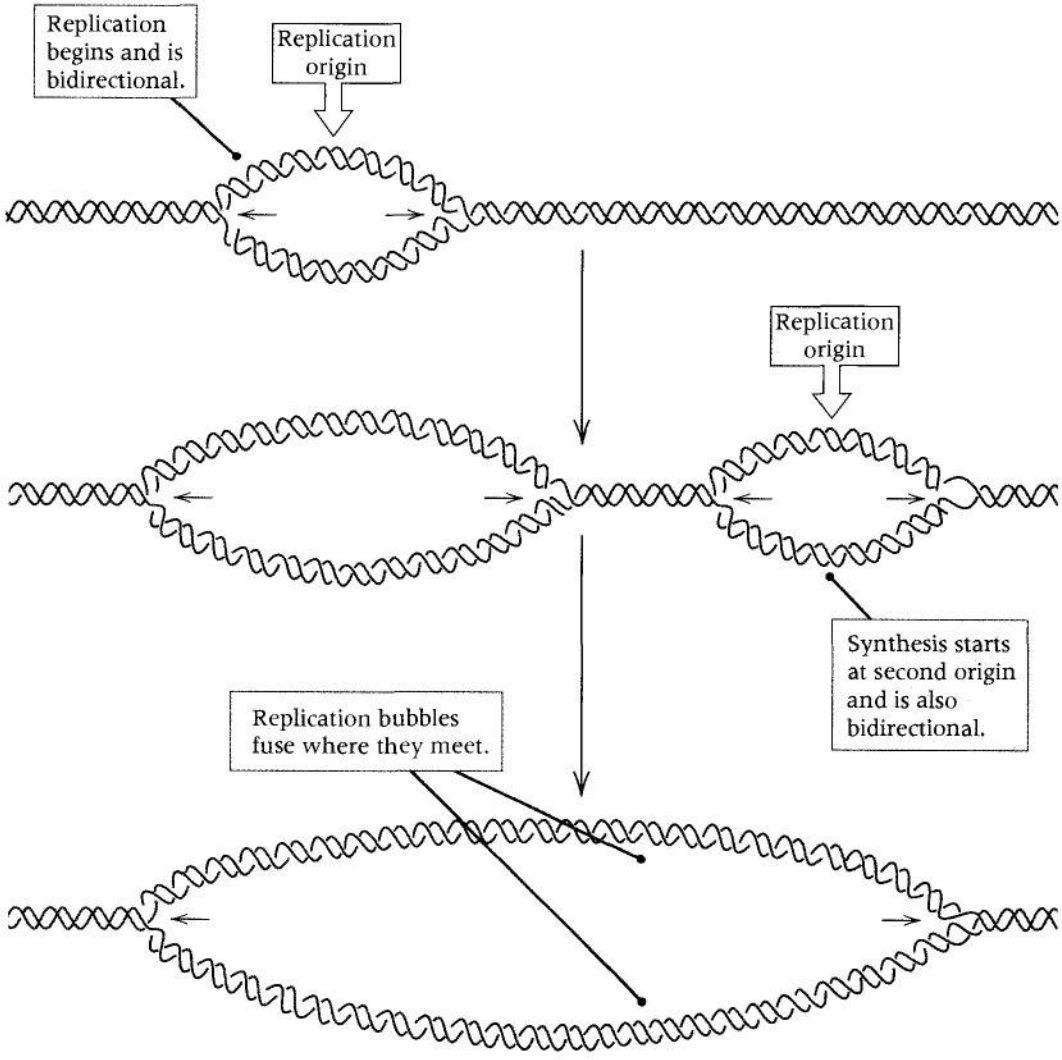


Figure 6.12 How loops enlarge and merge in bidirectional replication. The small arrows indicate the direction of movement of the replication forks.

6.3 Unwinding, Stabilization, and Stress Relief

Some of the main molecular players in DNA replication are illustrated in **Figure 6.13.** Unwinding the double helix to separate the parental strands requires a **helicase** protein that hydrolyzes ATP to drive the unwinding reaction. Most cells have several helicases specialized for different roles, such as replication, recombination, or repair. In *E. coli,* the principal helicase that acts at the replication fork is the product of the *dnaB* gene, which is denoted the DnaB helicase. The unwound strands of the double helix would tend to renature spontaneously, so they must be stabilized as single strands to serve as templates for DNA synthesis. This stabilization is a function of a **single-stranded DNA-binding protein (SSB).** The SSB binds single-stranded DNA tightly and cooperatively, and it has an affinity for single-stranded DNA at least 1000-fold greater than that for double-stranded DNA. It is this strong tendency for SSB to bind with single strands that stabilizes the templates for replication. In *E. coli,* apparently the same SSB is used in DNA replication, recombination, and repair.

Because the two strands of a replicating helix must make a full rotation to unwind each of the turns, some kind of swivel mechanism must exist to avoid the buildup of so much stress farther along the helix that strand separation would be brought to a halt. In *E. coli,* for example, only about 10 percent of the genome could be replicated before the torsional stress caused by unwinding became too great to continue. The swivel that relieves this stress is an enzyme called **gyrase,** which cleaves both strands of a DNA duplex, swivels the ends of the broken strands to relieve the torsional stress, and then rejoins the strands (**Figure 6.14**). Enzymes capable of catalyzing breakage and rejoining of DNA strands are known as *topoisomerases.* Gyrase is called a topoisomerase II because it makes a double-stranded break.

As the helix is being unwound by the helicase, the template strands stabilized by SSB, and the torsional stress relieved by the gyrase, the primer oligonucleotides are also synthesized in preparation for elongation of the new daughter strands. Primer synthesis is considered next.

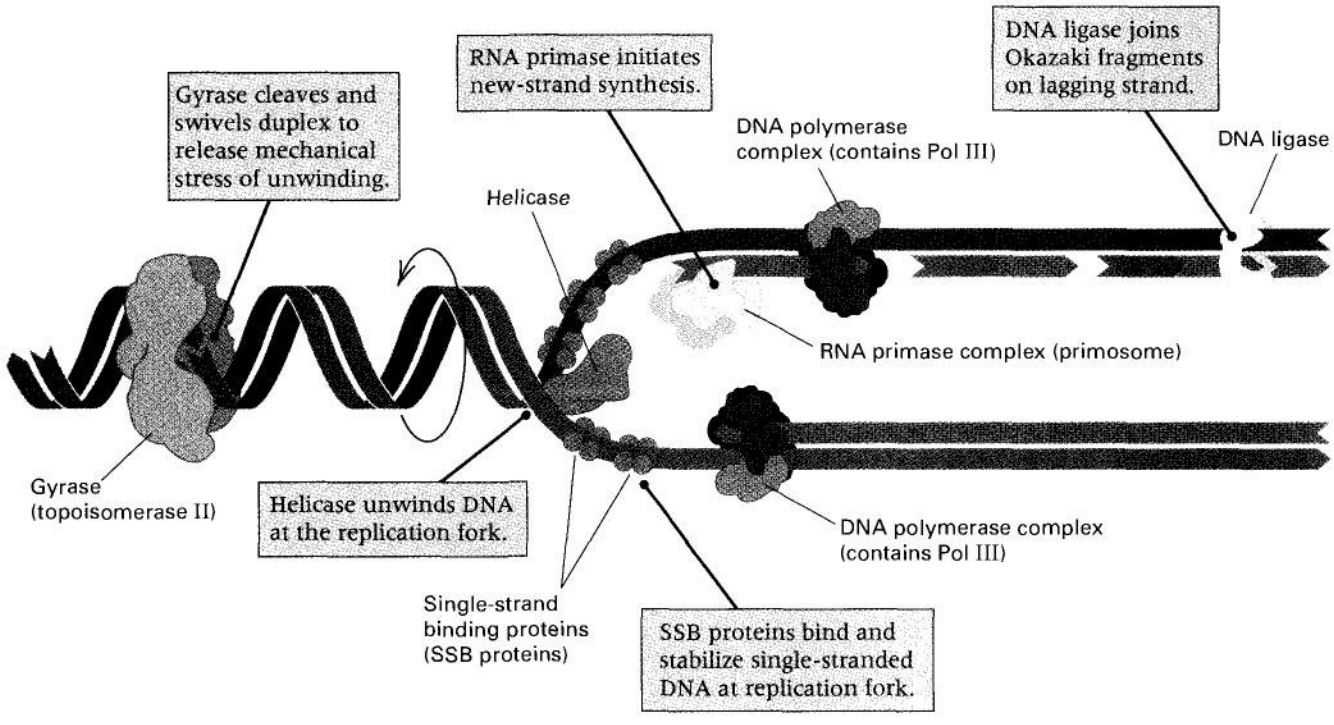


Figure 6.13 Role of some of the key proteins in DNA replication. The DNA polymerase III complex and the primase complex are both composed of multiple different polypeptide subunits. DNA polymerase I, which joins precursor fragments where they meet, is not shown.

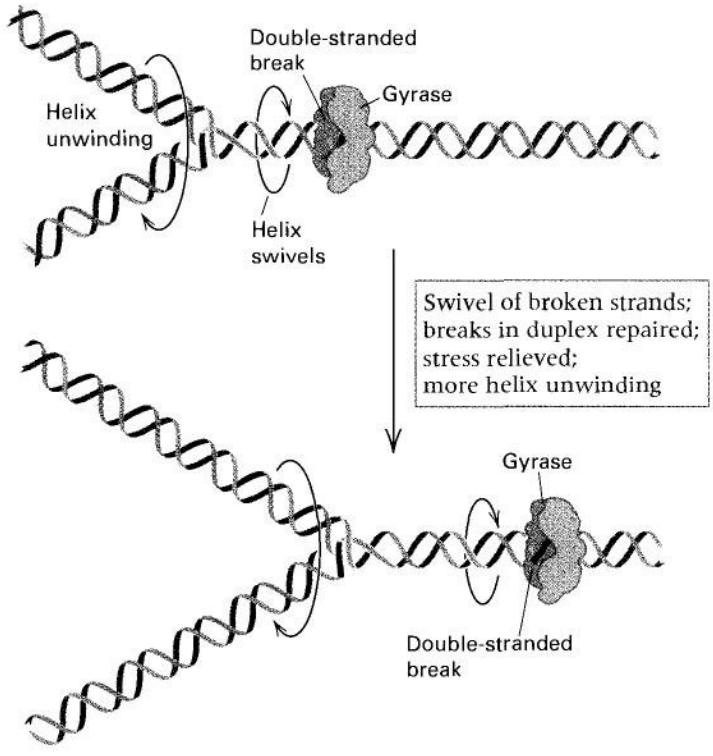


Figure 6.14 DNA gyrase introduces a double-stranded break ahead of the replication fork and swivels the cleaved ends around the central axis to relieve the stress of helix unwinding.

6.4 Initiation by a Primosome Complex

No known DNA polymerase is able to initiate new polynucleotide chains, so the presence of a primer chain with a free 3' hydroxyl group is absolutely essential for the initiation of replication. In living cells, the primer is a short segment of RNA; in cell-free replication *in vitro*, the primer may be either RNA or DNA. In most organisms, initiation is accomplished by a special type of RNA polymerase known as **primase.** In *E. coli*, the primase is DnaG, the product of the *dnaG* gene.

Like most other RNAs, the primer RNA is a single-stranded nucleic acid consisting of four types of nucleotides joined by 3'-to-5' phosphodiester bonds. The chemical differences that distinguish RNA from DNA are shown in **Figure 6.15.** The first difference is in the sugar component: RNA contains **ribose,** which is identical to the deoxyribose of DNA except for the presence of an −OH group on the 2' carbon atom. The second difference is in one of the four bases: The thymine found in DNA is replaced by the closely related pyrimidine *uracil* (*U*) in RNA. In RNA synthesis, a DNA strand is used as a template to form a complementary strand in which the bases in the DNA are paired with those in the RNA. The primase and other RNA polymerases differ from DNA polymerases in that they can initiate the synthesis of RNA chains without a primer.

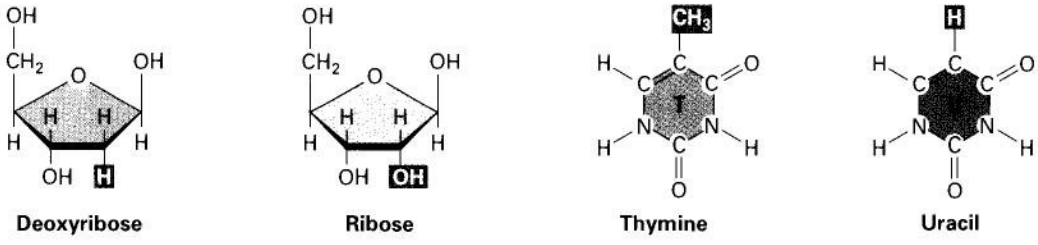


Figure 6.15 Differences between DNA and RNA. The chemical groups that are highlighted are the distinguishing features of deoxyribose and ribose and of thymine and uracil.

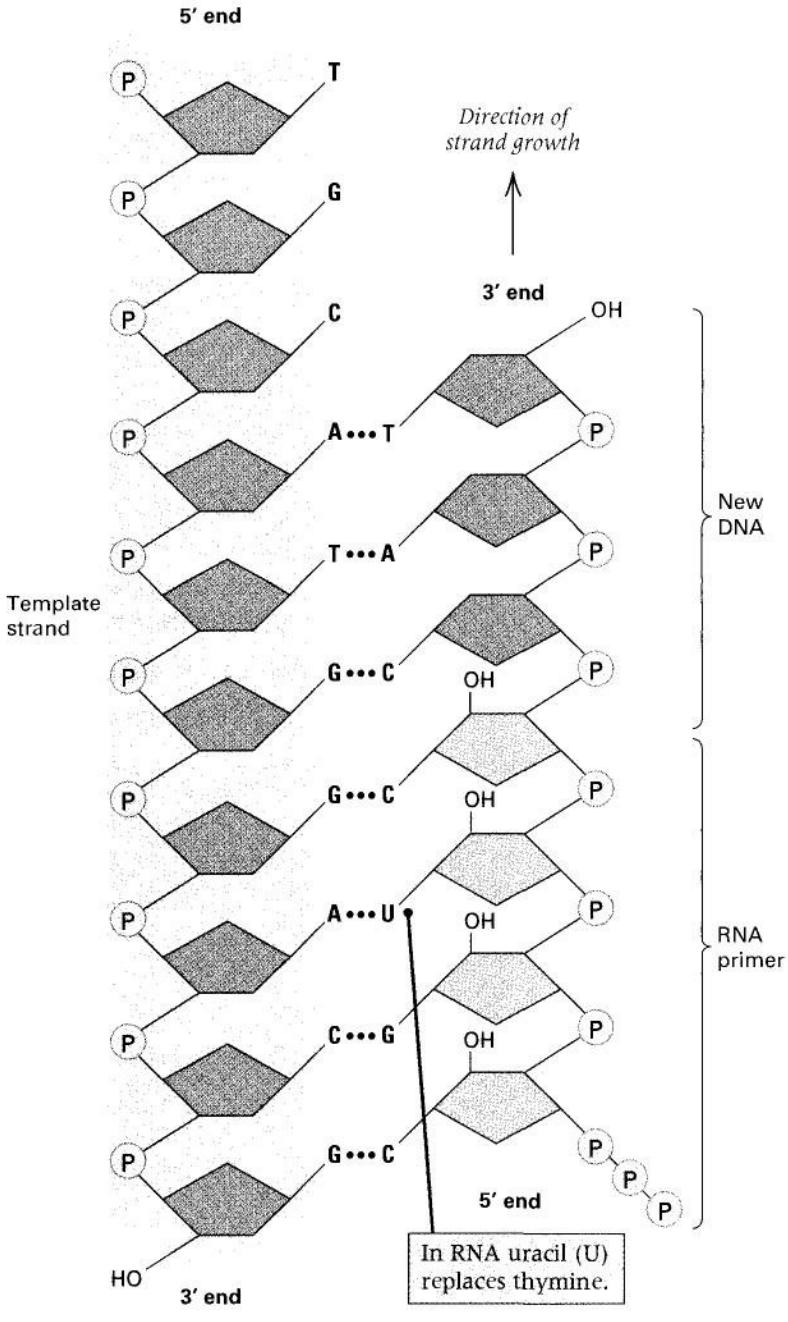


Figure 6.16 (above) Priming of DNA synthesis with an RNA segment. The DNA template strand is shown in blue; newly synthesized DNA in beige. The RNA segment, made by an RNA polymerase (primase), is shown in green. In the first step, DNA polymerase adds a deoxyribonucleotide to the 3' end of the primer. Successive deoxyribonucleotides are added to the 3' end of the growing chain.

The primer is synthesized with a short stretch of the unwound, single-stranded DNA as the template. The length of the primer differs according to the initiation event. In *E. coli*, the length is typically 2–5 nucleotides; in eukaryotic cells, it is generally 5–8 nucleotides. This short stretch of RNA provides the substrate onto which a DNA polymerase can add deoxynucleotides (**Figure 6.16**). The primase is usually found in a multienzyme complex composed of 15 to 20 polypeptide chains that is called a **primosome.** While it is being synthesized, each daughter strand has the molecular structure shown in **Figure 6.17**, with the primer at the 5' end and elongation taking place at the 3' end.

6.5 Chain Elongation and Proofreading

The elongation reaction catalyzed by DNA polymerase is the formation of a phosphodiester bond between the free 3'-OH group of the chain being extended and the innermost phosphorus atom of the nucleoside triphosphate being incorporated at the 3' end (**Figure 6.18**). The result is as follows:

> DNA synthesis proceeds by the elongation of primer chains, *always in the 5'-to-3' direction.*

Recognition of the appropriate incoming nucleoside triphosphate in replication depends on base pairing with the opposite nucleotide in the template chain. DNA polymerase will usually catalyze the polymerization reaction that incorporates the new nucleotide at the primer terminus only when the correct base pair is present, but errors of incorporation nevertheless occur.

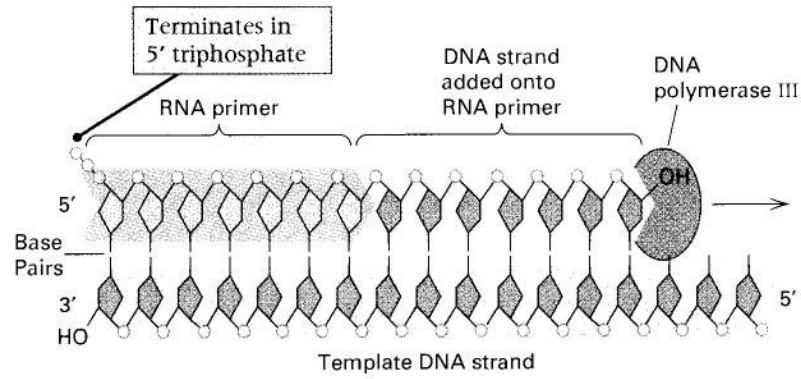


Figure 6.17 (at right) Prior to their being joined, each precursor fragment in the lagging strand has the structure shown here. The short RNA primer is shown in green.

Two DNA polymerases are needed for DNA replication in *E. coli*. The major replication enzyme is **DNA polymerase III (Pol III)**. In the cell, Pol III exists in a large complex with a molecular mass of about 1×10^6 daltons that is called the **Pol III holoenzyme**. The holoenzyme contains two Pol III proteins (each composed of three different polypeptide chains), plus at least seven other polypeptides that function in assembly of the holoenzyme and in the formation of a molecular clamp that holds and moves along the template strand. Although each cell has only 10–20 copies of the Pol III holoenzyme, it is responsible for the addition of the first deoxyribonucleotide to the

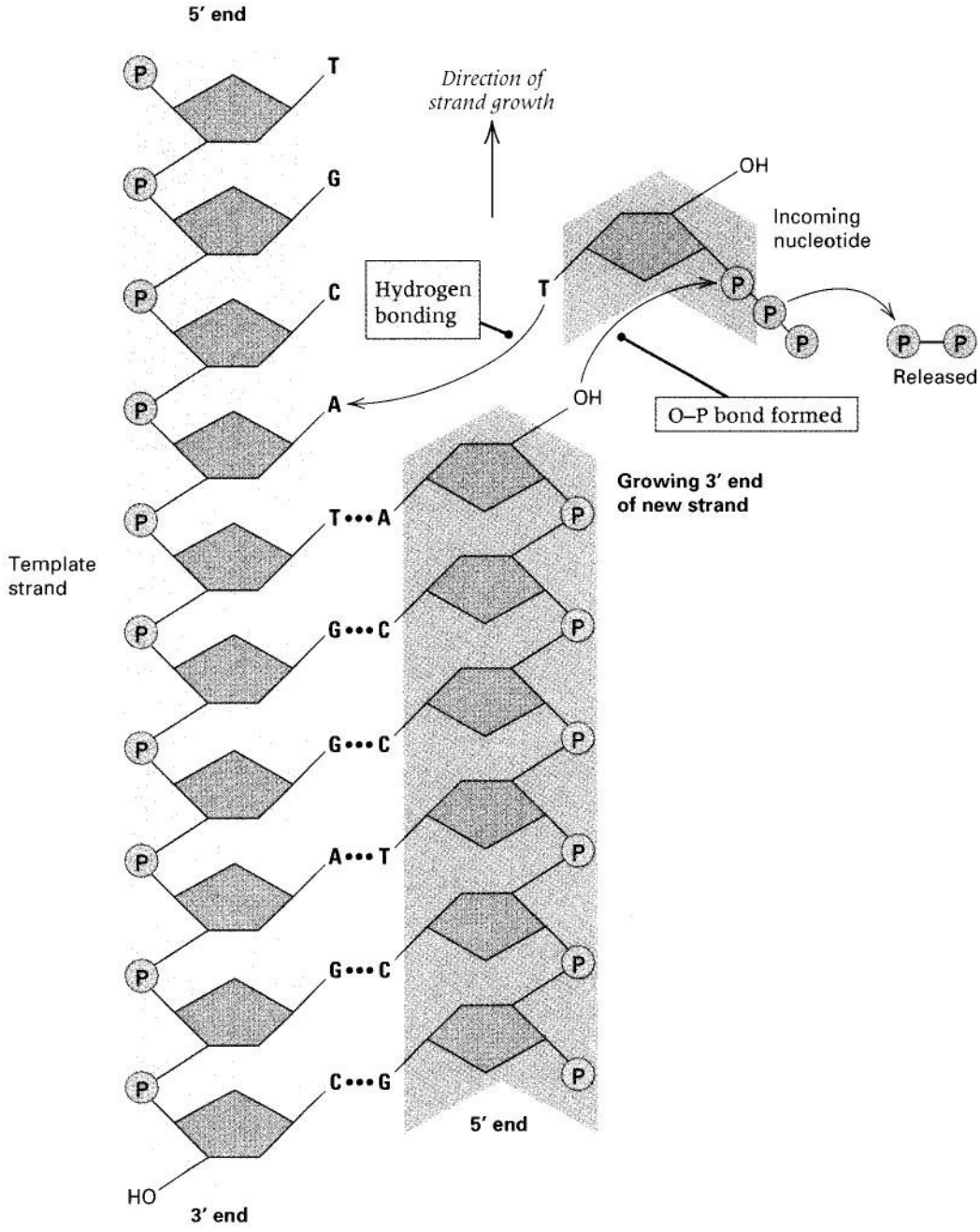


Figure 6.18 Addition of a nucleotide to the 3'-OH terminus of a primer strand. The recognition step is shown as the formation of hydrogen bonds between the A and the T. In the polymerization reaction, the 3'-OH group of the 3' end of the growing chain attacks the innermost phosphate group of the incoming trinucleotide.

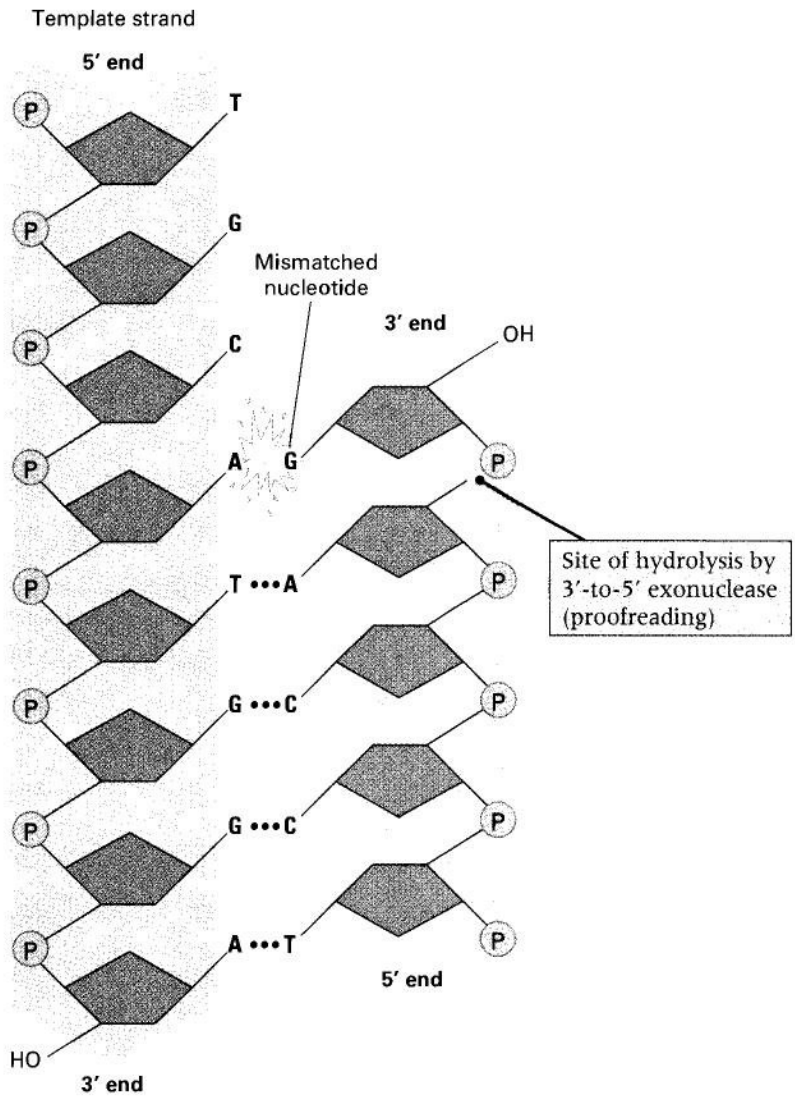


Figure 6.19 The 3'-to-5' exonuclease activity of the proofreading function. The growing strand is cleaved to release a mispaired nucleotide, in this case containing the base G, which does not pair with the base A in the template strand.

RNA primer, as well as for subsequent elongation of the daughter strand. The second polymerase is **DNA polymerase I (Pol I)**, which plays an essential role in removing the RNA primer. This activity will be described in a later section. Eukaryotic cells also contain several DNA polymerases. The major enzymes for chromosomal DNA replication are *polymerase* α (alpha) and *polymerase* δ (delta). Mitochondrial DNA is replicated by a special mitochondrial enzyme, *polymerase* γ (gamma).

In addition to their ability to polymerize nucleotides, most DNA polymerases also have an *exonuclease* activity that breaks a phosphodiester bond in the sugar–phosphate backbone of a nucleic acid chain and removes the terminal nucleotide from the 3' end. This kind of exonuclease is called a 3'-to-5' exonuclease. This exonuclease activity is extremely important genetically, because it provides a built-in mechanism for correcting incorporation errors that happen during polymerization. At a frequency of about one nucleotide per 10^5 replicated nucleotides, the polymerase incorporates an incorrect nucleotide onto the end of the growing daughter strand, where by "incorrect" nucleotide we mean one that cannot form a proper Watson–Crick base pair with the base in the template strand. The presence of an unpaired base activates the *3'-to-5' exonuclease activity*, which cleaves the unpaired nucleotide from the 3'-OH end of the growing chain (**Figure 6.19**). Because it cleaves off the incorrect nucleotide and gives the polymerase a second chance to get it right, the 3'-to-5' exonuclease activity of DNA polymerase is the *proofreading function*. The proofreading function can "look back" only one base (the one added last). Nevertheless,

> The genetic significance of the proofreading function is that it is an error-correcting mechanism that serves to reduce the frequency of mutation resulting from the incorporation of incorrect nucleotides in DNA replication.

Because DNA polymerase can elongate a newly synthesized DNA strand only at its 3' end, one of the daughter strands is made in short fragments, which are then joined together. The reason for this mechanism and the properties of these fragments are described next.

6.6 Discontinuous Replication of the Lagging Strand

Because DNA strands can be elongated only at the 3' end, the daughter strand that has its 5' end nearest the replication fork cannot be elongated in the same direction in which the replication fork is moving. This strand is the *lagging strand*. As shown in Figure 6.13, the lagging strand is replicated in short pieces that are later attached together to make a continuous daughter strand.

Fragments in the Replication Fork

Figure 6.20 illustrates the replication fork and the leading and lagging strands during DNA replication. The short fragments in which the lagging strand is synthesized are known as **precursor fragments** or **Okazaki fragments,** after their discoverer. The size of the precursor fragments is from 1000 to 2000 base pairs in prokaryotic cells and from 100 to 200 base pairs in eukaryotic cells. Because synthesis of the lagging strand is initiated only at intervals, at least one single-stranded region of the parental duplex is always present on one side of the replication fork (Figure 6.20). Such single-stranded regions can be seen in high-resolution electron micrographs of replicating DNA molecules, as indicated by the arrows in **Figure 6.21** in part A. Another implication of discontinuous replication of one strand, also apparent in Figure 6.20, is that the 3' end of the continuously replicated strand must always be "ahead of" (closer to) the replication fork than the 5' end of the discontinuously replicated strand. This difference is the physical basis of the terms *leading strand* and *lagging strand* that are used to distinguish the replicating strands.

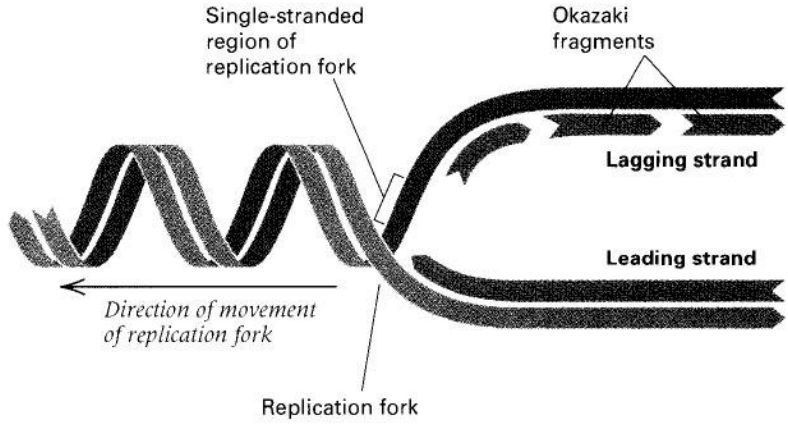


Figure 6.20 Short fragments in the replication fork. For each tract of base pairs, the lagging strand is synthesized later than the leading strand.

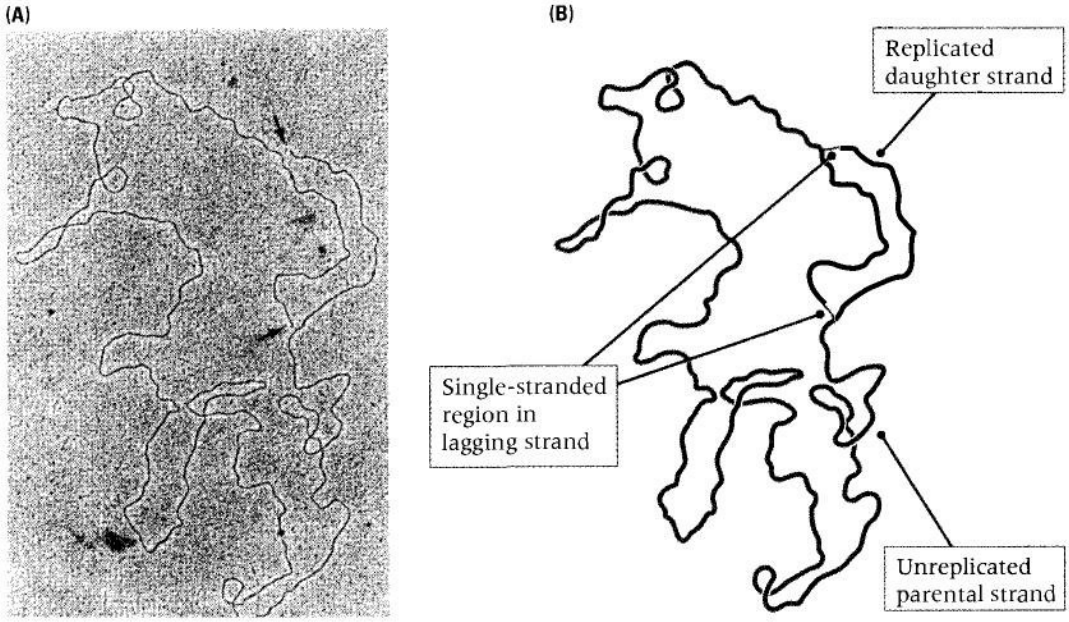


Figure 6.21 Electron micrograph (A) and an interpretive drawing (B) of a replicating θ molecule of phage λ DNA. In the electron micrograph, each arrow points to a short, single-stranded region of unreplicated DNA near the replication fork. The single-stranded regions are somewhat difficult to see, but they are apparent on close inspection. [Electron micrograph courtesy of Manuel Valenzuela.]



Separation of single-stranded and double-stranded DNA fragments by gel electrophoresis is one of the methods used to study the molecular biology of DNA replication. [Simon Fraser/Science Photo Library/Photo Researchers, Inc.]

The Joining of Precursor Fragments

The precursor fragments are ultimately joined to yield a continuous strand of DNA. This strand contains no RNA sequences, so the final stitching together of the lagging strand requires

- Removal of the RNA primer
- Replacement of the primer with a DNA sequence
- Joining where adjacent DNA fragments come into contact

Primer removal and replacement are accomplished in *E. coli* by DNA polymerase I (Pol I). Both prokaryotes and eukaryotes carry out the joining reaction by means of the enzyme **DNA ligase,** which can covalently join adjacent 3'-OH and 5'-P groups at a nick.

Details of how the primer is removed and replaced are shown in **Figure 6.22**. Pol III extends the growing strand until the RNA of the primer of the previously synthesized precursor fragment is reached. Where the DNA and RNA segments meet, there is a single-strand interruption, the equivalent of a nick in a continuous strand. The *E. coli* DNA ligase cannot seal the nick because a triphosphate is present (it can ligate only a 3'-OH and a 5'-*mono*phosphate). Here is where DNA polymerase I takes over. This enzyme has an exonuclease activity that can remove nucleotides from the 5' end of a base-paired fragment. It is effective with both DNA and RNA. This activity is a *5'-to-3' exonuclease activity.* Pol I acts at the nick and displaces it in the 5'-to-3' direction by removing RNA nucleotides, one by one, and successively adding DNA nucleotides to the 3' end of the DNA strand. When all of the RNA nucleotides have been removed, DNA ligase covalently joins the 3'-OH group to the terminal 5'-P of the precursor fragment. By this sequence of events, the precursor fragment is assimilated into the lagging strand. When the next precursor fragment reaches the RNA primer of the fragment just joined, the process of exonuclease degradation and resynthesis begins again. Polymerase I is essential for DNA replication, because the 5'-to-3' exonuclease activity required for removing the RNA primer and joining the precursor fragments is not present in Pol III.

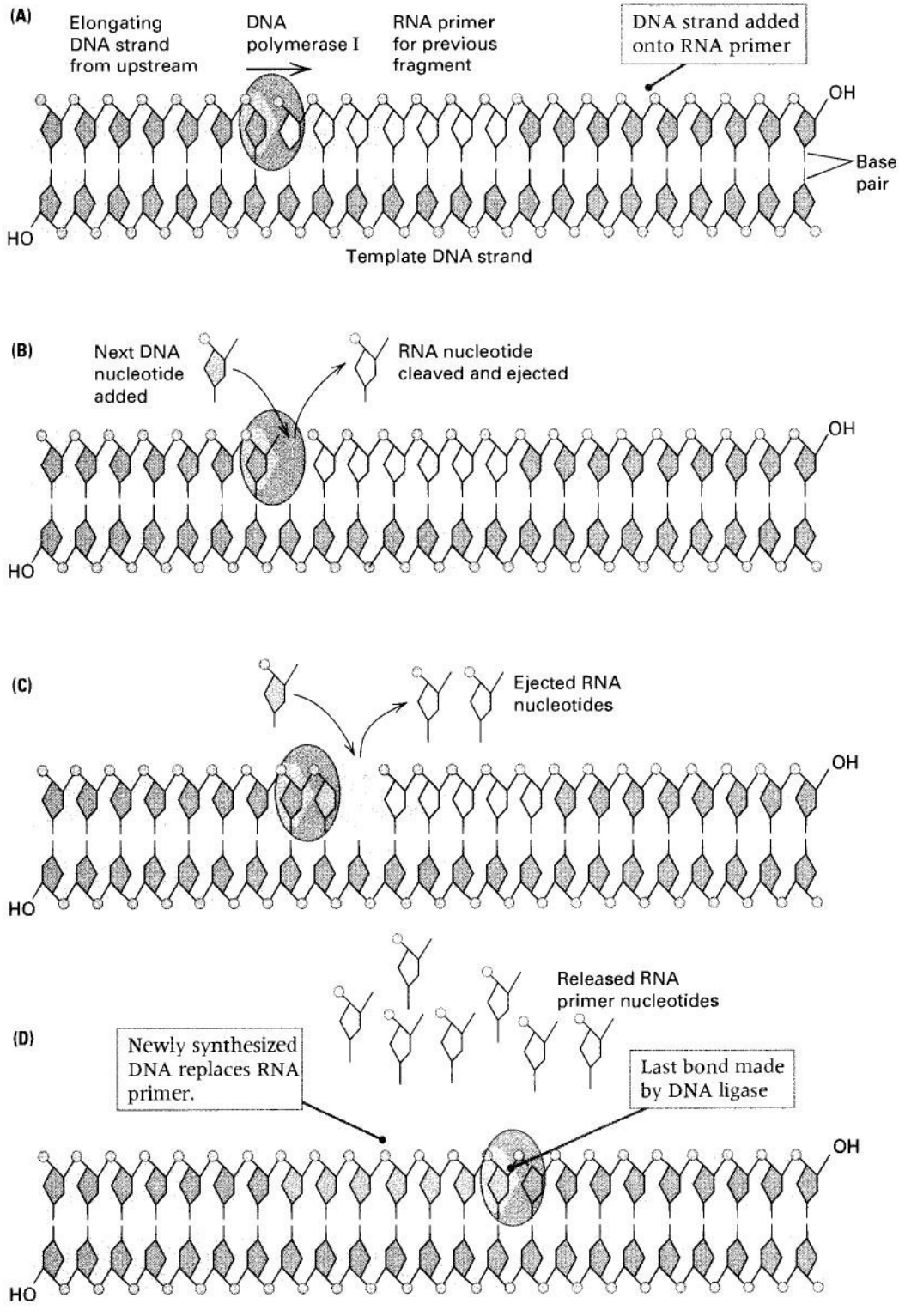


Figure 6.22 Sequence of events in the joining of adjacent precursor fragments. (A) The DNA polymerase of the upstream fragment meets the RNA primer from the next precursor fragment downstream. (B) The RNA nucleotide immediately adjacent is excised and replaced with a DNA nucleotide. (C) Each RNA nucleotide is cleaved and replaced in turn. (D) When all the RNA nucleotides have been replaced, the adjacent ends of the DNA strand are joined.

6.7 Terminator Sequencing of DNA

As we have seen, DNA polymerase elongates a polynucleotide chain through a reaction in which the 3' hydroxyl on the deoxyribose sugar of the growing chain attacks the 5' phosphate of the incoming nucleotide, forming a new phosphodiester bond and adding a new nucleotide to the chain. What would happen if an incoming nucleotide lacked a 3' hydroxyl group? It could still be attached to the growing chain by means of its 5' phosphate, but then, owing to the absence of a 3' hydroxyl group, further elongation would be stopped dead in its tracks. The ability of nucleotides lacking a 3'-OH to terminate strand elongation is a principle that allows the determination of the nucleotide sequence in a DNA strand.

No technique can determine the sequence of bases in an entire chromosome in a single experiment, so chromosomes are first cut into fragments a few hundred base pairs long, a size that can be sequenced easily. To obtain the sequence of a long stretch of DNA, a set of overlapping fragments is prepared, the sequence of each is determined, and all sequences are then combined. The procedures are straightforward, and DNA sequences have accumulated at such a rapid rate that large computer databases are necessary to manage the billions of nucleotides of DNA sequence already determined for a variety of genes from many organisms. A still larger database has been required to handle the DNA in human sperm or eggs, which consists of 3×10^9 nucleotide pairs. We will first describe the manual method of DNA sequencing still in use in many research laboratories; then we will explain how most DNA-sequencing machines work.

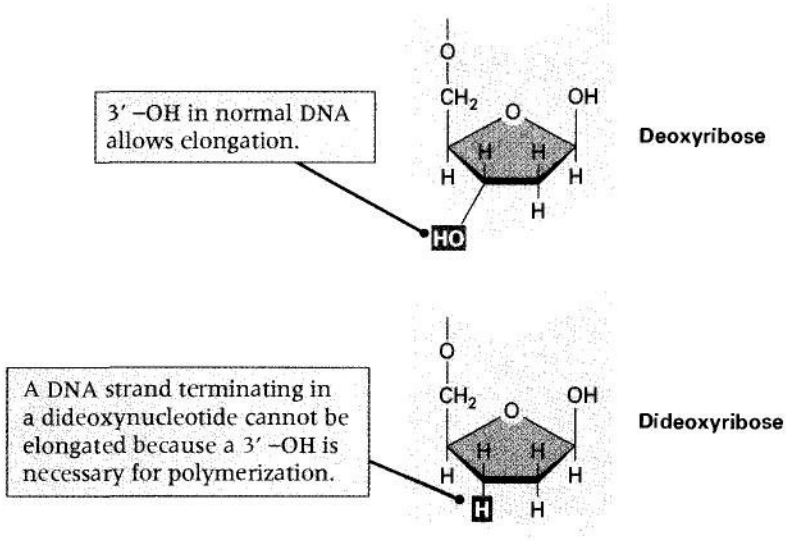


Figure 6.23 Structures of normal deoxyribose and the dideoxyribose sugar used in DNA sequencing. The dideoxyribose has a hydrogen atom (red) attached to the 3' carbon, in contrast with the hydroxyl group (red) at this position in deoxyribose. Because the 3' hydroxyl group is essential for the attachment of the next nucleotide in line in a growing DNA strand, the incorporation of a dideoxynucleotide immediately terminates synthesis.

The **dideoxy sequencing method** employs DNA synthesis in the presence of small amounts of nucleotides that contain the sugar **dideoxyribose** instead of deoxyribose (**Figure 6.23**). Dideoxyribose lacks the 3'-OH group, which is essential for attachment of the next nucleotide in a growing DNA strand, so incorporation of a dideoxynucleotide instead of a deoxynucleotide immediately terminates further elongation of the strand. To sequence a DNA strand, four DNA synthesis reactions are carried out. Each reaction contains the single-stranded DNA template to be sequenced, a single oligonucleotide primer complementary to a stretch of the template strand, all four deoxyribonucleoside triphosphates, and a small amount of *one* of the nucleoside triphosphates in the dideoxy form. Each reaction produces a set of fragments that terminates at the point at which a dideoxynucleotide was randomly incorporated in place of the normal deoxynucleotide. Therefore, in each of the four reactions, the lengths of the fragments are determined by the positions in the daughter strand at which the particular dideoxynucleotide present in that reaction was incorporated. The sizes of the fragments produced by chain termination are determined by gel electrophoresis, and the base sequence is then determined by the following rule:

> If a fragment containing n nucleotides is generated in the reaction containing a particular dideoxynucleotide, then position n in the *daughter strand* is occupied by the base present in the dideoxynucleotide. The numbering is from the 5' nucleotide of the primer.

For example, if a 93-base fragment is present in the reaction containing the dideoxy form of dATP, then the 93rd base in the daughter strand produced by DNA synthesis must be an adenine (A). Because most native duplex DNA molecules consist of complementary strands, it does not matter whether the sequence of the template strand or that of the daughter strand is determined. The sequence of the template strand can be deduced from the daughter strand because their nucleotide sequences are complementary. However, in practice, both strands of a molecule are usually sequenced independently and compared in order to eliminate most of the errors that can be made by misreading the gels.

With short, single-stranded fragments, molecules that differ in size by a single base can be separated by electrophoresis in an acrylamide gel. For example, if a reaction mixture contains a set of DNA fragments consisting of 20, 21, 22, . . . , 100 nucleotides, then electrophoresis will yield separate bands—one containing the fragment of size 20, the next the fragment of size 21, and so forth. This extraordinary sensitivity to size is the basis of the DNA-sequencing procedure.

The Sequencing Procedure

The manual procedure for dideoxy sequencing of a DNA fragment is diagrammed in **Figure 6.24**. In this example, the primer length is 20 nucleotides. Termination at G produces fragments of 21, 24, 27, and 29 nucleotides; termination at A produces fragments of 22 and 30 nucleotides; termination at T produces a fragment of 26 nucleotides; and termination at C produces fragments of 23, 25, and 28 nucleotides. Typically, either the primer or one of the deoxynucleotides used contains a radioactive atom, so after electrophoresis to separate the fragments by size, the fragments can be visualized as bands on photographic film placed over the gel. After electrophoresis and autoradiography, the positions of the bases in the daughter strand can be read directly from the gel, starting with position 21, as 5'-GACGCTGCGA-3'. This corresponds to the template strand sequence 3'-CTGCGACGCT-5'.

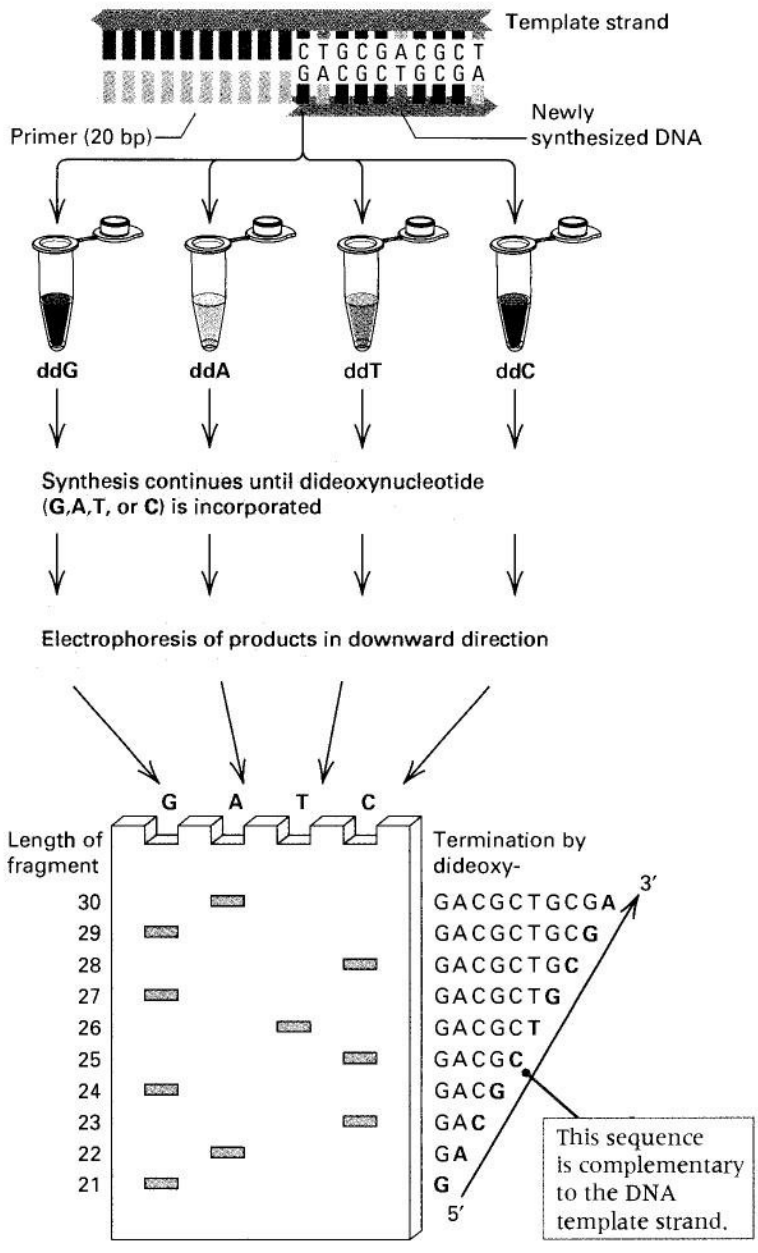


Figure 6.24 Dideoxy method of DNA sequencing. Four DNA synthesis reactions are carried out in the presence of all normal nucleoside triphosphates plus a small amount of *one* of the dideoxynucleoside triphosphates containing G, A, T, or C. Synthesis continues along the template strand until a dideoxynucleotide is incorporated. The products that result from termination at each dideoxynucleotide are indicated at the right. The fragments are separated by size via electrophoresis, and the positions of the nucleotides are determined directly from the gel. In this example, the length of the primer needed to initiate DNA synthesis is 20 nucleotides. Lengths of the DNA fragments are shown at the left of the gel. The sequence of the daughter strand is read from the bottom of the gel as 5'-GACGCTGCGA-3'.

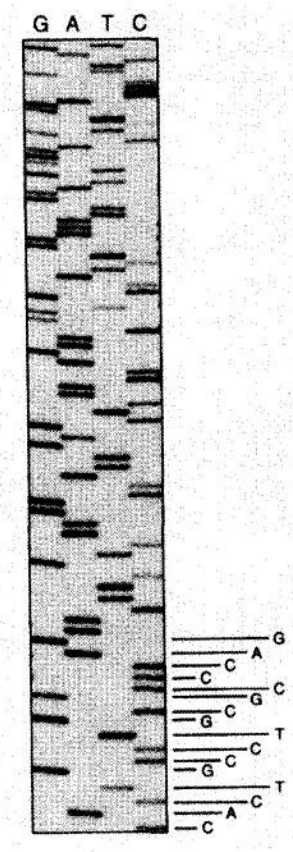


Figure 6.25 A section of a dideoxy sequencing gel. The sequence is read from the bottom to the top. Each horizontal row represents a single nucleotide position in the DNA strand synthesized from the template. The vertical columns result from termination by the dideoxy forms of G, A, T, or C. The sequence from the lower part of the gel is indicated.

Figure 6.25 is an autoradiogram of an actual sequencing gel. The shortest fragments are those that move the fastest and farthest. Each fragment contains the primer fragment at the 5' end of the daughter strand. The sequence can be read directly from the bottom to the top of the gel. The sequence of the first 16 bases in the segment in this gel, read from bottom to top, is

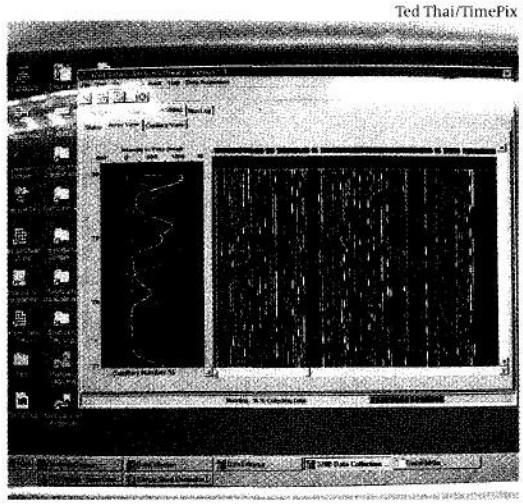
5'-CACTGCCTGCGCCCAG-3'

and so forth.

Shotgun Sequencing

Most DNA sequencing is carried out with machines that carry out electrophoresis and read the sequence automatically. The principle behind automated DNA sequencing is shown in **Figure 6.26**. On the left is a conventional sequencing gel obtained from the manual dideoxy procedure. On the right is another gel in which the dideoxynucleotides that terminate synthesis are each labeled with a different fluorescent dye (G, black; A, green; T, red; C, blue). Because the colors distinguish the products of DNA synthesis that terminate with each dideoxynucleotide, the products of all the synthesis reactions can be mixed together in the same tube and separated by electrophoresis in a single lane. In principle, the sequence could be read manually from the gel, as shown by the nucleotide sequence at the left of Figure 6.26B. However, automated sequencing machines work by continuing the electrophoresis until each band, in turn, drops off the bottom of the gel. As each band comes off the bottom of the gel, the fluorescent dye that it contains is excited by laser light, and the color of the fluorescence is read automatically by a photocell and recorded in a computer. Figure 6.26C is a trace of the fluorescence pattern that would emerge at the bottom of the gel in Figure 6.26B after continued electrophoresis. The nucleotide sequence is read directly from the colors of the alternating peaks along the trace. When used to maximum capacity, the fastest automated sequencing instruments can generate more than 100 Mb of nucleotide sequence per year. The actual amount of finished sequence is considerably smaller, because in a sequencing project, each long region of

Ted Thai/TimePix



A computer automatically keeps track of the DNA fragments produced in automated DNA sequencing as the fragments come off the bottom of the sequencing gel and are scanned by a laser to determine which fluorescent label they contain. Each color represents a different nucleotide incorporated into the growing DNA strand, and hence an individual nucleotide in the DNA sequence of interest.

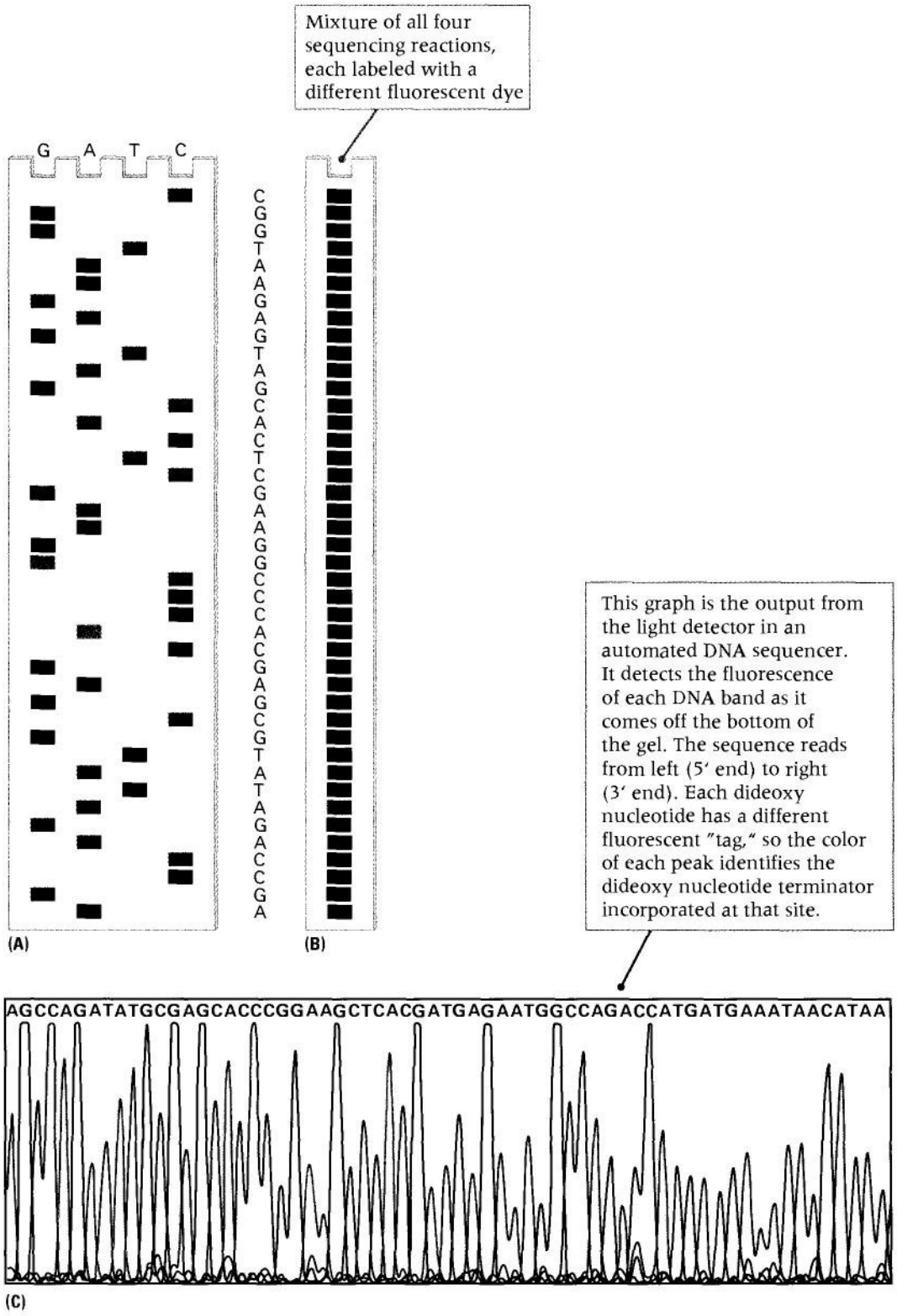


Figure 6.26 Automated DNA sequencing. (A) Conventional sequencing gel obtained from the dideoxy procedure; the DNA sequence can be determined directly from photographic film according to the positions of the bands. (B) Banding pattern obtained when each of the terminating nucleotides is labeled with a different fluorescent dye and the bands are separated in the same lane of the gel. (C) Trace of the fluorescence pattern obtained from the gel in part B by automated detection of the fluorescence of each band as it comes off the bottom of the gel during continued electrophoresis.

duplex DNA needs to be sequenced approximately 10 times for the sake of assembling the data into a continuous whole and minimizing sequencing errors.

A high redundancy in DNA sequencing is required, because the sequence of the target molecule is obtained in the form of many small single-stranded fragments cloned from random positions that need to be assembled into the finished sequence. This sequencing strategy is known as **shotgun sequencing,** and it is outlined in **Figure 6.27**. The double-stranded target molecule whose sequence is to be determined is shown in the middle. The target is first sheared into small, random fragments that are cloned into special vectors used for template preparation. In the shotgun sequencing phase (Figure 6.27A), a large number of templates are chosen at random and sequenced, each yielding a stretch of single-stranded sequence, typically 500–1000 bp in length. In the figure, sequences obtained using the top strand as template are shown as arrows pointing left across the top, and those obtained using the bottom strand as template are shown as arrows pointing right across the bottom. Although these short sequences originate from random positions and from either target strand, if there are enough of them, the overlapping regions can be identified. The overlaps allow each of the shotgun sequences to be arranged according to template strand and positioned within the target molecule, as indicated by the overlapping arrows. With a sufficient number of overlaps the complete sequence of the target molecule can be assembled, even if, as shown in the figure, there are a few small gaps in the sequence of one or the other strand of the target molecule. The sequencing redundancy not

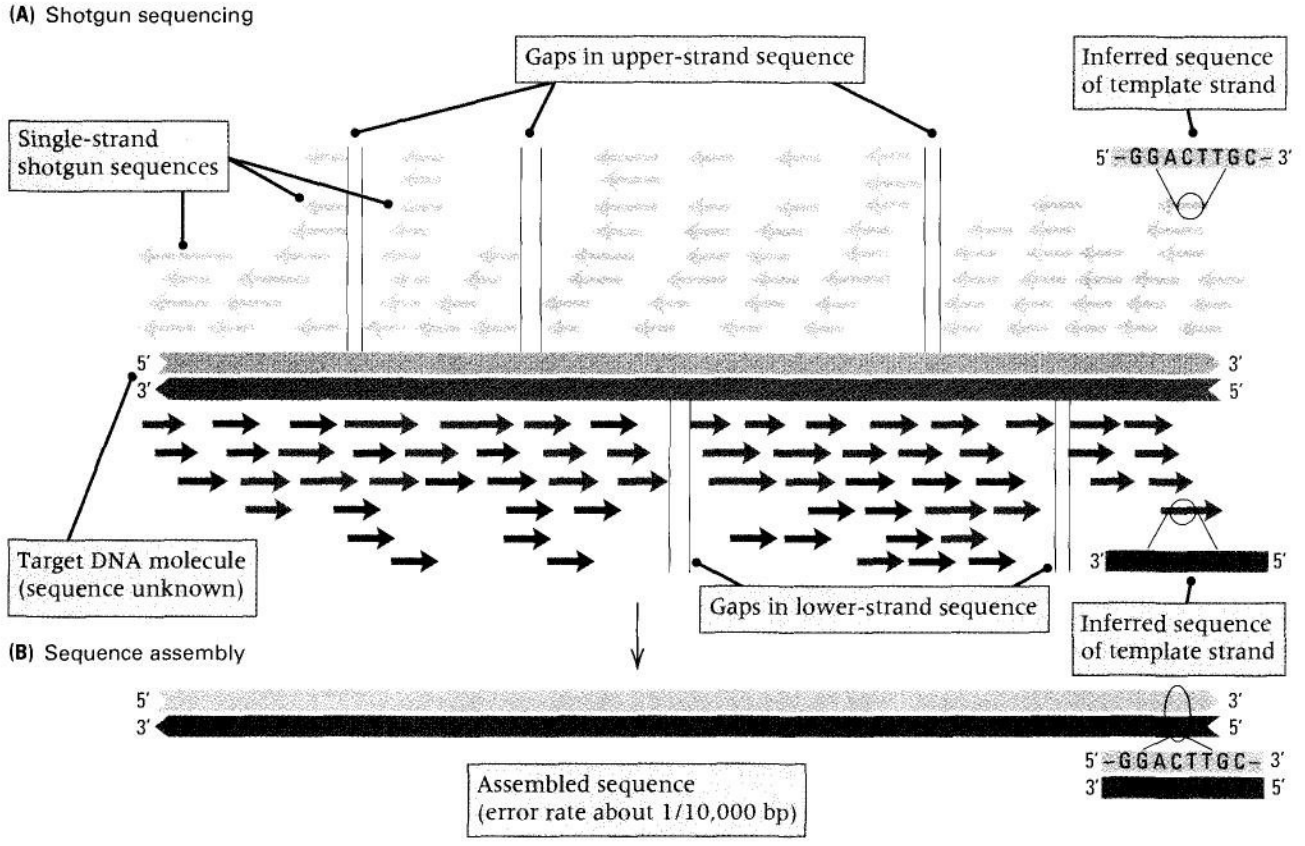
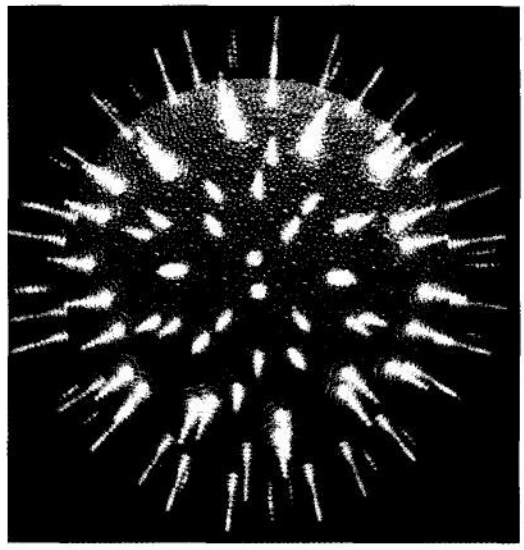


Figure 6.27 Shotgun sequencing strategy. (A) The DNA to be sequenced is sheared into many small fragments that are cloned and sequenced at random. The arrows indicate the individual position, length, and direction of the individual sequences. (B) With sufficient depth of coverage, computer analysis can be used to recognize the overlaps and join the individual sequences into a continuous sequence with (ideally) no gaps.

only aids in the final assembly of the finished sequence but also increases the accuracy, because multiple sequences of the same regions are compared, and miscalled nucleotides can be detected. The assembly of a large target molecule entails the analysis of so many short sequences that computational methods are essential. By these means, shotgun sequencing can be used to determine the entire sequence of genomes as large or larger than that of *Drosophila*.

Clinical Use of Dideoxynucleoside Analogs

The knowledge of DNA structure and replication has applications not only in procedures for the manipulation of DNA but also in the development of drugs for clinical use. An example is in one approach to the treatment of AIDS. A number of dideoxynucleoside analogs are effective at inhibiting replication of the viral genetic material. A few of these are illustrated in **Figure 6.28**. Recall that a *nucleoside* is a base attached to a sugar without a phosphate. A nucleoside *analog* is a molecule similar, but not identical, in structure to a nucleoside. In Figure 6.28, ddC is the normal dideoxyribocytidine nucleoside. It is effective against AIDS, as are the dideoxynucleoside analogs AZT, D4T, and ddI (and other such analogs). The nucleoside, rather than the nucleotide, is used in therapy because the nucleotide, having a highly charged phosphate group, cannot cross the cell membrane as easily.



Michael Freeman/Corbis

When the RNA genome of the Human Immunodeficiency Virus (HIV) is released into the cytoplasm of an infected cell, the enzyme reverse transcriptase (a product of a viral gene) synthesizes double-stranded DNA using the viral RNA as a template. This DNA molecule can integrate into the host DNA and be used to transcribe new molecules of HIV RNA.

The emergence of these drugs from our basic knowledge of DNA structure and replication is one of the prime examples of the fact that "pure" science may have many unforeseen practical applications. The basic experiments on DNA were carried out long before the recognition of AIDS as a distinct infectious disease and the discovery that HIV is the causative agent.

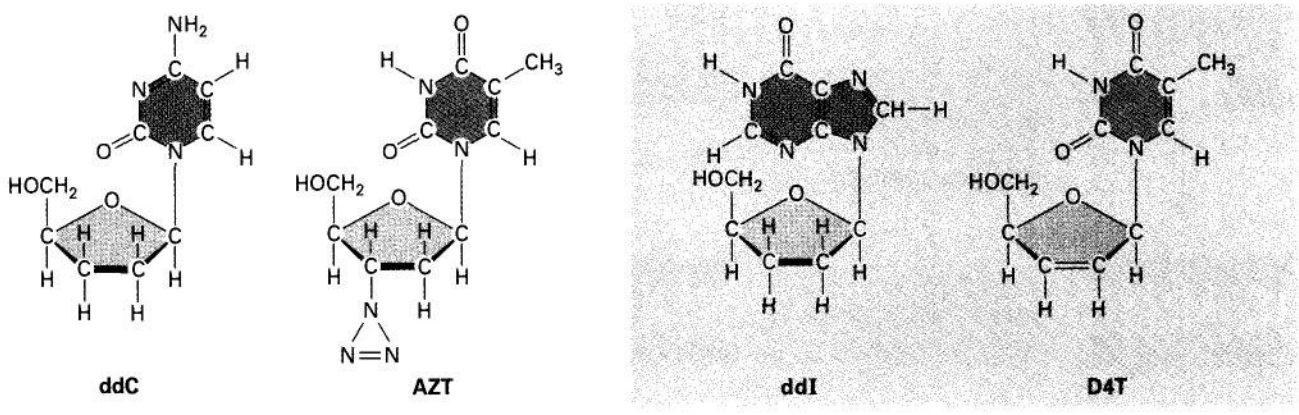


Figure 6.28 A few of the drugs that have been found to be effective in the treatment of AIDS by interfering with the replication of HIV virus. The technical names of the substances are ddI, 2′, 3′-dideoxyinosine; ddC, 2′,3′-dideoxycytidine; AZT, 3′-azido-2′,3′-dideoxythymidine; and D4T, 2′, 3′-didehydro-2′,3′-dideoxythymidine.

6.8 Molecular Mechanisms of Recombination

The molecular basis of genetic linkage mapping, discussed in Chapter 5, is recombination between two different DNA molecules; it is a process of breakage and rejoining. In eukaryotes, the process takes place early in meiosis after each molecule has replicated. For genetic studies of recombination, tetrad analysis in fungi such as yeast or *Neurospora* are particularly useful because all four products of any meiosis ("tetrads") can be recovered in a four-spore (budding yeast) or eight-spore (*Neurospora*) ascus (Section 5.5). Most asci from heterozygous *Aa* diploids contain ratios of

2 *A* : 2 *a* in four-spored asci, or
4 *A* : 4 *a* in eight-spored asci

because of normal Mendelian segregation. Occasionally, however, aberrant ratios are also found, such as

3 *A* : 1 *a* or 1 *A* : 3 *a* in four-spored asci, and
5 *A* : 3 *a* or 3 *A* : 5 *a* in eight-spored asci

Different types of aberrant ratios can also occur. The aberrant asci are said to result from **gene conversion** because it appears as if one allele has "converted" the other allele into a form like itself. Gene conversion is frequently accompanied by recombination between genetic markers on either side of the conversion event, even when the flanking markers are tightly linked. This implies that gene conversion can be one consequence of the recombination process.

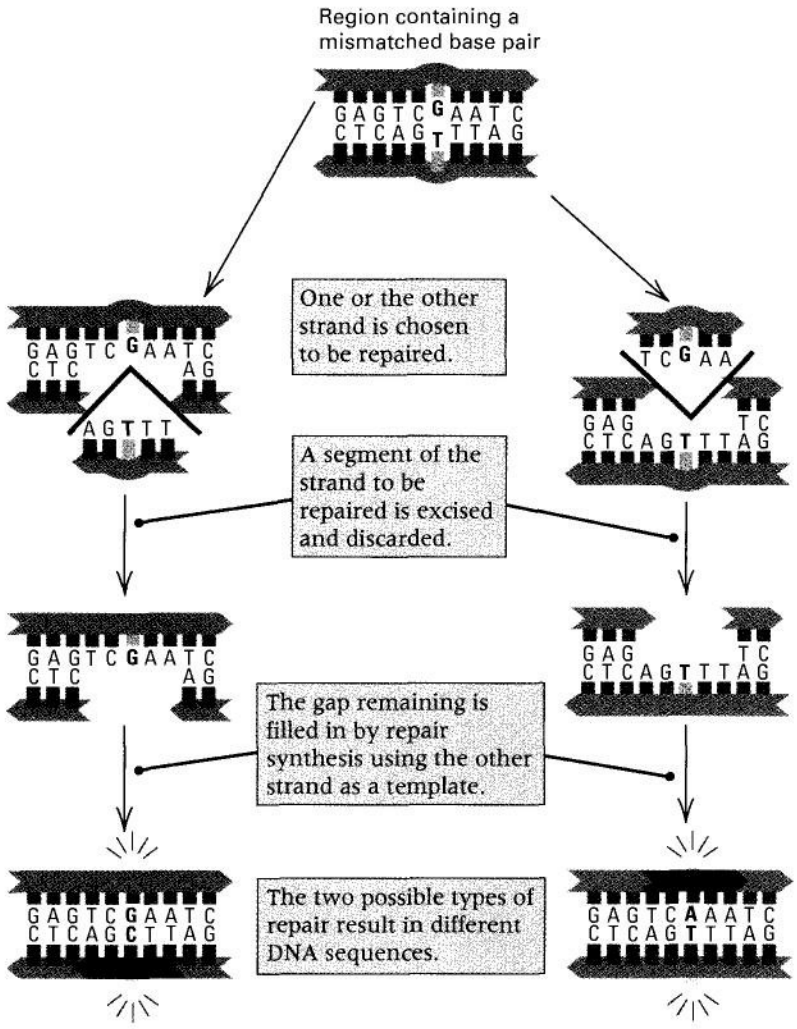


Figure 6.29 Mismatch repair consists of the excision of a segment of a DNA strand containing a base mismatch followed by repair synthesis. Either strand can be excised and corrected. In this example, the G–T mismatch is corrected to either G–C (left) or A–T (right).

Gene Conversion and Mismatch Repair

Gene conversion results from a normal DNA repair process in the cell known as **mismatch repair.** In this process, an enzyme recognizes any base pair in a DNA duplex in which the paired bases are mismatched—for example, G paired with T, or A paired with C. When such a mismatch is found in a molecule of duplex DNA, a small segment of one strand is excised and replaced with a new segment synthesized with the remaining strand as a template. In this manner the mismatched base pair is replaced. **Figure 6.29** shows an example in which a mismatched G–T pair is being repaired. The strand that is excised could be either the strand containing T or the one containing G, and the newly synthesized (repaired) segment, shown in red, would contain either C or A, respectively. The two possible products of repair differ in DNA sequence.

The role of mismatch repair in gene conversion is illustrated in **Figure 6.30.** The row of DNA duplexes across the top represents the DNA molecules in two pairs of sister chromatids in pachytene of meiosis,

immediately prior to recombination. One pair of sister chromatids carries the *A* allele, the sequence of which contains the G–C base pair highlighted in color, whereas the other pair of sister chromatids contains the *a* allele, in which the same base pair is A–T. In the process of recombination, the participating DNA duplexes at first exchange a single strand, each of which pairs with the nonexchanged strand in the other duplex. The second row shows the situation after the middle two duplexes have exchanged the bottom strand. In each molecule, the exchange of pairing partners creates a **heteroduplex** region in which any bases that are not identical in the parental duplexes become mismatched. In this example, one heteroduplex contains a G–T base pair and the other an A–C base pair. At this point the mismatch repair system comes into play and corrects the mismatches. Each mismatch can be repaired in either of two ways, so there are four possible ways in which the mismatches can be repaired. Two of the possibilities are shown. In one, the G–T mismatch is repaired to G–C, and the A–C mismatch is repaired to G–C also. Because the G–C pair defines the *A* allele, the result is that both middle duplexes carry *A*. After meiosis, when each sister chromatid is segregated to a different ascospore, the ascus will contain 3 *A* : 1 *a* and show an apparent $a \rightarrow A$ gene conversion. The other possibility for mismatch repair shown in Figure 6.30 is one in which both mismatches are repaired to A–T, which defines the *a* allele. The ascus that results from this type of correction will contain 1 *A* : 3 *a* and show an apparent $A \rightarrow a$ gene conversion.

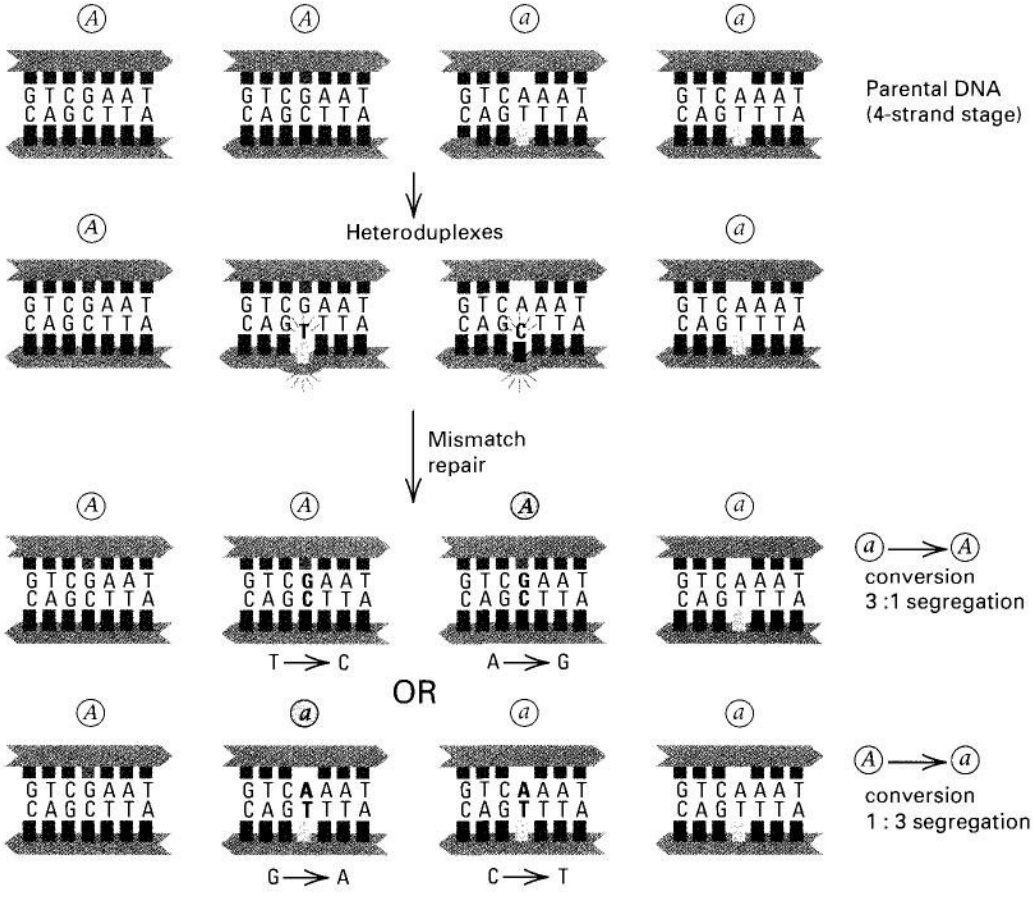


Figure 6.30 Mismatch repair resulting in gene conversion. Only a small part of the heteroduplex region is shown. Each heteroduplex containing a mismatched base pair can be repaired to give an *A* allele or an *a* allele. The patterns of repair leading to 3 *A* : 1 *a* and 1 *A* : 3 *a* segregation are shown.

The Holliday Model of Recombination

By what process are the heteroduplexes in Figure 6.30 formed as an intermediate in recombination? The first molecular model for recombination was suggested by Robin Holliday in 1964, and it illustrates why gene conversion in a small region is so frequently associated with the recombination of genetic markers flanking the region. Although this model has been superseded by more realistic versions to be examined in a few moments, it is worthwhile to examine the original Holliday model anyway, not only because of its simplicity but also because current models incorporate many of its key features.

In the **Holliday model** of recombination, each of the participating DNA duplexes (part A, Figure 6.31) initially undergoes a symmetrically positioned single-stranded nick in strands of the same polarity (Figure 6.31B). The nicked strands unwind in the region of the nicks, switch pairing partners, and form a heteroduplex region in each molecule (Figure 6.31C). The juxtaposed free ends of the DNA strands are joined together (Figure 6.31D), and further unwinding of the DNA duplexes and exchange of pairing partners increase the length of the heteroduplex region (this process is called *branch migration*). Two more nicks and rejoinings are necessary to resolve the structure in part D. As the configuration is drawn, these breaks can be in the inner strands, resulting in molecules that are *nonrecombinant* for markers flanking the heteroduplex region, or they can be in the outer strands, resulting in molecules that are *recombinant* for outside markers (Figure 6.31E).

Of the two possibilities for resolution of the structure in Figure 6.31D, one results in recombination of the outside markers and the other does not. These outcomes are equally likely, because there is no topological difference between the inner strands and the outer strands. To understand why, note that an end-for-end rotation of the

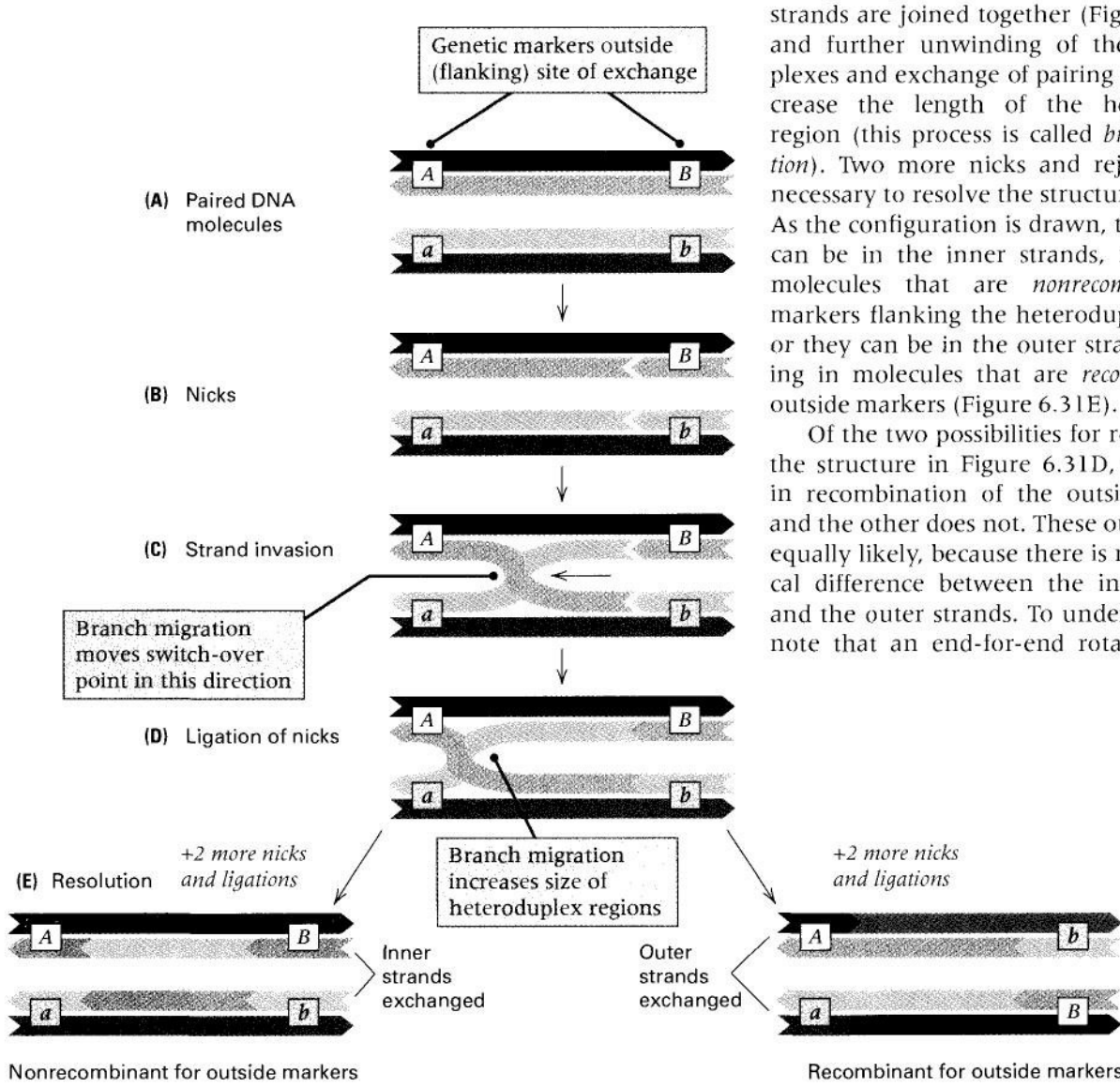


Figure 6.31 Holliday model of recombination. (A) Paired DNA molecules. (B) The exchange process is initiated by single-stranded breaks in strands of the same polarity. (C) Strands exchange pairing partners ("invasion"). (D) Ligation and branch migration. (E) Exchange is resolved either by the breaking and rejoining of the inner strands (resulting in molecules that are nonrecombinant for outside markers) or by the breaking and rejoining of the outer strands (resulting in recombinant molecules). Either type of resolution gives a heteroduplex region of the same length in each participating molecule.

lower duplex in Figure 6.31D gives the configuration shown in **Figure 6.32**. This type of configuration is known as a **Holliday junction.** Drawing it in this manner makes it clear that breakage and rejoining of the east–west strands and that of the north–south strands are topologically equivalent. However, north–south resolution results in recombination of the outside markers (*A b* and *a B* gametes), whereas east–west resolution gives nonrecombinants (*A B* and *a b* gametes).

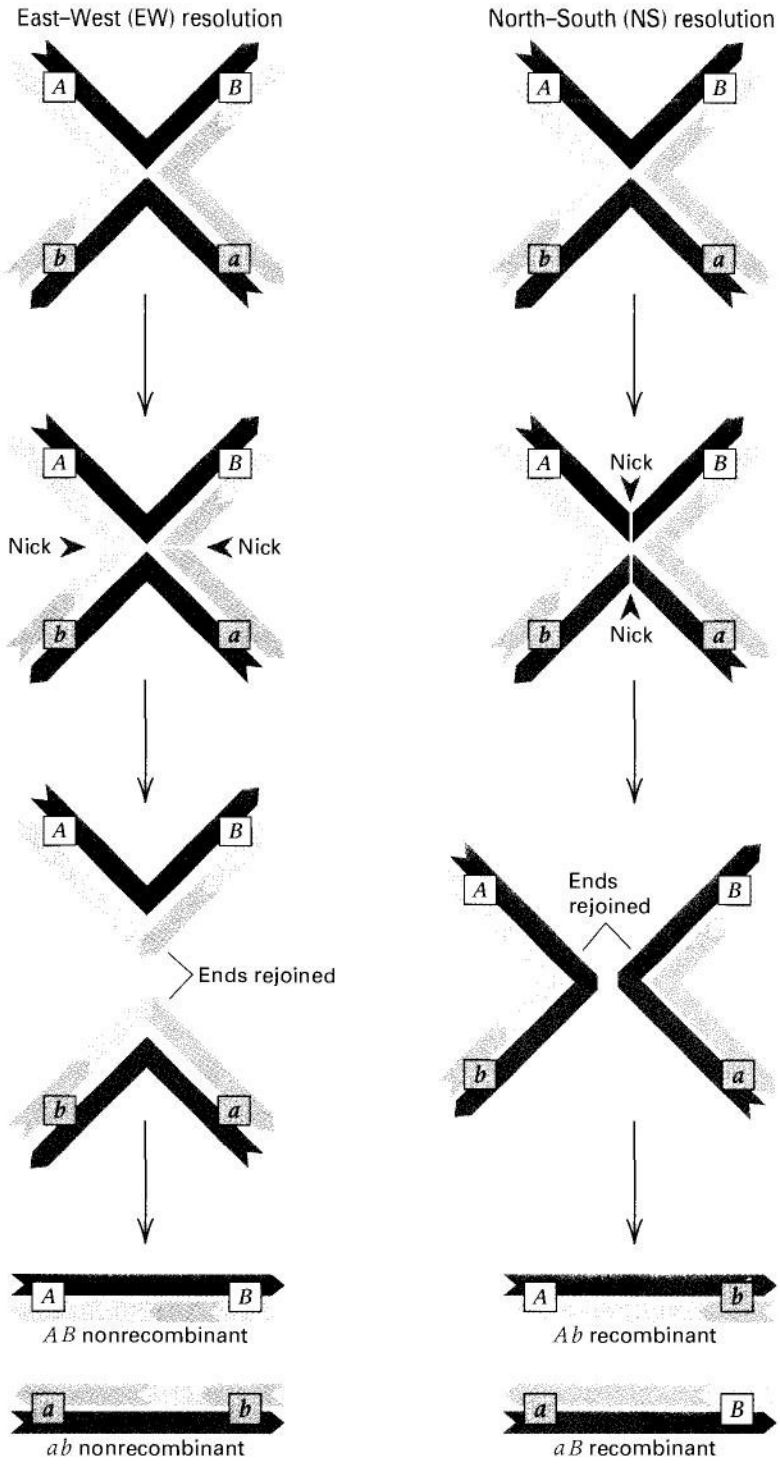


Figure 6.32 The cross-shaped configuration shows the topological equivalence of outer-strand and inner-strand resolution of the Holliday junction. East–west resolution corresponds to inner-strand breakage and rejoining; north–south resolution corresponds to outer-strand breakage and rejoining.

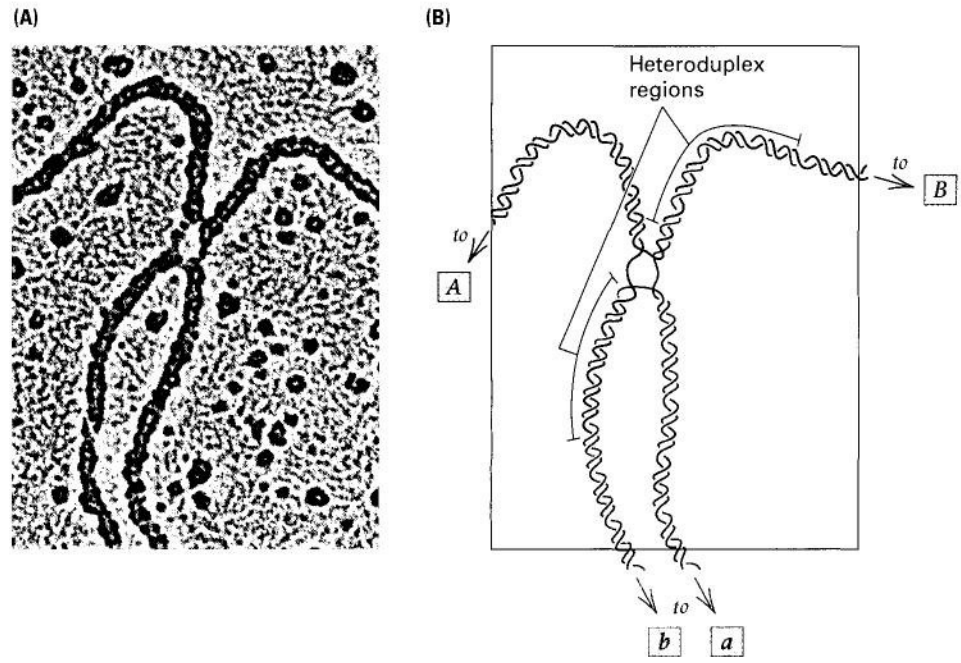


Figure 6.33 (A) Electron micrograph of a Holliday structure. Note the short, single-stranded region of DNA on each side of the oval-shaped structure where the duplexes come together. (B) Interpretive drawing of a Holliday structure in terms of double-helical molecules. Note the single-stranded regions where the duplexes come together, as well as the short heteroduplex region in two of the duplexes. For ease of comparison with Figure 6.32, the positions of the alleles *A*, *a* and *B*, *b* are indicated. Breakage and rejoining of the pair of single strands at the left and right yield duplexes that are nonrecombinant for the outside markers (*A B* and *a b*). Breakage and rejoining of the pair of single strands at the top and bottom yield duplexes that are recombinant for the outside markers (*A b* and *a B*).

An electron micrograph of a Holliday structure formed between two DNA duplexes is shown in **Figure 6.33**, part A. This is the physical structure of DNA that corresponds to the diagram in Figure 6.32. An interpretation of the structure that preserves the double-helical structure of duplex DNA is shown in Figure 6.33B, in which the colors match those at the top in Figure 6.32. The locations of the genetic markers (*A a* and *B b*) flanking the heteroduplex region are also indicated. The key feature of the Holliday structure is that there are four short, single-stranded regions where the duplexes come together. The Holliday structure can be resolved either by breakage and reunion in the single-stranded regions on the top and bottom (called NS in Figure 6.32) or by breakage and reunion in the single-stranded regions on the left and right (called EW in Figure 6.32). The left–right breakage and rejoining result in two duplexes that are nonrecombinant for the outside markers (*A B* and *a b*), whereas the top–bottom breakage and rejoining result in two duplexes that are recombinant for the outside markers (*A b* and *a B*).

Several types of enzymes are required for recombination. In *E. coli*, the exchange of strands is mediated by the RecA protein. Another enzyme, called the RecBCD nuclease because it contains three protein subunits (B, C, and D), has been implicated in strand exchange and branch migration. The resolution of the Holliday structure is also an enzymatic function, which is carried out by the **Holliday junction-resolving enzyme,** denoted RuvC. The key feature of the Holliday model is that heteroduplex regions are formed in the region of the exchange. Mismatch repair in this region can result in gene conversion, and the equal likelihood of north–south and east–west resolution (the RuvC function) implies that gene conversion will frequently be accompanied by recombination of flanking markers.

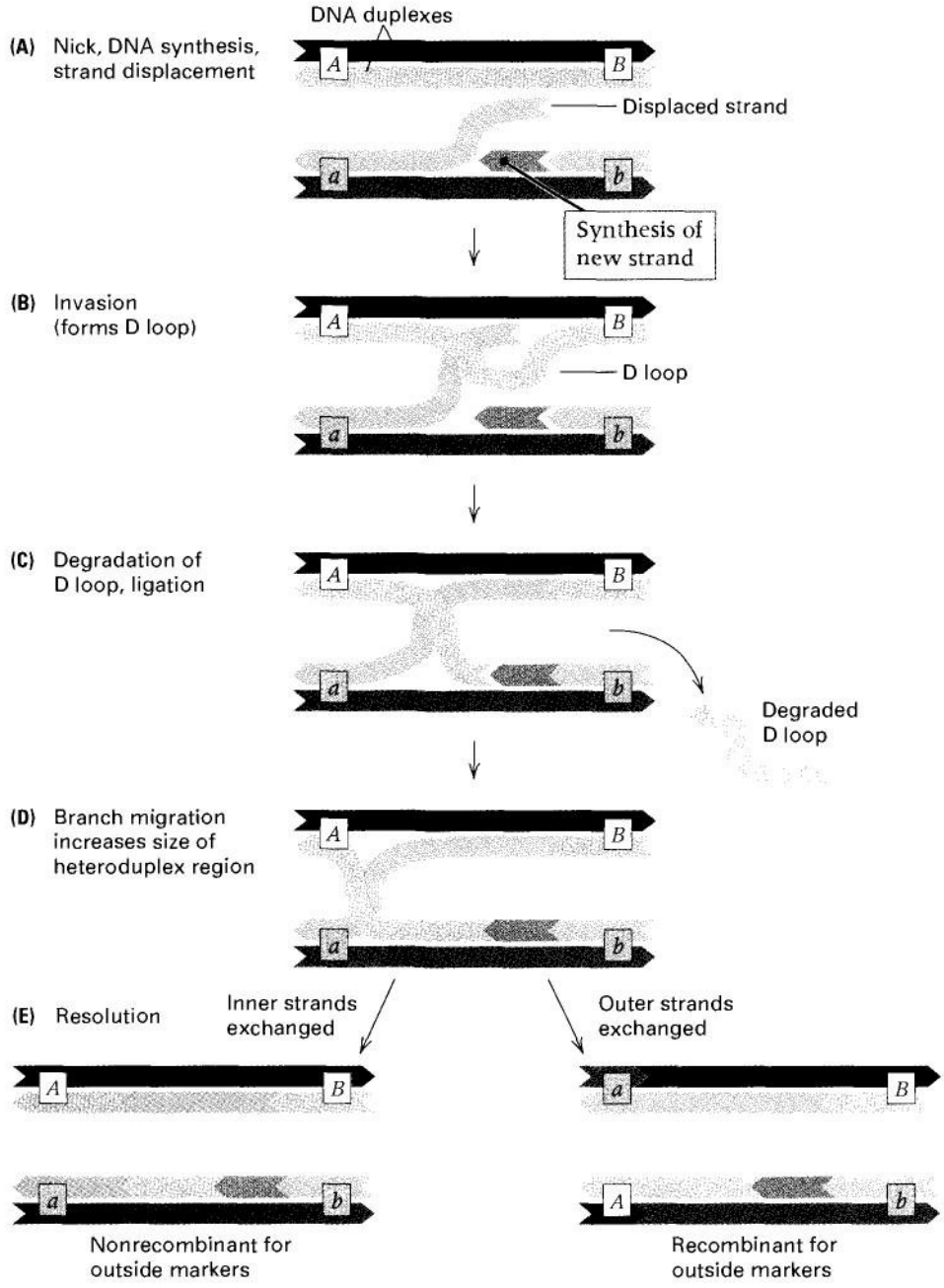


Figure 6.34 Model of recombination initiated by a break in one strand of a pair of DNA molecules. (A) A freed DNA strand is replaced by new synthesis and invades the partner molecule. (B) Displacement of a strand in the partner molecule forms a D (displacement) loop. (C) The D loop is degraded, and free DNA ends are ligated. (D) Branch migration increases the length of the heteroduplex region. (E) Resolution occurs in either of two topologically equivalent ways. In this model, the heteroduplex regions are not of equal length.

Single-Strand Break and Repair Model

Observed patterns of gene conversion indicate that the heteroduplex region is not always the same length in the DNA duplexes participating in an exchange. This is a difficulty for the Holliday model, in which the heteroduplex regions are symmetrical. However, in the **single-strand break and repair model** (Figure 6.34), the heteroduplex regions can be asymmetrical. Starting with a single-strand nick (Figure 6.34A), progressive unwinding and new DNA synthesis (red strand) liberate one DNA strand from a duplex. This strand can "invade" the other duplex (Figure 6.34B), resulting in displacement of the original strand and formation of a **D loop** (**displacement loop**). Nucleases degrade the D loop, and the invading strand is ligated in place (Figure 6.34C). Branch migration (Figure 6.34D) forms a second heteroduplex region, but the heteroduplex regions in the participating

CoNnection

Happy Holliday

Robin Holliday 1964
John Innes Institute, Bayfordbury, Hertfordshire, United Kingdom
A Mechanism for Gene Conversion in Fungi

When Holliday wrote this paper, gene conversion was a well-known but rather mysterious phenomenon. The favored explanation was the "copy-choice" model, in which, during DNA replication, the growing point of a daughter strand could switch its template between DNA strands of the same polarity in the parental DNA duplexes. A 3 : 1 conversion ratio could therefore be explained by postulating that there had been a template *switch in the region of the conversion. The copy-choice model predicted that DNA replicates conservatively rather than semiconservatively, that chromosome pairing must take place prior to DNA replication, and that multiple crossovers should always involve the same pair of chromatids. None of these predictions was supported by the experimental evidence, but no satisfactory alternative hypothesis was available. The Holliday model, although it is now known to be too simplistic, nevertheless solved the mystery in proposing, as the mechanism of gene conversion, the cross-pairing of DNA strands (rather than template switching) to create heteroduplex regions in which any mismatched bases could be enzymatically repaired.*

Tetrad analysis [in many fungi] has demonstrated that any heterozygous locus normally gives the expected 2 : 2 Mendelian ratio; but that aberrant tetrads also occasionally occur, where one allele is represented three times and the other once. . . . In spite of its difficulties the copy-choice hypothesis is still favored by a number of authors. This is so because it is assumed that conversion is impossible without some genetic replication. . . . In this paper an origin for conversion events which does not involve genetic replication will be discussed. . . . The model uses the complementarity of the two strands of DNA in order to explain specific chromatid pairing at the molecular level. After genetic replication and general genetic pairing, at certain points DNA molecules from opposite homologous chromatids unravel to form single strands, and these then anneal or coil up with the complementary strands from the other chromatid. Thus specific or effective pairing over short regions could occur. There are a number of possibilities of varying complexity as to how the DNA might get into this configuration. . . . If the annealed region spans a point of heterozygosity—a mutant site—mispairing of bases will occur. . . . Repair mechanisms may operate, and these, by adjusting the base sequences in order to restore normal base pairing, would bring about gene conversion in the absence of any genetic replication. The model indicates how a precise breakage and rejoining of chromatids could occur in the vicinity of the conversion, so that conversion would frequently be accompanied by the recombination of outside markers.

The model indicates how a precise breakage and rejoining of chromatids could occur.

Source: Genetical Research, Cambridge 5: 282–304

duplexes are of different lengths. As with the Holliday model, the configuration in Figure 6.34D can be resolved by RuvC in either the east–west or the north–south direction, yielding nonrecombinant or recombinant products, respectively.

Double-Strand Break and Repair Model

The **double-strand break and repair model** of homologous recombination is shown in Figure 6.35. The participating molecules are a broken duplex and an intact duplex. The broken duplex that initiates the process will typically contain a gap with missing sequences and strands of unequal length at the ends because of exonuclease activity. The ends are free to invade the homologous, intact duplex, forming a D loop (Figure 6.35B), which increases in extent until it bridges the gap (Figure 6.35C). Repair synthesis and ligation in both duplexes give the configuration in Figure 6.35D, which can be resolved in four ways because it contains two Holliday junctions, each of which can be resolved in the east–west or the north–south orientation. Two resolutions yield nonrecombinant duplexes and

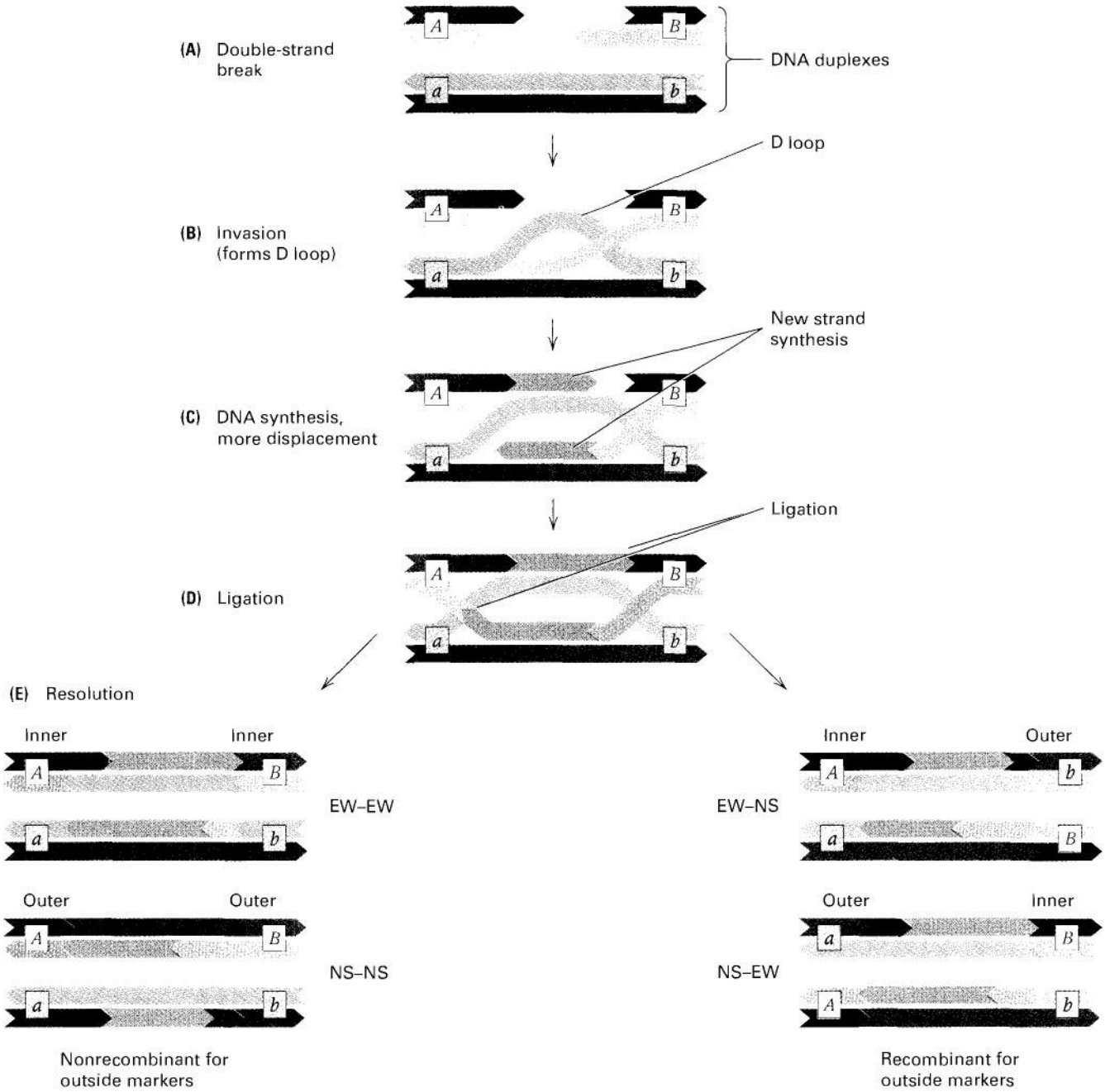


Figure 6.35 Model of recombination initiated by a double-stranded break. (A) The double-stranded break is trimmed back asymmetrically by nucleases. (B) One of the strands invades the partner molecule, forming a displacement loop. (C) The invading strand is elongated by new synthesis, and the displacement loop serves as a template for new synthesis in the other duplex. (D) Free DNA ends are ligated. (E) Two Holliday junctions are formed, which can be resolved in any one of four topologically equivalent ways (two of which result in recombination of outside markers). In this model, the heteroduplex regions are unequal in length, and all of them contain some newly synthesized DNA.

two yield recombinant duplexes, so genetic recombination is a frequent outcome. The double-strand break and repair model is currently favored as the principal mechanism of homologous recombination, in part because of the finding that in budding yeast, duplex DNA molecules with double-stranded breaks are very efficient at initiating recombination with a homologous unbroken duplex.

Chapter Summary

Replication is semiconservative in that each parental single strand, which serves as a template strand, is found in one of the double-stranded progeny molecules. Semiconservative replication was first demonstrated in the Meselson–Stahl experiment, in which equilibrium density-gradient centrifugation was used to separate DNA molecules containing two ^{15}N-labeled strands, two ^{14}N-labeled strands, or one of each. DNA molecules of prokaryotes usually have a single origin of replication; eukaryotic DNA molecules usually have many origins. Replication requires many proteins, including a helicase to unwind the helix, a gyrase to relieve the torsional stress farther along, and single-stranded DNA-binding proteins to stabilize the individual template strands. New-strand DNA synthesis always requires a primer, which in cells is a short RNA fragment made by an RNA polymerase (primase) located within a protein complex knows as a primosome; the RNA primer is removed at later stages of replication. Primer elongation and daughter-strand synthesis proceed by DNA polymerase first bringing in a nucleotide triphosphate with the correct base (the Watson–Crick pairing partner of the base in the template strand) and then allowing the free 3'-OH group of the growing strand to attach to the 5'-P group of the incoming nucleotide. (In this process the terminal P–P from the nucleotide triphosphate is cleaved and released.) Because double-stranded DNA is antiparallel, only one strand (the leading strand) grows in the direction of movement of the replication fork. The other strand (the lagging strand) is synthesized in the opposite direction as short fragments (Okazaki fragments) that are subsequently joined together.

The base sequence of a DNA molecule can be determined by dideoxynucleotide terminator sequencing. In this method, the DNA is isolated in discrete fragments containing several hundred nucleotide pairs. Complementary strands of each fragment are sequenced, and the sequences of overlapping fragments are combined to yield the complete sequence. The dideoxy sequencing method uses dideoxynucleotides to terminate daughter strand synthesis. The length of the terminated fragment reveals the position in the polynucleotide strand at which the dideoxynucleotide was incorporated.

Genetic recombination results from breakage and rejoining of DNA strands. Three models of recombination differ in the types of DNA breaks that initiate the process and in predictions about the types and frequencies of gene-conversion events. In gene conversion, one allele becomes converted into a homologous allele, which is detected by aberrant segregation such as 3 : 1 or 1 : 3 in fungal asci. Gene conversion is the outcome of mismatch repair in heteroduplexes. All recombination models include the creation of heteroduplexes in the region of the exchange and predict that 50 percent of gene conversions will be accompanied by recombination of genetic markers flanking the conversion event. The Holliday model assumes that single-strand nicks occur at homologous positions in the participating duplexes. Another model assumes initiation with a single-stranded nick in one of the duplexes. The double-strand break and repair model is the currently favored molecular model of homologous recombination, largely on the basis of evidence from budding yeast. In this model, recombination is initiated when a broken duplex invades an intact duplex, forming two Holliday junctions.

Key Terms

bidirectional replication
dideoxyribose
dideoxy sequencing method
displacement loop (D loop)
DNA ligase
DNA polymerase I (Pol I)
DNA polymerase III (Pol III)
double-strand break and repair model
equilibrium density-gradient centrifugation
gene conversion
gyrase
helicase
heteroduplex
Holliday junction
Holliday junction-resolving enzyme
Holliday model
lagging strand
leading strand
mismatch repair
Okazaki fragments
origin of replication
Pol III holoenzyme
precursor fragment
primase
primer
primosome
proofreading function
replication fork
replicon
ribose
rolling-circle replication
semiconservative replication
shotgun sequencing
single-strand break and repair model
single-stranded DNA-binding protein (SSB)
sister-chromatid exchange
terminus of replication
Θ replication
ultracentrifuge

Review the Basics

- What chemical feature at the 3' end of a DNA strand in the process of being synthesized is essential for elongation? Can the strand also be elongated at the 5' end?
- In DNA replication, what features distinguish the leading strand from the lagging strand?
- If the paired strands in duplex DNA were parallel rather than antiparallel, would replication still involve a leading strand and a lagging strand? Explain.
- In rolling-circle replication, does the newly synthesized strand correspond to the leading strand or the lagging strand in conventional replication? Explain your answer.
- Explain the "proofreading" function of DNA polymerase III (Pol III) and its eukaryotic counterparts.
- What feature of DNA replication guarantees that the incorporation of a dideoxynucleotide will terminate strand elongation?
- What is "shotgun sequencing"? Why does the method require that most regions be sequenced multiple times in different, independent clones?
- What unusual types of tetrads provide evidence that a region of heteroduplex DNA is created in the process of recombination?
- Why does the resolution of a Holliday junction result half the time in molecules that are recombinant for outside markers and half the time in molecules that are nonrecombinant for outside markers?
- Why are there two Holliday junctions in the double-strand break and repair model of recombination?

Guide to Problem Solving

Problem 1 A single-stranded DNA molecule is replicated *in vitro* by extending a primer oligonucleotide in the presence of nucleoside triphosphates whose two outermost phosphate groups (called the γ and β phosphates) are labeled with radioactive ^{32}P but whose innermost phosphate (the α phosphate) is not labeled. Is the resulting double-stranded DNA labeled or unlabeled? Explain.

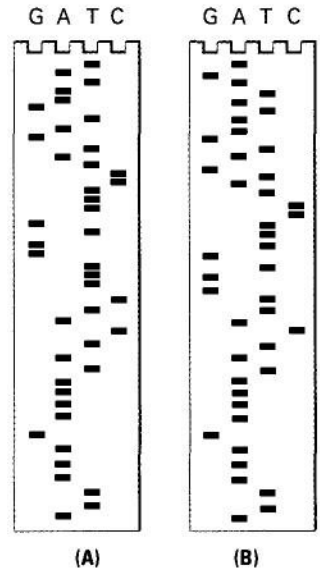
Answer Recall that the addition of each new nucleotide in strand elongation requires a nucleoside triphosphate, but in the reaction itself the covalent bond connecting the α and β phosphates is cleaved, releasing the still-connected β and γ phosphates into the medium. This means that none of the ^{32}P from each labeled triphosphate will be incorporated into the phosphodiester bond of the growing DNA strand, so the daughter duplex will be unlabeled.

Problem 2 The human genome contains approximately 3×10^9 base pairs of DNA. If the DNA molecules in each chromosome were placed end to end, what would the total length be, in meters? One angstrom unit (Å) = 10^{-8} centimeter.

Answer The standard B form of DNA has 3.4 Å per base pair (10 base pairs and 34 Å per complete turn). Hence the total length would be $(3 \times 10^9 \text{ bp})(3.4 \text{ Å/bp})(10^{-8} \text{ cm/Å})(10^{-2} \text{ m/cm}) = 1.02 \text{ m}$.

Problem 3 The DNA sequencing gels diagrammed in parts A and B of the accompanying illustration were obtained from human DNA by the dideoxy sequencing method. Part A comes from a person homozygous for the normal allele of the cystic fibrosis gene, part B from an affected person homozygous for a mutant allele. Use the bands in the gel to answer the following questions.

(a) Why are there discrete bands in the gel?
(b) Which end of the gel (top or bottom) corresponds to the 5' end of the DNA sequence?
(c) What nucleotide sequence is indicated by the bands in the gels?
(d) What is the sequence of the complementary strand in duplex DNA?
(e) Compare the sequences in parts A and B. What is special about the three nucleotides that give the bands indicated in red?



Answer

(a) Each band results from a fragment of DNA whose replication was terminated by the incorporation of a dideoxy nucleotide. The fragments differ by one nucleotide in length, and they are arranged by size with the smallest fragments at the bottom, the largest at the top.

(b) Because DNA replication elongates at the 3' end, each fragment of increasing length has an additional 3' nucleotide. Therefore, the DNA sequence is oriented with the 5' end at the bottom of the gel and the 3' end at the top.

(c) The sequences, read directly from the gel from bottom to top, are (part A)

5'-ATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATAT-3'

and (part B),

5'-ATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATAGA-3'

(d) The complementary strands in duplex DNA are antiparallel and have A paired with T and G paired with C. For part A, the complementary strand is

3'-TAATTTCTTTTATAGTAGAAACCACAAAGGATACTACTTATA-5'

and, for part B, the complementary strand is

3'-TAATTTCTTTTATAGTAACCACAAAGGATACTACTTATATCT-5'

(e) The three nucleotides in red in the wildtype gene are missing in the mutant gene. This three-base deletion is found in about 70 percent of the mutations in cystic fibrosis patients, and it results in a missing amino acid at position 508 in the corresponding protein.

Analysis and Applications

6.1 Consider a culture of *E. coli* cells grown for many generations in a ^{15}N-containing medium. The cells are washed and transferred to a ^{14}N-containing medium. After exactly two chromosome replications in the second medium, the DNA is extracted without any breakage whatsoever. What density classes would result and in what proportions?

6.2 What is meant by the statement that the DNA replication fork is asymmetrical?

6.3 The double-stranded DNA molecule of a newly discovered virus was found by electron microscopy to have a length of 68 micrometers (68×10^4 angstrom units, Å).

(a) How many base pairs are present in the viral genome?

(b) How many complete turns of the double helix are present in such a molecule?

6.4 Why does the incorporation of a dideoxynucleotide stop strand elongation dead in its tracks?

6.5 Specify the function or enzymatic activity of the following proteins that participate in DNA replication: helicase, single-stranded DNA binding protein, gyrase, polymerase III (Pol III), polymerase I (Pol I), DNA ligase, primase.

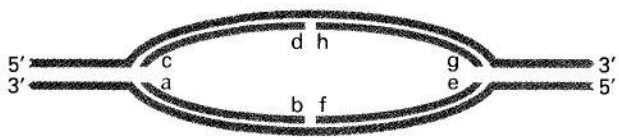
6.6 What is the chemical difference between the groups joined by DNA polymerase and DNA ligase?

6.7 What are three enzymatic activities of DNA polymerase I?

6.8 Why can RNA polymerases and primases initiate DNA replication, whereas DNA polymerases cannot?

6.9 A human chromosome of average size contains 1.3×10^8 nucleotide pairs. How long would it take to replicate this molecule if replication started at one end and proceeded to the other, assuming a typical eukaryotic rate of DNA synthesis of 50 nucleotide pairs per second? (The answer explains why DNA molecules in eukaryotes must have many origins of replication.)

6.10 For the replication bubble illustrated here, indicate the leading strand and the lagging strand at each replication fork and identify the ends as 3' or 5'.



6.11 Consider a hypothetical phage whose DNA replicates exclusively by rolling-circle replication. A phage with radioactive DNA in both strands infects a bacterium and is allowed to replicate in a nonradioactive medium. Assume that only daughter DNA from the elongating branch ever gets packaged into progeny phage particles.

(a) What fraction of the parental radioactivity will appear in progeny phage?

(b) How many progeny phage will contain radioactive DNA, and how will the number be affected by recombination?

6.12 A friend brings you three samples of nucleic acid and asks you to determine each sample's chemical identity (whether DNA or RNA) and whether the molecules are double-stranded or single-stranded. You use powerful nucleases to completely degrade each sample to its constituent nucleotides and then determine the approximate relative proportions of nucleotides. The results of your assay follow. What can you tell your friend about the nature of these samples?

Sample 1: dGMP 13% dCMP 14% dAMP 36% dTMP 37%

Sample 2: dGMP 12% dCMP 36% dAMP 47% dTMP 5%

Sample 3: GMP 22% CMP 47% AMP 17% UMP 14%

6.13 What is the fundamental difference between the initiation of θ replication and that of rolling-circle replication?

6.14 When the base composition of DNA from the bacterium *Mycobacterium tuberculosis* was determined, 18 percent of the bases were found to be adenine.

(a) What is the percentage of cytosine?

(b) What is the entire base composition of the DNA and the [G] + [C] content?

6.15 Explain why the following statements are true for double-stranded DNA. The symbol R stands for purine (A or G) and Y stands for pyrimidine (C + T).

(a) A = T and G = C

(b) A + C = G + T

(c) A + G = T + C

(d) R = Y

6.16 If DNA replication were conservative (old strands stay together, new strands stay together), what result would be expected in the Meselson–Stahl experiment after:

(a) One round of replication?

(b) Two rounds of replication?

(c) Three rounds of replication?

What were the actual observations?

6.17 The double-stranded DNA molecule of a newly discovered virus was found by electron microscopy to have a length of 34 μm.

(a) How many nucleotide pairs are present in one of these molecules?

(b) How many complete turns of the two polynucleotide chains are present in such a double helix?

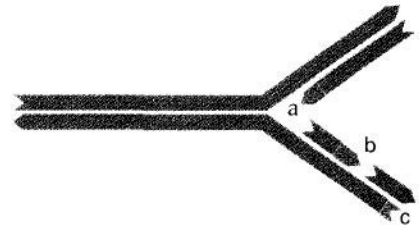
6.18 A "light" (^{14}N-labeled) DNA duplex forms a Holliday junction with a "heavy" (^{15}N-labeled) DNA duplex exactly in the middle. Assuming that the heteroduplex region has a negligible effect on the density of the resulting molecules, what would you expect the density of the resulting molecules to be after resolution of the Holliday junction?

6.19 An elegant combined chemical and enzymatic technique enables one to identify "nearest neighbors" of bases (adjacent bases in a DNA strand). For example, if the single-stranded tetranucleotide 5'-AGTC-3' were treated in this way, the nearest neighbors would be AG, GT, and TC (they are always written with the 5' terminus at the left). Before techniques were available for determining the complete base sequence of DNA, nearest-neighbor analysis was used to determine sequence relations. Nearest-neighbor analysis also indicated that complementary DNA strands are antiparallel, a phenomenon that you are asked to examine in this problem by predicting some nearest-neighbor frequencies. Assume that you have determined the frequencies of the following nearest neighbors: AG, 0.15; GT, 0.03; GA, 0.08; TT, 0.10.

(a) What are the nearest-neighbor frequencies of CT, AC, TC, and AA?

(b) If DNA had a parallel (rather than an antiparallel) structure, what nearest-neighbor frequencies could you deduce from the observed values?

6.20 What is the chemical group (3'-P, 5'-P, 3'-OH, or 5'-OH) at the sites indicated at the positions labeled a, b, and c in the accompanying figure?



6.21 Identify the chemical group that is at the indicated terminus of the daughter strand of the extended branch of the rolling circle in the accompanying diagram.



6.22 The following sequence of bases is present along one chain of a DNA duplex that has opened up at a replication fork, and synthesis of an RNA primer on this template begins by copying the base in red.

3'- · · · TCTGATATCAGTACG · · · -5'

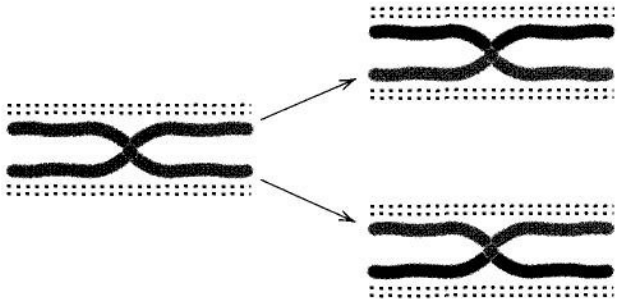
(a) If the RNA primer consists of eight nucleotides, what is its base sequence?

(b) In the intact RNA primer, which nucleotide has a free hydroxyl (-OH) terminus and what is the chemical group on the nucleotide at the other end of the primer?

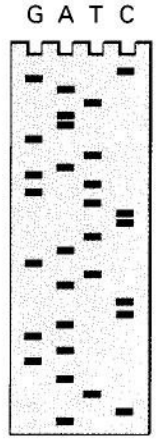
(c) If replication of the other strand of the original DNA duplex proceeds continuously (with few or no intervening RNA primers), is the replication fork more likely to move in a left-to-right or a right-to-left direction?

6.23 A 3.1-kilobase linear fragment of DNA was digested with *Pst*I and produced a 2.0-kb fragment and a 1.1-kb fragment. When the same 3.1-kb fragment was cut with *Hin*dIII, it yielded a 1.5-kb fragment, a 1.3-kb fragment, and a 0.3-kb fragment. When the 3.1-kb molecule was cut with a mixture of the two enzymes, fragments of 1.5, 0.8, 0.5, and 0.3 kb resulted. Draw a map of the original 3.1-kb fragment, and label the restriction sites and the distances between these sites.

6.24 The diagrams that follow depict the fluorescence patterns of chromosomes in mitotic metaphase after one and two rounds of DNA replication in the presence of BUdR, and the dotted lines represent the DNA strands in the DNA duplex present in each chromatid. Depict the BUdR labeling of each chromatid by (1) making the line solid if the strand is labeled with BUdR or (2) leaving it dashed if it is not labeled with BUdR.



6.25 The illustrated DNA sequencing gel at right was obtained by the dideoxy sequencing method. What is the nucleotide sequence of the strand synthesized in the sequencing reactions? What is the nucleotide sequence of the complementary template strand? Label each end as 3′ or 5′.



6.26 For the following restriction endonucleases, calculate the average distance between restriction sites in an organism whose DNA has a random sequence and equal proportions of all four nucleotides. The symbol R means any purine (A or G), and Y means any pyrimidine (T or C), but an R−Y pair must be either A−T or G−C.

*Taq*I	5′-TCGA-3′
	3′-AGCT-5′
*Bam*HI	5′-GGATCC-3′
	3′-CCTAGG-5′
*Hae*II	5′-RGCGCY-3′
	3′-YCGCGR-5′

6.27 For two linked genes, the existence of tetratype asci (spores *A B*, *a B*, *A b*, and *a b*) implies that

(a) Recombination takes place before DNA replication (the "two-strand" stage).

(b) Recombination takes place after DNA replication (the "four-strand" stage).

(c) Recombination must accompany DNA replication.

(d) None of the above.

6.28 In *Saccharomyces cerevisiae*, the wildtype gene *ADE-2* and one of its mutant *ade-2* alleles differ by a single base pair in the DNA; where the wildtype allele has an A−T base pair, the mutant allele has a G−C base pair. If the region in which the alleles differ is included in a heteroduplex, and both mismatches are repaired, how would one explain an ascus containing:

(a) 3 *ADE-2* : 1 *ade-2* spores?

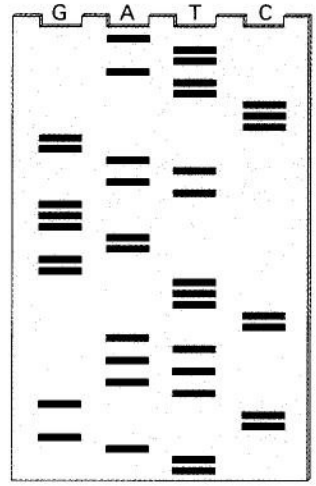
(b) 1 *ADE-2* : 3 *ade-2* spores?

(c) 2 *ADE-2* : 2 *ade-2* spores?

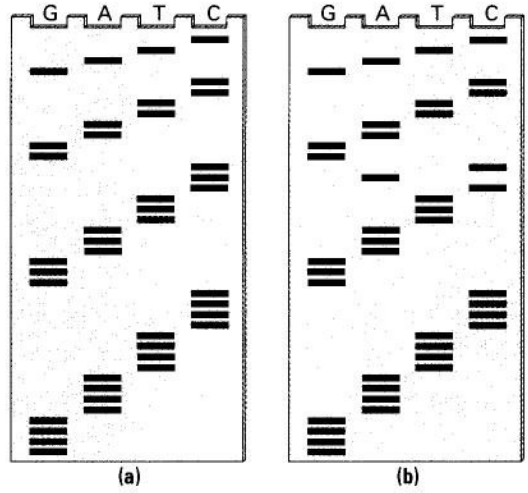
6.29 From the following sequencing gel, deduce the sequence of:

(a) The newly synthesized DNA strand.

(b) The template strand.

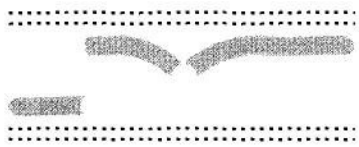


6.30 How do the DNA sequences shown below differ from each other?



Challenge Problems

6.31 The mitotic chromosome illustrated here was observed after two rounds of DNA replication in the presence of BUdR. The dotted lines represent the DNA strands in the DNA duplex present in each chromatid. Explain what process resulted in such a pattern of fluorescence, and depict the BUdR labeling of each chromatid by (1) making the line solid if the strand is labeled with BUdR or (2) leaving it dashed if it is not labeled with BUdR.



6.32 In *Saccharomyces cerevisae* a cross involving four linked genes is made, *A B C D* × *a b c d*, in which *B* and *C* are extremely close together in the DNA. Among 200 asci, one is found to contain the spores *A b c d*, *a b c D*, *A B C D*, and *a b c d*.

(a) What is unusual about this ascus?

(b) Can it be explained by a single event?

(c) What does the type of tetrad with respect to the *A*, *a* and *D*, *d* allele pairs indicate about this event?

(d) Could an ascus containing *A b c D*, *a b c d*, *A B C D*, and *a b c d* have the same explanation?

6.33 Tetrads are analyzed from a cross involving three linked genes in *Neurospora crassa*, his^+ co^+ leu^+ × *his co leu*. A small percentage of asci showed 3 : 1 segregation of each of the markers. Explain the following results.

(a) Among those with 3 : 1 segregation of *his*, the *co–leu* pair showed PD >> NPD (linkage).

(b) Among those with 3 : 1 segregation of *co*, the *his–leu* pair showed PD = NPD (no linkage!).

(c) Among those with 3 : 1 segregation of *leu*, the *his–co* pair showed PD >> NPD(linkage).

Further Reading

Bauer, W. R., F. H. C. Crick, and J. H. White. 1980. Supercoiled DNA. *Scientific American*, July.

Blow, J. J., ed. 1996. *Eukaryotic DNA Replication*. New York: IRL Press.

Cairns, J. 1966. The bacterial chromosome. *Scientific American*, January.

Camerini-Otero, R. D., and P. Hsieh. 1995. Homologous recombination proteins in prokaryotes and eukaryotes. *Annual Review of Genetics* 29: 509.

Davies, J. 1995. Vicious circles: Looking back on resistance plasmids. *Genetics* 139: 1465.

DePamphilis, M. L., ed. 2000. *Concepts in Eukaryotic DNA Replication*. Cold Spring Harbor NY: Cold Spring Harbor Laboratory Press.

Donovan, S., and J. F. X. Diffley. 1996. Replication origins in eukaroytes. *Current Opinion in Genetics & Development* 6: 203.

Eggleston, A. K., and S. C. West. 1996. Exchanging partners: Recombination in *E. coli*. *Trends in Genetics* 12: 20.

Falaschi, A. 2000. Eukaryotic DNA replication: A model for a fixed double replisome. *Trends in Genetics* 16: 88.

Flores-Rozas, H., and R. Kolodner. 2000. Links between replication, recombination and genome instability in eukaryotes. *Trends in Biochemical Sciences* 25: 196.

Haber, James E. 2000. Recombination: A frank view of exchanges and vice versa. *Current Opinion in Cell Biology* 12: 286.

Hubscher, U., and J. M. Sogo. 1997. The eukaryotic DNA replication fork. *News in Physiological Sciences* 12: 125.

Kelley, T., ed. 1988. *Eukaryotic DNA Replication*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Kornberg, A. 1995. *DNA Replication*. 2d ed. New York: Freeman.

Kornberg, R. D., and A. Klug. 1981. The nucleosome. *Scientific American*, February.

Kucherlapati, R. 1988. *Genetic Recombination*. Washington, DC: American Society for Microbiology.

Marians, K. 2000. Replication and recombination intersect. *Current Opinion in Genetics and Development* 10: 151.

GeNETics on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- The human gene for DNA polymerase α, *POLA*, is present in the short arm of the X chromosome. It is one of a small number of X-linked genes that violate "Ohno's law" that the set of genes present in the X chromosome is the same in all mammals. In marsupials, such as the kangaroo, and in monotremes, such as the platypus, **POLA** and the other exceptional genes are autosomal. In the evolution of mammals, the monotreme lineage branched off first, 150–170 million years ago, and the marsupial lineage branched off later, about 120–150 million years ago. The nonmarsupial branch evolved into the placental mammals, which includes ourselves and most of the furry creatures with which we are familiar. What does this evolutionary history imply about the evolution of the X chromosome? You can find out at this keyword site and also learn which other genes violate Ohno's law.

- Large-scale genome sequencing has resulted in the complete, or almost complete, genome sequences of a large number of bacteria, the yeast *Saccharomyces cerevisiae*, the nematode *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, and the human *Homo sapiens*—and a large number of other genomic sequencing projects are in progress. The keyword **Sanger Centre** will connect you with one of the world's leading sequencing centers and provide access to a cornucopia of information about various genome projects. The Sanger Centre, located in Hinxton, Cambridgeshire, is named after Fred Sanger, who in 1980 shared a

Neidhardt, F. C., R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, eds. 1996. Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (2 volumes). 2d ed. Washington, DC: American Society for Microbiology.

Roeder, G. S. 1997. Meiotic chromosomes: It takes two to tango. *Genes in Development* 15: 2600.

Singer, M., and P. Berg. 1991. *Genes & Genomes.* Mill Valley, CA: University Science Books.

Smith, P. J., and C. J. Smith. eds. *DNA Recombination and Repair.* 1999. New York: Oxford University Press.

Stahl, F. 1996. Meiotic recombination in yeast: Coronatin of the double-strand break–repair model. *Cell* 87: 965.

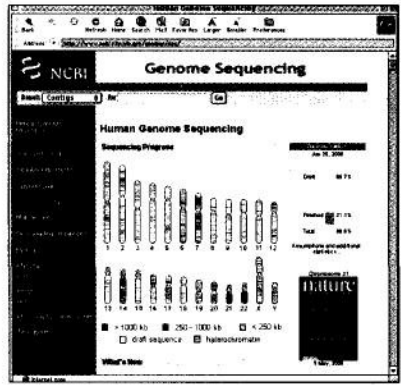
Stahl, F. 1998. Recombination in phage lamda; One geneticists's historical perspective. *Gene* 223: 95.

Nobel Prize in chemistry for his contributions to the determination of base sequences in nucleic acids. (It was his second; the first was awarded in 1958 for his work on the determination of amino acid sequences in proteins.)

- The **Holliday model** revolutionized thinking about recombination because it was the first model that suggested how recombination takes place at the molecular level and incorporated new concepts such as mismatch repair in heteroduplex regions. An animated version of the model can be found at this keyword site. Note the key stages of strand invasion and branch migration. Be sure to view the three-dimensional animation as well, because it shows why the two possible types of resolution of the Holliday junction are topologically equivalent.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 6 and you will be linked automatically.

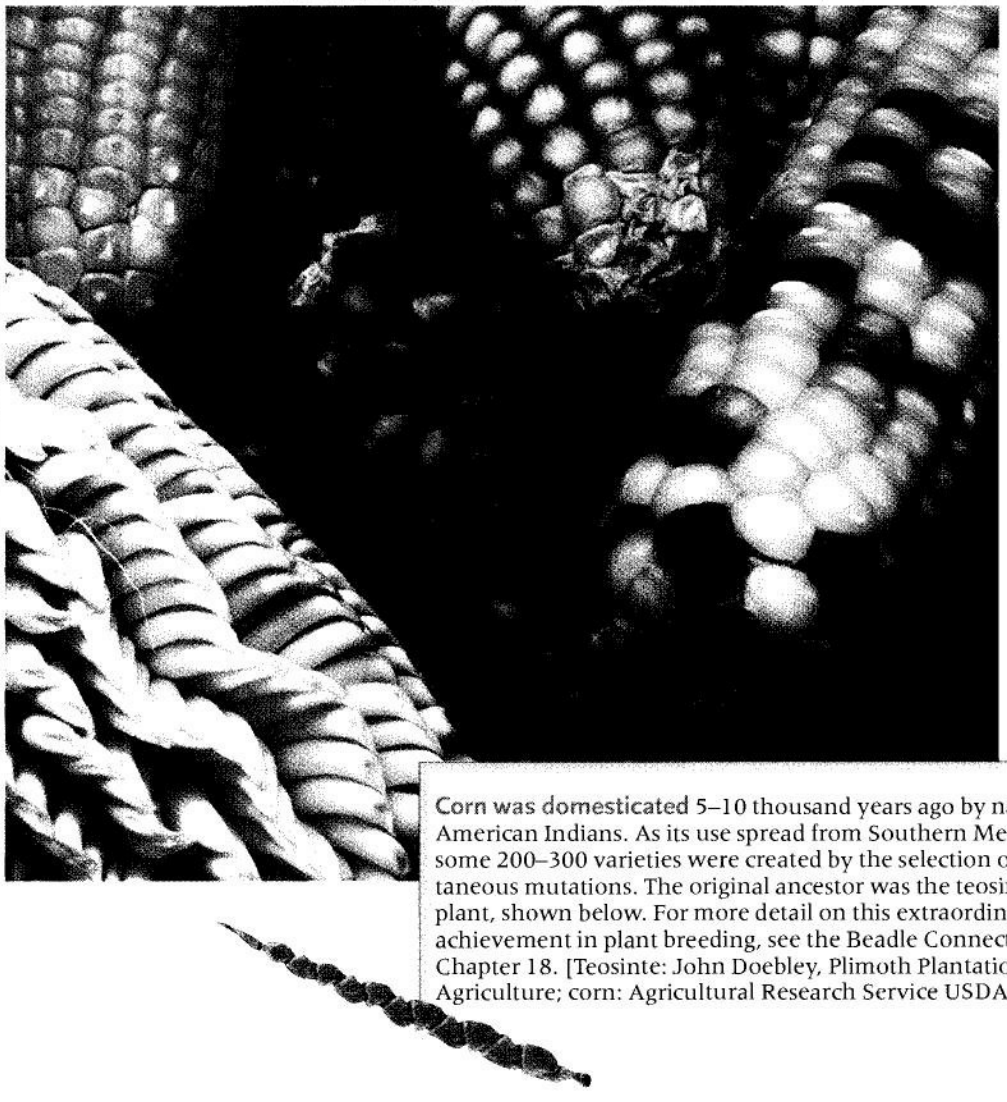
- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **Pic Site** for Chapter 6.



[Courtesy National Center for Biotechnology Information]

CHAPTER 7

Molecular Mechanisms of Mutation and DNA Repair



Corn was domesticated 5–10 thousand years ago by native American Indians. As its use spread from Southern Mexico, some 200–300 varieties were created by the selection of spontaneous mutations. The original ancestor was the teosinte plant, shown below. For more detail on this extraordinary achievement in plant breeding, see the Beadle Connection in Chapter 18. [Teosinte: John Doebley, Plimoth Plantation Agriculture; corn: Agricultural Research Service USDA.]

CHAPTER OUTLINE

7.1 Types of Mutations
Germ-Line and Somatic Mutations
Conditional Mutations
Classification by Function

7.2 The Molecular Basis of Mutation
Nucleotide Substitutions
Missense Mutations: The Example of Sickle-Cell Anemia
Insertions, Deletions, and Frameshift Mutations
Dynamic Mutation of Trinucleotide Repeats

7.3 Transposable Elements
Molecular Mechanisms of Transposition
Transposable Elements as Agents of Mutation
RIP: A Defense Against Transposons

7.4 Spontaneous Mutation
The Nonadaptive Nature of Mutation
Estimation of Mutation Rates
Hot Spots of Mutation

7.5 Mutagens
Base-Analog Mutagens
Chemical Agents That Modify DNA
Misalignment Mutagenesis
Ultraviolet Irradiation
Ionizing Radiation
Genetic Effects of the Chernobyl Nuclear Accident

7.6 Mechanisms of DNA Repair
Mismatch Repair
AP Repair
Photoreactivation
Excision Repair
Postreplication Repair
The SOS Repair System

7.7 Reverse Mutations and Suppressor Mutations
Intragenic Suppression
Intergenic Suppression
The Ames Test for Mutagen/Carcinogen Detection

PRINCIPLES

- Mutations can be classified by type of change at the molecular level, by conditions of expression, by effects on gene function, and in other ways.
- Many mutations occur in noncoding DNA sequences. Among those that do occur in coding sequences, missense (nonsynonymous) mutations are nucleotide substitutions resulting in an amino acid replacement, synonymous (silent) substitutions change a codon into a synonymous codon, and nonsense mutations create a translation-terminating codon. Single-base insertions or deletions result in frameshift mutations, in which the translational reading frame is offset. Expansion (growth in copy number) of trinucleotide repeats is an important mechanism of mutation associated with several human genetic diseases, including the fragile-X syndrome and Huntington disease.
- Transposable elements are DNA sequences that are able to change their location within a chromosome or to move between chromosomes. Insertion of transposable elements is an important mechanism of spontaneous mutation.
- Mutations can be induced by various agents, including base analogs, intercalating agents, highly reactive chemicals, and x rays; most mutagens are also carcinogens.
- Cells contain enzymatic pathways for the repair of different types of damage to DNA. Among the most important repair mechanisms are mismatch repair of mispaired nucleotides, AP repair of apurinic or apyrimidinic sites, excision repair, and postreplication repair.
- Suppressor mutations diminish or eliminate the phenotypic effects of a known mutation.

CONNECTIONS

Her Feeling for the Organism
Barbara McClintock 1950
The Origin and Behavior of Mutable Loci in Maize

X-Ray Daze
Hermann J. Muller 1927
Artificial Transmutation of the Gene

Replication Slippage in Unstable Repeats
Micheline Strand, Tomas A. Prolla, R. Michael Liskay, and Thomas D. Petes 1993
Destabilization of Tracts of Simple Repetitive DNA in Yeast by Mutations Affecting DNA Mismatch Repair

Any heritable change in a gene is a **mutation.** In this chapter, we examine the nature of mutations at the molecular level. We examine how mutations occur and what effects they have on genes and gene expression. Although mutations take place spontaneously, they can also be induced by radiation and a variety of chemical agents that damage nucleotides or that produce breaks in polynucleotide chains. Most types of DNA damage are corrected by special repair enzymes almost immediately after they occur. The new mutations that are transmitted in any generation therefore represent the minority of changes in DNA that have escaped repair.

7.1 Types of Mutations

Mutations can happen at any time and in any cell. The phenotypic effects can range from minor alterations that are detectable only by molecular methods (for example, DNA polymorphisms) to drastic changes in essential processes that cause death of the cell or organism.

Germ-Line and Somatic Mutations

Mutations can be classified in a variety of ways. In multicellular organisms, one important distinction is based on the type of cell in which the mutation first occurs. Those that arise in cells that ultimately form gametes are **germ-line mutations,** all others are **somatic mutations.** A somatic mutation yields an organism that is genotypically, and for many dominant mutations phenotypically, a mixture (*mosaic*) of normal and mutant tissue. Because reproductive cells are not affected, a somatic mutation cannot be transmitted to the progeny and may not be amenable to genetic analysis. In higher plants, however, somatic mutations can often be propagated by vegetative means without going through seed production; these means include grafting and the rooting of stem cuttings. Vegetative propagation is typical of many commercially important fruits, such as the "Delicious" apple and the "Florida" navel orange.

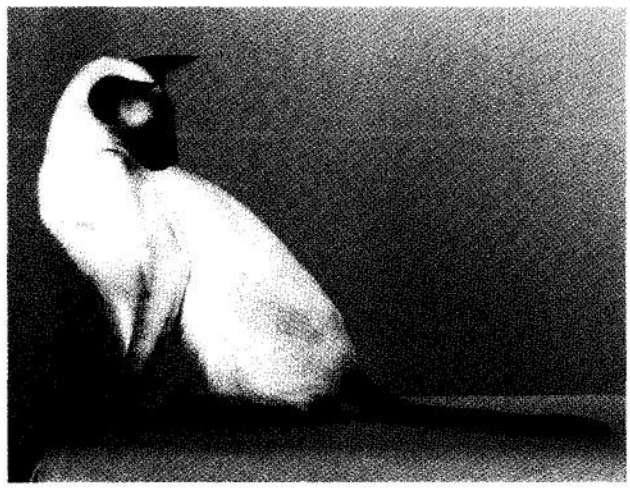


Figure 7.1 Siamese cat showing the characteristic pattern of pigment deposition. [Courtesy of Chanon.]

Conditional Mutations

Among the mutations that are most useful for genetic analysis are those whose effects can be turned on or off at will. These are called **conditional mutations** because they produce changes in phenotype in one set of environmental conditions (called the **restrictive conditions**) but not in another (called the **permissive conditions**). For example, a **temperature-sensitive mutation** is a conditional mutation whose expression depends on temperature. Usually, the restrictive temperature is high (in *Drosophila,* 29°C), and the organism exhibits a mutant phenotype above this critical temperature; the permissive temperature is lower (in *Drosophila,* 18°C), and under permissive conditions the phenotype is wildtype or nearly wildtype. Proteins containing amino acid replacements are often temperature-sensitive: The protein folds properly and functions nearly normally under permissive conditions, but it is unstable and denatures under restrictive conditions. Temperature-sensitive amino acid replacements are frequently used to block particular biochemical pathways under restrictive conditions, in order to test the importance of the pathways in various cellular processes, such as DNA replication.

An example of temperature sensitivity is found in the Siamese cat, with its black-tipped paws, ears, and tail (**Figure 7.1**). In this breed, an enzyme in the pathway for deposition of the black pigment melanin is temperature-sensitive. The pathway is blocked at normal body temperature, and

consequently the pigment is absent in the hair over most of the body where the temperature is warmest. The tips of the legs, ears, snout, and tail are cooler than the rest of the body, so the pigment is deposited in these areas.

Classification by Function

Mutations can also be classified according to the manner in which they affect gene expression. The major categories are described below. Classical terms for the categories were introduced by Herman J. Muller in 1931, but most of them have been superseded by modern equivalents. We will use the modern terms.

- A mutation that results in complete gene inactivation or in a completely nonfunctional gene product is a **loss-of-function mutation.** Examples of loss-of-function mutations include a deletion of all or part of a gene, and an amino acid replacement that inactivates the protein. Loss-of-function mutations are also called **null mutations** or **knockout mutations.** In Muller's terminology they are known as *amorphic mutations.*
- A mutation that reduces, but does not eliminate, the level of expression of a gene or the activity of the gene product is called a **hypomorphic mutation.** Hypomorphic mutations typically result from nucleotide substitutions that prevent a protein from being produced at the normal level, or from an amino acid replacement that impairs protein function. This class of mutations is sometimes referred to as *leaky.* The basis of the term is that, because the level of expression or activity differs from individual to individual by chance, a few individuals have enough enzyme activity to "leak through" to produce a quasi-normal phenotype. The opposite of a hypomorphic mutation is a **hypermorphic mutation.** As the prefix *hyper* implies, a hypermorphic mutant produces a greater-than-normal level of gene expression, typically because the mutation changes the regulation of the gene so that the gene product is overproduced. Less commonly, an amino acid replacement may increase the activity of an enzyme or other protein.
- A **gain-of-function mutation** is one that qualitatively alters the action of a gene. For example, a gain-of-function mutation may cause a gene to become active in a type of cell or tissue in which the gene is not normally active. Or it may result in the expression of a gene in development at a time during which the wildtype gene is not normally expressed. Whereas most loss-of-function and hypomorphic mutations are recessive, many gain-of-function mutations are dominant. Muller's term for a gain-of-function mutation is *neomorphic mutation.*

  Expression of a wildtype gene in an abnormal location is also called **ectopic expression.** For example, expression of the wildtype gene product of the *Drosophila* gene *eyeless* in tissues that do not normally form eyes results in the development of parts of compound eyes, complete with eye pigments, in abnormal locations (**Figure 7.2**). The locations can be in any tissues in which the gain-of-function is expressed, including on the legs or mouthparts, in the abdomen, or on the wings.
- Another type of mutation is a **dominant-negative mutation,** whose mutant product antagonizes the normal product of the gene. A typical dominant-negative mutation is an amino acid replacement that so alters the polypeptide conformation that when it is present in a multimeric molecule with wildtype subunits, it eliminates or reduces the activity of the complex. Most normal proteins function as multimers. Phenylalanine

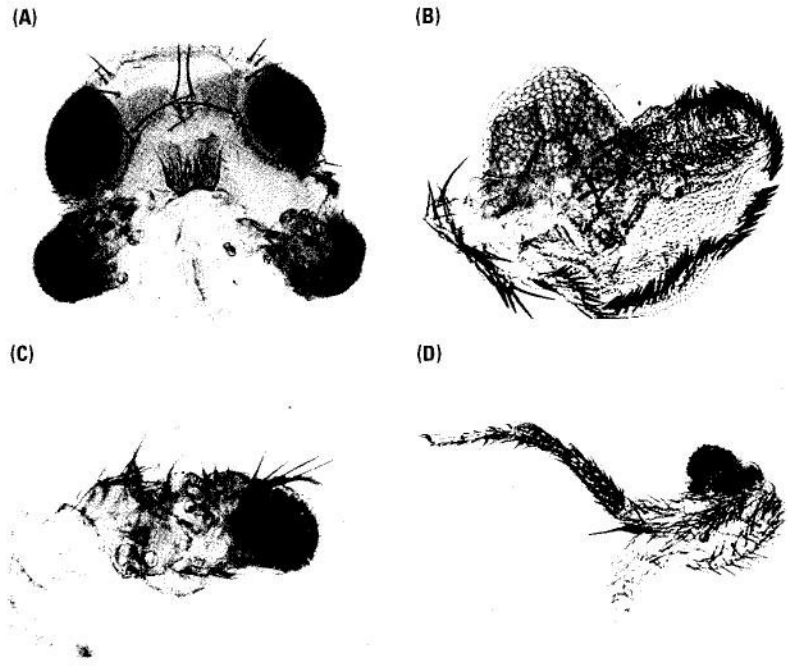


Figure 7.2 Ecoptic expression of the wildtype allele of the *eyeless* gene in *Drosophila* results in misplaced eye tissue. (A) An adult head in which both antennae form eye structures. (B) A wing with eye tissue growing out from it. (C) A single antenna in which most of the third segment consists of eye tissue. (D) Middle leg with an eye outgrowth at the base of the tibia. [Courtesy of G. Halder and W. J. Gehring. Reprinted with permission from G. Halder, P. Callaerts, and W. J. Gehring. *Science* 267: 1788. Copyright 1995 American Association for the Advancement of Science.]

hydroxylase is a tetramer of four identical polypeptide chains, for example, and hemoglobin is a tetramer composed of two alpha-globin and two beta-globin polypeptide chains. Dominant-negative mutations form "poison subunits" that are not only defective in themselves but also impair the activity of multimers in which they are present. The classical term for a dominant-negative mutation is *antimorphic mutation*.

7.2 The Molecular Basis of Mutation

All mutations result from changes in the nucleotide sequence of DNA or from deletions, insertions, or rearrangement of DNA sequences in the genome. The major types of chromosomal rearrangements are discussed in Chapter 9. Here we examine mutations at the molecular level.

Nucleotide Substitutions

At the molecular level, the simplest type of mutation is a **nucleotide substitution,** in which a nucleotide pair in a DNA duplex is replaced with a different nucleotide pair. For example, in an A → G substitution, an A is replaced with a G in one of the DNA strands. This substitution temporarily creates a mismatched G–T base pair; at the very next replication, the mismatch is resolved as a proper G–C base pair in one daughter molecule and as a proper A–T base pair in the other daughter molecule. In this case, the G–C base pair is the mutant and the A–T base pair is nonmutant. Similarly, in an A → T substitution, an A is replaced with a T in one strand, creating a temporary T–T mismatch, which is resolved by replication as T–A in one daughter molecule and A–T in the other. In this example, the T–A base pair is mutant and the A–T base pair is nonmutant. The T–A and the A–T are not equivalent, as can be seen by considering the nucleotide polarity of the strands: 5'-T-3'/3'-A-5' is the base pairing in one case, 5'-A-3'/3'-T-5' in the other.

Some nucleotide substitutions replace one pyrimidine base with the other or one purine base with the other. These are called **transition mutations.** The possible transition mutations are

$$\text{T} \rightarrow \text{C} \quad \text{or} \quad \text{C} \rightarrow \text{T}$$
$$\text{(pyrimidine} \rightarrow \text{pyrimidine)}$$

$$\text{A} \rightarrow \text{G} \quad \text{or} \quad \text{G} \rightarrow \text{A}$$
$$\text{(purine} \rightarrow \text{purine)}$$

Other nucleotide substitutions replace a pyrimidine with a purine or the other way around. These are called **transversion mutations.** The possible transversion mutations are

$$\text{T} \rightarrow \text{A}, \ \text{T} \rightarrow \text{G}, \ \text{C} \rightarrow \text{A}, \quad \text{or} \quad \text{C} \rightarrow \text{G}$$
$$\text{(pyrimidine} \rightarrow \text{purine)}$$

$$\text{A} \rightarrow \text{T}, \ \text{A} \rightarrow \text{C}, \ \text{G} \rightarrow \text{T}, \quad \text{or} \quad \text{G} \rightarrow \text{C}$$
$$\text{(purine} \rightarrow \text{pyrimidine)}$$

Altogether, there are four possible transitions and eight possible transversions. Therefore, if nucleotide substitutions were strictly random, then one would expect a 1 : 2 ratio of transitions to transversions. However:

> Spontaneous nucleotide substitutions are often biased in favor of transitions. Among spontaneous nucleotide substitutions in the human genome, the ratio of transitions to transversions is approximately 2 : 1.

Examination of the genetic code (Table 1.1 on page 20) shows that the bias toward transitions has an important implication for nucleotide substitutions in the third position of codons. In all codons with a pyrimidine in the third position, it does not matter which pyrimidine is present; likewise, in most codons that end in a purine, either purine will do. This means that most transition mutations in the third codon position do not change the amino acid in the encoded protein. Mutations that change the nucleotide sequence without changing the amino acid sequence are called **synonymous mutations** or **silent mutations,** because they change one codon into a "synonym" for the same amino acid.

Mutational changes in nucleotides that are outside of coding regions can also be silent. Most DNA polymorphisms are in this category. In noncoding regions, which include regulatory regions and the DNA between genes, the precise nucleotide sequence is often not critical. These sequences can undergo base substitutions, small deletions or additions, insertions of transposable elements, and other rearrangements, and yet the mutations may have no detectable

effect on phenotype. On the other hand, some noncoding sequences do have essential functions, and mutations in these sequences often do have phenotypic effects.

Missense Mutations: The Example of Sickle-Cell Anemia

Most nucleotide substitutions in coding regions do result in changed amino acids; these are called **missense mutations** or **nonsynonymous mutations**. A change in the amino acid sequence of a protein may alter the biological properties of the protein. The classic example of a phenotypic effect of a single amino acid change is the change responsible for the human hereditary disease *sickle-cell anemia*. The molecular basis of the sickle-cell mutation is shown in **Figure 7.3**. It is an A–T → T–A transversion in the second codon position for the sixth amino acid in the β-globin chain of hemoglobin, causing the normal glutamic acid (Glu or E) to be replaced with valine (Val or V). The amino acid replacement is therefore denoted as either Glu6Val or E6V. The sickle-cell mutation is recessive, and when homozygous, the defective β-polypeptide chain gives the hemoglobin protein a tendency to form long, needle-like polymers. Red blood cells in which polymerization takes place to become deformed into crescent, sickle-like shapes. Some of the deformed red blood cells are destroyed immediately (reducing the oxygen-carrying capacity of the blood and causing the anemia), whereas others

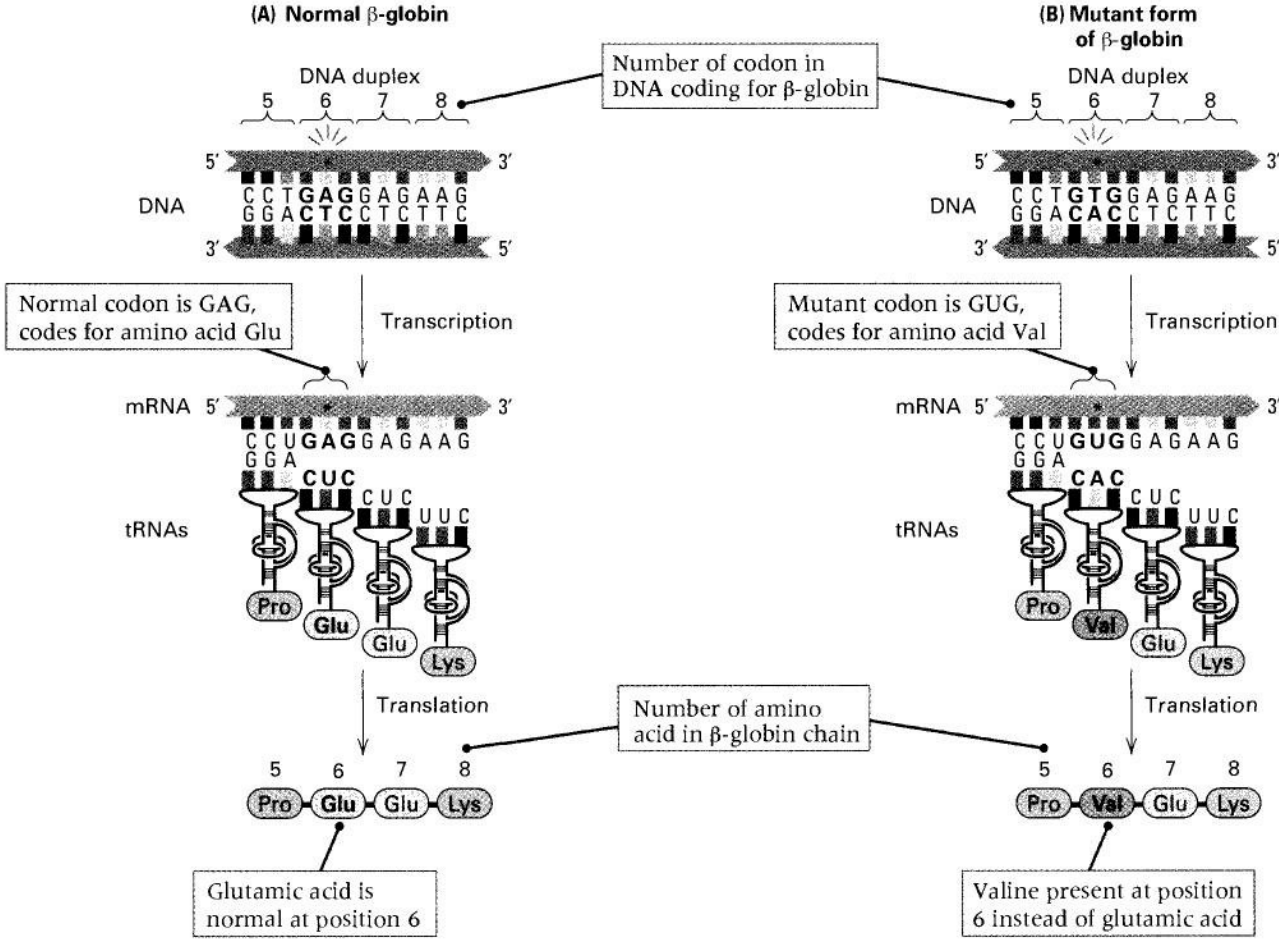


Figure 7.3 Molecular basis of sickle-cell anemia. (A) Part of the DNA in the normal β-globin gene is transcribed into a messenger RNA coding for the amino acid sequence Pro–Glu–Glu–Lys. The T in the marked A–T base pair is transcribed as the A in the GAG codon for Glu (glutamic acid). (B) Mutation of the normal A–T base pair to a T–A base pair results in the codon GUG instead of GAG. The codon GUG codes for Val (valine), so the polypeptide sequence in this part of the molecule is Pro–Val–Glu–Lys. The resulting hemoglobin is defective and tends to polymerize at low oxygen concentration.

become hung up in capillary vessels and cause partial or complete blocks in capillary circulation.

The consequences of the Glu6Val replacement are a profound set of pleiotropic effects. All of these effects are related to the breakdown of red blood cells, to the decreased oxygen-carrying capacity of the blood, or to physiological adjustments the body makes to try to compensate for the disease (such as enlargement of the spleen). Patients with sickle-cell anemia suffer bouts of severe pain. The anemia causes impaired growth, weakness, and jaundice. Affected people are so generally weakened that they are susceptible to bacterial infections, which are the most common cause of death in children with the disease.

Although sickle-cell anemia is a severe genetic disease that often results in premature death, it is relatively frequent in areas of Africa and the Middle East in which a type of malaria caused by the protozoan parasite *Plasmodium falciparum* is widespread. The association between sickle-cell anemia and malaria is not coincidental. The association results from the ability of the mutant β-hemoglobin to afford some protection against malarial infection. In the life cycle of the parasite, it passes from a mosquito to a human through the mosquito's bite. The initial stages of infection take place in cells in the liver, where specialized forms of the parasite are produced that are able to infect and multiply in red blood cells. Widespread infection of red blood cells impairs the ability of the blood to carry oxygen, causing the weakness, anemia, and jaundice characteristic of malaria. However, in people who are heterozygous for the sickle-cell allele, the infected blood cells undergo sickling and are rapidly removed from circulation. The proliferation of the parasite among the red blood cells is thereby checked, and the severity of the malarial infection is reduced. The resistance of heterozygous sickle-cell genotypes creates a genetic balancing act between the prevalence of the genetic disease sickle-cell anemia and that of the parasitic disease malaria. If the mutant sickle-cell hemoglobin becomes too frequent, more lives are lost from sickle-cell anemia than are gained by the protection against malaria; on the other hand, if the mutant sickle-cell hemoglobin becomes too rare, fewer lives are lost from sickle-cell anemia but the gain is offset by more deaths from malaria. The end result of this kind of genetic balancing act is discussed in quantitative terms in Chapter 17.

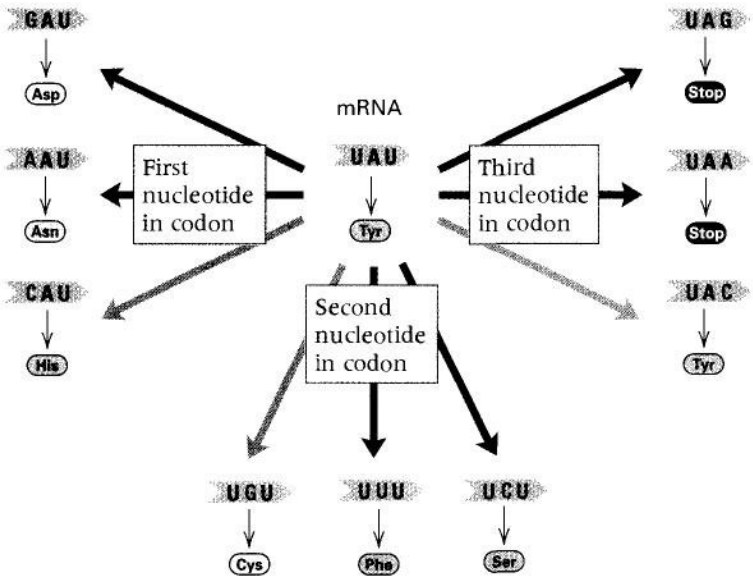


Figure 7.4 The nine codons that can result from a single base change in the tyrosine codon UAU. Blue arrows indicate transversions, gray arrows transitions. Tyrosine codons are in boxes. Two possible stop ("nonsense") codons are shown in red. Altogether, the codon UAU allows for six possible missense mutations, two possible nonsense mutations, and one silent mutation.

In contrast to the Glu6Val replacement, some amino acid replacements in hemoglobin have negligible effects on phenotype. Amino acid replacements are often tolerated in other proteins as well. The key issue is whether a particular amino acid replacement affects protein folding, function, or stability, and this depends on the protein, the amino acid and its precise role in the structure and function of the protein, and the nature of the amino acid replacement. For example, the replacement of one hydrophilic amino acid with another is expected to have a lesser effect than the replacement of a hydrophilic amino acid with a hydrophobic amino acid. Any change in the active site of an enzyme usually decreases enzymatic activity.

Figure 7.4 illustrates the nine possible codons that can result from a single nucleotide substitution in the UAU codon for tyrosine, organized according to whether the first, second, or third nucleotide in the codon is mutant. One mutation is silent (box), and six are missense mutations that change the amino acid inserted in the

polypeptide chain. The other two mutations create a stop codon that results in the premature termination of translation and the production of a truncated polypeptide. A nucleotide substitution that creates a new stop codon is called a **nonsense mutation.** Because nonsense mutations cause premature chain termination, the remaining polypeptide fragment is nearly always nonfunctional.

Insertions, Deletions, and Frameshift Mutations

Insertions or deletions of nucleotides can also occur in DNA, but at a rate considerably lower than that of nucleotide substitution. Most insertions and deletions in mammalian genomes are small (typically 1–6 base pairs), but some are large enough to eliminate entire genes or part of a chromosome (Chapter 9). If they occur outside of coding regions, small insertions or deletions may have negligible effects on phenotype unless they affect the expression of nearby genes.

Within a coding region, the insertion or deletion of exactly three nucleotides results in the addition or elimination of one amino acid from the encoded polypeptide chain, because three nucleotides is the length of a codon. An insertion or deletion of six nucleotides adds or eliminates two amino acids, and that of nine nucleotides adds or eliminates three amino acids. In general, provided that the number of nucleotides inserted or deleted is an exact multiple of three, an insertion or deletion adds or eliminates some number of amino acids from the polypeptide chain.

If the length of an insertion or deletion is not an exact multiple of three nucleotides, the mutation shifts the phase in which the ribosome reads the triplet codons and, consequently, alters all of the amino acids downstream from the site of the mutation. Such mutations are called **frameshift mutations** because they "shift" the reading frame of the codons in the mRNA. A common type of frameshift mutation is a single-nucleotide insertion or deletion. An example of a single-nucleotide insertion in the human β-globin gene is illustrated in **Figure 7.5**, where (A) is the wildtype sequence and (B) is the mutant sequence with a C–G nucleotide pair inserted at the position of the sunburst. The result of the insertion is that, starting with the codon in which the insertion occurs, the entire amino acid sequence of the polypeptide is altered because the reading frame is shifted, in this example by +1. Translation continues in the

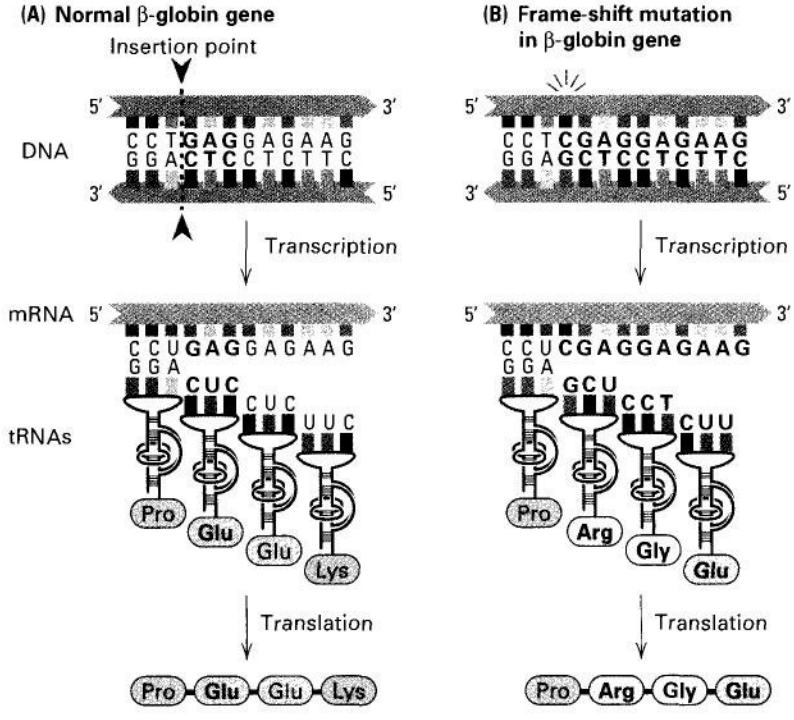


Figure 7.5 (A) A small part of human β-globin gene that encodes amino acids 5–8 in the polypeptide chain. (B) Result of a frameshift mutation (in this example, one that inserts a C–G base pair at the position indicated by the "burst"). All codons downstream from the site of the mutation are changed because of the shifted reading frame, so the entire amino acid sequence differs from that point on.

shifted reading frame until a termination codon is encountered. (Translation of a completely random nucleotide sequence would be expected to continue for only about 20 amino acids before running into a termination codon.) Any insertion or deletion that is not a multiple of three nucleotides results in a shifted reading frame, and all amino acids downstream from the insertion are different from the original. Unless it is very near the carboxyl terminus of a protein, a frameshift mutation almost always results in the synthesis of a nonfunctional protein.

Dynamic Mutation of Trinucleotide Repeats

Genetic studies of an X-linked form of mental retardation revealed an unexpected class of mutations called **dynamic mutations** because of the extraordinary genetic instability of the region of DNA involved. The X-linked condition, one of at least 12 genetic disorders associated with dynamic mutation, is associated with a class of X chromosomes that tends to fracture in cultured cells that are starved for DNA precursors, such as nucleotides. The position of the fracture is in region Xq27.3, near the end of the long arm. The X chromosomes containing this site are called *fragile-X* chromosomes, and the associated form of mental retardation is the *fragile-X syndrome*. The fragile-X syndrome affects about 1 in 2500 children. It accounts for about half of all cases of X-linked mental retardation and is second only to Down syndrome (Chapter 9) as a cause of inherited mental impairment.

A hint of something unusual was the paradoxical pattern of inheritance of the fragile-X syndrome, key features of which are illustrated in **Figure 7.6**. Approximately 1 in 5 males who carry the fragile X chromosome are themselves phenotypically normal and also have phenotypically normal children. The oddity is that the heterozygous daughters of such a "transmitting male" often have affected children of both sexes. In Figure 7.6, the transmitting male denoted I-2 is not affected, but the X chromosome that he transmits to his daughters (II-2 and II-5) somehow becomes altered in the female germ line in such a way that sons and daughters in the next generation (III) are affected. Both affected and normal granddaughters of the transmitting male may have affected progeny (generation IV).

The molecular basis of the fragile-X chromosome has been traced to a **trinucleotide repeat** of the form GGC (or, equivalently, CCG on the other strand) present in the DNA at the site where the breakage takes place. Normal X chromosomes have 6–54 tandem copies of the repeating unit, with an average of about 30, whereas affected persons have 230–2300 or more copies of the repeat. The trinucleotide repeat in the X chromosome in transmitting males is called the *premutation* and has an

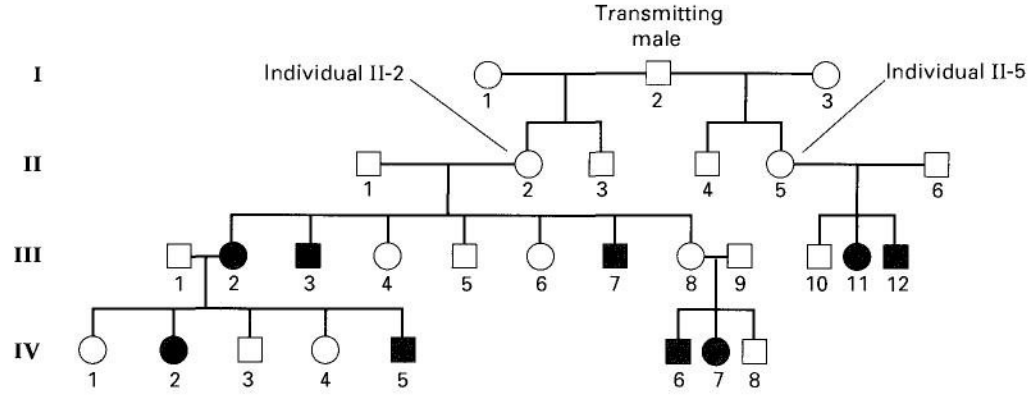


Figure 7.6 Pedigree showing transmission of the fragile-X syndrome. Male I-2 is not affected, but his daughters (II-2 and II-5) have affected children and grandchildren. [After C. D. Laird. 1987. *Genetics* 117: 587.]

intermediate number of copies, ranging from 52 to 230. The unprecedented feature of the trinucleotide premutation is that in the germline of females (and only in females), it often increases in copy number (called *trinucleotide expansion*) to a level of 230 copies or more, at which stage the chromosome causes the fragile-X syndrome. The functional basis of the disorder is that an excessive number of copies of the $(5'\text{-CCG-}3')_n$ repeat cause loss of function of a gene designated *FMR1* (fragile-site mental retardation-1), which is located immediately downstream of the repeat. Most fragile-X patients exhibit no *FMR1* messenger RNA, whereas normal persons and carriers do show expression.

Other genetic diseases associated with dynamic mutation include the neurological disorders myotonic dystrophy (chromosome 19, with an unstable repeat of CTG), Kennedy disease (X chromosome, repeat CAG), Friedreich ataxia (chromosome 9, GAA), spinocerebellar ataxia type 1 (chromosome 6, CAG), and Huntington disease (chromosome 4, CAG). Expansions occur primarily or exclusively in the female germ line for the fragile-X syndrome and myotonic dystrophy, primarily or exclusively in the male germ line for spinocerebellar ataxia type 1 and Huntington disease, and in the germ line of both sexes for some other dynamic mutations. Some dynamic mutations also show a phenomenon called **anticipation,** in which the number of repeats gradually increases as the trinucleotide repeat is transmitted through several generations. This results in a more severe disease with a progressively earlier age of onset through time.

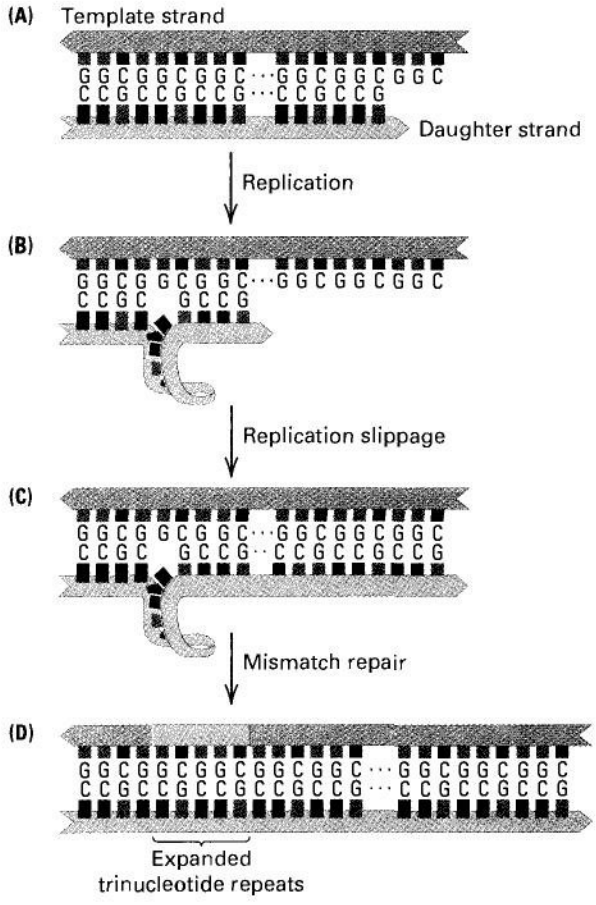


Figure 7.7 Model of replication slippage. (A) A trinucleotide repeat in the act of replication. (B) The 3' end of the growing strand momentarily "breathes" (detaches from the template) and reanneals to the template at a point upstream from its original location. (C) Continued replication duplicates the region between the points of detachment and reannealing. (D) Mismatch repair of the shorter strand creates a duplex with a trinucleotide expansion.

The mechanism of trinucleotide expansion is illustrated in **Figure 7.7**. The process is called **replication slippage.** As replication is proceeding along a template strand containing the repeats (A), the replication complex momentarily dissociates from the template strand. In reassociating with the template, the 3' end of the new strand backtracks along the template and pairs with an upstream set of repeats (B). Replication continues normally from this point (C), but some of the repeats will be replicated twice (expanded), the level of expansion depending on how far the replication complex backtracked in reassociating. The template and the daughter strand cannot pair properly because they have a different number of repeats, but this situation is corrected by the mismatch repair system (Chapter 6), and one outcome of repair is that the expanded region is introduced into the template strand (D). Although the mechanism of dynamic mutation is known, it is not known why some trinucleotide repeats in the genome are genetically unstable whereas other are stable, or why the trinucleotide premutation state is uniquely prone to expansion whereas chromosomes that may have only somewhat fewer copies are genetically stable.

CoNnection

Her Feeling for the Organism

Barbara McClintock 1950
Cold Spring Harbor Laboratory,
Cold Spring Harbor, New York
The Origin and Behavior of Mutable Loci in Maize

Many geneticists regard McClintock's papers on transposable elements as difficult. Her discoveries were completely novel. Genes that could move from one place to another in the genome were unheard of. She had no terminology with which to discuss such things, so she had to adapt the conventional terminology to describe a unique situation. What we now call a transposable element, and believe to be universal among organisms, McClintock calls a "chromatin element." McClintock was a superb geneticist and cytogeneticist, perhaps the best of her generation. In tracking down transposable elements by genetic means, McClintock had to use considerable ingenuity in designing crosses that were sometimes quite complex. Her writing style is also uniquely McClintock.

The excerpt that follows is a good example. It is a relentless marshaling of observation and hypothesis, experiment and result, interpretation and deduction. McClintock was awarded a Nobel Prize in 1983. The following passage deals with the discovery of Ds (Dissociation) *and* Ac (Activator). *In modern terminology, we call* Ac *an autonomous transposable element because it codes for all the proteins it needs for its own transposition;* Ds *is a nonautonomous transposable element because it requires* Ac *to provide the proteins necessary for it to move.*

Transposition of *Ac* takes place from one position in the chromosomal complement to another—very often from one chromosome to another.

In the course of an experiment designed to reveal the genic composition of the short arm of chromosome 9, a phenomenon of rare occurrence (or recognition) in maize began to appear with remarkably high frequencies in the cultures. The terms mutable genes, unstable genes, variegation, mosaicism, [and] mutable loci have been applied to this phenomenon. Its occurrence in a wide variety of organisms has been recognized. . . . A fortunate discovery was made early in the study of the mutable loci which proved to be of singular importance in showing the kinds of events that are associated with their origin and behavior. A locus was found in the short arm of chromosome 9 at which breaks were occurring in somatic cells. The time and frequency of the breakage events occurring at this *Ds* (*Dissociation*) locus appeared to be the same as the time and frequency of the mutation-producing mutable loci. An extensive study of the *Ds* locus has indicated the reason for this relationship and has produced the information required to interpret



In the corn hollow, at Cornell University in 1929, are, standing left to right, Charles Burnham, Marcus Rhoades, Rollins Emerson, and Barbara McClintock. Kneeling with the dog is George Beadle.

7.3 Transposable Elements

In a 1940s study of the genetics of kernel mottling in maize (**Figure 7.8**), Barbara McClintock discovered a genetic element that not only regulated the mottling but also caused chromosome breakage. She called this element *Dissociation (Ds)*. Genetic mapping showed that the chromosome breakage always occurs at or very near the location of *Ds*. McClintock's critical observation was that *Ds* does not have a constant location but occasionally moves to a new position (**transposition**), causing chromosome breakage at the new site. Furthermore, *Ds* moves only if a second element, called *Activator (Ac)*, is also present in the same genome. In addition, *Ac* itself moves within the genome and can also cause modification in the expression of genes at

the events occurring at mutable loci. It has been concluded that the changed phenotypic expression of such loci is related to changes in a chromatin element other than that composing the genes themselves, and that mutable loci arise when such chromatin is inserted adjacent to the genes that are affected. . . . The *Ds* locus is composed of this kind of material. Various types of alterations are observed as a consequence of events occurring at the *Ds* locus. . . . They involve chromosome breakage and fusion. The breaks are related, however, to events occurring at this one specific locus in the chromosome—the *Ds* locus. . . . [Among the known events is] transposition of *Ds* activity from one position to another in the chromosome complement with or without an associated gross chromosomal rearrangement. . . . It is from the transpositions of *Ds* that some of the new mutable loci may arise. . . . In one case the transposed *Ds* locus appeared in a single gamete of a plant carrying chromosome 9 with a dominant *C* (colored aleurone) allele. . . . The new position of *Ds* corresponded to the known location of *C*. . . . Significantly, the appearance of *Ds* activity at this new position was correlated with the disappearance of the normal action of the *C* locus. The resulting phenotype was the same as that produced by the known recessive allele *c*. . . . That the *c* phenotype in this case was associated with the appearance of *Ds* at the *C* locus was made evident because mutations from *c* to *C* occurred [in which] *Ds* action concomitantly disappeared. . . . The origin and behavior of this mutation at the *c* locus have been interpreted as follows: Insertion of the chromatin composed of *Ds* adjacent to the *C* locus is responsible for complete inhibition of the action of *C*. Removal of this foreign chromatin can occur. In many cases, the mechanism associated with the removal results in restoration of former genic organization and action. . . . The mutation-producing mechanisms involve only *Ds*. No gene mutations occur at the *C* locus; the restoration of its action is due to the removal of the inhibiting *Ds* chromatin. . . . [The movement] of *Ds* requires an activator. This activator has been designated *Ac*. . . . *Ac* shows a very important characteristic not exhibited in studies of the usual genetic factors. This characteristic is the same as that shown by *Ds*. Transposition of *Ac* takes place from one position in the chromosomal complement to another—very often from one chromosome to another. Again, as in *Ds*, changes in state may occur at the *Ac* locus. . . . It should be emphasized that when no *Ac* is present in a nucleus, no mutation-producing events occur at *Ds*.

Source: Proceedings of the National Academy of Sciences of the USA 36: 344–355

or near its insertion site. Since McClintock's original discovery, many other **transposable elements** have been discovered. They can be grouped into "families" based on similarity in DNA sequence. The genomes of most organisms contain multiple copies of each of several distinct families of transposable elements. Once situated in the genome, transposable elements can persist for long periods and undergo multiple mutational changes. Up to 35 percent of the human genome consists of evolutionary remnants of transposable elements.

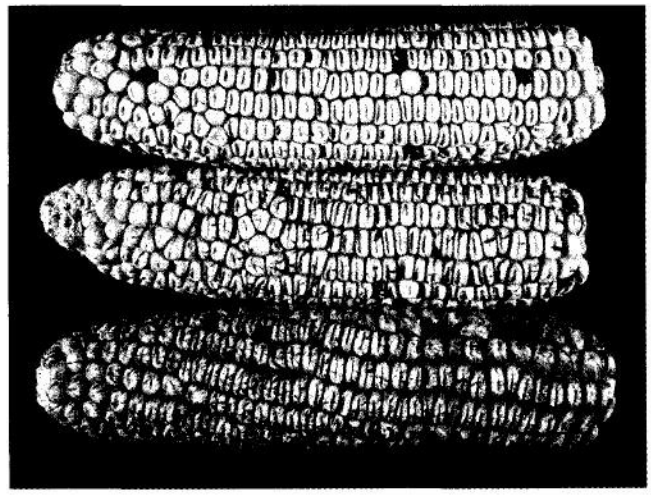


Figure 7.8 Sectors of purple and yellow tissue in the endosperm of maize kernels resulting from the presence of the transposable elements *Ds* and *Ac*. The different level of sectoring in some ears results from dosage effects of *Ac*. [Courtesy of Jerry L. Kermicle.]

Molecular Mechanisms of Transposition

The molecular mechanism of *Ds* transposition is illustrated in **Figure 7.9**. In this example the insertion goes into the wildtype

shrunken gene in maize chromosome 9, causing a knockout mutation. To initiate the process, the target site for insertion is cleaved with a staggered cut, leaving a 3' overhang of 8 nucleotides on each strand. The overhanging 3' ends are ligated with the 5' ends of the *Ds* element to be inserted, leaving an 8-nucleotide gap in each strand. When the gap is filled by repair enzymes, the result is a new insertion of *Ds* flanked by an 8-base-pair duplication of the target sequence. The *Ds* element can insert in either orientation.

The presence of a target-site duplication is characteristic of most transposable element insertions, and it results from asym-

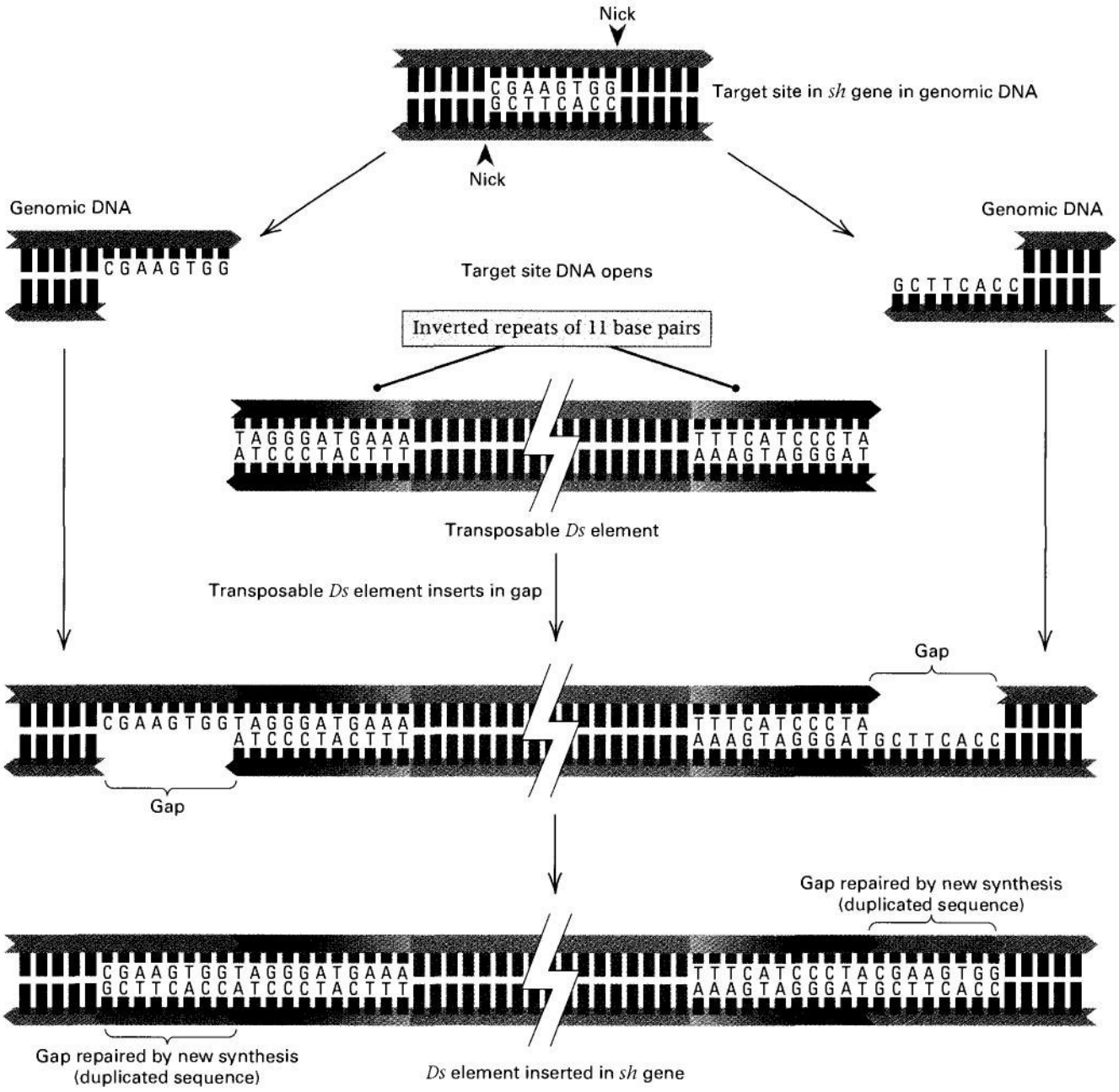


Figure 7.9 The sequence arrangement of a cut-and-paste transposable element (in this case the *Ds* of maize) and the changes that take place when it inserts into the genome. *Ds* is inserted into the maize *sh* gene at the position indicated. In the insertion process, a sequence of eight base pairs next to the site of insertion is duplicated and flanks the *Ds* element.

metrical cleavage of the target sequence. For elements like *Ds*, target-site cleavage is a function of a transposase protein that catalyzes transposition. Each family of transposable elements has its own **transposase** that may differ in the distance between the cuts made in the target DNA strands. Depending on the particular transposable element, the distance may be 1–12 base pairs, and this determines the length of the target-site duplication. Most transposable elements have many potential target sites scattered throughout the genome, and they usually show little or no sequence similarity from one site to the next.

The *Ds* element is one of a large class of transposable elements called **cut-and-paste elements** that insert via a mechanism like that in Figure 7.9. Characteristic of these elements is the presence of *terminal inverted repeats*, a sequence repeated in inverted orientation at each end of the element. In the case of *Ds*, the terminal inverted repeats are 11 bp in length (Figure 7.9), but in other families of transposable elements, they can be up to a few hundred base pairs long. The terminal repeats are usually essential for efficient transposition, because they contain binding sites for the transposase that allow the element to be recognized and ligated into the cleaved target site. Many transposable elements encode their own transposase in sequences located in the central region between the terminal inverted repeats, so these elements are able to promote their own transposition. Elements in which the transposase gene has been deleted or inactivated by mutation are transposable only if another member of the family, encoding a functional transposase, is present in the genome to provide this activity. The inability of the maize *Ds* element to transpose without *Ac* results from the absence of a functional transposase gene in *Ds*. The presence of an *Ac* element provides *trans-activation* that enables a *Ds* element to transpose.

Another large class of transposable elements possess terminal direct repeats, typically 200–500 bp in length, called *long terminal repeats*, or *LTRs*. As the name implies, terminal direct repeats are present in the same orientation at both ends of the element (**Figure 7.10**, part A), whereas terminal inverted repeats are present in reverse orientation (Figure 7.10B). These elements are called **LTR retrotransposons** because they transpose using an RNA transcript as an intermediate. A typical LTR retrotransposon is the *copia* element

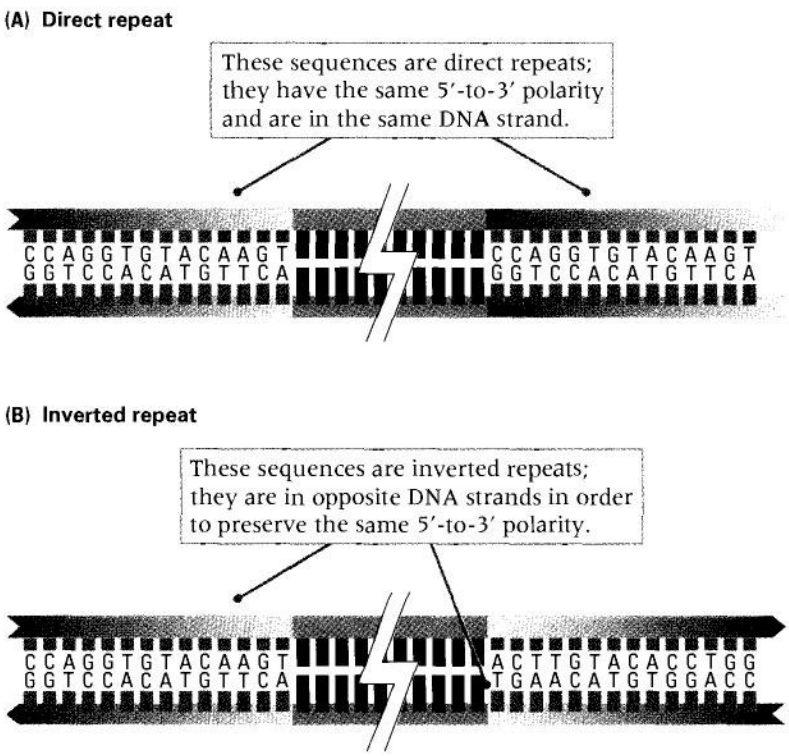


Figure 7.10 (A) In a direct repeat, a DNA sequence is repeated in the same left-to-right orientation. (B) In an inverted repeat, the sequence is repeated in the reverse left-to-right orientation *in the opposite strand*. The opposite strand is stipulated in order to maintain the correct 5'-to-3' polarity.

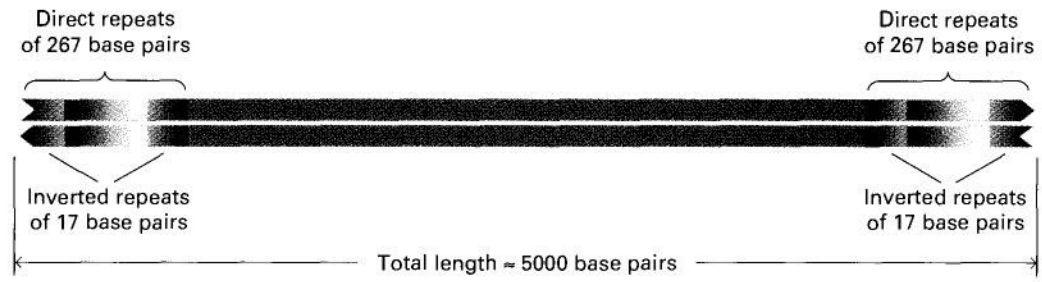


Figure 7.11 Sequence organization of a *copia* retrotransposable element of *Drosophila melanogaster.*

from *Drosophila* illustrated in **Figure 7.11**; in this case each LTR is itself flanked by short terminal inverted repeats. Transposition of a retrotransposon begins with transcription of the element into an RNA copy. Among the encoded proteins is an enzyme known as **reverse transcriptase,** which can "reverse-transcribe," using the RNA transcript as a template for making a complementary DNA daughter strand. A primer is needed for reverse transcription. For retrotransposons it is usually a cellular transfer RNA molecule whose 3' end is complementary to part of the LTR. The reverse transcriptase adds successive deoxyribonucleotides to the 3' end of the tRNA, using the original RNA transcript as a template. Single-stranded cleavage of the RNA template by an element-encoded RNase provides a primer for second-strand DNA synthesis using the first DNA strand as a template. In this way a double-stranded DNA copy is made of the RNA transcript, and this is inserted into the target site in a fashion analogous to the mechanism shown in Figure 7.9, using yet another element-encoded protein. (We have glossed over many details, such as how the primers are removed and replaced with DNA.)

Some retrotransposable elements have no terminal repeats and are called **non-LTR retrotransposons.** This class includes elements denoted *LINE* (long interspersed elements) and *SINE* (short interspersed elements). LINE and SINE elements are the most abundant types of transposable elements in mammalian genomes, although cut-and-paste and LTR retrotransposons are also found. An example of a SINE in the human genome is a set of related sequences called the *Alu* family because its members contain a characteristic restriction site for the restriction endonuclease *Alu*I. The *Alu* sequences are about 300 base pairs in length and are present in approximately 500,000 copies in the human genome. The *Alu* family alone accounts for about 5 percent of human DNA. In many organisms, transposable elements of various families constitute a significant part of the total genome size.

Transposable Elements as Agents of Mutation

Transposable elements are responsible for many mutations in all organisms in which they occur. For example, in some genes in *Drosophila,* approximately half of all spontaneous mutations that have visible phenotypic effects result from insertions of transposable elements. We have already seen (in Figure 7.9) an example of mutation associated with the insertion of *Ds* into the *shrunken* gene in maize. The wrinkled-seed mutation in Mendel's peas is another good example. In this case, the transposable element is a cut-and-paste element related to the maize *Ac* element and also produces an 8-bp target-site duplication. The insertion site is in the gene for starch-branching enzyme I (SBEI), and the insertion creates a loss-of-function allele. Most transposable elements are present in nonessential regions of the genome and usually cause no detectable phenotypic change. But when an element transposes, it can insert into an essential region and disrupt that region's function. If transposition inserts an element into a coding region of DNA, then the inserted element interrupts the coding region. Because most transposable elements contain coding regions of their own, either

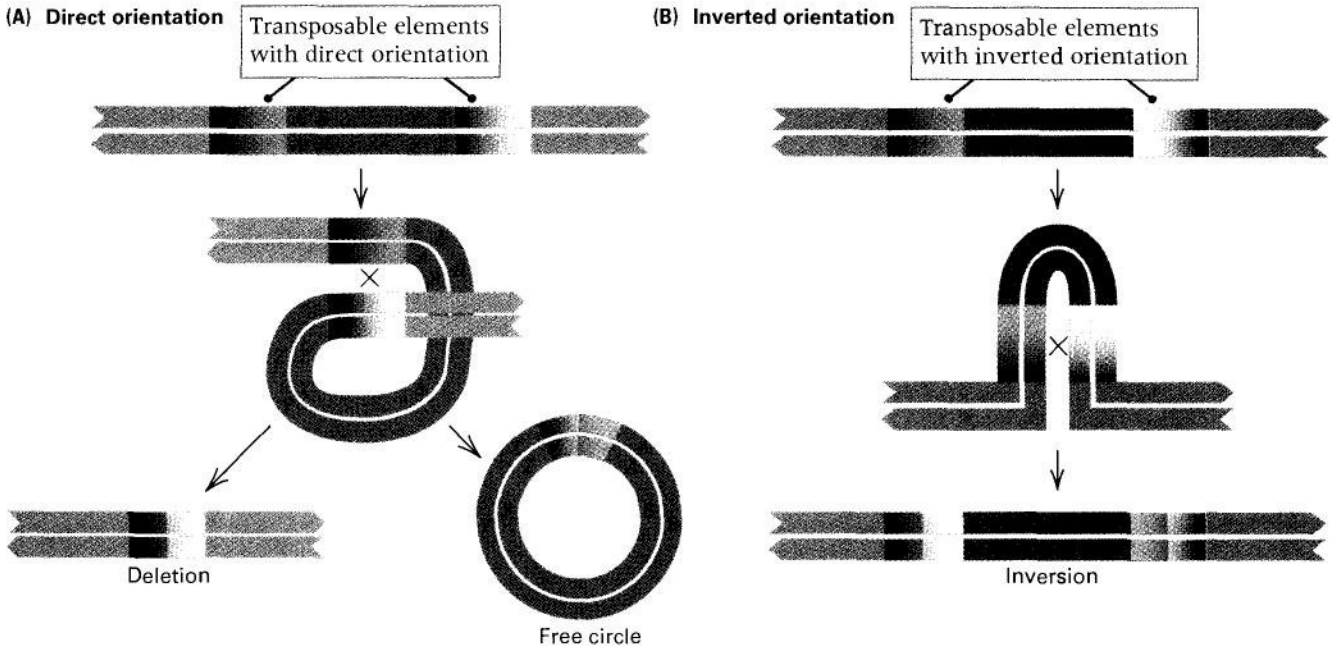
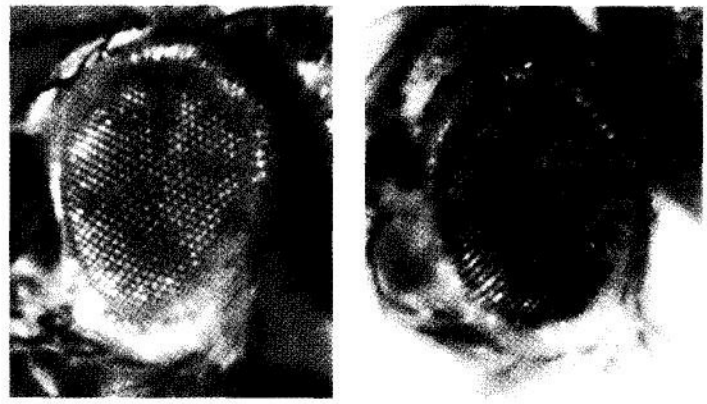


Figure 7.12 Recombination between transposable elements (or other repeated sequences) in the same chromosome. (A) If the repeats are in direct orientation, then recombination results in a deleted duplex and a circular molecule containing the deleted region. (B) If the repeats are in inverted orientation, then recombination results in an inversion of the region between them.

transcription of the transposable element interferes with transcription of the gene into which it is inserted, or transcription of the gene terminates within the transposable element. The insertion therefore causes a knockout mutation. Even if transcription proceeds through the element, the phenotype will be mutant because the coding region then contains incorrect sequences.

Genetic aberrations can also be caused by recombination between different (nonallelic) copies of a transposable element, which is called **ectopic recombination.** **Figure 7.12** illustrates two possible outcomes of recombination between copies present in the same DNA molecule. In part A the copies are present in direct orientation. In this case, pairing between the copies forms a loop. Recombination between the copies results in the formation of a free circle of DNA that contains the region between the elements, whereas the remaining part of the molecule has a deletion of this region. In part B the copies are present in inverted orientation. In this case, pairing between the copies creates a *foldback* or *hairpin* structure instead of a loop, and recombination between the copies results in an *inversion* in which the order of the genes between the elements is reversed.



In these *Drosophila* eyes, the patches of red are caused by excision of a transposable element restoring normal eye color to an otherwise peach-colored eye. The eye with the uniform peach color shows the phenotype of the mutant in a genetic background in which transposition cannot occur.

Ectopic recombination between copies of a transposable element in different DNA

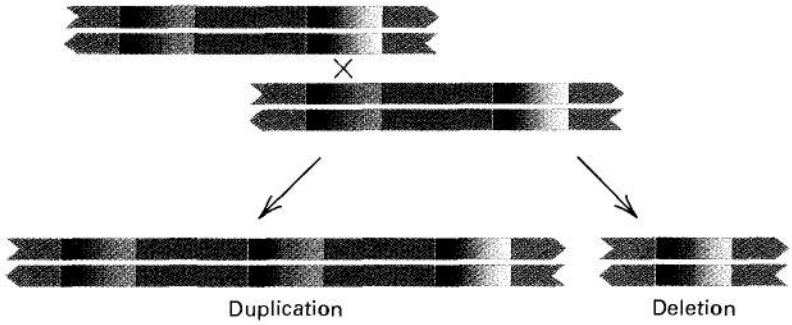


Figure 7.13 Unequal crossing-over between homologous transposable elements present in the same orientation in different chromatids results in products that contain either a duplication or a deletion of the DNA sequences between the elements. Crossing-over takes place at the four-strand stage of meiosis, but only the two strands participating in the crossover are shown.

molecules is shown in **Figure 7.13**. In this case, the DNA molecules are assumed to be those in homologous chromosomes. Recombination between nonallelic copies produces one product with a duplication of the region between the copies and a reciprocal product with a deletion of the same region. A similar picture could be drawn for ectopic recombination between copies in nonhomologous chromosomes. In this case, the result would be an interchange of terminal segments between nonhomologous chromosomes, which is called a *reciprocal translocation* (Chapter 9).

RIP: A Defense Against Transposons

Transposable elements are like molecular parasites in that they invade and multiply within the genome. Organisms have evolved various defense mechanisms to attack sequences that are present in multiple copies in the genome. Most of these are mechanisms of "silencing," in which the sequences remain intact but their transcription is prevented or their transcripts are destroyed. These mechanisms are discussed in the context of gene regulation in Chapter 12. However, some organisms have evolved defense mechanisms that attack repeated sequences by means of mutation. Notable among these is a process known as **RIP (repeat-induced point mutation)**, which occurs in the filamentous fungus *Neurospora crassa* and related species.

RIP takes place in haploid cells undergoing the sexual cycle. Because the cells are haploid, each gene should be present in only a single copy. Sequences that are longer than a few hundred base pairs and that are repeated in the haploid genome are subjected to RIP. The mechanisms of repeat recognition and mutagenesis are poorly understood, but the result of RIP is that the repeated sequences undergo multiple transition mutations in which G–C nucleotide pairs are substituted with A–T nucleotide pairs. The hypermutation creates multiple missense and nonsense mutations in coding sequences that may be present in the repeated sequences, effectively eliminating a functional gene product. In addition, many of the remaining C-bearing nucleotides in the repeats become methylated, which decreases transcription. (The role of cytosine methylation in gene regulation is discussed in Chapter 12.) RIP is so efficient at detecting and mutating repeats that it affords a convenient approach to obtaining new mutations in single-copy genes. The procedure is to introduce extra partial or complete copies of the gene of interest into the genome by DNA transformation, after which RIP rips into the repeats, including the original single-copy gene that is present. The *Neurospora* genome does, however, include some sequences that are tandemly repeated, as they are in other organisms, such as those transcribed into ribosomal RNA and those encoding the histone chromosomal proteins; these sequences are protected from RIP by an unknown mechanism.

7.4 Spontaneous Mutation

Mutations are statistically random events. There is no way of predicting when, or in which cell, a mutation will take place. However, every gene mutates spontaneously at a characteristic rate, so it is possible to assign probabilities to particular mutational events. The various kinds of mutational alterations in DNA differ substantially in complexity, so their probabilities of realization are quite different. The mutational process is also random in the sense that whether a particular mutation happens is unrelated to any adaptive advantage it may confer on the organism in its environment.

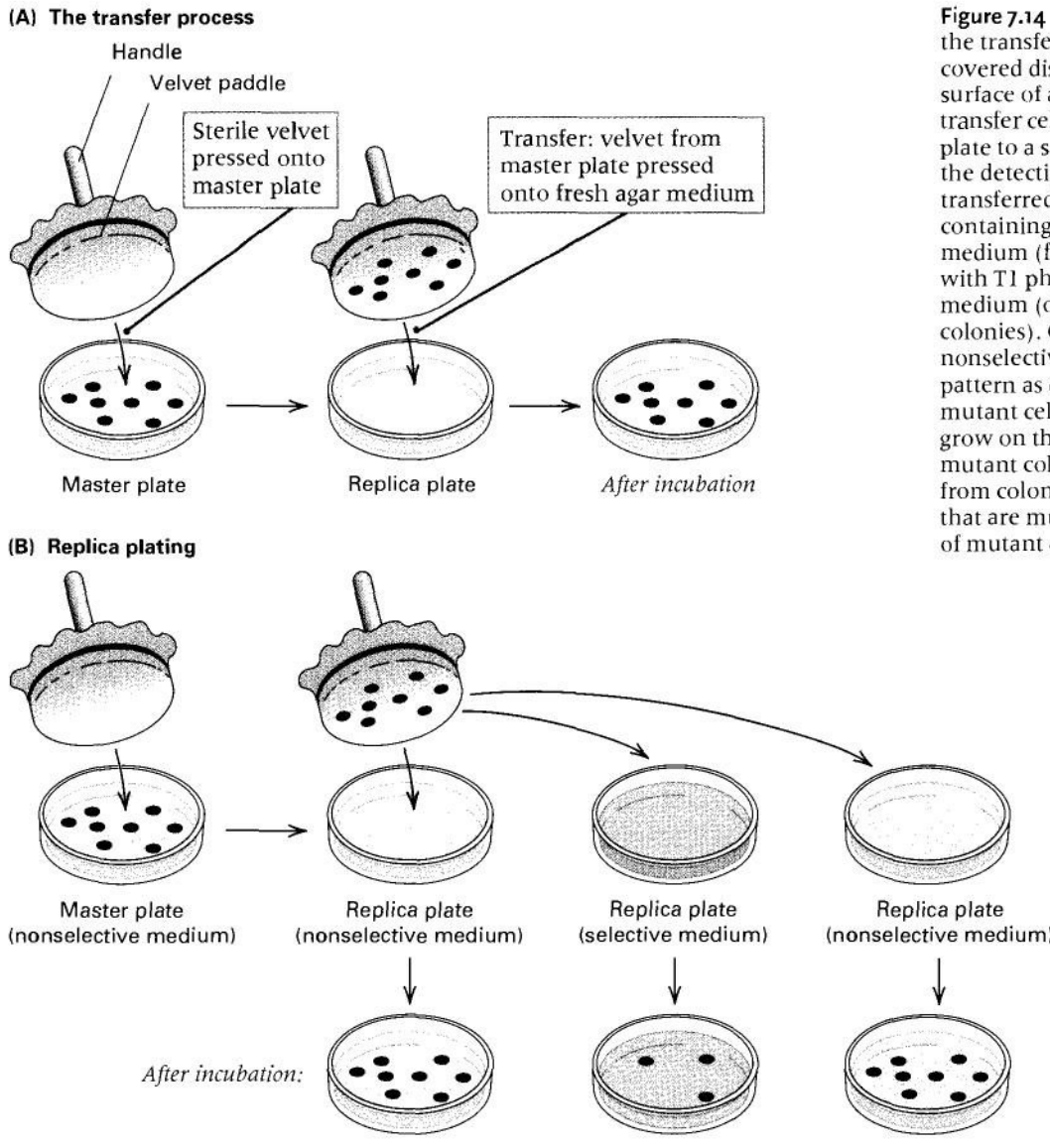


Figure 7.14 Replica plating. (A) In the transfer process, a velvet-covered disk is pressed onto the surface of a master plate in order to transfer cells from colonies on that plate to a second medium. (B) For the detection of mutants, cells are transferred onto successive plates containing either a selective medium (for example, one spread with T1 phages) or a nonselective medium (on which all cells form colonies). Colonies form on the nonselective plate in the same pattern as on the master plate. Only mutant cells (for example, T1-r) can grow on the selective plate; the mutant colonies that form derive from colonies on the master plate that are mutant. Colonies consisting of mutant cells are shown in red.

A potentially favorable mutation does not arise *because* the organism has a need for it. The experimental basis for this conclusion is presented in the next section.

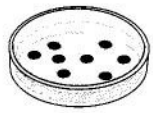
The Nonadaptive Nature of Mutation

The concept that mutations are spontaneous, statistically random events unrelated to adaptation was not widely accepted until the late 1940s. Before that time, it was believed that mutations occurred in bacterial populations *in response to* particular selective conditions. The basis for this belief was the observation that when antibiotic-sensitive bacteria are spread on a solid growth medium containing the antibiotic, some colonies form that consist of cells that have an inherited resistance to the drug. The initial interpretation of this observation (and similar ones) was that these adaptive variations were *induced* by the selective agent itself.

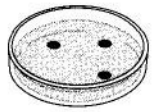
Several types of experiments showed that adaptive mutations take place spontaneously and hence were present at low frequency in the bacterial population even *before* it was exposed to the antibiotic. One experiment utilized a technique developed by Joshua and Esther Lederberg called **replica plating** (Figure 7.14). In this procedure, a suspension of bacterial cells is spread on a solid medium. After colonies



Velvet paddle



Replica plate (nonselective medium)



Replica plate (selective medium)

have formed, a piece of sterile velvet mounted on a solid support is pressed onto the surface of the plate. Some bacteria from each colony stick to the fibers, as shown in Figure 7.3A. Then the velvet is pressed onto the surface of fresh medium, transferring some of the cells from each colony, which give rise to new colonies that have positions identical to those on the first plate. Figure 7.3B shows how this method was used to demonstrate the spontaneous origin of mutant *E. coli* that are resistant to the bacteriophage T1 (T1-r mutants). A master plate containing about 10^7 cells growing on nonselective medium (lacking T1 phage) was replica-plated onto a series of plates that had been spread with about 10^9 T1 phages. After incubation for a time sufficient for colony formation, a few colonies of phage-resistant bacteria appeared in the same positions on each of the selective replica plates. This meant that the T1-r cells that formed the colonies must have been transferred from corresponding positions on the master plate. Because the colonies on the master plate had never been exposed to the phage, the mutations to resistance must have been present, by chance, in a few original cells not exposed to the phage.

The replica-plating experiment illustrates that

> Selective techniques merely select mutants that preexist in a population.

This principle is the basis for understanding how natural populations of rodents, insects, and disease-causing bacteria become resistant to the chemical substances used to control them. A familiar example is the high level of resistance to insecticides, such as DDT, that now exists in many insect populations. The resistance is due to selection for spontaneous mutations affecting behavioral, anatomical, and enzymatic traits that enable the insect to become tolerant to the chemical. Similar problems are encountered in controlling plant pathogens. For example, the introduction of a new variety of a crop plant that is resistant to a particular strain of disease-causing fungus results in only temporary protection against the disease. The resistance inevitably breaks down because of the occurrence of spontaneous mutations in the fungus that enable it to attack the new plant genotype. Such mutations have a clear selective advantage, and the mutant alleles rapidly become widespread in the fungal population.

Estimation of Mutation Rates

Spontaneous mutations tend to be rare (except for unusual phenomena such as dynamic mutation), and the methods used to estimate the frequency with which they arise require large populations and (often) special techniques. The **mutation rate** is the probability that a gene undergoes mutation in a single generation or in forming a single gamete. Measurement of mutation rates is important in population genetics, in studies of evolution, and in analysis of the effects of environmental mutagens.

One of the earliest techniques to be developed for measuring mutation rates is the ***ClB* method.** *ClB* refers to a special X chromosome of *Drosophila melanogaster.* This X chromosome has a large inversion (*C*) that (for reasons described in Chapter 9) prevents the recovery of recombinant chromosomes in the progeny from a female heterozygous for the chromosome; a recessive lethal (*l*); and the dominant marker Bar (*B*), which reduces the normal round eye to a bar shape. The presence of a recessive lethal in the X chromosome means that males with that chromosome and females homozygous for it cannot survive. The technique is designed to detect mutations arising in a normal X chromosome.

In the *ClB* procedure, females heterozygous for the *ClB* chromosome are mated with males that carry a normal X chromosome (**Figure 7.15**). Females with the Bar phenotype are chosen from the progeny and then individually mated with normal males. (The presence of the Bar phenotype indicates that the females are heterozygous for the *ClB* chromosome and the normal X chromosome from the male parent.) A ratio of two females to one male is expected among the progeny from such a cross, as shown in the lower part of the illustration. The critical observation in determining the mutation rate is the fraction of males produced in the second cross. Because the *ClB* males die, all of the males in this generation must contain an X chromosome derived from the X chromosome of the initial normal male (top row of illustration). Furthermore, it must be a nonrecombinant X

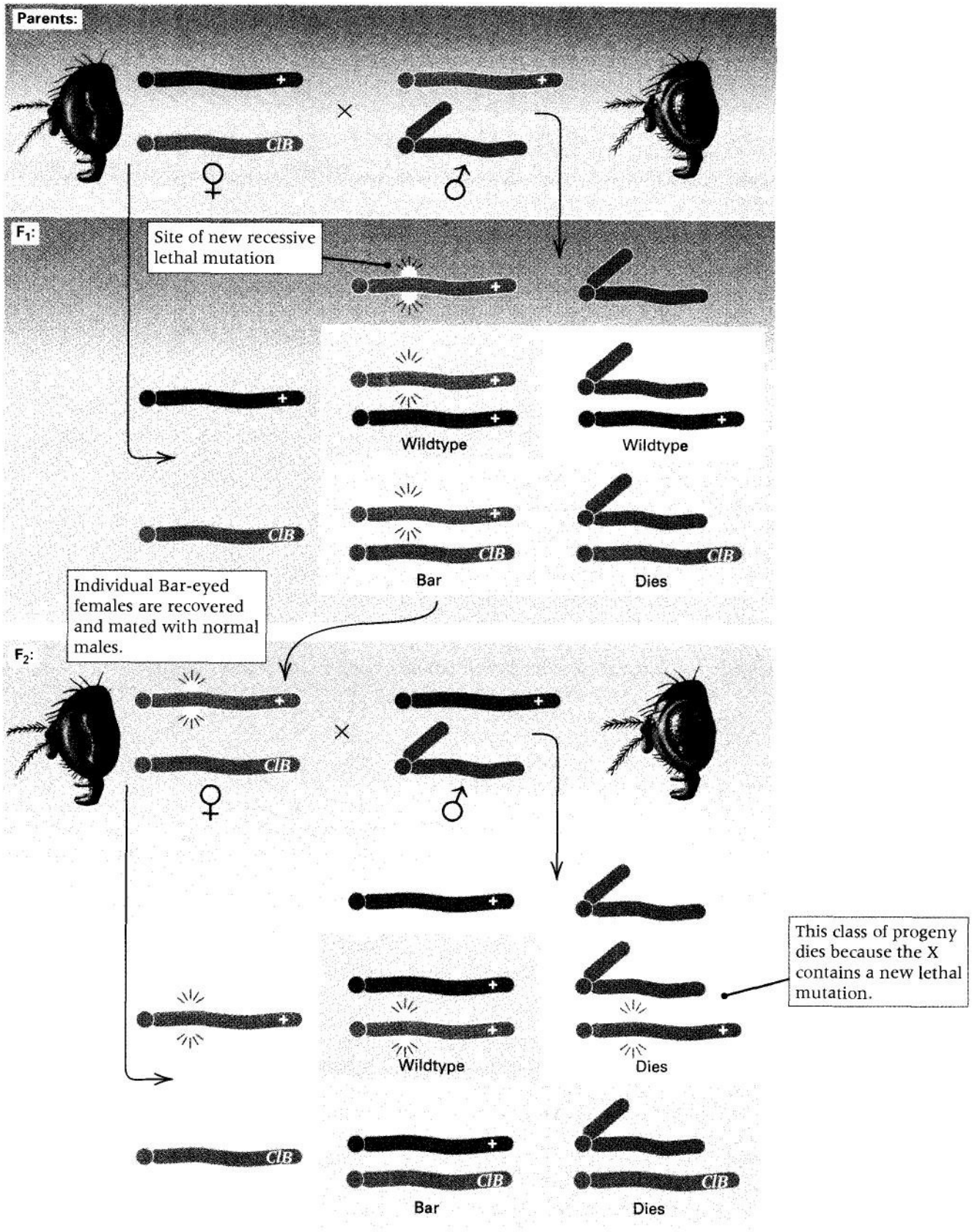


Figure 7.15 The *ClB* method for estimating the rate at which spontaneous recessive lethal mutations arise on the *Drosophila* X chromosome.

chromosome because of the inversion (*C*) in the *ClB* chromosome. Occasionally, the progeny include no males, and this means that the X chromosome present in the original sperm underwent a mutation somewhere along its length to yield a new recessive lethal. The method provides a quantitative estimate of the rate at which mutation to a recessive lethal allele occurs *at any of the large number of genes* in the X chromosome. About 0.15 percent of the X chromosomes acquire new recessive lethals in spermatogenesis; this means that the mutation rate is 1.5×10^{-3} recessive lethals per X chromosome per generation. Note that the *ClB* method tells us nothing about the mutation rate for a particular gene, because the method does not reveal how many genes on the X chromosome would cause lethality if they were mutant. Since the time that the *ClB* method was devised, a variety of other methods have been developed for determining mutation rates in *Drosophila* and other organisms. Of significance is the fact that mutation rates vary widely from one gene to another. For example, the yellow-body mutation in *Drosophila* occurs at a frequency of 10^{-4} per gamete per generation, whereas mutations to streptomycin resistance in *E. coli.* occur at a frequency of 10^{-9} per cell per generation. Furthermore, within a single organism, the frequency can vary enormously; in *E. coli* it ranges from 10^{-5} for some genes to 10^{-9} for others.

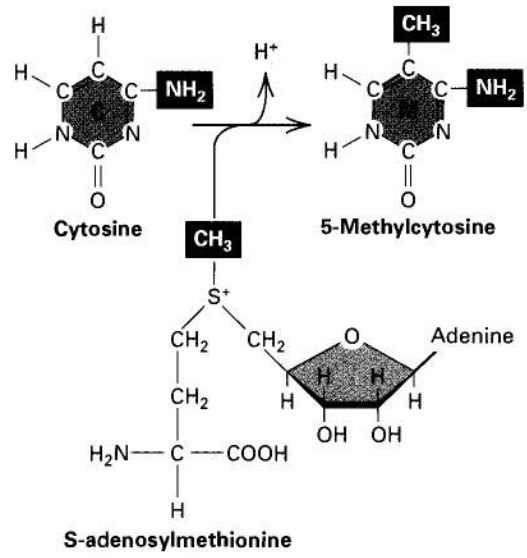


Figure 7.16 Methylation of cytosine at the number-5 position in the base. The methyl donor is S-adenosylmethionine.

Hot Spots of Mutation

Certain DNA sequences are called mutational **hot spots** because they are more likely to undergo mutation than others. Mutational hot spots may include monotonous runs of a single nucleotide or tandem repeats of a short sequence, such as an unstable trinucleotide repeat, that may expand or contract by replication slippage or other mechanisms. Hot spots are found at many sites throughout the genome and within genes. For genetic studies of mutation, the existence of hot spots means that a relatively small number of sites accounts for a disproportionately large fraction of all mutations. For example, in the analysis of the *rII* gene in bacteriophage T4 (Chapter 10), one extreme hot spot accounted for approximately 20 percent of all the mutations obtained.

In many organisms, including bacteria, maize, and mammals (but not yeast, fruit flies, or nematodes), at least 1 percent of the cytosine bases are methylated at the carbon-5 position, yielding 5-methylcytosine instead of ordinary cytosine (**Figure 7.16**). The methylation is carried out in duplex DNA and catalyzed by a *cytosine methylase* that uses S-adenosylmethionine as the donor of the methyl group ($-CH_3$). The methylation is not random but, rather, occurs preferentially at certain target sequences that contain cytosine—in mammals at cytosines in CG dinucleotides. In DNA replication, the 5-methylcytosine pairs with guanine and replicates normally. The genetic function of cytosine methylation is discussed in Chapter 12. Relative to mutation, cytosine methylation is an important contributor to mutational hot spots because the methylated cytosines are usually highly mutable. The mutations are usually G–C → A–T transitions, for the reason illustrated in **Figure 7.17**. Both cytosine and 5-methylcytosine are subject to occasional loss of an amino group. For cytosine, this loss yields *uracil.* Because uracil pairs with adenine instead of guanine, replication of a

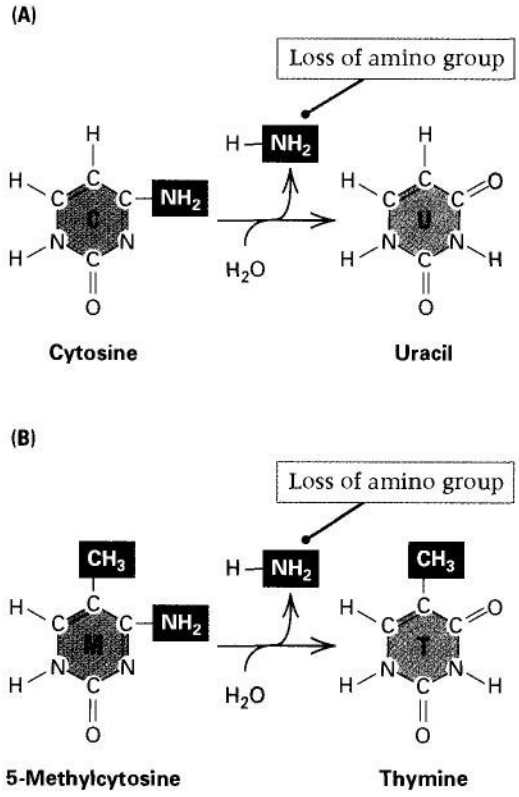


Figure 7.17 (A) Spontaneous loss of the amino group of cytosine to yield uracil. (B) Spontaneous loss of the amino group of 5-methylcytosine to yield thymine.

molecule containing a G–U base pair would ultimately lead to substitution of an A–T pair for the original G–C pair, because in successive rounds of replication the affected base pair becomes G–U → A–U → A–T. However, cells possess a special repair enzyme that specifically removes uracil from DNA. The repair process is initiated as shown in **Figure 7.18**. Figure 7.18A shows deamination of cytosine leading to the presence of a uracil-containing base. The uracil is recognized by an enzyme called **DNA uracil glycosylase,** which recognizes the incorrect G–U base pair and cleaves the offending uracil from the deoxyribose sugar to which it is attached (Figure 7.18B). This enzyme works by scanning along duplex DNA until a uracil nucleotide is encountered, at which point a specific arginine residue in the enzyme intrudes into the DNA through the minor groove. The intrusion compresses the DNA backbone flanking the uracil and results in the uracil flipping out so that it becomes accessible to cleavage by the uracil glycosylase. This leaves behind a deoxyribose sugar in the duplex DNA that lacks any base (Figure 7.18C). The base-less sugar is then corrected by a DNA repair mechanism called *AP repair,* which is discussed in Section 7.5. The end result of the repair is that the original C → U change rarely leads to mutation because the U is removed and replaced with C again, restoring the original C–G base pair. In contrast, loss of the amino group of 5-methylcytosine yields *thymine* (Figure 7.17B), which is a normal DNA base and hence is not removed by the DNA uracil glycosylase. In this case, deamination changes the original G–C pair into a G–T pair, which in the next round of replication gives G–C (wildtype) and A–T (mutant) daughter molecules.

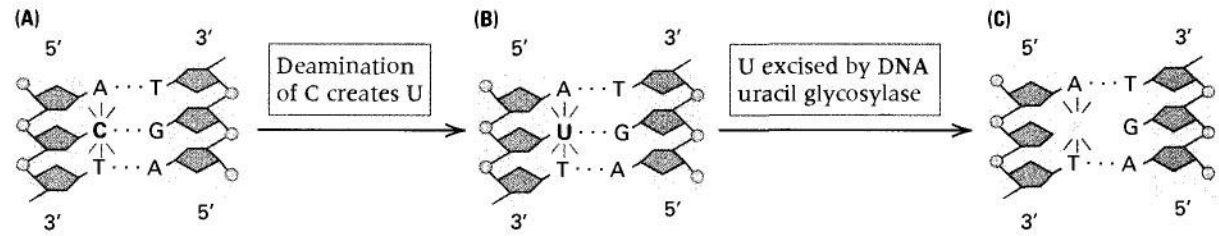


Figure 7.18 Mechanism by which uracil-containing nucleotides are formed in DNA and removed. (A) Cytosine present in a DNA duplex. (B) Deamination of the cytosine produces uracil. (C) Uracil is cleaved from the deoxyribose sugar by DNA uracil glycosylase. The deoxyribose with the uracil detached is then excised from the DNA backbone by another enzyme (AP endonuclease), and the gap is repaired.

7.5 Mutagens

A **mutagen** is any agent that causes an increase in the rate of mutation above the spontaneous background. The first evidence that external agents could increase the mutation rate was presented in 1927 by Hermann Muller, who showed that x rays are mutagenic in *Drosophila*. Since then, a large number of physical agents and chemicals have been shown to increase the mutation rate. Because some environmental contaminants are mutagenic, as are numerous chemicals found in tobacco products, mutagens are of great importance in public health.

Base-Analog Mutagens

A **base analog** is a molecule sufficiently similar to one of the four DNA bases that it can be incorporated into a DNA duplex in the course of normal replication. Such a substance must be able to pair with a normal base in the template strand. Some base analogs are mutagenic because they are more prone to mispairing than are the normal bases. This mechanism of mutagenesis can be illustrated with 5-bromouracil (Bu), a commonly used base analog that is efficiently incorporated into the DNA of bacteria and viruses.

The base 5-bromouracil is an analog of thymine, and the bromine atom is about the same size as the methyl group of thymine (**Figure 7.19**, part A). If cells are grown in a medium containing 5-bromouracil, a thymine is sometimes replaced by a 5-bromouracil during replication, resulting in the formation of an A-Bu pair at a normal AT site (Figure 7.19B). The subsequent mutagenic activity of the incorporated 5-bromouracil stems from a rare shift in the configuration of the base, which is promoted by the bromine atom. In the mutagenic configuration, 5-bromouracil pairs preferentially with guanine (Figure 7.19C), so in the next round of replication, one daughter molecule has a G–C pair at the altered site, yielding an A–T → G–C transition (**Figure 7.20**).

Experiments also suggest that 5-bromouracil is mutagenic in another way. The concentration of the nucleoside triphosphates in most cells is regulated by the concentration of thymidine triphosphate (dTTP). This regulation results in appropriate relative amounts of the four triphosphates for DNA synthesis. One aspect of the regulatory process is inhibition of the syn-

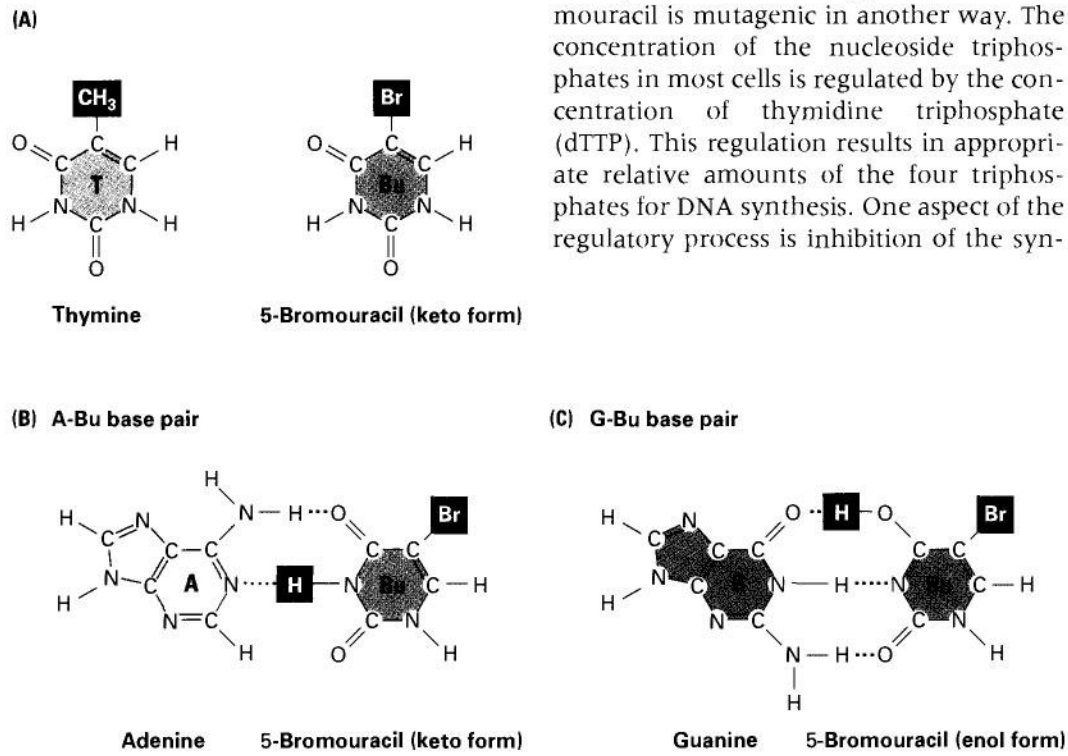


Figure 7.19 Mispairing mutagenesis by 5-bromouracil. (A) Structures of thymine and 5-bromouracil. (B) A base pair between adenine and the keto form of 5-bromouracil. (C) A base pair between guanine and the rare enol form of 5-bromouracil. One of 5-bromouracil's hydrogen atoms changes position to create the keto form.

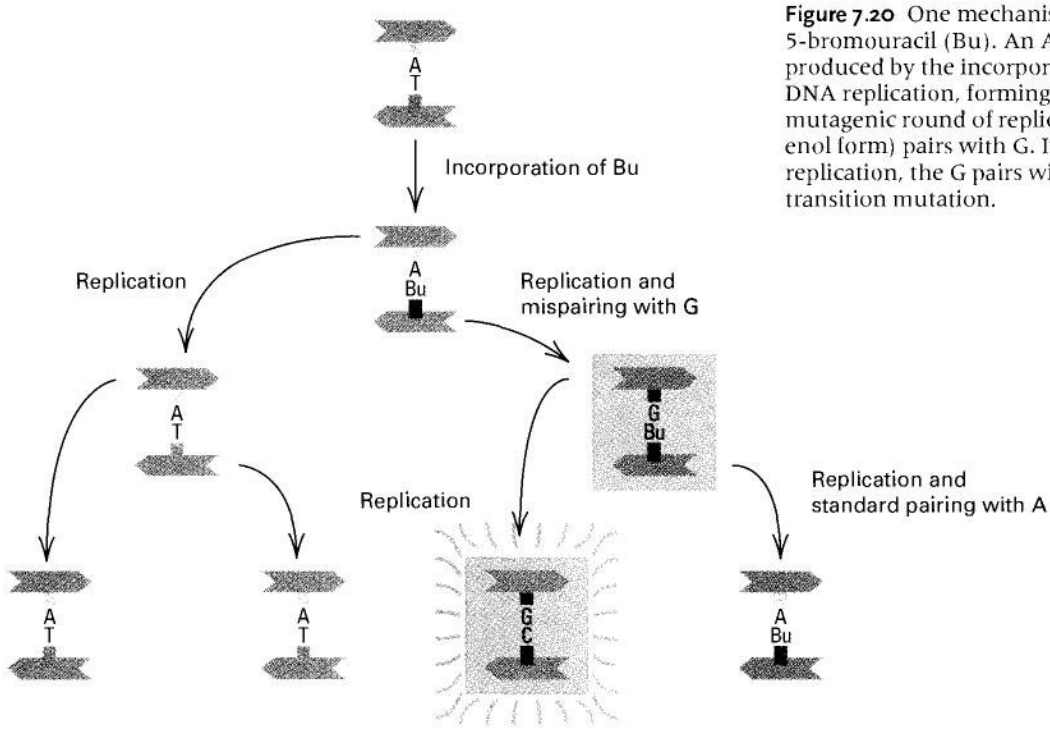
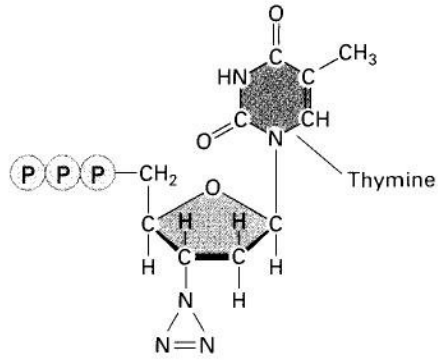


Figure 7.20 One mechanism for mutagenesis by 5-bromouracil (Bu). An AT → GC transition is produced by the incorporation of 5-bromouracil in DNA replication, forming an A-Bu pair. In the mutagenic round of replication, the Bu (in its rare enol form) pairs with G. In the next round of replication, the G pairs with C, completing the transition mutation.

thesis of deoxycytidine triphosphate (dCTP) by excess dTTP. The 5-bromouracil nucleoside triphosphate also inhibits production of dCTP. When 5-bromouracil is added to the growth medium, dTTP continues to be synthesized by cells at the normal rate, whereas the synthesis of dCTP is significantly reduced. The dTTP : dCTP ratio then becomes quite high, and the frequency of misincorporation of T opposite G increases. Although repair systems can remove some incorrectly incorporated thymine, in the presence of 5-bromouracil the rate of misincorporation can exceed the rate of correction. An incorrectly incorporated T pairs with A in the next round of DNA replication, yielding a G–C → A–T transition in one of the daughter molecules. Thus 5-bromouracil usually induces transitions in both directions: A–T → G–C by the route in Figure 7.20, and G–C → A–T by the misincorporation route.

Just as a base analog is incorporated into DNA in place of a normal base, a **nucleotide analog** can be incorporated in place of a normal nucleotide. Modern DNA-sequencing methods are based on the use of dideoxy nucleotide analogs to terminate strand elongation at predetermined sites (Chapter 6). Dideoxy nucleotide analogs are also used in the clinical treatment of various diseases, including cancer and viral infections (Section 6.7). One of the most important drugs in the treatment of AIDS is the dideoxy analog AZT (**Figure 7.21**), which inhibits replication of the HIV virus because the polymerase of the virus incorporates AZT in preference to the normal deoxythymidine.



3'-Azido-2',3'-deoxythymidine (AZT)

Figure 7.21 Chemical structure of AZT, which is used in the clinical treatment of AIDS.

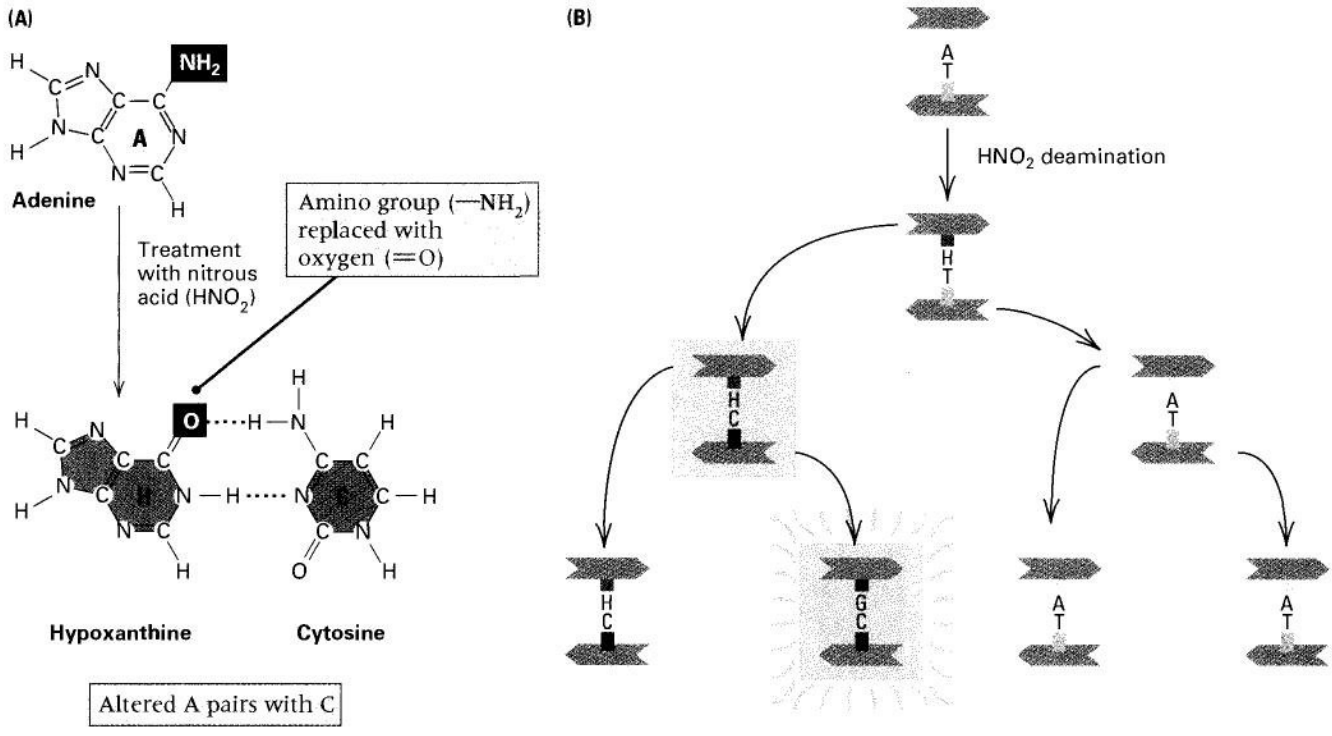


Figure 7.22 Nitrous acid mutagenesis. (A) Conversion of adenine into hypoxanthine, which pairs with cytosine. (The cytosine → uracil conversion is shown in Figure 7.17.) (B) Production of an AT → GC transition. In the mutagenic round of replication, the hypoxanthine (H) pairs with C. In the next round of replication, the C pairs with G, completing the transition mutation.

Chemical Agents That Modify DNA

Many mutagens are chemicals that react with DNA and change the hydrogen-bonding properties of the bases. These mutagens are active on both replicating and nonreplicating DNA, in contrast to the base analogs, which are mutagenic only when DNA replicates. Several of these chemical mutagens, of which nitrous acid is a well-understood example, are highly specific in the changes they produce. Others—for example, the alkylating agents—react with DNA in a variety of ways and produce a broad spectrum of effects.

Nitrous acid (HNO_2) acts as a mutagen by converting amino (NH_2) groups of the bases adenine, cytosine, and guanine into keto (=O) groups. This process of *deamination* alters the hydrogen-bonding specificity of each base. Deamination of adenine yields a base, hypoxanthine, that pairs with cytosine rather than thymine, resulting in an A–T → G–C transition (**Figure 7.22**). As discussed earlier, deamination of cytosine yields uracil (Figure 7.17A), which pairs with adenine instead of guanine, producing a G–C → A–T transition.

In genetic research, chemical mutagens such as nitrous acid (and many others) are exceedingly useful in prokaryotic systems but are not particularly useful as mutagens in higher eukaryotes because they are difficult to introduce into eukaryotic cells. However, alkylating agents are highly effective in eukaryotes, as well as prokaryotes. **Alkylating agents,** such as ethyl methanesulfonate (EMS) and nitrogen mustard (the structures of which are shown in **Figure 7.23**), are potent mutagens that have been used extensively in genetic research. These agents add to the DNA bases various chemical groups that either alter their base-pairing properties or cause struc-

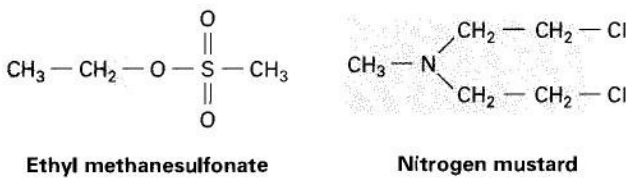
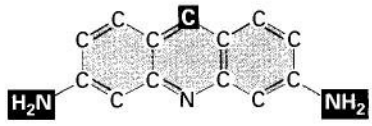


Figure 7.23 The chemical structures of two highly mutagenic alkylating agents; the alkyl groups are shown in red.



Proflavin

Figure 7.24 The structure of proflavine, an acridine derivative. Other mutagenic acridines have additional atoms on the NH_2 group and on the C of the central ring. Hydrogen atoms are not shown.

tural distortion of the DNA molecule. Alkylation of either guanine or thymine causes mispairing, leading to the transitions A–T → G–C and G–C → A–T. EMS reacts less readily with adenine and cytosine.

Misalignment Mutagenesis

The **acridine** molecules, of which proflavin is an example (**Figure 7.24**), are planar, three-ringed molecules whose dimensions are roughly the same as those of a purine–pyrimidine pair. These substances insert between adjacent base pairs of DNA, a process called **intercalation.** The effect of the intercalation of an acridine molecule is to cause the adjacent base pairs to move apart by a distance roughly equal to the thickness of one pair (**Figure 7.25**). When DNA that contains intercalated acridines is replicated, the template and daughter DNA strands can misalign at the sites of intercalation, with the result that a nucleotide is either added or deleted in the daughter strand. In a coding region, the result of a single-base addition or deletion is a frameshift mutation (Section 7.2).

Ultraviolet Irradiation

Ultraviolet (UV) light is a mutagen in all viruses and cells. The effects are caused by chemical changes in the bases resulting from absorption of the energy of the light. The major products formed in DNA after UV irradiation are covalently joined pyrimidines (**pyrimidine dimers**), primarily thymine (**Figure 7.26**, part A), that are adjacent in the same polynucleotide strand. This chemical linkage brings the bases closer together, causing a distortion of the helix (Figure 7.26B), which blocks transcription and transiently blocks DNA replication. Pyrimidine dimers can be repaired in ways discussed later in this chapter.

The autosomal recessive disease *xeroderma pigmentosum* is the result of a defect in a system that repairs ultraviolet-damaged DNA. Persons with this disease are extremely sensitive to sunlight, which results in excessive skin pigmentation and the development of numerous skin lesions that frequently become cancerous. Removal of pyrimidine dimers does not occur in the DNA of cells cultured from patients with the disease, and the cells are sensitive to much lower doses of ultraviolet light than are cells from normal persons.

Even in normal people, excessive exposure to the UV rays in sunlight increases the risk of skin cancer. Ozone, a form of oxygen (O_3) present high in the atmosphere (at a height of 15–30 kilometers), absorbs UV rays and ameliorates, to some extent, the damaging effects of sunlight. The thinning of the ozone layer (the "ozone

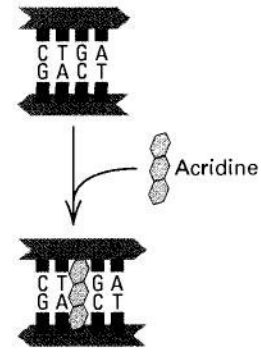
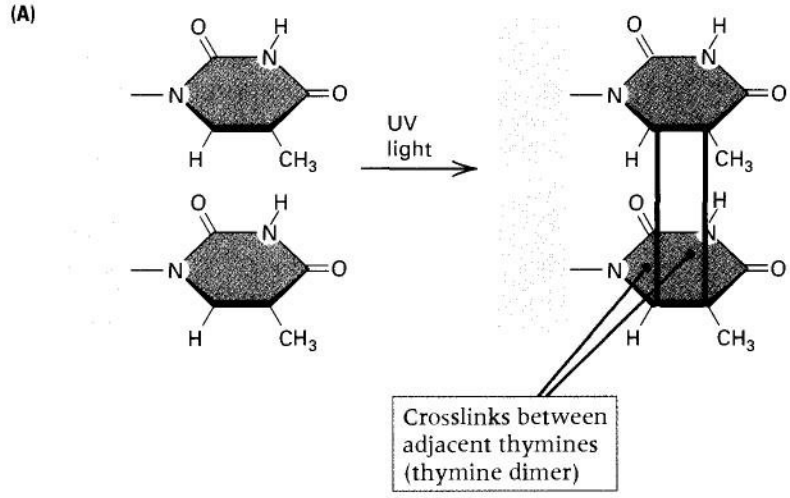


Figure 7.25 Separation of two base pairs caused by intercalation of an acridine molecule.



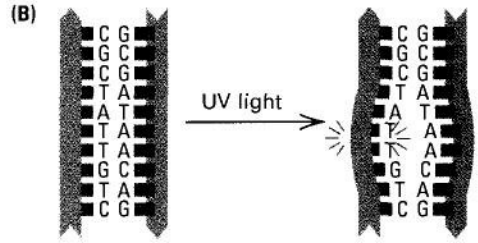


Figure 7.26 (A) Structural view of the formation of a thymine dimer. Adjacent thymines in a DNA strand that have been subjected to ultraviolet (UV) irradiation are joined by formation of the bonds shown in red. Other types of bonds between the thymine rings also are possible. Although not drawn to scale, these bonds are considerably shorter than the spacing between the planes of adjacent thymines, so the double-stranded structure becomes distorted. The shape of each thymine ring also changes. (B) The distortion of the DNA helix caused by two thymines moving closer together when joined in a dimer.

hole"), which is especially pronounced near the Earth's poles, greatly increases the effective exposure to UV and correspondingly increases the rate of skin cancer. The thinning of the ozone layer is attributed in part to the destructive effects of chlorofluorocarbons (CFCs) used as pressurizers in aerosol spray cans and as refrigerants, when these agents ascend to the upper atmosphere.

Ionizing Radiation

Ionizing radiation includes x rays and the particles and radiation released by radioactive elements (α and β particles and γ rays). When x rays were first discovered late in the nineteenth century, their power to pass through solid materials was regarded as a harmless entertainment and a source of great amusement:

> By 1898, personal x rays had become a popular status symbol in New York. The *New York Times* reported that "there is quite as much difference in the appearance of the hand of a washerwoman and the hand of a fine lady in an x-ray picture as in reality." The hit of the exhibition season was Dr. W. J. Morton's full-length portrait of *"the x-ray lady,"* a "fashionable woman who had evidently a scientific desire to see her bones." The portrait was said to be a "fascinating and coquettish" picture, the lady having agreed to be photographed without her stays and corset, the better to satisfy the "longing to have a portrait of well-developed ribs." Dr. Morton said women were not afraid of x rays: "After being assured that there is *no danger* they take the rays without fear."
>
> The titillating possibility of using x rays to see through clothing or to invade the privacy of locked rooms was a familiar theme in popular discussions of x rays and in cartoons and jokes. Newspapers carried advertisements for *"x-ray proof underclothing"* for those seeking to protect themselves from x-ray inspection.
>
> The luminous properties of radium soon produced a full-fledged radium craze. A famous woman dancer performed *radium dances* using veils dipped in fluorescent salts containing radium. *Radium roulette* was popular at New York casinos, featuring a "roulette wheel washed with a radium solution, such that it glowed brightly in the darkness; an unseen hand cast the ball on the turning wheel and sparks marked its course as it bounded from pocket to glimmery pocket." A patent was issued for a process for making women's gowns luminous with radium, and Broadway producer Florenz Ziegfeld snapped up the rights for his stage extravaganzas.
>
> Even while the unrestrained use of x rays and radium was growing, evidence was accumulating that the new forces might not be so benign after all. Hailed as tools for fighting cancer, they could also cause cancer. Doctors using x rays were the first to learn this bitter lesson. (Quoted from S. Hilgartner, R. C. Bell, and R. O'Connor, *Nukespeak* (San Francisco: Sierra Club Books, 1982).

Doctors were indeed the first to learn the lesson. Many suffered severe x-ray burns or required amputation of overexposed hands or arms. Many others died from radiation poisoning or from radiation-induced cancer. By the mid-1930s, the number of x-ray deaths had grown so large that a monument to the "x-ray martyrs" was erected in a hospital courtyard in Germany. Yet the full hazards of x-ray exposure were not widely appreciated until the 1950s.

When ionizing radiation interacts with water or with living tissue, highly reactive ions called *free radicals* are formed. The free radicals react with other molecules, including DNA, which results in the carcinogenic and mutagenic effects. The intensity of a beam of ionizing radiation can be described quantitatively in several ways. There are, in fact, a bewildering variety of units in common use (**Table 7.1**). Some of the units (becquerel, curie) deal with the number of disintegrations emanating from a material,

Table 7.1 Units of radiation

Unit (abbreviation)	Magnitude
becquerel (Bq)*	1 disintegration/second = 2.7×10^{-11} Ci
curie (Ci)	3.7×10^{10} disintegrations/second = 3.7×10^{10} Bq
gray (Gy)*	1 joule/kilogram = 100 rad
rad (rad)	100 ergs/gram = 0.01 Gy
rem (rem)	damage to living tissue done by 1 rad = 0.01 Sv
roentgen (R)	produces 1 electrostatic unit of charge per cubic centimeter of dry air under normal conditions of pressure and temperature. (By definition, 1 electrostatic unit repels with a force of 1 dyne at a distance of 1 centimeter.)
sievert (Sv)	100 rem

*Units officially recognized by the International System of Units as defined by the General Conference on Weights and Measures.

ConNection

X-Ray Daze

Hermann J. Muller 1927
University of Texas, Austin, Texas
Artificial Transmutation of the Gene

Mutagenesis, which means the deliberate induction of mutations, plays an important role in genetics because it makes possible the identification of genes that control biological processes, such as development and behavior. For the demonstration that x rays could be used to induce new mutations, Muller was awarded a Nobel Prize in 1946. His discovery also had important practical implications in regard to x rays (and other mutagenic agents) present in the environment. After World War II ended with the nuclear bombing of Hiroshima and Nagasaki, Muller became a leader in the fight against the above-ground testing of nuclear bombs because of the release of radioactivity into the atmosphere. He also condemned the casual use of x rays, such as for the fitting of children's shoes, and he campaigned for the engineering of safer x-ray machines for medical use. Muller's prose could be turgid at times, thanks to his use of subordinate clauses. This excerpt captures the style.

Most modern geneticists will agree that gene mutations form the chief basis of organic evolution, and therefore of most of the complexities of living things. Unfortunately for the geneticists, however, the study of these mutations, and, through them, of the genes themselves, has heretofore been very seriously hampered by the extreme infrequency of their occurrence under ordinary conditions, and by the general unsuccessfulness of attempts to modify decidedly, and in a sure and detectable way, this sluggish "natural" mutation rate. . . . On theoretical grounds, it has appeared to the present writer that radiations of short wave length should be especially promising for the production of mutational changes, and for this and other reasons a series of experiments concerned with this problem has been undertaken during the past year on the fruit fly, *Drosophila melanogaster*. . . . It has been found quite conclusively that treatment of the sperm with relatively heavy doses of x-rays induces the occurrence of true "gene mutations" in a high proportion of the treated germ cells. Several hundred mutants have been obtained in this way in a short time and considerably more than a hundred have been followed through three, four or more generations. They are (nearly all of them, at any rate) stable in their inheritance. . . . Regarding the types of mutations produced, it was found that the recessive lethals greatly outnumbered the nonlethals producing a visible morphological abnormality. . . . In addition to gene mutations, it was found that there is also caused by x-ray treatment a high proportion of rearrangements in the linear order of the genes. . . . The transmuting action of x-rays on the genes is not confined to the sperm cells, for treatment of the unfertilized females causes mutations about as readily as treatment of the males. . . . In conclusion, the attention of those working along classical genetic lines may be drawn to the opportunity, afforded them by the use of x-rays, of creating in their chosen organisms a series of artificial races for use in the study of genetics. If, as seems likely on genetic considerations, the effect is common to most organisms, it should be possible to produce, "to order," enough mutations to furnish respectable genetic maps.

Treatment of the sperm with relatively heavy doses of x-rays induces the occurrence of true "gene mutations" in a high proportion of the treated germ cells.

Source: Science 66: 84–87

others (roentgen) with the number of ionizations the radiation produces in air, still others (gray, rad) with the amount of energy imparted to material exposed to the radiation, and some (rem, sievert) with the effects of radiation on living tissue. The types of units have proliferated through the years in attempts to encompass different types of radiation, including nonionizing radiation, in a common frame of reference. The units in Table 7.1 are presented only for ease of interpreting the multitude of units found in the literature on radiation genetics. They need not be memorized.

Genetic studies of ionizing radiation support the following general principle:

> Over a wide range of x-ray doses, the frequency of mutations induced by x rays is proportional to the radiation dose.

One type of evidence supporting this principle is the frequency with which X-chromosome recessive lethals are induced

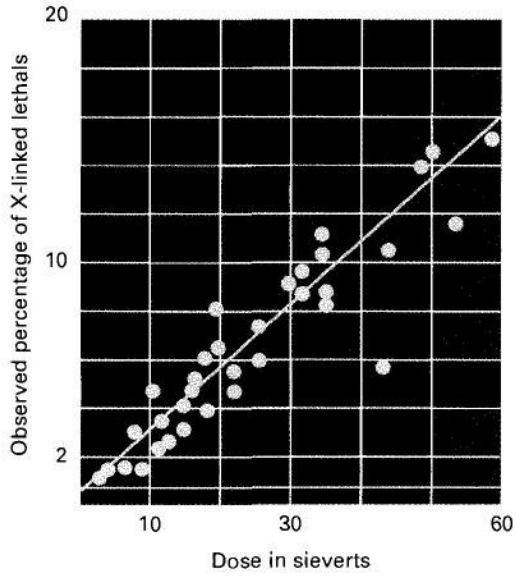


Figure 7.27 The relationship between the percentage of X-linked recessive lethals in *D. melanogaster* and x-ray dose, obtained from several experiments. The frequency of spontaneous X-linked lethal mutations is 0.15 percent per X chromosome per generation.

Table 7.2 Annual exposure of human beings in the United States to various forms of ionizing radiation

Source	Dose (in millisieverts*)
Natural radiation	
Radon gas	2.06
Cosmic rays	0.27
Natural radioisotopes in the body	0.39
Natural radioisotopes in the soil	0.28
Total natural radiation	3.00
Other radiation sources	
Diagnostic x rays	0.39
Radiopharmaceuticals	0.14
Consumer products (x rays from TV, radioisotopes in clock dials) and building materials	0.10
Fallout from weapon tests	<0.01
Nuclear power plants	<0.01
Total from non-natural sources	0.63
Total from all sources	3.63

*One millisievert (mSv) equals 1/1000 sievert, or 0.1 rem.

Source: Reprinted with permission from *The Committee on the Biological Effects of Ionizing Radiations, Health Effects of Exposure to Low Levels of Ionizing Radiation.* Copyright 1990 by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, D.C.

in *Drosophila* (**Figure 7.27**). The mutation rate increases linearly with increasing x-ray dose. For example, an exposure of 10 sieverts increases the frequency from the spontaneous value of 0.15 percent to about 3 percent. The mutagenic and lethal effects of ionizing radiation at low to moderate doses result primarily from damage to DNA. Three types of damage in DNA are produced by ionizing radiation: single-strand breakage (in the sugar–phosphate backbone), double-strand breakage, and alterations in nucleotide bases. The single-strand breaks are usually efficiently repaired, but the other damage is responsible for mutation and lethality. In eukaryotes, ionizing radiation also results in chromosome breaks. Although mechanisms exist for repairing the breaks, the repair often leads to chromosome rearrangements such as translocations (Chapter 9). In human cells in culture, a dose of 0.2 sievert results in an average of one visible chromosome break per cell.

Although the effects of extremely low levels of radiation are extremely difficult to measure because of the background of spontaneous mutation, most experiments support the following principle:

> There appears to be no threshold exposure below which mutations are not induced. Even very low doses of radiation induce some mutations.

Ionizing radiation is widely used in tumor therapy. The basis for the treatment is the increased frequency of chromosomal breakage (and the consequent lethality) in cells undergoing mitosis compared with cells in interphase. Tumors usually contain many more mitotic cells than most normal tissues, so more tumor cells than normal cells are destroyed. Because not all tumor cells are in mitosis at the same time, irradiation is carried out at intervals of several days to allow interphase tumor cells to enter mitosis. Over a period of time, most tumor cells are destroyed.

Table 7.2 gives representative values of doses of ionizing radiation received by human beings in the United States in the course of a year. The unit of measure is the millisievert, which equals 0.1 rem. The exposures in Table 7.2 are on a yearly basis, so over the course of a generation, the total exposure is approximately 100 millisiev-

erts. Note that, with the exception of diagnostic x rays, which yield important compensating benefits, most of the total radiation exposure comes from natural sources, particularly radon gas. Less than 20 percent of the average radiation exposure comes from artificial sources. Nevertheless, there are dangers inherent in any exposure to ionizing radiation, particularly in an increased risk of leukemia and certain others cancers in the exposed persons. With regard to increased genetic diseases in future generations, the risk that a small amount of additional radiation will have mutagenic effects is low enough that most geneticists are currently more concerned about the effects of the many chemical mutagens (as well as carcinogens) that are introduced into the environment from a variety of sources.

The National Research Council of the National Academy of Sciences (USA) regularly updates the estimated risks of radiation exposure. The latest estimates are summarized in **Table 7.3**. The message is that an additional 10 millisieverts of radiation per generation (about a 10 percent increase in the annual exposure) is expected to cause a relatively modest increase in diseases that are wholly or partly due to genetic factors. The most common conditions in the table are heart disease and cancer. No estimate for the radiation-induced increase is given for either of these traits because the genetic contribution to the total is still very uncertain.

Genetic Effects of the Chernobyl Nuclear Accident

On April 26, 1986, a nuclear power plant exploded near the city of Chernobyl (also spelled Chornobyl) in the Ukraine. The immediate area was heavily contaminated,

Table 7.3 Estimated genetic effects of an additional 10 millisieverts per generation

Type of disorder	Current incidence per million liveborn	Additional cases per million liveborn per 10 mSv per generation	
		First generation	At equilibrium
Autosomal dominant			
Clinically severe	2500	5–20	25
Clinically mild	7500	1–15	75
X-linked	400	<1	<5
Autosomal recessive	2500	<1	Very slow increase
Chromosomal			
Unbalanced translocations	600	<5	Very little increase
Trisomy	3800	<1	<1
Congenital abnormalities			
(multifactorial)	20,000–30,000	10	10–100
Other multifactorial disorders			
Heart disease	600,000	Unknown	Unknown
Cancer	300,000	Unknown	Unknown
Others	300,000	Unknown	Unknown

Source: Reprinted with permission from *The Committee on the Biological Effects of Ionizing Radiations, Health Effects of Exposure to Low Levels of Ionizing Radiation.* Copyright 1990 by the National Academy of Sciences. Courtesy of the National Academy Press, Washington, D.C.

and clouds of radioactive debris disseminated over long distances. It was the largest publicly acknowledged nuclear accident in history. It is estimated to have released 50–200 million curies of radiation, along with a wide variety of heavy-metal and chemical pollutants. Iodine-131 and cesium-137 were the principal radioactive contaminants. People living in the area were evacuated almost immediately, but many were heavily exposed to radiation. Within a short time there was a notable increase in the frequency of thyroid cancer in children, and other health effects of radiation exposure were also detected.

At first, little attention was given to possible genetic effects of the Chernobyl disaster, because relatively few people were acutely exposed and because the radiation dose to people outside the immediate area was considered too small to worry about. In the district of Belarus, some 200 kilometers north of Chernobyl, the average exposure to iodine-131 was estimated at approximately 0.185 sievert per person. This is a fairly high dose, but the exposure was brief because the half-life of iodine-131 is only 8 days. Radiation from the much longer-lived cesium-137 (half-life of 30 years) was estimated at less than 5 millisieverts per year. On the basis of data from laboratory animals and studies of the survivors of the Hiroshima and Nagasaki atomic bombs, little detectable genetic damage was expected from these exposures.

Nevertheless, 10 years after the meltdown, studies of people living in Belarus did indicate a remarkable increase in the mutation rate. The observations focused on STRP polymorphisms because of their intrinsically high mutation rate due to replication slippage and other factors. Each STRP was examined in parents and their offspring to detect any DNA fragments that increased or decreased in size, indicative of an increase or decrease in the number of repeating units contained in the DNA fragment. Five STRP polymorphisms were studied, with the results summarized in **Figure 7.28**. Two of the loci (blue dots) showed no evidence for an increase in the mutation rate. This is the expected result. The unexpected finding was that three of the loci (red dots) did show a significant increase, and the level of increase was consistent with an approximate doubling of the mutation rate. It is still not known whether the increase was detected because replication slippage is more sensitive to radiation than other types of mutations or because the effective radiation dose sustained by the Belarus population was much higher than originally thought.

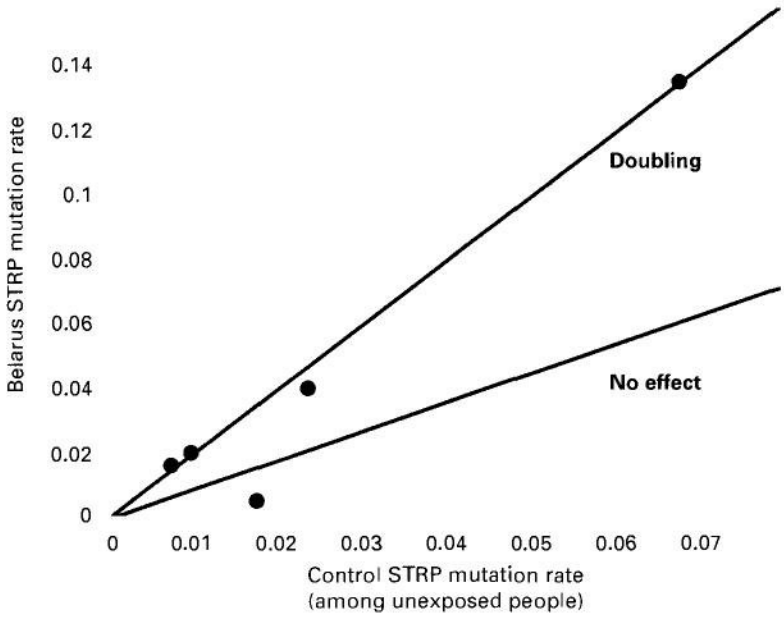


Figure 7.28 Mutation rates of five STRP polymorphisms among people of Belarus exposed to radiation from Chernobyl and among unexposed British people. Three of the loci (red dots) show evidence of an approximately twofold increase in the mutation rate. [Data from Y. E. Dubrova, V. N. Nesterov, N. G. Krouchinsky, V. A. Ostapenko, R. Neumann, D. L. Neil, and A. J. Jeffreys. 1996. *Nature* 380: 183.]

The finding of increased mutation among the people of Belarus is paralleled by observations carried out in two species of voles, small rodents related to hamsters and gerbils, that live in the immediate vicinity of Chernobyl. These animals live in areas still highly contaminated by radiation, heavy metals, and chemical pollutants. The animals eat food that is so heavily contaminated that they themselves become extremely radioactive. Yet they survive and reproduce in this harsh environment.

Among the voles of Chernobyl, the rate of mutation in the cytochrome *b* gene present in mitochondrial DNA was estimated at 2×10^{-4} mutations per nucleotide per generation. This rate is at least 200 times

higher than that found among animals living in a relatively uncontaminated area just 30 kilometers to the southwest of Chernobyl. It is so high that, on average, each mitochondrial DNA molecule of approximately 17 kb undergoes three new mutations in each generation. It is as yet unknown whether this high rate of mutation is due to radiation exposure, to nonradioactive pollutants, or to the synergistic effects of both.

7-6 Mechanisms of DNA Repair

Spontaneous damage to DNA in human cells takes place at a rate of approximately 1 event per billion nucleotide pairs per minute (or, per nucleotide pair, at a rate of 1×10^{-9} per minute). This may seem like quite a low rate, but it implies that every 24 hours, in every human cell, the DNA is damaged at approximately 10,000 different sites. Under ordinary circumstances, much of the damage done to DNA by spontaneous chemical reactions in the nucleus, as well as that caused by chemical mutagens and radiation, can be repaired. We have already seen (Section 7.4) one example of a repair enzyme: DNA uracil glycosylase, which acts to remove uracil nucleotides that get into DNA, such as by the deamination of cytosine. In this section, we examine other key molecular mechanisms for the repair of aberrant or damaged DNA.

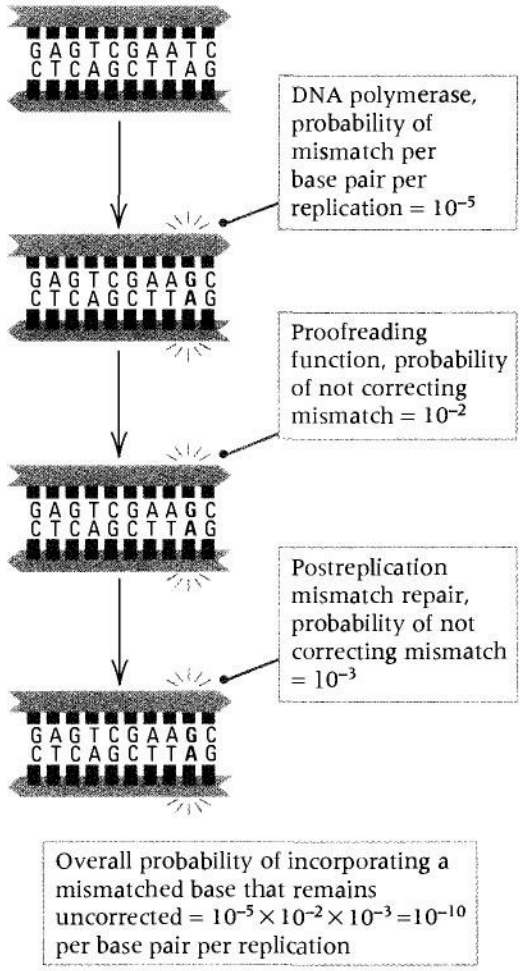


Figure 7.29 Summary of rates of error in DNA polymerization, proofreading, and postreplication mismatch repair. The initial rate of nucleotide misincorporation is 10^{-5} per base pair per replication. The proofreading function of DNA polymerase corrects 99 percent of these errors, and of those that remain, postreplication mismatch repair corrects 99.9 percent. The overall rate of misincorporated nucleotides that are not repaired is 10^{-10} per base pair per replication.

Mismatch Repair

We have already encountered the mismatch repair system in the context of gene conversion and genetic recombination (Section 6.8). However, the most important role of mismatch repair is as a "last chance" error-correcting mechanism in replication. During DNA replication, mismatched nucleotides are incorporated at the rate of about 10^{-5} per template nucleotide per round of replication. As shown in **Figure 7.29**, approximately 99 percent of these are immediately corrected by the proofreading function (3'-to-5' exonuclease activity) of the major replication polymerase. This leaves a mismatch rate of 10^{-7} per template nucleotide per round of replication; 99.9 percent of the remaining mismatches are corrected by the mismatch repair system, yielding an overall mismatch rate of

$$10^{-5} \times 10^{-2} \times 10^{-3} = 10^{-10}$$

The operation of **mismatch repair** is illustrated in **Figure 7.30**. When a mismatched base is detected, one of the strands is cut in two places and a region around the mismatch is removed. The excised region is

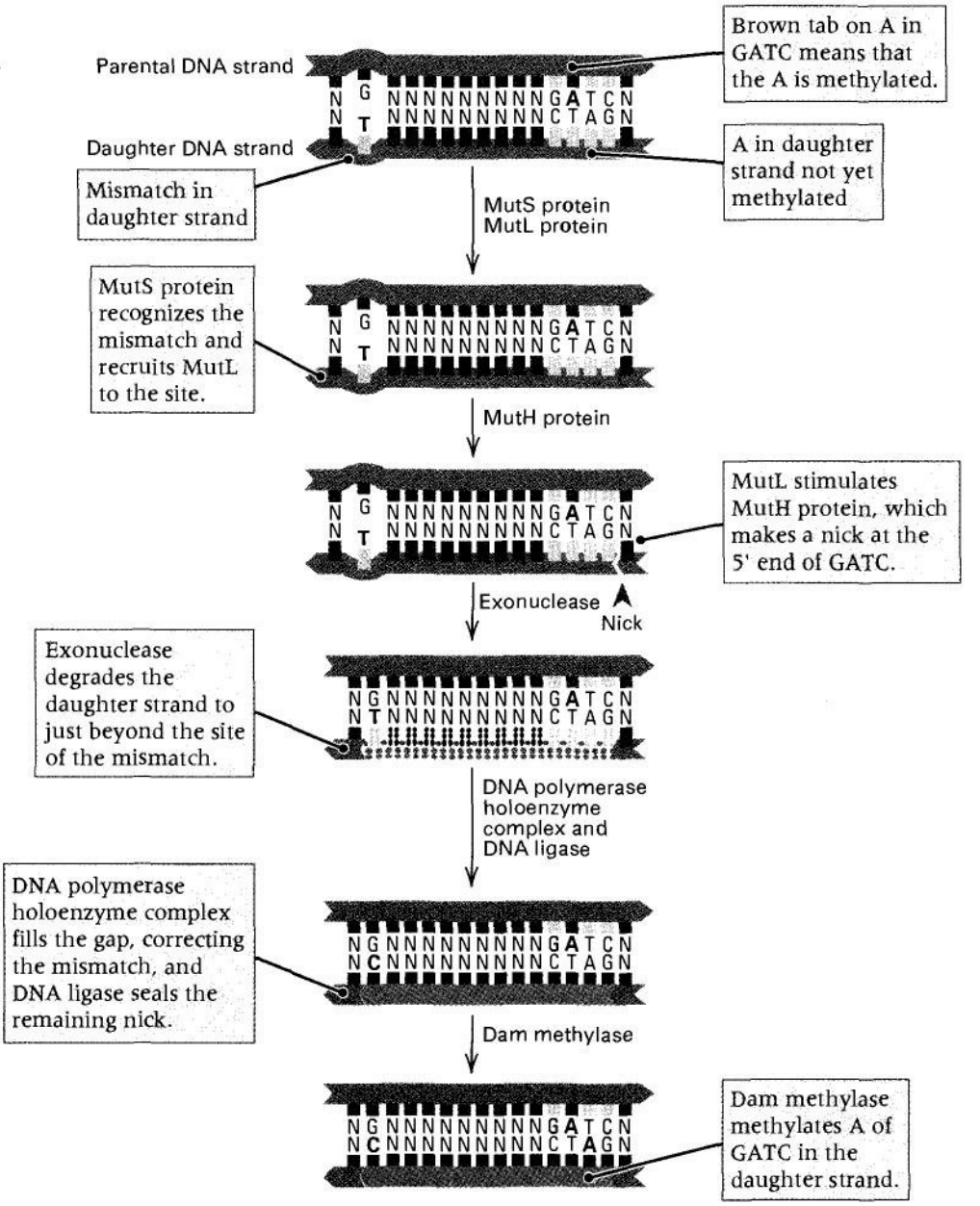


Figure 7.30 Mismatch repair consists of the excision of a segment of a DNA strand that contains a base mismatch, followed by repair synthesis. In *E. coli*, cleavage takes place at the nearest methylated GATC sequence in the unmethylated strand. An exonuclease removes successive nucleotides until just past the mismatch, and the resulting gap is repaired. Either strand can be excised and corrected, but in newly synthesized DNA, methylated bases in the template strand often direct the excision mechanism to the newly synthesized strand that contains the incorrect nucleotide.

variable in size. In *E. coli*, an enzyme first cleaves the DNA at the nearest GATC sequence on the unmethylated (newly synthesized) strand, provided that the site is within about 1 kb of the mismatch. Then an exonuclease degrades the cleaved DNA strand to a position on the other side of the mismatch (the excision step), creating a single-stranded gap. After excision, DNA polymerase fills in the gap by using the remaining strand as a template, thereby eliminating the mismatch. The mismatch-repair system also corrects most small insertions or deletions.

If the DNA strand that is removed in mismatch repair is chosen at random, then

CoNnection

Replication Slippage in Unstable Repeats

Micheline Strand,[1] Tomas A. Prolla,[2] R. Michael Liskay,[2] and Thomas D. Petes[1] 1993

[1]University of North Carolina, Chapel Hill, North Carolina

[2]Yale University, New Haven, Connecticut

Destabilization of Tracts of Simple Repetitive DNA in Yeast by Mutations Affecting DNA Mismatch Repair

One of the most important principles of genetics is that many fundamental processes are carried out in much the same way in many different organisms. Basic cellular processes are similar in all eukaryotic cells, and the basic mechanisms of DNA replication and repair are fundamentally the same in eukaryotes and prokaryotes. The similarity of mismatch repair among organisms motivated this paper. The authors report that yeast cells deficient in mismatch repair are very deficient in their ability to replicate faithfully tracts of short repeating sequences. Because this type of instability also characterizes certain human hereditary colon cancers, the authors correctly predicted that the mutant genes causing the colon cancer would be genes involved in mismatch repair.

The genomes of all eukaryotes contain tracts of DNA in which a single base or a small number of bases is repeated. Expansions of such tracts have been associated with several human disorders including the fragile-X syndrome. In addition, simple repeats are unstable in certain forms of colorectal cancer, suggesting a defect in DNA replication of repair. . . . One mechanism [for changes in repeat number] is DNA polymerase slippage: during replication of the tract, the primer and template strand transiently dissociate and then reassociate in a misaligned configuration. If the unpaired bases are in the primer strand, continued synthesis results in an elongation of the tract whereas unpaired bases in the template strand result in a deletion. . . . The most common simple repeat in most eukaryotes is poly(GT)$_{10-30}$. We examined the genetic control of the stability of poly(GT) tracts. . . . Three mutations affecting mismatch repair in yeast were examined, *pms1* and *mlh1* (homologs of *MutL*) and *msh2* (a homolog of *MutS*). . . . In two different genetic backgrounds using two different instability assay systems, all three mutations lead to 100- to 700-fold elevated levels of tract instability. . . . Yeast contains two DNA polymerases that have 3′–5′ exonuclease ("proofreading") activities. In strains with mutations in the nuclease portion of these genes, neither affects tract stability as dramatically as mutations that reduce mismatch repair. . . . The results show that the instability of poly(GT) tracts in yeast is increased by mutations in the mismatch repair genes but is relatively unaffected by mutations affecting the proofreading functions of DNA polymerase. These results indicate that DNA polymerase *in vivo* has a very high rate of slippage on templates containing simple repeats, but most of the errors are corrected by cellular mismatch repair systems. Thus, the instability of simple repeats observed for some human diseases may be a consequence of either an increased rate of DNA polymerase slippage or a decreased efficiency of mismatch repair.

The instability of poly(GT) tracts in yeast is increased by mutations in the mismatch repair genes.

Source: Nature 365: 274–276

the repair process sometimes creates a mutant molecule by cutting the strand that contains the correct base and using the mutant strand as a template. However, this is prevented from happening in newly synthesized DNA because the daughter strand is less methylated than the parental strand. The mismatch-repair system recognizes the degree of methylation of a strand and *preferentially excises nucleotides from the undermethylated strand.* This helps ensure that incorrect nucleotides incorporated into the daughter strand in replication will be removed and repaired. The daughter strand is always the undermethylated strand because its methylation lags somewhat behind the moving replication fork, whereas the parental strand was fully methylated in the preceding round of replication.

The mechanism of mismatch repair has been studied extensively in bacteria. Mutants defective in the process were identified as having high rates of spontaneous mutation. The products of two genes, *mutL* and *mutS,* recognize and bind to a mismatched base pair. This triggers the excision

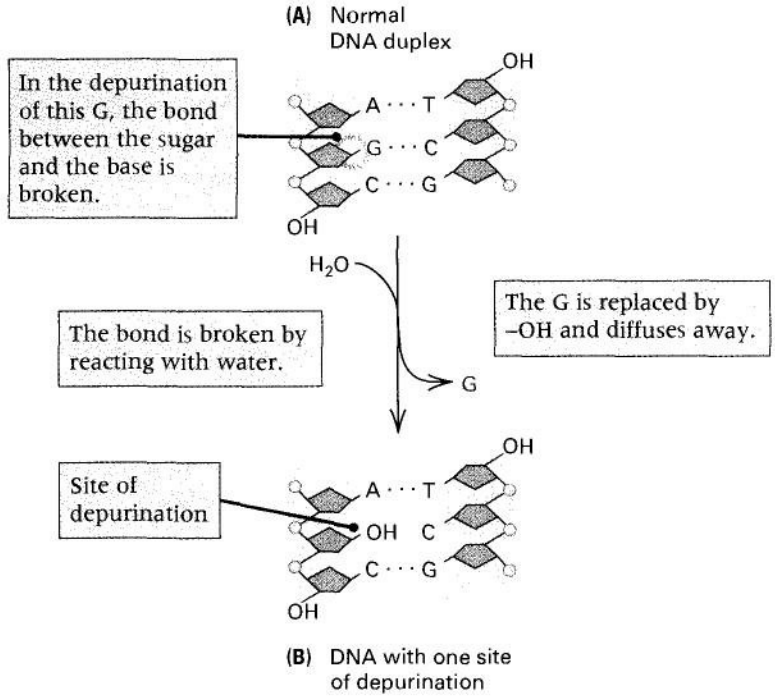


Figure 7.31 Deoxyribose–purine bonds are somewhat unstable and prone to undergo spontaneous reaction with water (hydrolysis), which results in loss of a purine base from the DNA (depurination). (A) Part of a DNA molecule prior to depurination. The bond between the labeled G and the deoxyribose to which it is attached is about to be hydrolyzed. (B) Hydrolysis of the bond releases the G purine, which diffuses away from the molecule and leaves a hydroxyl (−OH) in its place in the depurinated DNA.

of a tract of nucleotides from the newly synthesized strand. Experiments carried out in yeast by Thomas Petes and colleagues revealed that tandem repeats of short nucleotide sequences are a thousandfold less stable in mutants deficient in mismatch repair than in wildtype yeast. By that time it was already known that some forms of human hereditary colorectal cancer result in decreased stability of simple repeats, and the yeast researchers suggested that these high-risk families might be segregating for an allele causing a mismatch-repair deficiency. Within less than two years, scientists in several laboratories had identified four human genes homologous to *mutL* or *mutS*, any one of which, when mutated, results in hereditary nonpolyposis colorectal cancer (HNPCC). Most cases of this type of cancer may be caused by mutations in one of these four mismatch-repair genes.

AP Repair

AP repair acts on nucleotides that have lost their base and become either *apurinic* (loss of a purine, A or G) or *apyrimidinic* (loss of a pyrimidine, T or C). Spontaneous depurination is an important source of mutation, because the sugar–purine bonds are relatively labile and subject to hydrolysis. The process is illustrated in **Figure 7.31**. The rate of spontaneous depurination is approximately 3×10^{-9} depurinations per purine nucleotide per minute, which is at least tenfold greater than any other single source of spontaneous degradation. If DNA replication takes place before the apurinic site is repaired, then an adenine nucleotide is almost always inserted in the daughter strand opposite the apurinic site. After another round of replication, therefore, the original G−C pair becomes a T−A pair, an example of a transversion mutation.

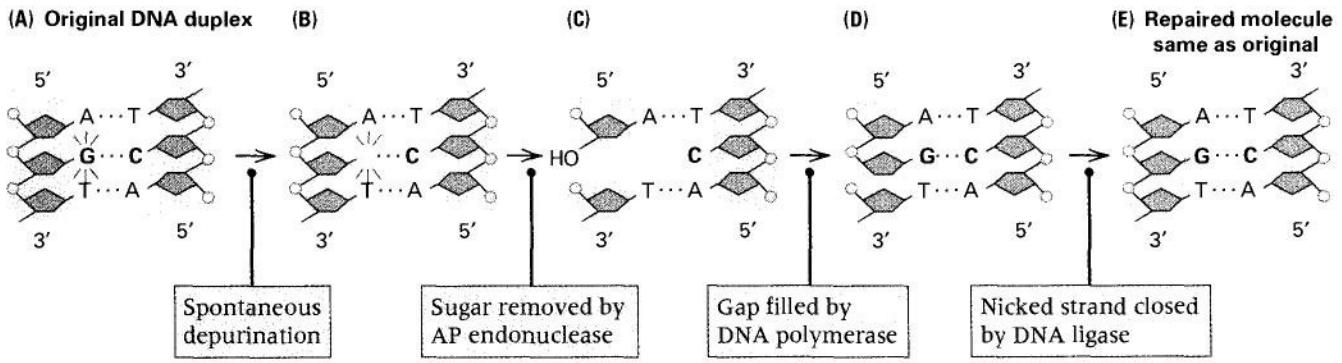


Figure 7.32 Action of AP endonuclease. (A) Original DNA duplex. (B) Spontaneous hydrolysis of guanine results in loss of the base. (C) AP endonuclease excises the empty deoxyribose from the DNA strand. (D) DNA polymerase fills the gap using the continuous strand as a template. (E) The remaining nick is closed by DNA ligase, restoring the original sequence.

Loss of a base is not always mutagenic because AP repair can restore the original nucleotide present at the site. The process is illustrated in **Figure 7.32**. The key enzyme is the *AP endonuclease*, which cleaves the baseless sugar from the DNA, leaving a single-stranded gap that is repaired by DNA polymerase. The gap filling leaves one nick remaining in the repaired strand, which is closed by DNA ligase, completing the repair.

Photoreactivation

Various enzymes can recognize and catalyze the direct reversal of specific DNA damage. A classic example is the reversal of UV-induced pyrimidine dimers by **photoreactivation,** in which an enzyme breaks the bonds that join the pyrimidines in the dimer and thereby restores the original bases. The enzyme binds to the dimers in the dark but then utilizes the energy of blue light to cleave the bonds. Another important example is the reversal of guanine methylation in O^6-methyl guanine by a methyl transferase enzyme, which otherwise would pair like adenine in replication.

Excision Repair

Excision repair is a ubiquitous multistep enzymatic process by which a stretch of a damaged DNA strand is removed from a duplex molecule and replaced by resynthesis using the undamaged strand as a template (**Figure 7.33**). The system was originally discovered through its role in the repair of pyrimidine dimers caused by UV light, but the substrate for repair can be any distortion in the duplex molecule. In excision repair, a repair endonuclease recognizes the distortion produced by the DNA damage and makes two cuts in the sugar–phosphate backbone. The excised region is usually quite precise. In prokaryotes, the cleavage sites are 8 nucleotides from the 5' end and 5 nucleotides from the 3' end of the damage; and in eukaryotes, the cleavage sites are 24 nucleotides from the 5' end and 5 nucleotides from the 3' end of the damage. A free 3'-OH group is produced at the 5' cut, which DNA polymerase uses as a primer to synthesize a new strand while displacing the DNA segment that contains the damage. The final step of the repair process is the joining of the newly synthesized segment to the contiguous strand by DNA ligase.

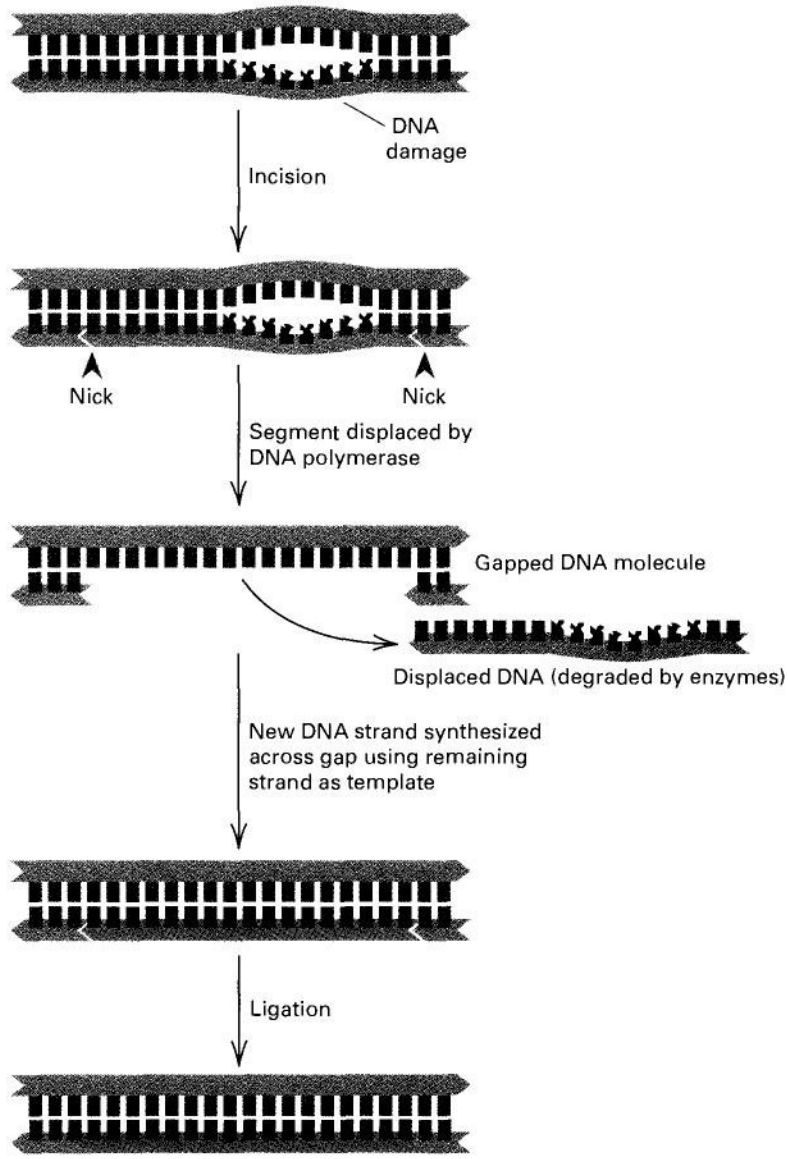


Figure 7.33 Mechanism of excision repair of DNA damage.

Postreplication Repair

Sometimes DNA damage persists rather than being reversed or removed, but its harmful effects can be minimized. This often requires replication across damaged areas, so the process is called **postreplication repair.** For example, when DNA polymerase reaches a damaged site (such as a pyrimidine dimer), it stops synthesis of the strand. However, after a brief time, synthesis is reinitiated beyond the damage and chain growth continues, producing a

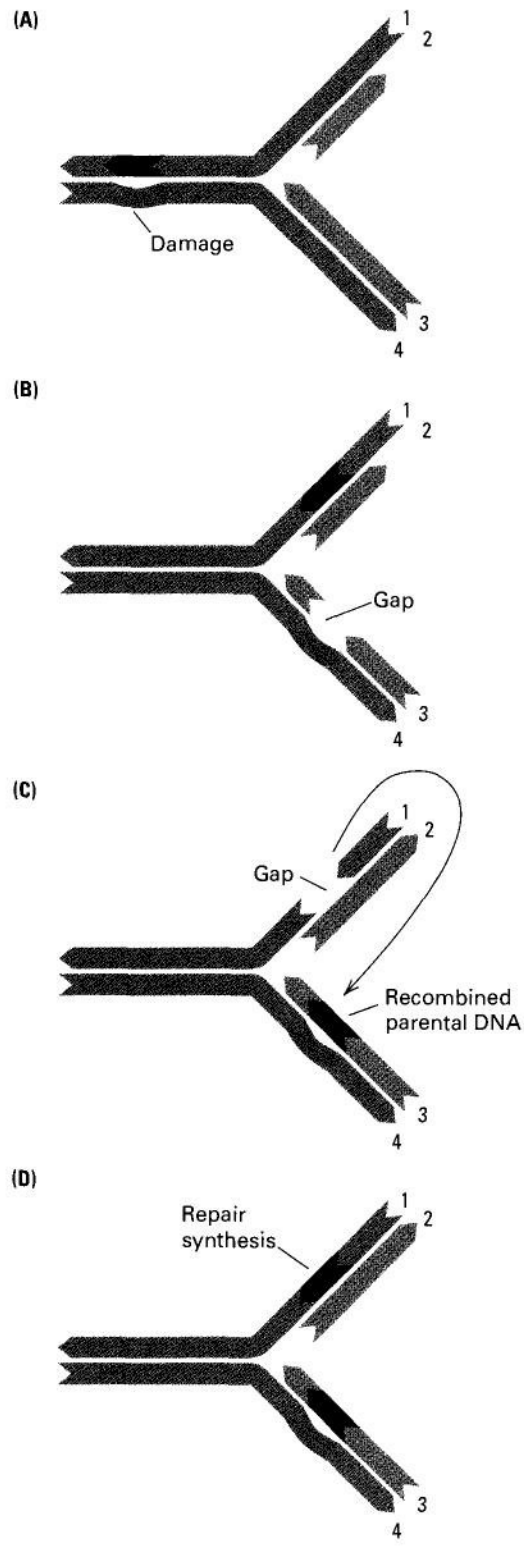


Figure 7.34 Postreplication repair. (A) A molecule with DNA damage in strand 4 is being replicated. (B) By reinitiation of synthesis beyond the damage, a gap is formed in strand 3. (C) A segment of parental strand 1 is excised and inserted in strand 3. (D) The gap in strand 1 is next filled in by repair synthesis.

gapped strand with the damaged spot in the gap.



The gap can be filled by strand exchange with the parental strand that has the same polarity, and the secondary gap produced in that strand can be filled by repair synthesis (**Figure 7.34**). The products of this exchange and resynthesis are two intact single strands, each of which can then serve in the next round of replication as a template for the synthesis of an undamaged DNA molecule.

The SOS Repair System

SOS repair, which is found in *E. coli.* and related bacteria, is a complex set of processes that includes a bypass system that allows DNA replication to take place across pyrimidine dimers or other DNA distortions, but at the cost of the fidelity of replication. Even though intact DNA strands are formed, the strands are often defective. SOS repair is said to be *error-prone*. A significant feature of the SOS repair system is that it is not always active but is induced by DNA damage. Once activated, the SOS system allows the growing fork to advance across the damaged region, adding nucleotides that are often incorrect because the damaged template strand cannot be replicated properly. The proofreading system of DNA polymerase is relaxed to allow polymerization to proceed across the damage, despite the distortion of the helix.

7.7 Reverse Mutations and Suppressor Mutations

In most of the mutations we have considered so far, a wildtype (normal) gene is changed into a form that results in a mutant phenotype, an event called a **forward mutation.** Mutations are frequently reversible, and an event that restores the wildtype phenotype is called a **reversion.** A reversion may result from a **reverse mutation,** an exact reversal of the alteration in base sequence that occurred in the original forward mutation, restoring the

wildtype DNA sequence. A reversion may also result from the occurrence of a second mutation at some other site in the genome, which compensates for the effect of the original mutation. Reversion by the exact reversal mechanism is infrequent. The second-site mechanism is much more common, and a mutation of this kind is called a **suppressor mutation.** A suppressor mutation can occur at a different site in the gene containing the mutation that it suppresses (*intra*genic suppression) or in a different gene (*inter*genic suppression). Most suppressor mutations do not fully restore the wildtype phenotype, for reasons that will be apparent in the following discussion of the two kinds of suppression.

Intragenic Suppression

Reversion of frameshift mutations usually results from **intragenic suppression;** the mutational effect of the addition (or deletion) of a nucleotide pair in changing the reading frame of the mRNA is rectified by a compensating deletion (or addition) of a second nucleotide pair at a nearby site in the gene. In an ingenious experiment discussed in Chapter 11, intragenic frameshift suppressors were used to demonstrate that the genetic code is a triplet code.

A second type of intragenic suppression is observed when loss of activity of a protein, caused by one amino acid change, is at least partly restored by a change in a second amino acid in the same protein. An example of this type of suppression in the protein product of the *trpA* gene in *E. coli* is shown in Figure 7.35. The TrpA polypeptide, composed of 268 amino acids, is one of two polypeptides that make up the enzyme tryptophan synthase in *E. coli*. The mutation shown, one of many that inactivate the enzyme, causes a change of amino acid number 210 from glycine in the wildtype protein to glutamic acid. This glycine is not in the active site; rather, the inactivation is caused by a change in the folding of the protein, which indirectly affects the active site. The activity of the mutant protein is partly restored by a second mutation that results in a change in amino acid number 174 from tyrosine to cysteine. A protein in which only the tyrosine has been replaced by cysteine is inactive, again because of a change in the shape of the protein. This change at the second site, found repeatedly in different reversions of the original mutation, restores activity because the two regions that contain amino acids 174 and 210 interact fortuitously to produce a protein with the correct folding. This type of reversion can usually be taken to mean that the two regions of the protein interact, and studies of the amino acid se-

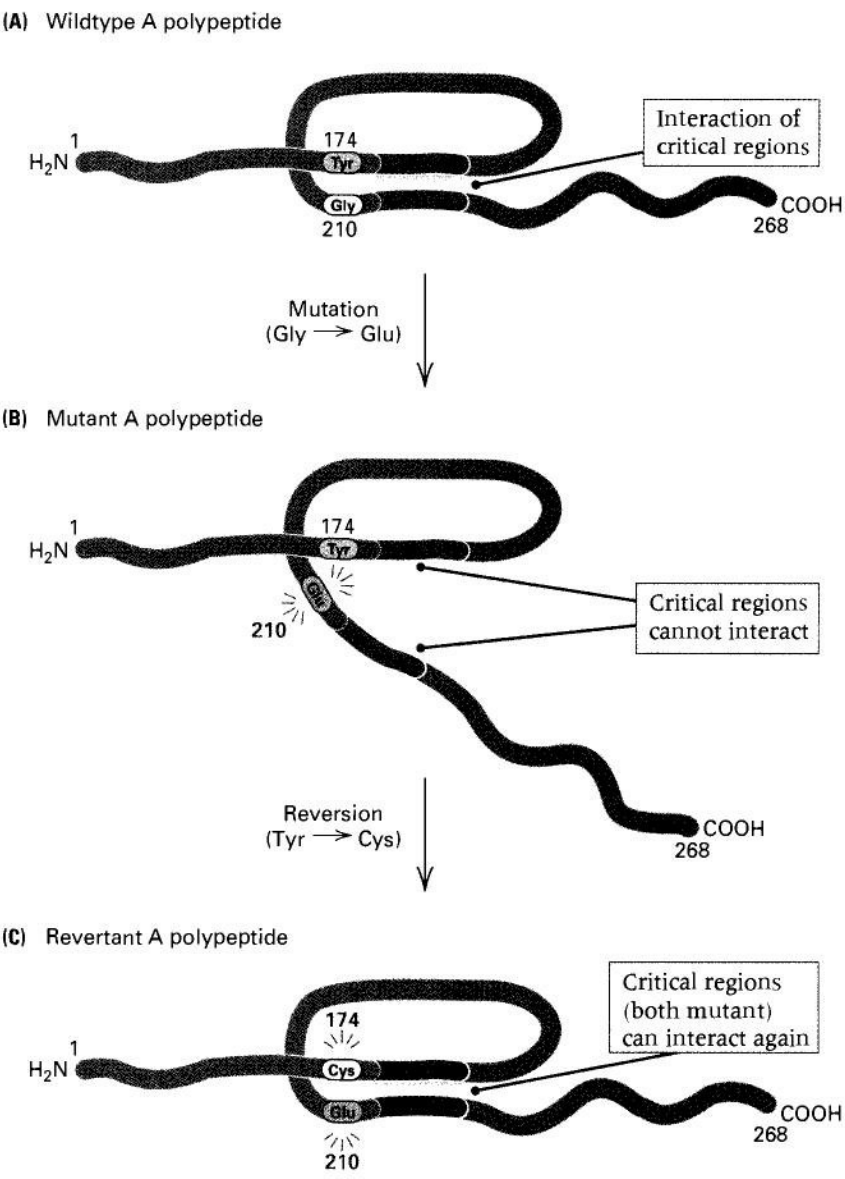


Figure 7.35 Model for the effect of mutation and intragenic suppression on the folding and activity of the A protein of tryptophan synthase in *E. coli*. (A) The wildtype protein in which two critical regions (orange segments) of the polypeptide chain interact. (B) Disruption of proper folding of the polypeptide chain by a substitution of amino acid 210 prevents the critical regions from interacting. (C) Suppression of the effect of the original forward mutation by a subsequent change in amino acid 174. The structure of the region that contains this amino acid is altered, bringing the critical regions together again.

quences of mutants and revertants are often informative in elucidating the structure of proteins.

Intergenic Suppression

Intergenic suppression refers to a mutational change *in a different gene* that eliminates or suppresses the mutant phenotype. Many intergenic suppressors are very specific in being able to suppress the effects of mutations in only one or a few other genes. For example, phenotypes resulting from the accumulation of intermediates in a metabolic pathway as a consequence of mutational inactivation of one step in the pathway can sometimes be suppressed by mutations in other genes that function earlier in the pathway, which reduces the concentration of the intermediate. These suppressors act only to suppress mutations that affect steps further along the pathway, and they suppress all alleles of these genes.

Other intergenic suppressors are more general in being able to suppress mutations in many functionally unrelated genes—and usually in only a subset of alleles of these genes. For example, certain intergenic suppressors in *Drosophila* affect transcription factors that control the transcription of particular families of transposable elements. Alleles that are mutant because they contain a transposable element whose transcription interferes with that of the gene can be suppressed by an intergenic suppressor that prevents transcription of the inserted transposable element.

In prokaryotes, the best-understood type of intergenic suppression is found in some tRNA genes, mutant forms of which can suppress the effects of particular mutant alleles of many other genes. These suppressor mutations change the anticodon sequence in the tRNA and thus the specificity of mRNA codon recognition by the tRNA molecule. Mutations of this type were first detected in certain strains of *E. coli*, which were able to suppress particular phage-T4 mutants that failed to form plaques on standard bacterial strains. These strains were also able to suppress particular alleles of numerous other genes in the bacterial genome. The suppressed mutations were, in each case, nonsense mutations—those in which a stop codon (UAA, UAG, or UGA) had been introduced within the coding sequence of a gene, with the result that polypeptide synthesis was prematurely terminated and only an amino-terminal fragment of the polypeptide was synthesized. The suppressors are therefore called **nonsense suppressors.** The mechanism of suppression is discussed in detail in Chapter 11, but an overview is that the suppressor mutation occurs in the anticodon of one of the normal tRNA molecules, allowing it to pair with a nonsense codon and insert an amino acid so that translation can continue. Such mutations are not lethal because most tRNA genes are present in several copies, so nonmutated copies of the tRNA are also present to allow correct translation of the codon.

The Ames Test for Mutagen/Carcinogen Detection

In view of the increased number of chemicals present as environmental contaminants, tests for the mutagenicity of these substances have become important. Furthermore, most carcinogens are also mutagens, so mutagenicity provides an initial screening for potential hazardous agents. One simple method of screening large numbers of substances for mutagenicity is a reversion test that uses nutritional mutants of bacteria. In the simplest type of reversion test, a compound that is a potential mutagen is added to solid growth medium, a known number of a mutant bacterium are plated, and the revertant colonies are counted. An increase in the reversion frequency significantly greater than that obtained in the absence of the test compound identifies the substance as a mutagen. However, simple tests of this type fail to demonstrate the mutagenicity of a large number of potent carcinogens. The explanation for this failure is that many substances are not directly mutagenic (or carcinogenic); rather, they require a conversion into mutagens by enzymatic reactions that take place in the livers of animals and that have no counterpart in bacteria. The normal function of these enzymes is to protect the organism from various naturally occurring harmful substances by converting them into soluble nontoxic substances that can be disposed of in the urine. However, when the enzymes encounter certain human-made and natural compounds, they convert

these substances, which may not be harmful in themselves, into mutagens or carcinogens. The enzymes of liver cells, when added to the bacterial growth medium, activate the compounds and allow their mutagenicity to be recognized. The addition of liver extract is one step in the Ames test for carcinogens and mutagens.

In the **Ames test,** histidine-requiring (His^-) mutants of the bacterium *Salmonella typhimurium*, containing either a base substitution or a frameshift mutation, are tested for reversion to His^+. In addition, the bacterial strains have been made more sensitive to mutagenesis by the incorporation of several mutant alleles that inactivate the excision-repair system and that make the cells more permeable to foreign molecules. Because some mutagens act only on replicating DNA, the solid medium used contains enough histidine to support a few rounds of replication but not enough to permit the formation of a visible colony. The medium also contains the potential mutagen to be tested and an extract of rat liver. If the test substance is a mutagen or is converted into a mutagen, some colonies are formed. A quantitative analysis of reversion frequency can also be carried out by incorporating various amounts of the potential mutagen in the medium. The reversion frequency generally depends on the concentration of the substance being tested and, for a known carcinogen or mutagen, correlates roughly with its carcinogenic potency in animals. The Ames test is simple, rapid, inexpensive, and exquisitely sensitive. Some chemicals can be detected to be mutagenic in amounts as small as 10^{-9} g, and a condensate of as little as 1/100 of a cigarette can be shown to be mutagenic. The test is also highly quantitative (**Figure 7.36**). Chemicals need not be classified simplistically as "mutagenic" or "nonmutagenic." They can be classified according to their potency as mutagens, because more than a millionfold range in potency can be detected.

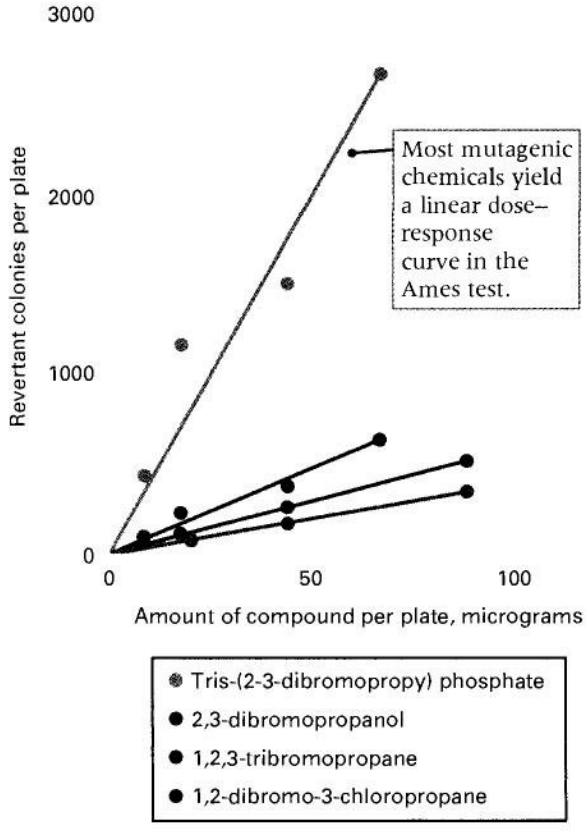


Figure 7.36 Linear dose–response relationships obtained with various chemical mutagens in the Ames test. [Data from B. N. Ames. 1974. *Science* 204: 587–593.]

Chapter Summary

Nucleotide substitution mutations may be transitions (A ⟷ G or T ⟷ C) or transversions (any other change). Within a coding region, a missense (nonsynonymous) substitution results in an amino acid replacement, a synonymous (silent) substitution results in a synonymous codon, and a nonsense substitution results in a chain-terminating codon (UAA, UAG, or UGA). Insertions or deletions that are not exact multiples of three nucleotides result in frameshift mutations, in which the translational reading frame is offset from the point of mutation onward. Conditional mutations are expressed under one set of conditions (restrictive conditions, such as high temperature) and are not expressed under another set (permissive conditions, such as low temperature). In organisms in which cytosines may be methylated, the sites of methylation are often hotspots of mutation because deamination of 5-methyl cytosine produces thymine, which results in a C–G to T–A transition. In the human genome, some trinucleotide repeats are prone to expansion (increase in copy number), and the probability of expansion of the premutation state can approach 100 percent.

Transposable elements are capable of moving from one location in the genome to another. The main classes of transposable elements are cut-and-paste elements, which typically have inverted repeats and transpose via a

transposase; LTR retrotransposable elements, which have long direct or inverted repeats and transpose via a reverse transcriptase; and non-LTR retrotransposable elements, which also transpose via a reverse transcriptase but have no direct or inverted repeats. Insertions of transposable elements are the cause of many visible mutations. Some organisms have evolved mutational mechanisms for silencing repeated sequences—for example, RIP (repeat-induced point mutation) in *Neurospora*.

Mutations can be induced chemically by direct alteration of DNA—for example, by nitrous acid, which deaminates bases. Base analogs, such as 5-bromouracil, are incorporated into DNA in replication. They undergo mispairing more often than the normal bases, giving rise to transition mutations. Alkylating agents react with bases and cause mispairing. Acridine molecules intercalate (interleaf) between base pairs of DNA and cause misalignment of parental and daughter strands in DNA replication, giving rise to frameshift mutations, usually of one or two bases. Ionizing radiation results in oxidative free radicals that cause a variety of alterations in DNA, including single-stranded and double-stranded breaks. Although the amount of genetic damage from normal background and other sources of radiation is believed to be low, studies of persons exposed to fallout from the Chernobyl nuclear meltdown indicate an increased mutation rate, at least for simple tandem repeat polymorphisms (STRPs).

A variety of systems exist for repairing damage to DNA. The mismatch-repair system corrects mispaired nucleotides missed by the proofreading function of DNA polymerase. Methylation of parental DNA strands and delayed methylation of daughter strands allows the mismatch-repair system to operate selectively on the daughter DNA strand. The AP repair system excises nucleotides that have lost a base (creating an apurinic or apyrimidinic site) as a result of spontaneous hydrolysis or chemical mutagenesis. In excision repair, a stretch of a distorted DNA strand is excised and replaced with a newly synthesized copy of the undamaged strand. In photoreactivation, pyrimidine dimers produced by ultraviolet radiation are restored by cleavage of the connecting bond. Damage to DNA that results in the formation of O^6-methyl guanine is repaired by a special methyl transferase. In all of these systems, the correct template is restored. Postreplication repair is an exchange process in which gaps in one daughter strand produced by aberrant replication across damaged sites are filled by nondefective segments from the parental strand of the other branch of the newly replicated DNA. Thus a new template is produced by this system. The SOS repair system in bacteria is error-prone (mutagenic) because it facilitates replication across stretches in the template DNA that are too damaged to serve as a normal template.

Mutant organisms sometimes revert to the wildtype phenotype. Reversion is normally due to an additional mutation at another site. Reversion by a second mutation is called suppression, and the secondary mutations are called suppressor mutations. These can be intragenic or intergenic. In intragenic suppression, a mutation that changes one region of a protein alters the folding of the protein, and a change in another amino acid causes correct folding again. Intergenic suppression takes place through a variety of mechanisms. For example, in nonsense suppression, a chain-termination mutation (a stop codon) is read by a mutant tRNA molecule with an anticodon that can hydrogen-bond with the stop codon, and an amino acid is inserted at the site of the stop codon. The Ames test measures reversion as an indicator of mutagens and carcinogens; it uses an extract of rat liver, which in mammals occasionally converts intrinsically harmless molecules into mutagens and carcinogens.

Key Terms

acridine
alkylating agent
Ames test
anticipation
AP repair
base analog
CIB method
conditional mutation
cut-and-paste elements
DNA uracil glycosylase
dominant-negative mutation
dynamic mutation
ectopic expression
ectopic recombination
excision repair
forward mutation
frameshift mutation
gain-of-function mutation
germ-line mutation
hot spot
hypermorphic mutation
hypomorphic mutation
intercalation
intergenic suppression
intragenic suppression
ionizing radiation
loss-of-function mutation
knockout mutation
LTR retrotransposons
mismatch repair
missense mutation
mutagen
mutation
mutation rate
nitrous acid
non-LTR retrotransposons
nonsense mutation
nonsense suppressor
nonsynonymous mutation
nucleotide analog
nucleotide substitution
null mutation
permissive conditions
photoreactivation
postreplication repair
pyrimidine dimer
replica plating
replication slippage
restrictive conditions
reverse mutation
reverse transcriptase

reversion
RIP (repeat-induced point mutation)
silent mutation
somatic mutation
SOS repair
suppressor mutation
synonymous mutation
temperature-sensitive mutation
transition mutation
transposable element
transposase
transposition
transversion mutation
trinucleotide repeat

Review the Basics

- What is a spontaneous mutation? an induced mutation? a conditional mutation?
- Define and give a hypothetical example of each of the following types of nucleotide substitutions: transition, transversion, missense (nonsynonymous), synonymous (silent), nonsense.
- For a conditional mutation, what determines whether the mutant phenotype will be expressed?
- How is it possible for an organism to be homozygous for a mutation that is a temperature-sensitive lethal?
- What is a trinucleotide expansion disease and how is the "expansion" related to replication slippage? Name one human disease associated with trinucleotide expansions.
- What is meant by the term *mutation rate?* In practice, what does it mean to say that a particular allele has a mutation rate of 10^{-6} per gene per generation?
- What are transposable elements? Explain how a transposable element can cause a new mutation.
- How was the technique of replica plating used to demonstrate that antibiotic-resistance mutations are present at low frequency in populations of bacteria even in the absence of the antibiotic?
- What is a hot spot of mutation? Why might a mutational hot spot coincide with the site of a methylated cytosine in the DNA?
- What is mismatch repair and how does it happen? Name one human disease associated with a defective mismatch-repair system.
- Spontaneous depurination is the most frequent cause of damage to DNA, but the overwhelming majority of such lesions are repaired. How does the repair take place?

Guide to Problem Solving

Problem 1 The molecule 2-aminopurine (Ap) is an analog of adenine that pairs with thymine. It also occasionally pairs with cytosine. What types of mutations will be induced by 2-aminopurine?

Answer In problems of this sort, first note the base that is replaced, and then follow what can happen in subsequent rounds of DNA replication when the base analog pairs normally or abnormally. In this example, the 2-aminopurine (Ap) is incorporated in place of adenine opposite thymine, forming an Ap–T base pair. In the following rounds of replication, the Ap will usually pair with T, but when it occasionally pairs with C, it forms an Ap–C pair. In the next round of replication, the Ap again pairs with T (forming an Ap–T pair), but the C pairs with G (forming a mutant G–C pair). Hence the type of mutation induced by 2-aminopurine is a change from an A–T pair to a G–C pair, which is a transition mutation.

Problem 2 Among genomes as diverse as the bacteriophages λ and T4, the bacterium *Escherichia coli*, and the fungi *Neurospora crassa* and *Saccharomyces cerevisiae*, the spontaneous mutation rate is approximately 0.003 mutation per genome per DNA replication. The genome sizes of λ, *E. coli*, and *S. cerevisiae* are approximately 50 thousand, 4.6 million, and 13.5 million base pairs, respectively.

(a) What proportion of genomes escape spontaneous mutation in each replication?

(b) What is the mutation rate per base pair per DNA replication in each of these genomes?

(c) What might the differences in mutation rate per base pair imply about the evolution of mutation rates?

Answer (a) If the mutation rate is 0.003 mutation per genome per DNA replication, then $1 - 0.003 = 99.7$ percent of the genomes contain no new mutations after one round of replication. **(b)** The rate of mutation per base pair per DNA replication for λ is $0.003/50{,}000 = 6 \times 10^{-8}$. For the other genomes, the values are calculated similarly and equal 6.5×10^{-10} for *E. coli* and 2.2×10^{-10} for *S. cerevisiae*. **(c)** The great variation in spontaneous mutation rate per base pair among these genomes suggests that the larger genomes have evolved lower mutation rates per nucleotide pair through either greater fidelity of DNA replication or greater efficiency of repair mechanisms.

Problem 3 The nontemplate DNA strand in an amino-acid–coding portion of a gene is shown here, along with the amino acid sequence of the polypeptide chain encoded in this region. This particular region of the polypeptide is quite tolerant of amino acid replacements, so many missense mutations in the region do not destroy function. Four frameshift mutations are isolated: −A is a deletion of the red A, +A is a duplication of the red A, −G is a deletion of the red G, and +G is a duplication of the red G. As expected, all four mutations lead to a nonfunctional protein. Recombination is used to combine the mutations, in the expectation that a "−" deletion of one nucleotide will be compensated for by a "+" addition slightly farther along, because the second mutation will shift the reading frame back to the normal reading frame, and changes in the amino acids between the mutations should not destroy function. An unexpected result was obtained. Although the combination +A with −G did, indeed, restore protein function, the combination −A with +G was still mutant. How can you account for this result?

```
GCTACAAAAATGACTGGCACTGCA
AlaThrLysMetThrGlyThrAla
```

Answer In this kind of problem, the best strategy is first to deduce the amino acid sequence of each mutant polypeptide chain. The answer often becomes clear immediately. In this case, the mutant amino acid sequences for all four mutants and the two types of double mutants are shown. Note that the −A mutation results in a termination codon (UGA), which terminates translation at that codon (indicated by the raised dot). The −A with +G combination also has this termination codon, so it remains mutant; even though the frameshift is corrected by the +G, the correction is ineffectual because translation is already terminated prior to this point. The +A with −G combination is nonmutant, because the frameshifts do compensate, and there is no termination codon in the region between them.

```
−A            GCUACAAAAUGACUGGCACUGCA
              AlaThrLys·

+A            GCUACAAAAAAUGACUGGCACUGCA
              AlaThrLysAsnAspTrpHisCys

−G            GCUACAAAAAUGACUGCACUGCA
              AlaThrLysMetThrAlaLeu

+G            GCUACAAAAAUGACUGGGCACUGCA
              AlaThrLysMetThrGlyHisCys

+A with −G    GCUACAAAAAAUGACUGCACUGCA
              AlaThrLysAsnAspCysThrAla

−A with +G    GCUACAAAAUGACUGGGCACUGCA
              AlaThrLys·
```

Analysis and Applications

7.1 Occasionally, a person is found who has one blue eye and one brown eye or who has a sector of one eye a different color from the rest. Can these phenotypes be explained by new mutations? If so, in what types of cells must the mutations occur?

7.2 A mutation is isolated that cannot be induced to revert. What types of molecular changes might be responsible?

7.3 There are 12 possible substitutions of one nucleotide pair for another (for example, A → G). Which changes are transitions and which transversions?

7.4 Weedy plants that are resistant to the herbicide atrazine have a single amino acid substitution in the gene *psbA* that results in the replacement of a serine with an alanine in the polypeptide. Is the base change in the *psbA* gene that results in this amino acid replacement a transition or a transversion?

7.5 What is the minimum number of single-nucleotide substitutions that would be necessary for each of the following amino acid replacements?

(a) Trp → Lys **(c)** Met → His
(b) Tyr → Gly **(d)** Ala → Asp

7.6 In the *rIIB* gene of bacteriophage T4, there is a small coding region near the beginning of the gene in which the particular amino acid sequence that is present is not essential for protein function. (Single-base insertions and deletions in this region were used to prove the triplet nature of the genetic code.) Under what condition would a +1 frameshift mutation at the beginning of this region not be suppressed by a −1 frameshift mutation at the end of the region?

7.7 What genetic characteristic is often associated with mutations caused by the insertion of a transposable element?

7.8 In a coding sequence made up of equal proportions of A, U, G, and C, what is the probability that a random nucleotide substitution will result in a chain-termination codon?

7.9 If a mutagen produces frameshift mutations, what kinds of mutations will it induce to revert?

7.10 Does the nucleotide sequence of a mutation tell you anything about its dominance or recessiveness?

7.11 What are two ways in which pyrimidine dimers are removed from DNA in *E. coli*?

7.12 For mutations that result from spontaneous missense substitutions, the rate of forward mutation (from wildtype allele to mutant allele) is always much greater than the rate of reverse mutation (from a mutant allele back to wildtype). Explain why.

7.13 Describe the phenomenon of repeat-induced point mutation (RIP). In what organisms does it take place? Explain why RIP is regarded as an evolved defense mechanism against transposable elements.

7.14 Spontaneous depurination of DNA occurs at the rate of approximately 3×10^{-9} depurinations per purine nucleotide per minute. A diploid human cell has a genome size of 6×10^{9} base pairs. Approximately how many spontaneous depurinations must be repaired in each cell per day?

7.15 The X chromosome of *Drosophila* contains about 2700 genes. In each generation, about 1 percent of all X chromosomes undergo spontaneous mutation to a new recessive lethal. What proportion of all X-linked genes are capable of mutating to a lethal if the rate of spontaneous mutation is 10^{-5} per gene per generation.

7.16 A *Drosophila* male carries an X-linked temperature-sensitive recessive allele that is viable at 18°C but lethal at 29°C. What sex ratio would be expected among the progeny if the progeny were reared at 29°C and the male were mated to:

(a) A normal XX female?
(b) An attached-X female?

7.17 This problem illustrates how conditional mutations can be used to determine the order of genetically controlled steps in a developmental pathway. A certain organ undergoes development in the sequence of stages A → B → C, and both gene *X* and gene *Y* are necessary for the sequence to proceed. A conditional mutation X' is sensitive to heat (the gene product is inactivated at high temperatures), and a conditional mutation Y' is sensitive to cold (the gene product is inactivated at cold temperatures). The double mutant X'/X' Y'/Y' is created and reared at either high or low temperatures. How far would development proceed in each of the following cases at the high temperature and at the low temperature?

(a) Both *X* and *Y* are necessary for the A → B step.
(b) Both *X* and *Y* are necessary for the B → C step.
(c) *X* is necessary for the A → B step, and *Y* for the B → C step.
(d) *Y* is necessary for the A → B step, and *X* for the B → C step.

7.18 A Lac^+ culture of *E. coli* gives rise to a Lac^- strain that results from a UGA nonsense mutation in the *lacZ* gene. The Lac^- strain gives rise to Lac^+ colonies at the rate of 10^{-8} per cell per generation, and 90 percent of the latter are caused by suppressor tRNA molecules. What is the rate of production of suppressor mutations in the original Lac^+ culture?

7.19 If a gene in a particular chromosome has a probability of mutation of 5×10^{-5} per generation, and if the allele in a particular chromosome is followed through successive generations.

(a) What is the probability that the allele does not undergo a mutation in 10,000 consecutive generations?
(b) What is the average number of generations before the allele undergoes a mutation?

7.20 In the DNA duplex at the bottom of the page suppose that the sequence 5'-GGG . . . AT . . . GGG-3' represents the sequence of a transposable element that is 2 kb in length. What is the sequence of the duplex that would result from homologous recombination between the two copies, assuming that the exchange takes place in the middle between the A and the T? Preserve the coloring of the complementary strands.)

7.21 In the mouse, a dose of approximately 1 gray (Gy) of x rays produces a rate of induced mutation equal to the rate of spontaneous mutation. Taking the spontaneous mutation rate into account, what is the total mutation rate at 1 Gy? What dose of x rays will increase the mutation rate by 50 percent? What dose will increase the mutation rate by 10 percent?

7.22 Among several hundred missense mutations in the gene for the A protein of tryptophan synthase in *E. coli*, fewer than 30 of the 268 amino acid positions are affected by one or more mutations. Explain why the number of positions affected by amino acid replacements is so low.

7.23 Human hemoglobin C is a variant in which a lysine in the β-hemoglobin chain is substituted for a particular glutamic acid. What single-base substitution can account for the hemoglobin-C mutation?

7.24 How many amino acids can substitute for tyrosine by a single-base substitution? (Do not assume that you know which tyrosine codon is being used.)

7.25 The accompanying diagrams show two double-stranded DNA molecules, each of which contains two copies of a transposable element in either (a) direct or (b) inverted orientation. In each case, diagram the result of homologous recombination between the transposable elements.

(a)



(b)



7.26 Which of the following amino acid replacements would be expected with the highest frequency among mutations induced by 5-bromouracil? (1) Met → Leu, (2) Met → Lys, (3) Leu → Pro, (4) Pro → Thr, (5) Thr → Arg.

7.27 Dyes are often incorporated into a solid medium to determine whether bacterial cells can utilize a particular sugar as a carbon source. For instance, on eosin-methylene blue (EMB) medium containing lactose, a Lac^+ cell yields a purple colony and a Lac^- cell yields a pink

```
5'-ATTAGTAGGG...AT...GGGTTTTTTTTTTTTTTTTCCC...AT...CCCATTAGTA-3'
3'-TAATCATCCC...TA...CCCAAAAAAAAAAAAAAAAGGG...TA...GGGTAATCAT-5'
```

Problem 7.20

GeNETics on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- DNA elements that could undergo **transposition** from one site in the genome to another were originally discovered by Barbara McClintock in maize. Her experiments began by studying a mutation in the *colored* (*C*) gene in chromosome 9 that is necessary for synthesis of a purple anthocyanin deposited in the outer layer of corn seed. The mutation was caused by the insertion of a transposable element (*Ds*) that could not transpose on its own but that could transpose in the presence of an "active" element (*Ac*). In genetic backgrounds carrying *Ac*, McClintock observed frequent excision of *Ds* from its insertion site in *C*, restoring the color in the cell lineage and resulting in mottled kernels. Visit this keyword site for a visualization of this process at the molecular and cellular levels.

- Some inherited defects in **DNA repair** are expressed early in life. These include ataxia-telangiectasia, Bloom syndrome, Fanconi anemia, and xeroderma pigmentosum. The incidence and genetic basis of each of these conditions, along with their symptoms and other manifestations, as well as possibilities for treatment, can be found in this database of over 550 childhood illnesses.

- The **Cherynobyl** (also spelled Chornobyl) nuclear accident was by far the most devastating in the history of nuclear power. It released into the Earth's atmosphere about 400 times more radioactive material than the atomic bomb dropped on Hiroshima, but less than 1 percent of the amount from atmospheric weapons tests conducted in the 1950s and 1960s. This keyword site summarizes what is known about the nuclear accident and its aftermath—what happened, how it happened, how many people were seriously exposed to radiation, the immediate health effects, and the anticipated long-term health and environmental effects of this environmental disaster.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 7 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **Pic Site** for Chapter 7.



colony. If a population of Lac^+ cells is treated with a mutagen that produces Lac^- mutants and the population is allowed to grow for many generations before the cells are placed on EMB-lactose medium, a few pink colonies are found among a large number of purple ones. However, if the mutagenized cells are plated on the medium immediately after exposure to the mutagen, some colonies appear that are *sectored* (half purple and half pink). Explain how the sectored colonies arise.

7.28 Mutant alleles of two genes, *a* and *b*, in a haploid organism interact as follows: Most ab^+ and a^+b combinations are mutant, but a few *ab* combinations are wildtype. Each *a* allele that is suppressed requires a particular *b* allele, and each *b* allele that is suppressed requires a particular *a* allele. However, most *a* and *b* alleles are not suppressible by any alleles of the other gene. What kind of interaction between the *a* and *b* gene products can explain these results?

7.29 The accompanying diagrams show two nonhomologous chromosomes, each containing a copy of a transposable element (shaded), which can be oriented either (a) in the same direction or (b) in opposite directions with respect to the centromere. For clarity, the length of the transposable element is greatly exaggerated relative to the length of the chromosome. (In reality, the average transposable element in *Drosophila* is about 0.01 percent of the length of a chromosome.) Draw diagrams illustrating the consequences of ectopic recombination between the transposable elements.

(a)



(b)



7.30 A *Neurospora* strain that is unable to synthesize arginine (and that therefore requires this amino acid in the growth medium) produces a revertant colony able to grow in the absence of arginine. A cross is made between the revertant and a wildtype strain. What proportion of the progeny from this cross would be arginine-independent if the reversion occurred by each of the following mechanisms?

(a) a precise reversal of the nucleotide change that produced the original arg^- mutant allele
(b) a mutation to a suppressor of the arg^- allele occurring in another gene located in a different chromosome
(c) a suppressor mutation occurring in another gene located 10 map units away from the arg^- allele in the same chromosome

Challenge Problems

7.31 Gene conversion results from mismatch repair in a heteroduplex DNA molecule that gives rise to fungal asci with ratios of alleles such as 3 *A* : 1 *a* and 1 *A* : 3 *a* , instead of 2 *A* : 2 *a* .

(a) Why is a ratio of 2 *A* : 2 *a* expected?
(b) Why, among a large number of gametes chosen at random, is the Mendelian segregation ratio of 1 *A* : 1 *a* still observed in spite of gene conversion?

7.32 Which of the following 8-spore asci might arise from gene conversion at the *m* locus? Describe how these asci might arise.

Ascus type	A	B	C	D	E	F
Spore 1	+	+	+	+	+	+
Spore 2	+	+	+	+	+	+
Spore 3	+	*m*	+	+	*m*	*m*
Spore 4	+	*m*	*m*	*m*	*m*	*m*
Spore 5	*m*	*m*	*m*	+	+	*m*
Spore 6	*m*	*m*	*m*	*m*	+	*m*
Spore 7	*m*	+	*m*	*m*	*m*	*m*
Spore 8	*m*	+	*m*	*m*	*m*	*m*

7.32 A strain of *E. coli* contains a mutant $tRNA^{Leu}$. Instead of recognizing the codon 5'-CUG-3', as it does in nonmutant cells, the mutant tRNA recognizes the codon 5'-GUG-3'. A missense mutation of another gene, affecting amino acid number 28 along the chain, is suppressed in cells with the mutant $tRNA^{Leu}$.

(a) Assuming normal Watson–Crick base pairs, determine the anticodons of the wildtype and mutant $tRNA^{Leu}$ molecules.
(b) What kind of mutation is present in the mutant $tRNA^{Leu}$ gene?
(c) What amino acid would be inserted at position 28 of the mutant polypeptide chain if the missense mutation were not suppressed?
(d) What amino acid is inserted at position 28 when the missense mutation is suppressed?

Further Reading

Ames, B. W. 1979. Identifying environmental chemicals causing mutations and cancer. *Science* 204: 587.

Berg, D. E., and M. M. Howe. 1989. *Mobile DNA*. American Association for Microbiology. Washington, DC.

Bhatia, P. K., Z. G. Wang, and E. C. Friedberg. 1996. DNA repair and transcription. *Current Opinion in Genetics & Development* 6: 146.

Buermeyer, A. B., S. M. Deschenes, S. M. Baker, and R. M. Liskay. 1999. Mammalian DNA mismatch repair. *Annual Review of Genetics* 33: 533–564.

Comfort, N. C. 1995. Two genes, no enzyme: A second look at Barbara McClintock and the 1951 Cold Spring Harbor Symposium. *Genetics* 140: 1161.

Drake, J. W. 1991. Spontaneous mutation. *Annual Review of Genetics* 25: 125.

Hartl, D. L., A. R. Lohe, and E. R. Lozovskaya. 1997. Modern thoughts on an ancient *mariner*: Function, evolution, regulation. *Annual Review of Genetics* 31: 337.

Hickson, I. D. 1997. *Base Excision Repair of DNA Damage*. Austin, TX: Landes Bioscience.

Hoeijmakers, J. H. J. 1993. Nucleotide excision repair I: From *E. coli* to yeast. *Trends in Genetics* 9: 173.

Jackson, S. P. 1996. The recognition of DNA damage. *Current Opinion in Genetics & Development* 6: 19.

McClintock, B. 1965. Control of gene action in maize. *Brookhaven Symposium on Quantitative Biology* 18: 162.

Nickoloff, J. A., and M. F. Hoekstra, eds. 1998. *DNA Damage and Repair*. Totowa, NJ: Humana Press.

Plasterk, R. H. A., Z. Izsvak, and Z. Ivics. 1999. Resident aliens: The Tc1/*mariner* superfamily of transposable elements. *Trends in Genetics* 15: 326.

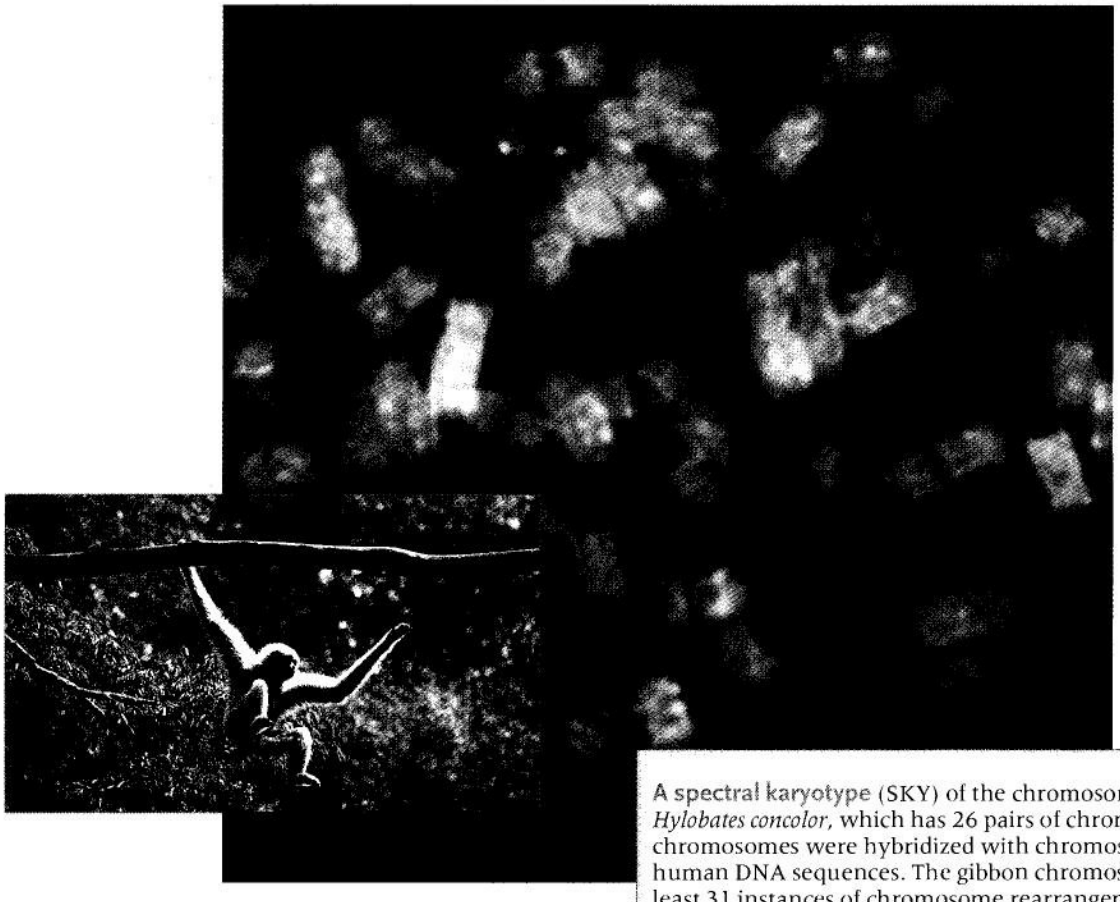
Shcherbak, Y. M. 1996. Confronting the nuclear legacy. I. Ten years of the Chernobyl era. *Scientific American*, April.

Smith, P. J., and C. J. Smith, eds. *DNA Recombination and Repair*. 1999. New York: Oxford University Press.

Sommer, S. S. 1995. Recent human germ-line mutation: Inferences from patients with hemophilia B. *Trends in Genetics* 11: 141.

CHAPTER 8

Molecular Organization of Chromosomes



A spectral karyotype (SKY) of the chromosomes of the gibbon *Hylobates concolor*, which has 26 pairs of chromosomes. The chromosomes were hybridized with chromosome-specific human DNA sequences. The gibbon chromosomes show at least 31 instances of chromosome rearrangement relative to the human genome; these are indicated by the multicoloring of the gibbon chromosomes. Other studies indicate that most of the changes occurred in the evolutionary lineage leading to the gibbon. The photograph shows a related species, *Hylobates lar*. [SKY: Courtesy of Johannes Wienberg; gibbon: J.C.Carton/ Bruce Coleman, Inc.]

CHAPTER OUTLINE

8.1 **Genome Size and Evolutionary Complexity: The C-Value Paradox**

8.2 **The Supercoiling Of DNA**
Topoisomerase Enzymes

8.3 **The Structure of Bacterial Chromosomes**

8.4 **The Structure of Eukaryotic Chromosomes**
The Nucleosome—The Structural Unit of Chromatin
The Nucleosome Core Particle
The Arrangement of Chromatin Fibers in a Chromosome

8.5 **Polytene Chromosomes**

8.6 **Repetitive Nucleotide Sequences in Eukaryotic Genomes**
Kinetics of DNA Renaturation
Analysis of Genome Size and Repetitive Sequences by Renaturation Kinetics

8.7 **Unique and Repetitive Sequences in Eukaryotic Genomes**
Unique Sequences
Highly Repetitive Sequences
Middle-Repetitive Sequences

8.8 **Centromere and Telomere Structure**
Molecular Structure of the Centromere
Molecular Structure of the Telomere

PRINCIPLES

- Most organisms have a genome composed of one or more molecules of double-stranded DNA. In eukaryotes the nuclear DNA molecules are usually linear, but in prokaryotes some DNA molecules are circular and others linear. Mitochondrial and chloroplast DNA is also circular. Circular DNA molecules are invariably supercoiled.
- Genome size varies greatly among organisms—and sometimes even among closely related organisms. Among metazoans, there is no correlation between the genome size of an organism and its evolutionary or metabolic complexity; this is known as the C-value paradox.
- DNA is rarely in a fully extended state but rather is folded in an intricate way to reduce its effective volume. In bacteria, circular chromosomal DNA is bound with certain proteins and folded to form a multiply looped structure of independently supercoiled domains called a nucleoid.
- Each eukaryotic chromosome contains a single, usually very long, DNA molecule. This molecule is combined with histone proteins to form chromatin. The elementary unit of chromatin is the nucleosomal core particle—an octamer of histone proteins around which is wrapped a length of duplex DNA.
- Most eukaryotic genomes, in addition to the "unique" DNA sequences that include the majority of genes, also contain DNA sequences that are highly repetitive or moderately repetitive.
- The centromere is a specialized DNA structure that functions as the center of chromosome movement in nuclear division.
- The telomere is another specialized DNA structure that serves to stabilize the chromosome tips from shortening through progressive loss of DNA.

CONNECTIONS

Post-Genomics Genetics
Shawn Ahmed and Jonathan Hodgkin, 2000
MRT-2 Checkpoint Protein is Required for Germline Immortality and Telomere Replication in C. elegans

Telomeres: The Beginning of the End
Carol W. Greider and Elizabeth H. Blackburn 1987
The Telomere Terminal Transferase of Tetrahymena *Is a Ribonucleoprotein Enzyme with Two Kinds of Primer Specificity*

Genome structure and genetic organization differ profoundly between prokaryotes and eukaryotes. Prokaryotes are unicellular organisms with relatively small genomes consisting of one or more DNA molecules that contain the essential genetic information. Usually the genome consists of just one molecule, but in some prokaryotes the essential genes are divided between two or more molecules. The cells may also contain one or more copies of accessory DNA molecules (*plasmids*) that contain nonessential information. The genome and the plasmids are often circular DNA molecules, but some prokaryotic genomes and plasmids are linear. Eukarotyic cells usually have larger genomes in the form of linear DNA molecules packaged with characteristic types of proteins (*histones*) into chromosomes, which are gathered together inside a nucleus enclosed in a nuclear envelope. (Circular chromosomes are occasionally found, but they are aberrations.) The DNA molecules in mitochondria and chloroplasts are also circular. The genetics of these organelles is the subject of Chapter 16. In this chapter we examine the molecular structures of prokaryotic genomes and of eukaryotic chromosomes.

8.1 Genome Size and Evolutionary Complexity: The C-Value Paradox

A summary of a small sample of genome sizes is shown in **Table 8.1**. Bacteriophage MS2 is one of the smallest viruses; it has only four genes in a single-stranded RNA molecule containing 3569 nucleotides.

Table 8.1 Genome size of some representative viral, bacterial, and eukaryotic genomes

Genome	Approximate length in thousands of nucleotides	Form
Viruses		
MS2	4	Single-stranded RNA
SV40	5	Circular double-stranded DNA
ϕX174	5	Circular single-stranded DNA; double-stranded replicative form
M13	6	
λ	50	Linear double-stranded DNA
Herpes simplex	152	
T2,T4,T6	165	
Smallpox	267	
Prokaryotes		
Methanococcus jannaschii	1600	Circular double-stranded DNA
Escherichia coli	4600	
Borrelia burgdorferi	910	Linear double-stranded DNA
Eukaryotes		Haploid chromosome number
Saccharomyces cerevisiae (yeast)	13,000	16
Caenorhabditis elegans (nematode)	97,000	6
Arabidopsis thaliana (wall cress)	100,000	5
Drosophila melanogaster (fruit fly)	180,000	4
Homo sapiens (human being)	3,000,000	23
Zea mays (maize)	4,500,000	10
Amphiuma sp. (salamander)	76,500,000	14

SV40 virus, which infects monkey and human cells, has a genetic complement of five genes in a circular double-stranded DNA molecule of about 5 kb (5000 nucleotide pairs). The more complex phages and animal viruses have as many as 250 genes and DNA molecules ranging from 50 to 300 kb. Prokaryotic genomes are substantially larger. Archaeal genomes (for example, *Methanococcus jannaschi*) are generally similar in size to bacterial genomes. Some prokaryotic genomes consist of linear DNA, others circular DNA. For example, the chromosome of the spirochete *Borrelia burgdorferi*, the agent of Lyme disease, is a linear DNA molecule of about 910 kb, and that of *Escherichia coli* is a circular DNA molecule of 4600 kb. The genomes of unicellular eukaryotes are even larger. The genome size of budding yeast, *Saccharomyces cerevisiae*, is 13,000 kb. The units of length of nucleic acids in which genome sizes are typically expressed are as follows:

1. ***kilobase (kb)*** 10^3 nucleotide pairs (double-stranded) or 10^3 nucleotides (single-stranded)
2. ***megabase (Mb)*** 10^6 nucleotide pairs (double-stranded) or 10^6 nucleotides (single-stranded)
3. ***gigabase (Gb)*** 10^9 nucleotide pairs (double-stranded) or 10^9 nucleotides (single-stranded)

In these terms, viral genomes are typically in the range 100–1000 kb, bacterial genomes typically in the range 1–10 Mb, and eukaryotic genomes typically in the range 100–1000 Mb. (The smallest eukaryotic genomes are about 10 Mb.)

Among multicellular eukaryotes, however, genome size often differs enormously among species with the same level of metabolic, developmental, and behavioral complexity. This lack of correlation is known as the **C-value paradox.** The differences are often quite incredible. Differences in genome size among species of protozoa are 5800-fold, among arthropods 250-fold, fish 350-fold, algae 5000-fold, and angiosperms 1000-fold (**Figure 8.1**). The term *paradox* is amply justified by observing that the genome size of *Amphiuma* salamanders is 90 Gb, whereas that of the human is 3 Gb—values that differ by a factor of 30.

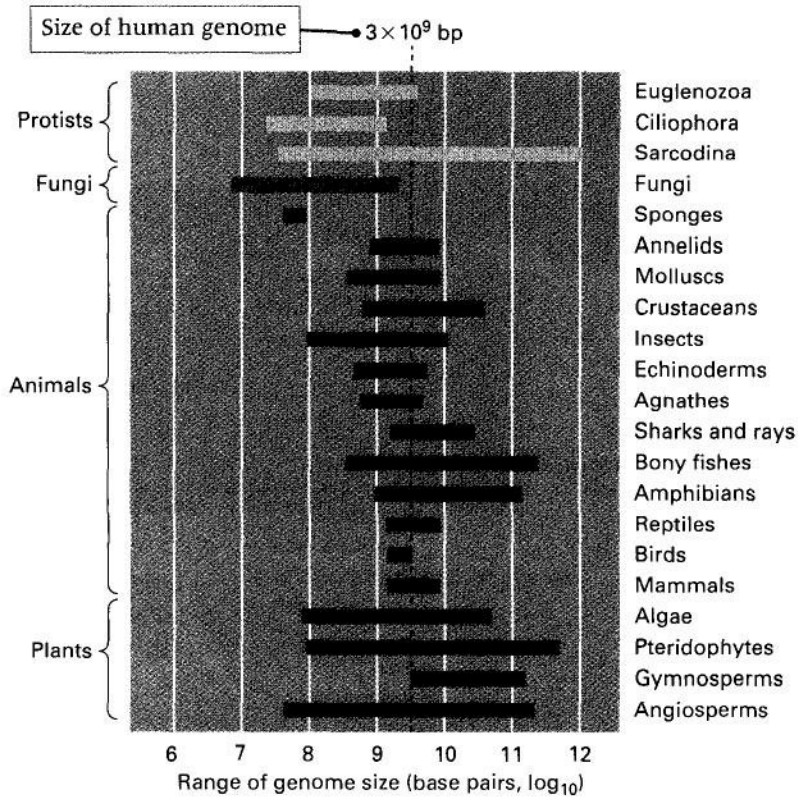
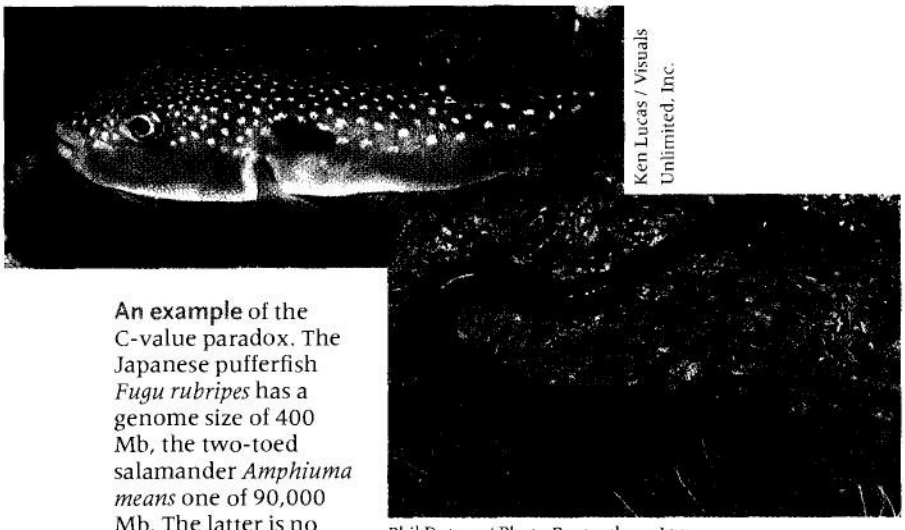


Figure 8.1 Genome size ranges over several orders of magnitude in some groups of organisms, and genome size is not correlated with developmental, metabolic, or behavioral complexity.

The C-value paradox is not due to differences in the number of genes. For example, the number of transcripts produced by the human genome is no more than about 10^5, which serves as a rough estimate of the number of genes. Considering the number and size of both transcripts and proteins produced in human cells, it appears that no more than about 4 percent of the human genome actually codes for proteins. To take an example from plants, rice and maize have about the same number of genes (transcripts and proteins), but the maize genome at 2.5 Gb is about six times larger than that of rice at 0.43 Gb. In nearly all higher animals and plants, the actual number of genes is much less than the theoretical maximum. The reason for the discrepancy is that in higher organisms, much of the DNA has functions other than coding for the amino acid sequence of proteins (as we shall see in Chapter 11), and



Ken Lucas / Visuals Unlimited, Inc.

Phil Dotson / Photo Researchers, Inc.

An example of the C-value paradox. The Japanese pufferfish *Fugu rubripes* has a genome size of 400 Mb, the two-toed salamander *Amphiuma means* one of 90,000 Mb. The latter is no more "complex" than the former.

most of the rest consists of transposable elements and other types of repetitive sequences discussed in Section 8.6.

8.2 The Supercoiling of DNA

The genome of *E. coli* is a circular DNA molecule about 1500 μm long compacted inside a cell only about 2 μm long and 1 μm in diameter. The DNA molecule must be compacted; otherwise it would not fit inside the cell. At the same time, the genes must be available for transcription as their expression is needed. Part of the compaction of the circular chromosome is due to its being **supercoiled,** which means that segments of double-stranded DNA are twisted around one another, in a manner analogous to the way a telephone cord can be twisted around itself. The geometry of supercoiling can be illustrated by a simple example (**Figure 8.2**). Consider first a linear duplex DNA molecule whose ends are joined in such a way that each strand forms a continuous circle. The covalently closed circular DNA molecule is said to be **relaxed** if no twisting is present other than the helical twisting (Figure 8.2A). The individual polynucleotide strands of a relaxed circle form the usual right-handed (positive) helical structure with ten nucleotide pairs per turn of the helix. Suppose you were to cut one strand in a relaxed circle and unwind it one complete rotation of 360° so as to undo one complete turn of the double helix. When the ends were rejoined again, the result would be a circular helix that is "underwound." Because a DNA molecule has a strong tendency to maintain its standard helical form with ten nucleotide pairs per turn, the circular molecule would respond to the underwinding in one of two ways:

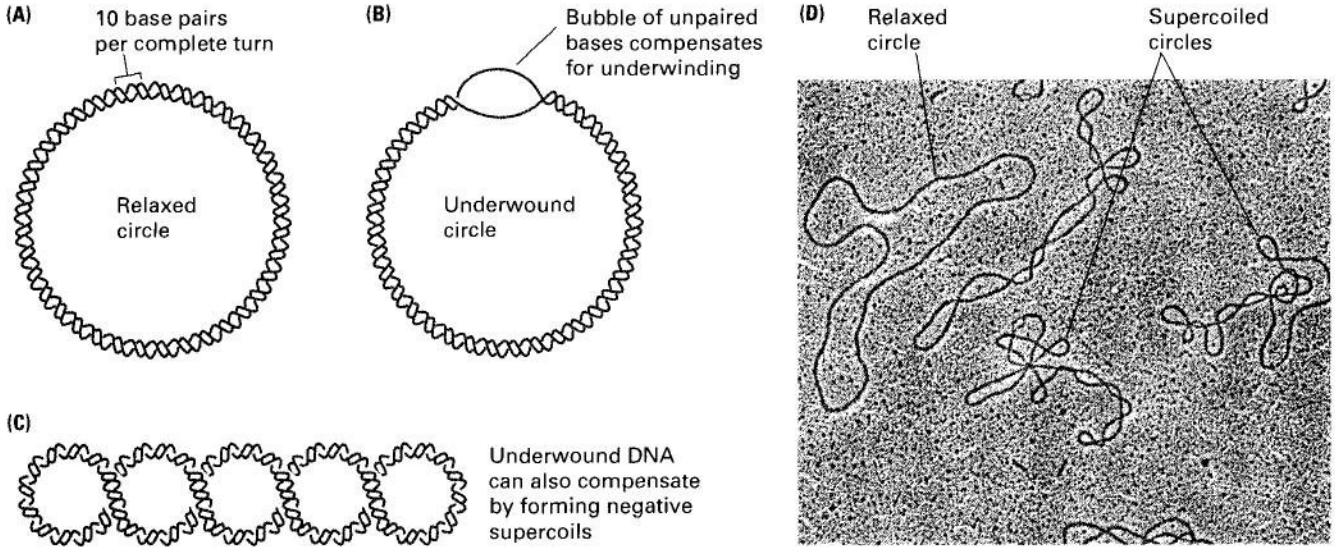
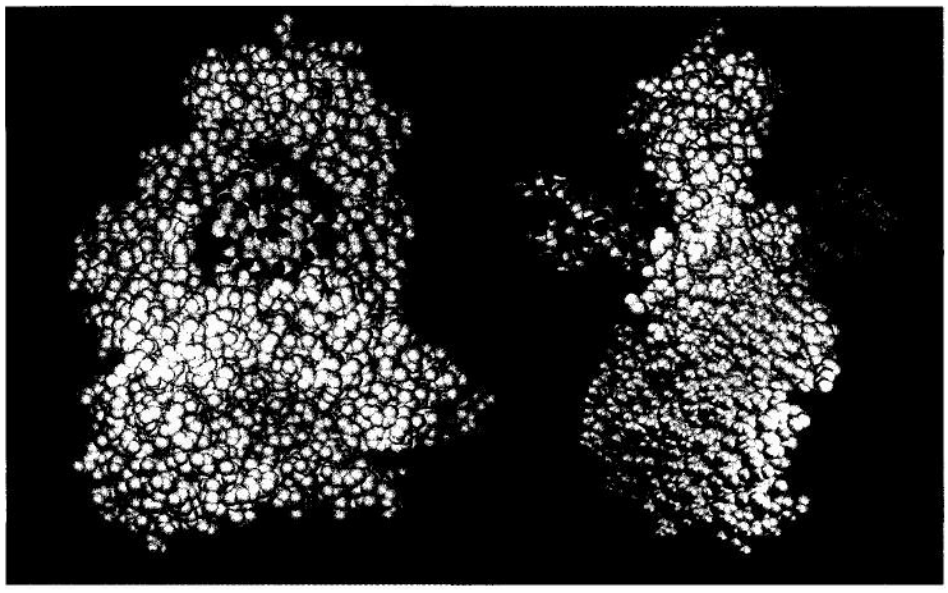


Figure 8.2 Different states of a covalent circle. (A) A nonsupercoiled (relaxed) covalent circle with 36 helical turns. (B) An underwound covalent circle with only 32 helical turns. (C) The molecule in part B, but with four twists to eliminate the underwinding. (D) Electron micrograph showing nicked circular and supercoiled DNA of phage PM2. Note that no bases are unpaired in part C. In solution, parts B and C would be in equilibrium. [Electron micrograph courtesy of K. G. Murti.]



Proposed structure of an association between duplex DNA and topoisomerase I in which the DNA passes through a hole formed in the enzyme. The two views are perpendicular to each other. The structure of the enzyme includes a pair of jaw-like projections that can open and close. The intermediate shown here is created when the enzyme closes its jaws around a DNA molecule, completely enclosing it. What happens then is that one DNA strand is cleaved, swiveled, and reconnected, whereupon the jaws open and the DNA molecule is released. [Courtesy of C. D. Lima, J. C. Wang, and A. Móndragon. Reprinted by permission from *Nature* 367: 138, C. D. Lima, J. C. Wang, and A. Móndragon. Copyright 1994 Macmillan Magazines Ltd.]

(1) by forming regions with "bubbles" in which the bases are unpaired (Figure 8.2B) or (2) by twisting the circular molecule in the opposite sense from the direction of underwinding. This twisting is the supercoiling, and a molecule with right-handed twists is **negatively supercoiled** (Figure 8.2C). Examples of supercoiled molecules are shown in Figure 8.2C and D. The two responses to underwinding are not independent, and underwinding is usually accommodated by a combination of the two processes: An underwound molecule contains some bubbles of unpaired bases and some supercoiling, and the supercoiling predominates.

Although *E. coli* and most other bacteria have negatively supercoiled DNA, most thermophiles and other members of the kingdom Archaea have positively supercoiled DNA, in which the double helix is relatively overwound and the twists in the supercoiled DNA are left-handed. What about supercoiling in eukaryotes? The DNA in eukaryotic chromosomes typically has only a few regions that are supercoiled, and these are usually around genes that are actively being transcribed.

Topoisomerase Enzymes

The supercoiling of natural DNA molecules is produced by DNA **topoisomerase** enzymes. The class of enzymes denoted *topoisomerase I* act by wrapping themselves completely around a DNA duplex and causing a single-stranded nick by breaking a phosphodiester bond in the backbone of one of the DNA strands. The nicked ends are swiveled around the intact strand, changing the helical twist (number of helical turns). Then the nick is resealed and the duplex released from the enzyme. Depending on conditions, topoisomerase I enzymes can either increase or decrease the amount of supercoiling.

A second class of topoisomerase enzymes is called *topoisomerase II*. These enzymes work by producing a double-stranded gap in one molecule through which another double-stranded molecule is passed. In this way topoisomerase II enzymes are able to pass one DNA duplex entirely through another or to separate two circular DNA molecules that are interlocked. The mechanism of action of topoisomerase II from budding yeast *Saccharomyces cerevisiae* is illustrated in **Figure 8.3**. The molecular structure of the en-

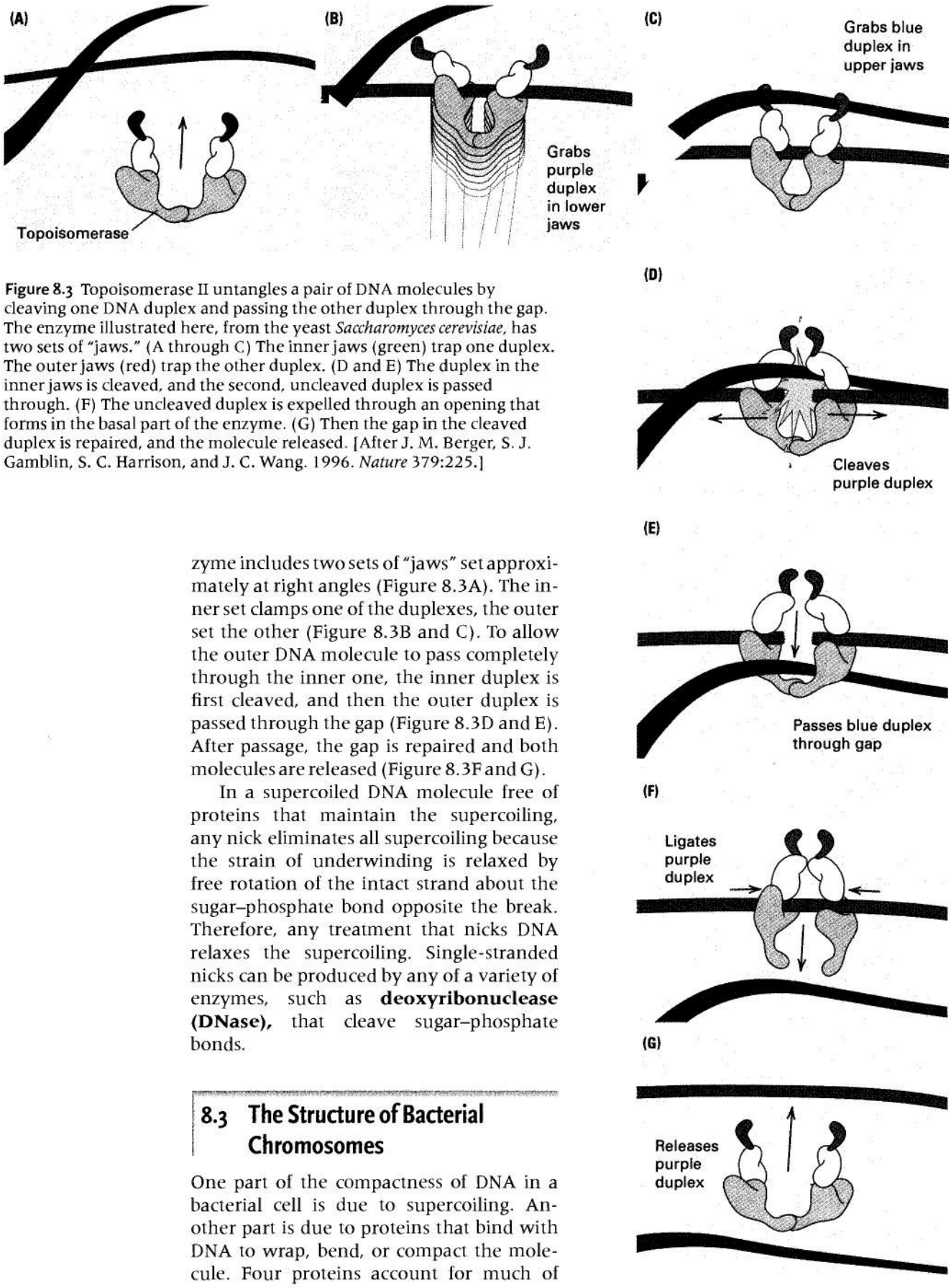


Figure 8.3 Topoisomerase II untangles a pair of DNA molecules by cleaving one DNA duplex and passing the other duplex through the gap. The enzyme illustrated here, from the yeast *Saccharomyces cerevisiae*, has two sets of "jaws." (A through C) The inner jaws (green) trap one duplex. The outer jaws (red) trap the other duplex. (D and E) The duplex in the inner jaws is cleaved, and the second, uncleaved duplex is passed through. (F) The uncleaved duplex is expelled through an opening that forms in the basal part of the enzyme. (G) Then the gap in the cleaved duplex is repaired, and the molecule released. [After J. M. Berger, S. J. Gamblin, S. C. Harrison, and J. C. Wang. 1996. *Nature* 379:225.]

zyme includes two sets of "jaws" set approximately at right angles (Figure 8.3A). The inner set clamps one of the duplexes, the outer set the other (Figure 8.3B and C). To allow the outer DNA molecule to pass completely through the inner one, the inner duplex is first cleaved, and then the outer duplex is passed through the gap (Figure 8.3D and E). After passage, the gap is repaired and both molecules are released (Figure 8.3F and G).

In a supercoiled DNA molecule free of proteins that maintain the supercoiling, any nick eliminates all supercoiling because the strain of underwinding is relaxed by free rotation of the intact strand about the sugar–phosphate bond opposite the break. Therefore, any treatment that nicks DNA relaxes the supercoiling. Single-stranded nicks can be produced by any of a variety of enzymes, such as **deoxyribonuclease (DNase),** that cleave sugar–phosphate bonds.

8.3 The Structure of Bacterial Chromosomes

One part of the compactness of DNA in a bacterial cell is due to supercoiling. Another part is due to proteins that bind with DNA to wrap, bend, or compact the molecule. Four proteins account for much of

this level of packaging (HU for wrapping, FIS and IHF for bending, and HNS for compaction), and together they constitute approximately 10 percent of the total protein in the cell. The result is that the chromosome of *E. coli* is a condensed unit called a folded chromosome or **nucleoid.** The term *chromosome* is technically a misnomer for this structure, because it is not a true chromosome in the sense of a eukaryotic chromosome. Nevertheless, the term is widely used. The most striking feature of the bacterial nucleoid is that the DNA is organized into a set of looped domains (**Figure 8.4**). As isolated from bacterial cells, the nucleoid contains small amounts of several proteins that are thought to be responsible for the multiply looped structure. The degree of condensation of the isolated nucleoid (that is, its physical dimensions) is affected by a variety of factors, and some controversy exists about the state of the nucleoid within the cell.

Figure 8.4 also shows loops of DNA in the *E. coli* chromosome that are supercoiled. Those loops that are not supercoiled result from the action of DNases during isolation. The presence of both supercoiled and relaxed loops indicates that the loops are in some way independent of one another. In the preceding section, we stated that supercoiling is generally eliminated in a DNA molecule by one single-strand break. However, such a break in the *E. coli* chromosome does not eliminate all supercoiling. If nucleoids, all of whose loops are supercoiled, are treated with a DNase and examined at various times after treatment, it is observed that supercoiling is relaxed in one loop at a time, not in all loops at once (**Figure 8.5**). The loops must be isolated from one another in such a way that rotation in one loop is not transmitted to other loops. The independence is probably the result of proteins that bind to the DNA in a way that prevents rotation of the helix.

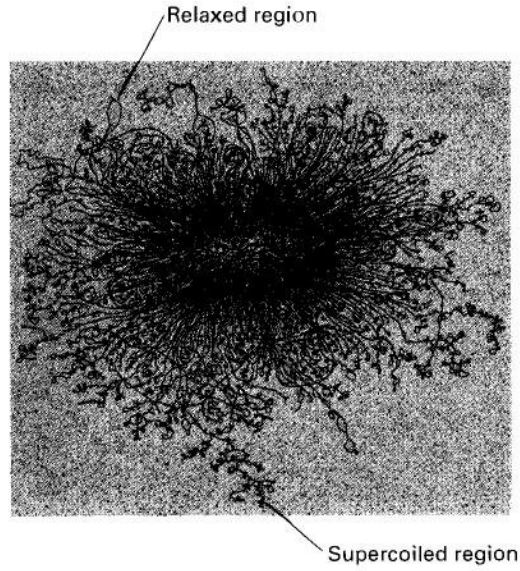


Figure 8.4 Electron micrograph of an *E. coli* chromosome, showing the multiple loops emerging from a central region.

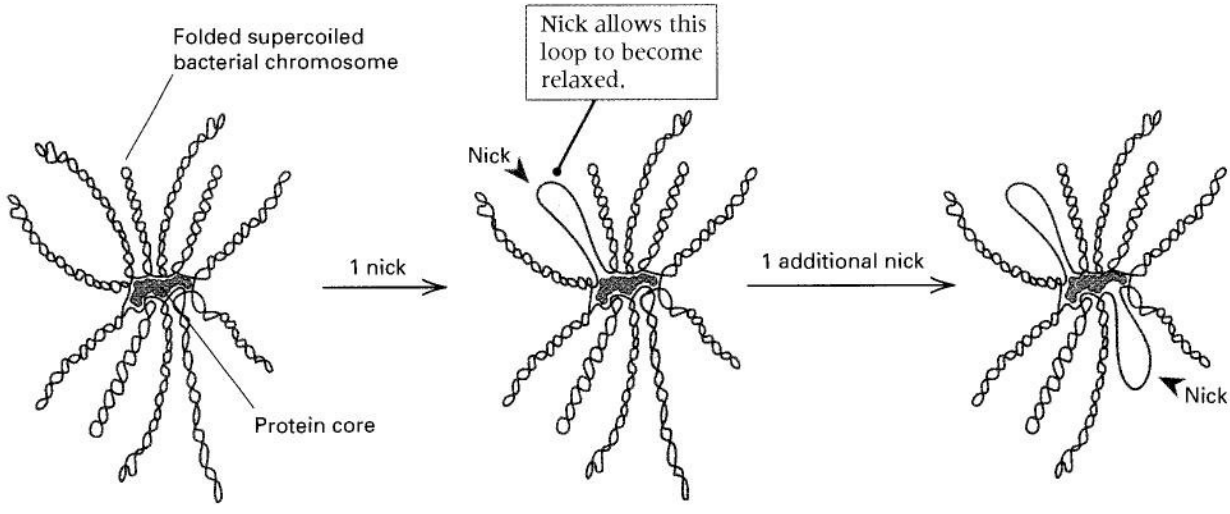


Figure 8.5 Schematic drawing of the folded supercoiled *E. coli* chromosome, showing 11 of the 40 to 50 loops attached to a protein core (blue shaded area) and the opening of loops by nicks.

8.4 The Structure of Eukaryotic Chromosomes

A remarkable feature of the genetic apparatus of eukaryotes is the feat of packaging in which an enormous length of genetic material is condensed into a relatively few small chromosomes. The largest of the 23 human chromosomes contains a DNA molecule that is 82 mm (8.2×10^4 μm) in length. However, at metaphase of a mitotic division, this DNA molecule is condensed into a compact structure about 10 μm long and less than 1 μm in diameter. An analogy may make it easier to appreciate the prodigious feat of packaging that such chromosome condensation represents. If the DNA molecule in human chromosome 1 (the largest chromosome) were a cooked spaghetti noodle 1 mm in diameter, it would stretch for 25 miles; in chromosome condensation, this noodle is gathered together, coil upon coil, until at metaphase it is a canoe-sized tangle of spaghetti 16 feet long and 2 feet wide. After cell division, the noodle is unwound again.

The Nucleosome—The Structural Unit of Chromatin

The DNA of all eukaryotic chromosomes is associated with numerous protein molecules in a stable, ordered aggregate called **chromatin.** Some of the proteins present in chromatin determine chromosome structure and the changes in structure that occur during the division cycle of the cell. Other chromatin proteins appear to play important roles in regulating chromosome functions.

The Nucleosome Core Particle

The simplest form of chromatin is present in nondividing eukaryotic cells, when chromosomes are not sufficiently condensed to be visible by light microscopy. Chromatin isolated from such cells is a complex aggregate of DNA and proteins. The major class of proteins comprises the **histone** proteins. Histones are largely responsible for the structure of chromatin. Five major types of histones—H1, H2A, H2B, H3, and H4—are present in the chromatin of nearly all eukaryotes in amounts about equal in mass to that of the DNA. Histones are small proteins (100–200 amino acids) that differ from most other proteins in that from 20 to 30 percent of the amino acids are lysine and arginine, both of which have a positive charge. (Only a few percent of the amino acids of a typical protein are lysine and arginine.) The positive charges enable histone molecules to bind to DNA, primarily by electrostatic attraction to the negatively charged phosphate groups in the sugar–phosphate backbone of DNA. Placing chromatin in a solution with a high salt concentration (for example, 2 molar NaCl) to eliminate the electrostatic attraction causes the histones to dissociate from the DNA. Histones also bind tightly to each other; both DNA–histone and histone–histone binding are important for chromatin structure.

The histone molecules from different organisms are remarkably similar, with the exception of H1. In fact, the amino acid sequences of H3 molecules from widely different species are almost identical. For example, the sequences of H3 of cow chromatin and pea chromatin differ by only 4 of 135 amino acids. The H4 proteins of all organisms also are quite similar; cow and pea H4 differ by only 2 of 102 amino acids. There are few other proteins whose amino acid sequences vary so little from one species to the next. When the variation between organisms is very small, we say that the sequence is highly **conserved**. The extraordinary conservation in histone composition through hundreds of millions of years of evolutionary divergence is consistent with the important role of these proteins in the structural organization of eukaryotic chromosomes.

In the electron microscope, chromatin resembles a regularly beaded thread (**Figure 8.6**). The bead-like units in chromatin are called **nucleosomes.** The organization of the nucleosomes in chromatin is illustrated in **Figure 8.7**, part A. Each unit has a definite composition, consisting of two molecules each of H2A, H2B, H3, and H4, a segment of DNA containing about 200 nucleotide pairs, and one molecule of histone H1. The complex of two subunits each of H2A, H2B, H3, and H4, as well as part of the DNA, forms each "bead," and the remaining DNA and histone H1 bridges between the beads.

Brief treatment of chromatin with certain DNases (for example, *micrococcal nucle-*

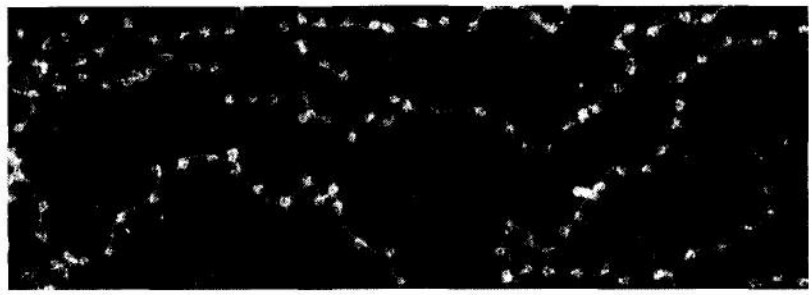


Figure 8.6 Dark-field electron micrograph of chromatin, showing the beaded structure at low salt concentration. The beads have diameters of about 100 Å. [Courtesy of Ada Olins.]

ase from the bacterium *Staphylococcus aureus*) yields a collection of small particles of quite uniform size consisting only of histones and DNA. The DNA fragments in these particles are of lengths equal to about 200 nucleotide pairs or small multiples of that unit size (the precise size varies with species and tissue). These particles result from cleavage of the linker DNA segments between the beads (Figure 8.7B). More

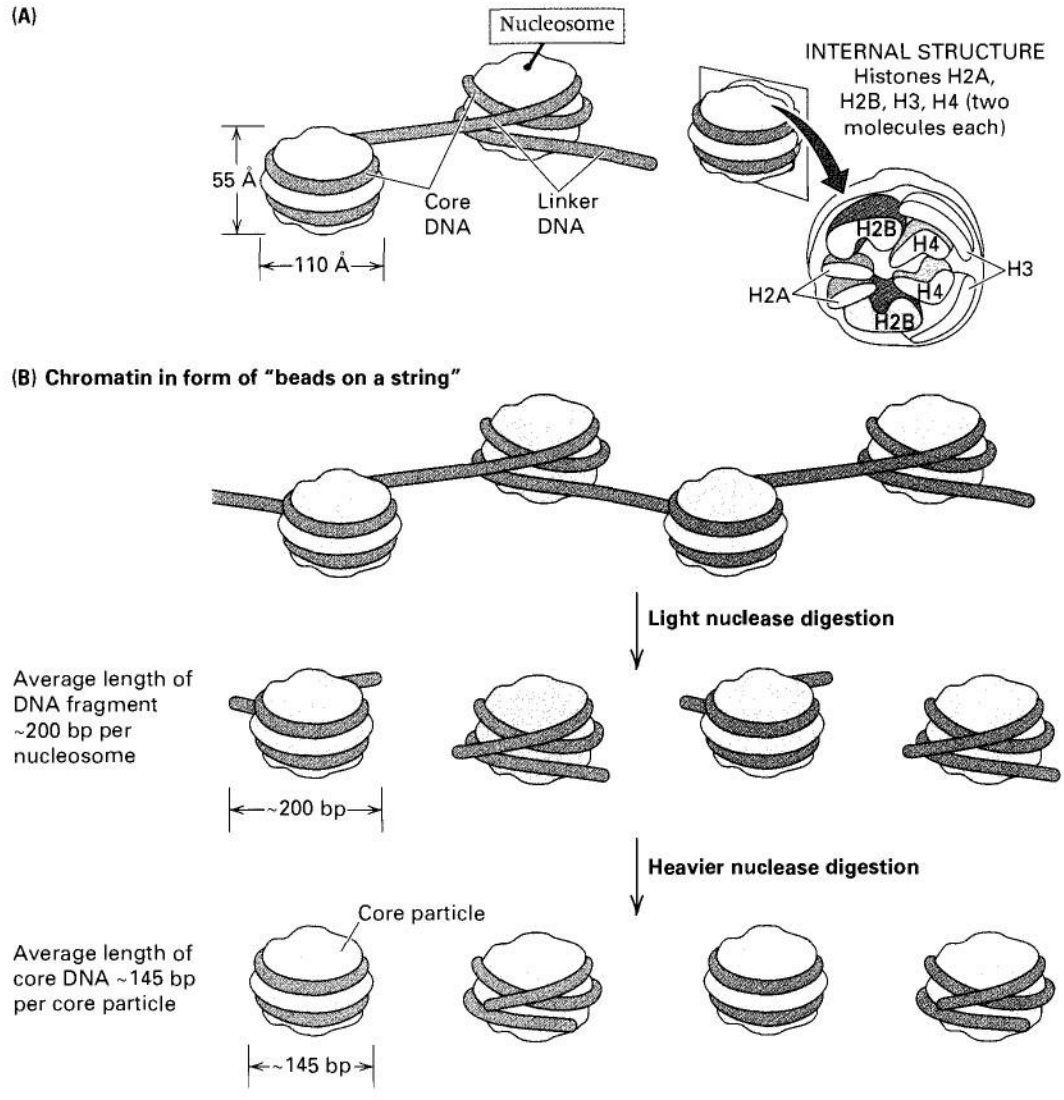


Figure 8.7 (A) Organization of nucleosomes. The DNA molecule is wound one and three-fourths turns around a histone octamer called the core particle. If H1 were present, it would bind to the octamer surface and to the linkers, causing the linkers to cross. (B) Effect of treatment with micrococcal nuclease. Brief treatment cleaves the DNA between the nucleosomes and results in core particles associated with histone H1 and approximately 200 base pairs of DNA. More extensive treatment results in loss of H1 and digestion of all but 145 base pairs of DNA in intimate contact with each core particle.

extensive treatment with DNase results in loss of the H1 histone and digestion of all the DNA except that protected by the histones in the bead. The resulting structure is called a **core particle,** the detailed structure of which is shown in **Figure 8.8**. It consists of two molecules each of H2A (yellow), H2B (red), H3 (blue), and H4 (green), around which is wound a segment of DNA of about 145 nucleotide pairs. Each nucleosome is composed of a core particle, the linker DNA between adjacent core particles (which is removed by extensive nuclease digestion), and one molecule of H1; the H1 binds to the histone octamer and to the linker DNA, causing the linkers that extend from both sides of the core particle to cross and draw nearer to the octamer, though some of the linker DNA does not come into contact with any histones. The size of the linker ranges from 20 to 100 nucleotide pairs for different species and varies even in different cell types in the same organism (55 nucleotide pairs is usually considered an average size). Little is known about the structure of the linker DNA or about whether it has a special genetic function, and the cause of the variation in its length is also unknown.

The Arrangement of Chromatin Fibers in a Chromosome

The DNA molecule of a chromosome is folded and refolded again in such a way that it is convenient to think of chromosomes as having several hierarchical levels of organization, each responsible for a particular degree of shortening of the enormously long strand (**Figure 8.9**). Assembly of DNA and histones can be considered the first level—namely, a sevenfold reduction in length of the DNA and the formation of a beaded flexible fiber 110 Å (11 nm) wide (Figure 8.9B), roughly five times the width of free DNA (Figure 8.9A). The structure of chromatin varies with the concentration of salts, and the 110-Å fiber is present only when the salt concentration is quite low. If the salt concentration is increased slightly, then the fiber becomes shortened somewhat by forming a zigzag arrangement of closely spaced beads between which the linking DNA is no longer visible in electron micrographs. If the salt concentration is further increased to that present in living cells, then a second level of compaction occurs: the organization of the 110-Å nucleosome fiber into a shorter, thicker fiber

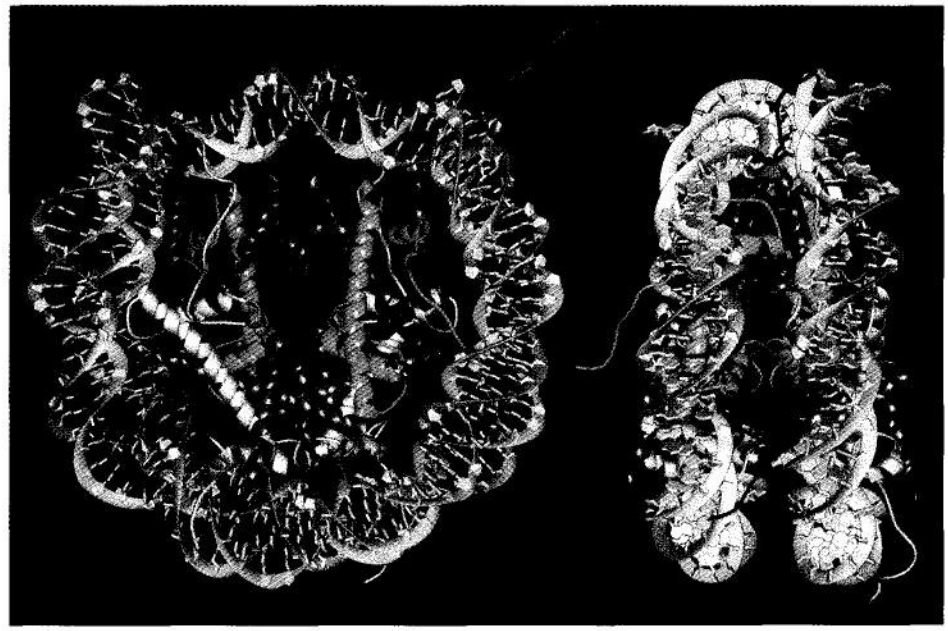


Figure 8.8 Subunit structure of the nucleosome core particle with the DNA duplex wrapped around it. The color coding for the histone monomers is H2A, yellow; H2B, red; H3, blue; and H4, green. [Courtesy of Timothy J. Richmond. Reprinted by permission from *Nature* 389: 251, K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Copyright 1997 Macmillan Maganizes Ltd.]

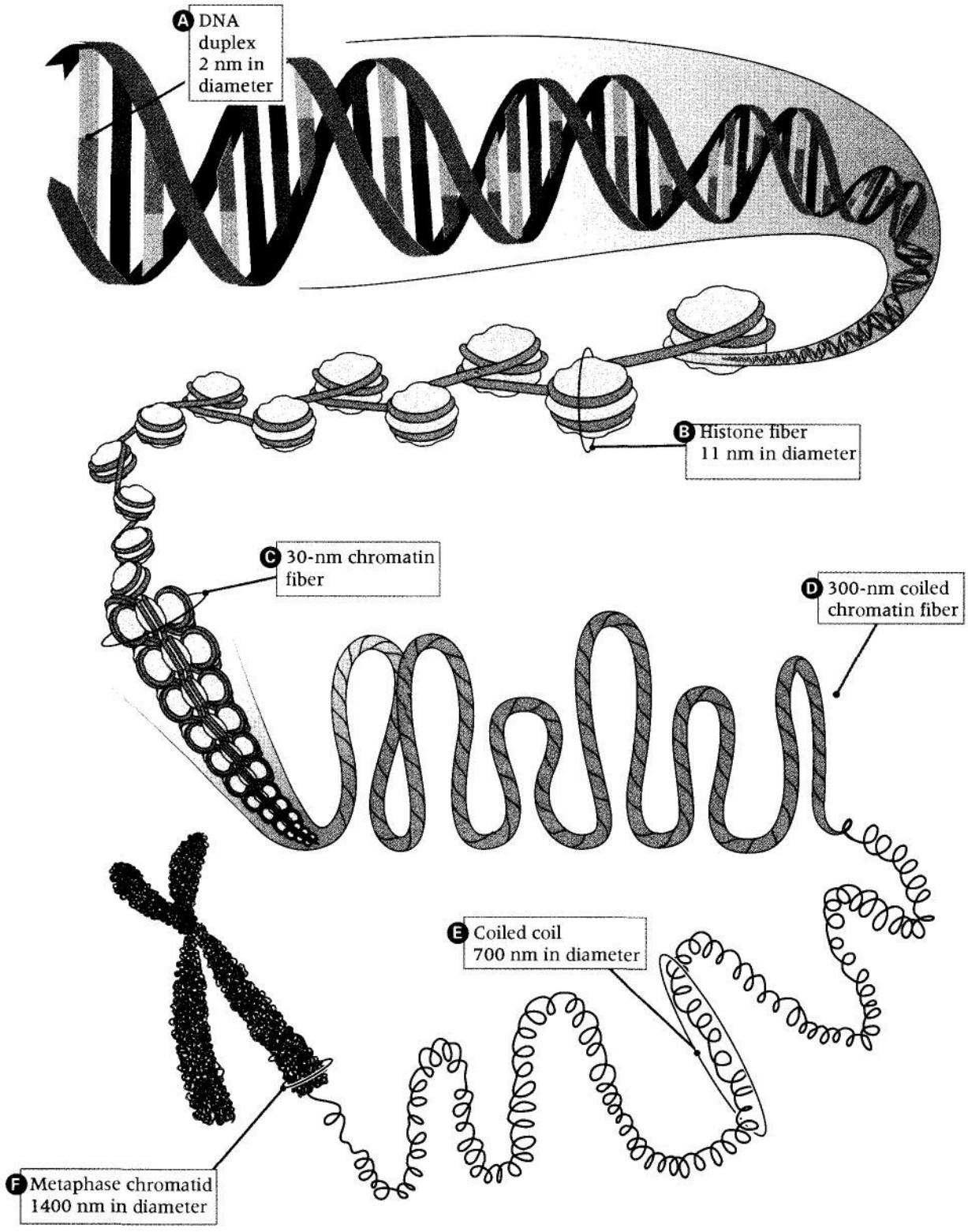


Figure 8.9 Various stages in the condensation of DNA (A) and chromatin (B through E) in forming a metaphase chromosome (F). The dimensions indicate known sizes of intermediates, but the detailed structures are hypothetical.

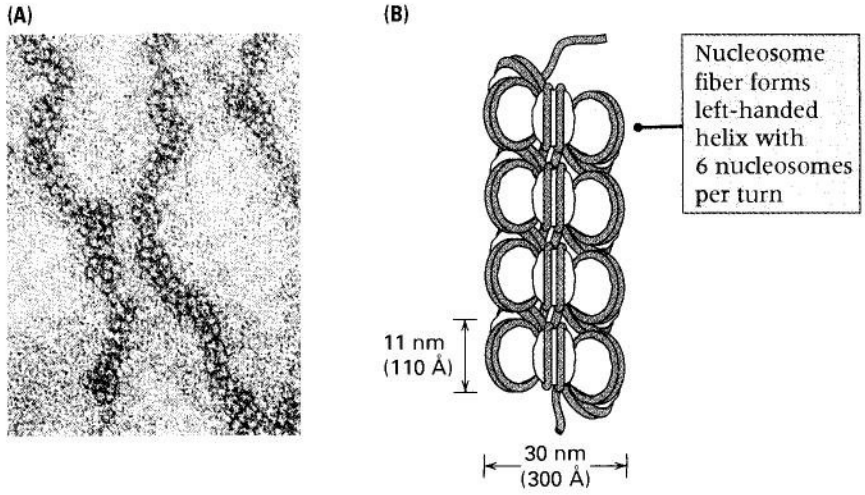


Figure 8.10 (A) Electron micrograph of the 30-nm component of mouse metaphase chromosomes. (B) A proposed solenoidal model of chromatin. The DNA (blue-gray) is wound around each nucleosome. It is unlikely that the real structure is so regular. [Electron micrograph courtesy of Barbara Hamkalo; drawing after J. T. Finch and A. Klug. 1976. *Proc. Nat. Acad. Sci. USA* 73: 1900.]

with an average diameter ranging from 300 to 350 Å; it is called the **30-nm fiber** (Figure 8.9C). In forming this structure, the 110-Å fiber apparently coils in a somewhat irregular left-handed superhelix or solenoidal supercoil with six nucleosomes per turn (**Figure 8.10**). Most intracellular chromatin is believed to have the solenoidal supercoiled configuration.

The final level of organization is that in which the 30-nm fiber condenses into a chromatid of the compact metaphase chromosome (Figure 8.9D through F). Little is known about this process other than that it seems to proceed in stages. In electron micrographs of isolated metaphase chromosomes from which histones have been removed, the partly unfolded DNA has the form of an enormous number of loops that seem to extend from a central core, or **scaffold,** composed of nonhistone chromosomal proteins (**Figure 8.11**). Electron microscopic studies of chromosome condensation in mitosis and meiosis suggest that the scaffold extends along the chromatid and that the 30-nm fiber becomes arranged into a helix of loops radiating from the scaffold. Details are not known

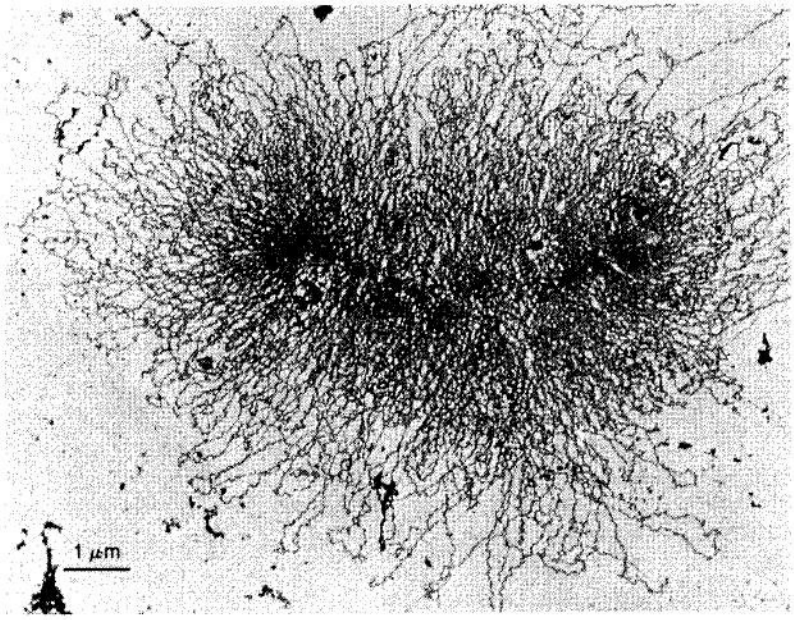


Figure 8.11 Electron micrograph of a partially disrupted anaphase chromosome of the milkweed bug *Oncopeltus fasciatus*, showing multiple loops of 30-nm chromatin at the periphery. [From V. Foe, H. Forrest, L. Wilkinson, and C. Laird. 1982. *Insect Ultrastructure* 1: 222.]

about the additional folding that is required of the fiber in each loop to produce the fully condensed metaphase chromosome.

The genetic significance of the compaction of DNA and protein into chromatin and ultimately into the chromosome is that it greatly facilitates the movement of the genetic material during nuclear division. Relative to a fully extended DNA molecule, the length of a metaphase chromosome is reduced by a factor of approximately 10^4 as a result of chromosome condensation. Without chromosome condensation, the chromosomes would become so entangled that there would be many more abnormalities in the distribution of genetic material into daughter cells.

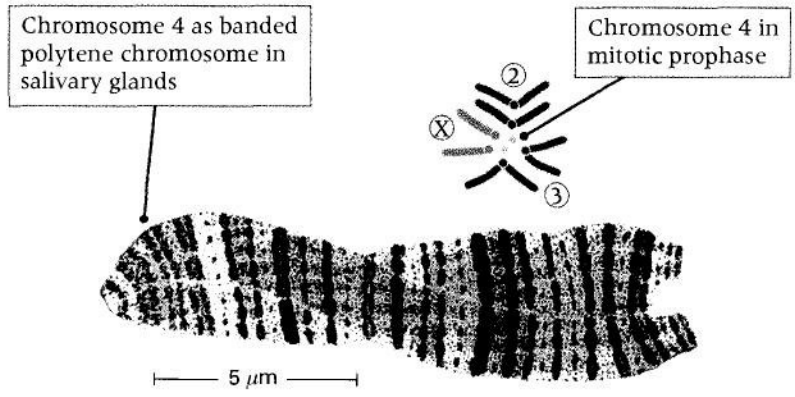


Figure 8.12 The polytene fourth chromosome of *Drosophila melanogaster.* The somatic chromosomes of *Drosophila*, drawn to scale with respect to the polytene fourth chromosome, are shown at the upper right as they appear in mitotic prophase. [From C. Bridges. 1935. *Journal of Heredity* 26: 60. Reproduced by permission of Oxford University Press.]

8.5 Polytene Chromosomes

Giant chromosomes called **polytene chromosomes** are found in the nuclei of cells of the salivary glands and certain other tissues of the larvae of *Drosophila* and other two-winged (dipteran) flies. These chromosomes contain about 1000 identical DNA molecules laterally aligned. Each has a length and cross-sectional diameter many times greater than those of the corresponding chromosome at mitotic metaphase in ordinary somatic cells, as well as a constant and distinctive pattern of transverse banding (**Figure 8.12**). The polytene structures are formed by repeated replication of the DNA in a closely synapsed pair of homologous chromosomes without separation of the replicated chromatin strands. Polytene chromosomes are atypical chromosomes and are formed in "terminal" tissues of the larva that are eliminated in the formation of the pupa. Although they do not contribute to the adult fly, the polytene tissues of larvae have been especially valuable in the genetics of *Drosophila*, as will become apparent in Chapter 9.

In polytene nuclei of *D. melanogaster* and other dipteran species, large blocks of heterochromatin adjacent to the centromeres are aggregated into a single compact mass called the **chromocenter.** Because the two largest chromosomes in *Drosophila* (chromosomes numbered 2 and 3) have centrally located centromeres, the chromosomes appear in the configuration shown in **Figure 8.13**. The paired X chromosomes (in a female), the left and right arms of chromosomes 2 and 3, and a short chromosome (chromosome 4) project from the chromocenter. In a male, the Y chromosome, which consists almost entirely of heterochromatin, is incorporated in the chromocenter.

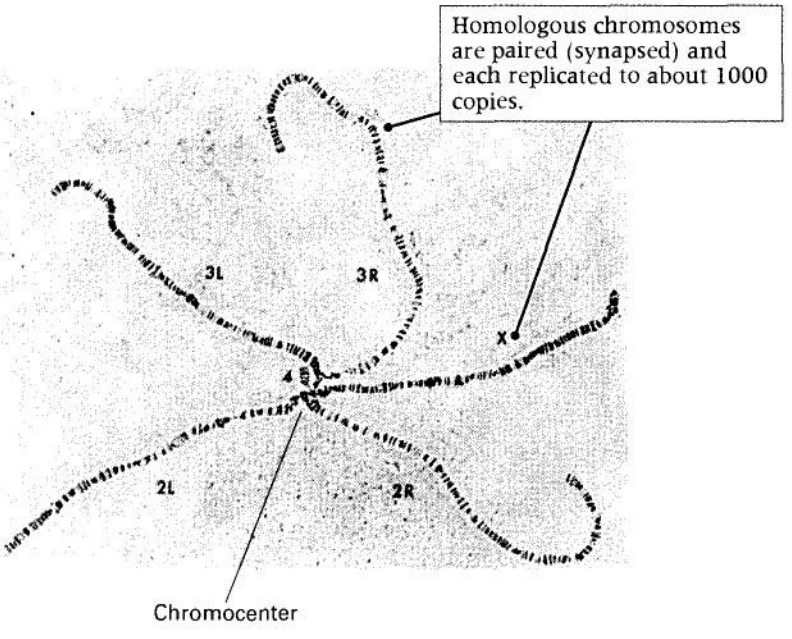


Figure 8.13 Polytene chromosomes from a larval salivary gland cell of *Drosophila melanogaster.* The chromocenter is the central region in which the centromeric regions of all chromosomes are united. [Courtesy of George Lefevre.]

The darkly staining transverse bands in polytene chromosomes have about a tenfold range in width. These bands result from the side-by-side alignment of tightly folded regions of the individual chromatin strands that are often visible in mitotic and meiotic prophase chromosomes as chromomeres. More DNA is present within the bands than in the interband (lightly stained) regions. About 5000 bands have been identified in the *D. melanogaster* polytene chromosomes. This linear array of bands, which has a pattern that is constant and characteristic for each species, provides a finely detailed **cytological map** of the chromosomes. The banding pattern is such that observers with sufficient training and experience can identify short regions in any of the chromosomes (Figure 8.13).

Because of their large size and finely detailed morphology, polytene chromosomes are exceedingly useful for a process called *in situ* nucleic acid hybridization. During the procedure of *in situ hybridization*, nuclei containing polytene chromosomes are squashed and the chromosomal DNA denatured, after which labeled probe DNA or RNA is added under conditions that favor renaturation. After washing, the only probe that remains in the chromosomes has formed hybrid duplexes with chromosomal DNA, and its position can be identified cytologically (**Figure 8.14**).



Figure 8.14 Fluorescent *in situ* hybridization with polytene chromosomes of *D. melanogaster*. The pink signals show the chromosomal locations of the gene for the enzyme alcohol dehydrogenase. This strain has two copies of the gene, one the endogenous copy and the other an ectopic copy introduced by germline transformation. [Courtesy of Mark L. Siegal.]

8.6 Repetitive Nucleotide Sequences in Eukaryotic Genomes

In bacterial genomes the variation in average base composition from one region to another is usually quite small. In eukaryotes, by contrast, some components of the genome have a base composition that is quite different from the average. For example, one component of crab DNA is only 3 percent G + C, compared to an average of approximately 50 percent G + C for the rest of the genome. The components with unusually low or unusually high G + C contents often consist of fairly short nucleotide sequences that may be tandemly repeated as many as a million times in a haploid genome. Such repeated sequences are known as **satellite DNA**. In the mouse, satellite DNA accounts for about 10 percent of the genome. Other **repetitive sequences** also are present in eukaryotic DNA. Because repetitive DNA consists of many highly similar or identical sequences, fragments of repetitive DNA renature more readily than fragments of nonrepetitive DNA. Information about the size of repeated sequences and the number of copies of a particular sequence can be obtained through studies of the rate of renaturation. The quantitative analysis of DNA renaturation is considered next.

Kinetics of DNA Renaturation

As we saw in Chapter 2, when separated DNA strands are present together in solution, the rate-limiting step in renaturation is the initial collision between pairs of complementary single strands. Because the chance of initial collision is concentration-dependent, the rate of reassociation in-

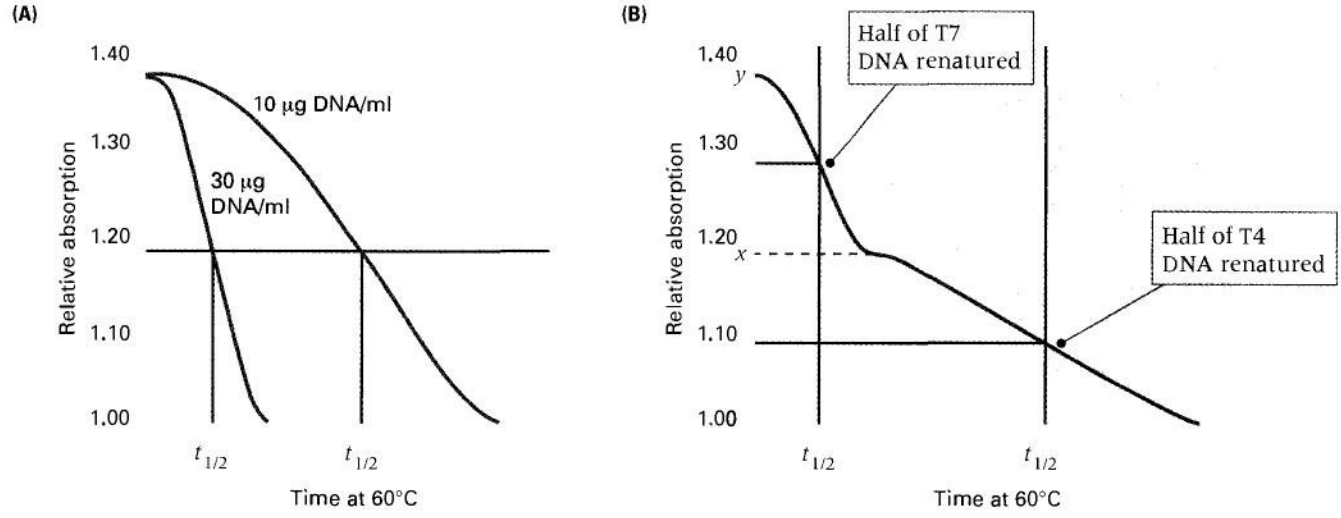


Figure 8.15 (A) Dependence of renaturation time on the concentration of phage T7 DNA. After a period at 90°C to separate the strands, the DNA was cooled to 60°C. Renaturation is complete when the relative absorption reaches 1. (B) Renaturation of a mixture of T4 and T7 DNA, each at the same temperature. Extrapolation (black dashed line) yields the early portion of the T4 curve; the ratio of the absorptions at points *x* and *y* yields the fraction of the total DNA that is T4 DNA. The times required for half-completion of renaturation, $t_{1/2}$, are obtained by drawing the red horizontal lines, which divide each curve equally in the vertical direction, and then extending the red vertical lines to the time axis.

creases with DNA concentration (part A, **Figure 8.15**). Any increase in DNA concentration results in a corresponding increase in the number of potential pairing partners for a given strand. The study of DNA renaturation has contributed greatly to our understanding of the number and types of DNA sequences present in various genomes.

To illustrate the analysis of reassociation, we begin by comparing the renaturation rates of solutions of DNA molecules from the bacteriophages T7 and T4. These molecules have no common nucleotide sequences, and they also differ in molecular weight because the T7 genome is smaller than that of T4. In solutions in which the number of grams of DNA per milliliter is the same, the molar concentration of T7 is greater than that of T4. Thus if each solution is separately denatured and renatured, the molecules of T7 will renature more rapidly than those of T4. If the two solutions are instead mixed, the T7 and T4 will renature independently of one another (because they are not homologous), and a curve such as that in Figure 8.15B will be obtained. Note that the curve consists of two steps: one for the more rapidly renaturing T7 molecules and the other for the T4 molecules. Each step in the curve accounts for half of the change in the absorption of the solution at 260 nanometers (A_{260}) because the initial concentration (in μg of DNA/ml, which is proportional to A_{260}) of each type of molecule was the same.

As is apparent in Figure 8.15A, the renaturation curve of a molecule that contains no repeating base sequences consists of only a single step, and the rate of renaturation is a function of the size of the molecule. If such molecules are fragmented into many components of equal size, then the molar concentration of each component will be the same as that of the unbroken molecule, so the renaturation rate should be unchanged by fragmentation. The rate actually decreases somewhat, but the important point is that *if a DNA molecule contains only a single nonrepetitive sequence of base pairs, then breakage of the molecule does not yield a renaturation curve consisting of steps.* In contrast, if a molecule with an overall sequence that includes both a unique component and several copies of different repeated sequences is fragmented, then *fragments containing the more numerous repeated sequences will renature more rapidly than the fragments containing*

portions of the unique sequence. For example, consider a molecule containing 50,000 base pairs and consisting of 100 copies of a tandemly repeated sequence of 500 base pairs. If the molecules are broken into about 100 fragments of roughly equal size, each fragment will be about 500 base pairs. Although the renaturation curve for the fragments will have a single step, the renaturation rate will be characteristic of molecules 500 nucleotides in length. If a genome contains multiple families of repeated sequences whose abundances differ, the renaturation curve will have steps—one step for each repeating sequence. This principle is the basis of the analysis of renaturation kinetics.

Renaturation kinetics can be described in a simple mathematical form, because the reaction is one in which the rate-limiting step is the initial collision of two molecules. In such a case, the fraction of single strands remaining denatured at a time t after the start of renaturation is given by the expression

$$\frac{C}{C_0} = \frac{1}{[1 + kC_0t]} \qquad (1)$$

in which C is the concentration of single-stranded DNA in moles of nucleotide per liter, C_0 is the initial concentration, and k is a constant. The expression C_0t is commonly called **Cot,** and a plot of C/C_0 versus C_0t is called a **Cot curve.** When renaturation is half completed ($t = t_{1/2}$), then $C/C_0 = 1/2$. Substitution of these values into Equation (1) leads, after some simplification, to

$$C_0t_{1/2} = 1/k \qquad (2)$$

The value of $1/k$ depends on experimental conditions, but for a particular set of conditions, the value is proportional to the number of nucleotides in the renaturing sequences. The longer the sequence, the greater the time to achieve half-complete renaturation for a particular starting concentration (because the number of molecules is smaller). Equations (1) and (2) apply to a single molecular species, a point to which we will return. If a molecule consists of several subsequences, then one needs to know C_0 for each subsequence, and a set of values of $1/k$ will be obtained (one for each step in the renaturation curve), each value depending on the length of the subsequence. What is meant by the "length" of the sequence that determines the rate of renaturation is best described by example. A DNA molecule containing only adenine in one strand (and thymine in the other) has a repeating length of 1. The repeating tetranucleotide . . . GACTGACT . . . has a repeating length of 4. A nonrepeating DNA molecule containing n nucleotide pairs has a unique length of n.

Experimentally, the number of nucleotides per repeating unit is not determined directly. Generally, renaturation curves for a series of molecules of known molecular weight *with no repeating elements in their sequences* (and hence yielding one-step renaturation curves) serve as standards. Molecules composed of short repeating sequences are also occasionally used. A set of curves of this kind is shown in **Figure 8.16**. Note that two of these simple curves represent the entire genomes of *E. coli* and phage T4. With the standard conditions for Cot analysis used to obtain this set of curves, the sequence length n (in base pairs) that yields a particular value of $C_0t_{1/2}$ is

$$n = (5 \times 10^5)C_0t_{1/2}$$

in which t is in seconds, C_0 is in nucleotides per liter, and 5×10^5 is a constant dependent on the conditions of renaturation. Through this formula, the experimentally determined value of Cot ($C_0t_{1/2}$) yields an estimate of the repeat length, n, of a repetitive sequence. Note again that C_0 is not the overall DNA concentration but the concentration of the individual sequence producing a particular step in a curve. How one obtains the necessary value of C_0 will become clear when we analyze a Cot curve. Such an analysis begins by first noting the number of steps in the curve (each step of which represents a sequence or class of sequences of a particular length) and the fraction of the material represented by each step. The observed value of $C_0t_{1/2}$ for each step must be corrected by first inferring the value of C_0 for each sequence class. The lengths of the sequences are then determined from these corrected values by comparison to standards, and the sequence lengths and sequence abundances, as a proportion of the total, are compared to obtain the number of copies of each sequence. This analysis is best understood by looking at an example, which we will do next.

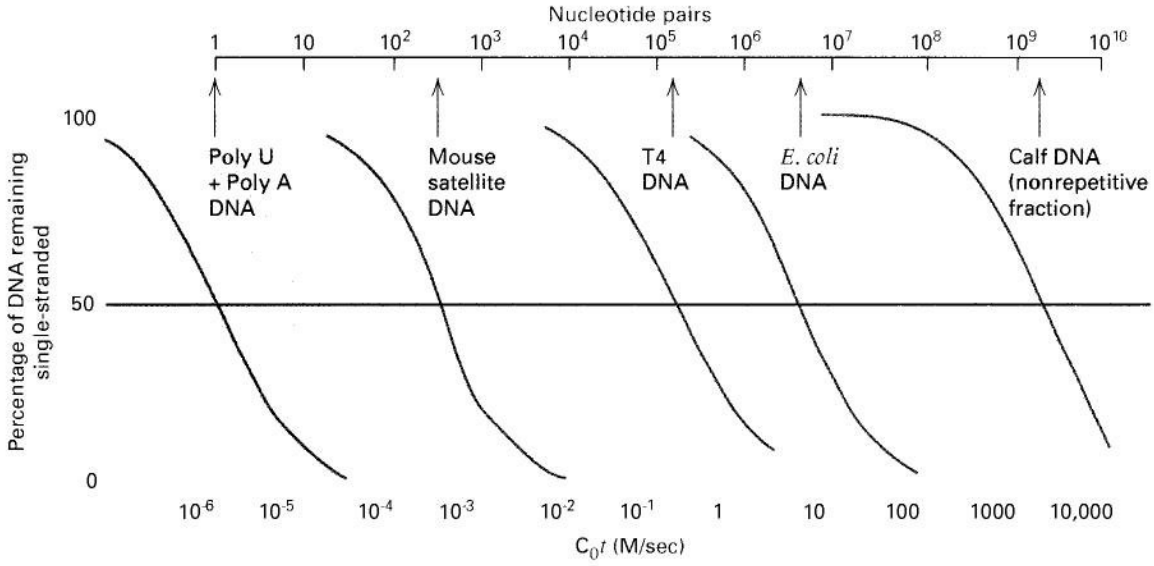


Figure 8.16 A set of Cot curves for various DNA samples. The black arrows pointing up to the red scale indicate the number of nucleotide pairs for each sample; they align with the intersection of each curve with the horizontal red line (the point of half-renaturation, or $C_0t_{1/2}$). The *y*-axis on this graph can be related to that on Figure 8.15 by noting that maximum absorption represents totally single-stranded DNA and minimum absorption represents totally double-stranded DNA.

Analysis of Genome Size and Repetitive Sequences by Renaturation Kinetics

Figure 8.17 shows a Cot curve typical of those obtained in analyses of the renaturation kinetics of eukaryotic genomes. Three discrete steps are evident: 50 percent of the DNA has $C_0t_{1/2} = 10^3$, 30 percent has $C_0t_{1/2} = 10^0 = 1$, and 20 percent has $C_0t_{1/2} = 10^{-2}$. The scale at the top of the figure was obtained from Cot analysis of molecules that have unique sequences of known lengths, as in Figure 8.16. The sequence sizes cannot be determined directly from the observed $C_0t_{1/2}$ values, because each value of C_0 used in plotting the horizontal axis in the body of the figure is the total DNA concentration in the renaturation mixture. Multiplying each $C_0t_{1/2}$ value by the fraction of the total DNA that it represents yields the necessary corrected $C_0t_{1/2}$ values (that is, 0.50×10^3, 0.30×1, and 0.20×10^{-2}). From the size scale at the top of Figure 8.17, the corresponding sequence sizes (repeat lengths) are approximately 3.0×10^8, 2.2×10^5, and 1×10^3 nucleotide pairs, respectively.

To determine the approximate number of copies of each sequence, we make use of the fact that the number of copies of a sequence having a particular renaturation rate is inversely proportional to $t_{1/2}$ and hence to the observed (uncorrected) $C_0t_{1/2}$ values for each class. Thus if the haploid genome

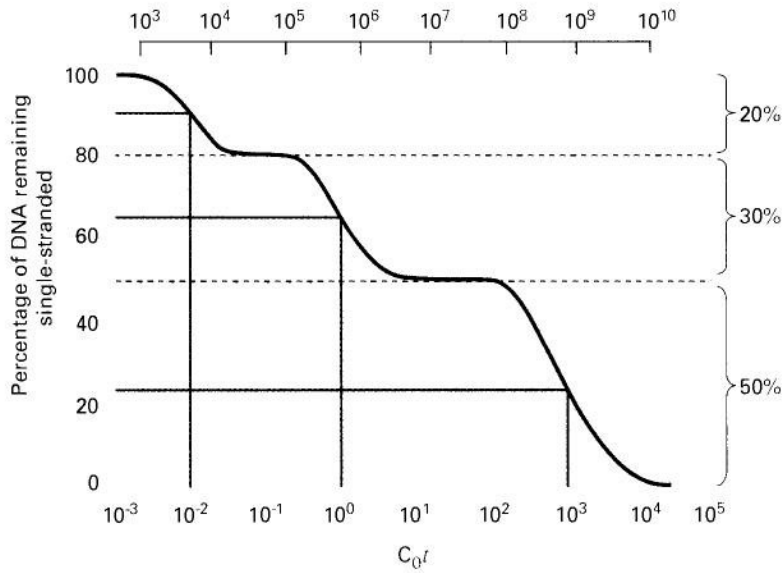


Figure 8.17 The Cot curve analyzed in the text. The scale at the top is the same as that in Figure 8.16, that at the bottom in moles (M) per second. The black dashed lines indicate the fractional contribution of each class of molecules to the total DNA.

Shawn Ahmed and Jonathan Hodgkin 2000
Medical Research Council Laboratory of Molecular Biology
Cambridge, United Kingdom
MRT-2 Checkpoint Protein is Required for Germline Immortality and Telomere Replication in C. elegans

This paper demonstrates that formal genetic analysis—mutation and mapping—still provides a uniquely powerful approach to understanding the biology of an organism, even when, as in this case, the complete DNA sequence of the organism's genome is known. The reasons are, first, that sequence does not always reveal function and, second, that many genes have two or more functions, only one of which is recognized. In this example, analysis of a mutation with the insidious phenotype of gradual germ-line extinction revealed a gene with an unexpected dual function in telomere replication and response to DNA damage. Of what use was the genomic sequence? Once the new mutation was genetically mapped, a likely candidate gene could immediately be identified by inspection of all the coding sequences in the region. The suspect was then quickly shown to be the culprit by sequencing the candidate gene in the mutant.

In most higher organisms, the germ line is responsible for perpetuation of the species . . . and can be thought of as an immortal cell lineage. . . . To investigate how the germ line achieves immortality, a screen was conducted for *Caenorhabditis elegans* mutants with mortal germ lines. . . . Sixteen independent *mortal germ-line mutants* were identified [among 400 mutagenized lines] that became effectively sterile between generations F4 and F16. . . . The unexpectedly large number of *mrt* mutants that were identified in our small pilot screen suggests a conservative estimate of ~50 genes that are specifically required for germ-line immortality in *C. elegans*. . . . For one of these mutants, *mrt-2*, the generation at sterility varies widely, from generation F10 to generation F28. . . . These results suggest that *mrt-2* accumulates some kind of damage that is inherited. . . . The *mrt-2* germ line is hypersensitive to x-rays, which can damage DNA by inducing double-stranded breaks. . . . To gain further insight into the genome stability defect, the chromosomes were examined. Early generation *mrt-2* worms contained the six pairs of chromosomes expected for wildtype, whereas late-generation *mrt-2* worms contained only three, four, or five chromosomes, suggesting the occurrence of chromosome fusions. . . . All chromosome fusions tested were homozygous viable, indicating that they were not missing any essential genes and were therefore probably end-to-end fusions. . . . The late-onset end-to-end chromosome fusion phenotype suggested a defect in telomere replication. . . . [Molecular studies showed that] the telomeres of *mrt-2* worms shorten at the rate of about 12 base pairs per cell division (about 125 base pairs per generation). . . . The *mrt-2* gene is a *C. elegans* homolog of the *Schizosaccharomyces pombe rad1*$^+$ and *Saccharomyces cerevisiae RAD17* checkpoint genes that are conserved from yeast to mammals. In yeast, these genes are required to delay cell-cycle progression in response to DNA damage or in response to a block in DNA replication. . . . The checkpoint protein described here, MRT-2, is required both for telomere replication and for response to DNA damage.

The late-onset end-to-end chromosome fusion phenotype suggested a defect in telomere replication

Source: Nature 403: 159–164

contains only one copy of the longest sequence (3.0×10^8 base pairs), it contains 10^3 copies of the sequence (or sequences) of length 2.2×10^5 base pairs and 10^5 copies of the sequence (or sequences) of length 1×10^3 base pairs. An estimate of the total number of base pairs per genome would in this case be

$$(3.0 \times 10^8) + (10^3 \times 2.2 \times 10^5) + (10^5 \times 1 \times 10^3) = 6.2 \times 10^8 \text{ nucleotide pairs}$$

The different sequence components of a eukaryotic DNA molecule can be isolated by procedures that recover the double-stranded molecules formed at different times during a renaturation reaction. The method used is to allow renaturation to proceed only to a particular $C_0t_{1/2}$ value. The reassociated molecules present at that point are then separated from the remaining single-stranded molecules, usually by passing the solution through a tube filled

with a form of calcium phosphate crystal (hydroxyapatite) that preferentially binds double-stranded DNA.

8.7 Unique and Repetitive Sequences in Eukaryotic Genomes

The nucleotide sequence composition of many eukaryotic genomes has been examined via analysis of the renaturation kinetics of DNA. The principal finding is that eukaryotic organisms differ widely in the proportion of the genome that consists of repetitive DNA sequences and in the types of repetitive sequences that are present. In most eukaryotic genomes, the DNA consists of three major components:

- **Unique, or single-copy, sequences** This is usually the major component, typically constituting 30–75 percent of the chromosomal DNA in most organisms.
- **Highly repetitive sequences** This component usually constitutes 5–45 percent of the genome, depending on the species. Some of these sequences are the satellite DNA referred to earlier. The sequences in this class are typically 5–300 nucleotide pairs per repeat and present in as many as 10^5 copies per genome.
- **Middle-repetitive sequences** This component typically constitutes 1–30 percent of a eukaryotic genome; it includes sequences that are usually present in 10–1000 copies per genome.

These different components can be identified not only by the kinetics of DNA reassociation but also by the number of bands that appear in Southern blots with the use of appropriate probes and by other methods. It should be clear from the preceding discussion of DNA reassociation that the dividing line between many middle-repetitive sequences and highly repetitive sequences is arbitrary.

Unique Sequences

Most gene sequences and the adjacent nucleotide sequences required for their expression are contained in the unique-sequence component. With minor exceptions (for example, the repetition of one or a few genes), the genomes of viruses and prokaryotes are composed entirely of single-copy sequences; in contrast, such sequences constitute only about 40 percent of the total genome in some sea urchin species, a little more than 50 percent of the human genome, and about 70 percent of the *D. melanogaster* genome.

Highly Repetitive Sequences

Many highly repetitive sequences are localized in blocks of tandem repeats, whereas others are dispersed throughout the genome. Dispersed highly repetitive sequences include some SINE and LINE retrotransposable elements that are found at high copy number in certain genomes, including the human genome (Chapter 7). Among the localized highly repetitive sequences, most are fairly short. They can differ greatly in abundance from one genome to the next: 6 percent in the human genome, 18 percent in the *D. melanogaster* genome, but 45 percent in the *D. virilis* genome. One of the simplest possible repetitive sequences makes up 25 percent of the genomes of certain species of land crabs; it is composed of an alternating AT sequence (· · · ATATATAT · · ·) with about 3 percent G + C interspersed. In the *D. virilis* genome, the major components of the localized highly repetitive class are three different but related sequences of seven base pairs rich in A + T:

5'-ACAAACT-3'
5'-ATAAACT-3'
5'-ACAAATT-3'

Blocks of satellite (highly repetitive) sequences in the genomes of several organisms have been located by *in situ* hybridization with metaphase chromosomes. The satellite sequences located by this method have been found to be in the regions of the chromosomes called **heterochromatin.** These are regions that condense earlier in prophase than the rest of the chromosome and are darkly stainable by many standard dyes used to make chromosomes visible (**Figure 8.18**); sometimes the heterochromatin remains highly condensed throughout the cell cycle. The **euchromatin,** which makes up most of the genome, is visible only in the mitotic cycle. The major heterochromatic regions are

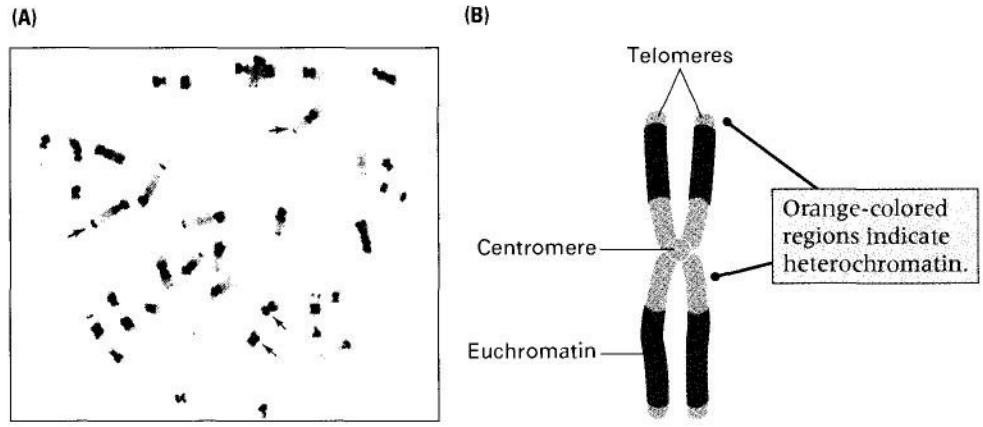


Figure 8.18 (A) Metaphase chromosomes of the ground squirrel *Ammospermophilus harrissi*, stained to show the heterochromatic regions near the centromere of most chromosomes (red arrows) and the telomeres of some chromosomes (black arrows). (B) An interpretive drawing. [Micrograph courtesy of T. C. Hsu.]

adjacent to the centromere; smaller blocks are present at the ends of the chromosome arms (the telomeres) and interspersed with the euchromatin. In many species, an entire chromosome, such as the Y chromosome in *D. melanogaster* and in mammals, is almost completely heterochromatic. Different highly repetitive sequences have been purified from *D. melanogaster*, and *in situ* hybridization has shown that each sequence has its own distinctive distribution among the chromosomes.

The genetic content of heterochromatin is summarized in the following generalization:

> The number of genes located in heterochromatin is small relative to the number in euchromatin.

The relatively small number of genes means that many large blocks of heterochromatin are genetically almost inert, or devoid of function. Indeed, heterochromatic blocks can often be rearranged in the genome, duplicated, or even deleted without major phenotypic consequences.

Middle-Repetitive Sequences

Middle-repetitive sequences constitute about 12 percent of the *D. melanogaster* genome and 40 percent or more of the human and other eukaryotic genomes. These sequences differ greatly in the number of copies and in their locations in the genome. They comprise many families of related sequences and include several groups of genes. For example, the two major types of ribosomal RNA are transcribed from a tandem pair of genes that is repeated several hundred times in most eukaryotic genomes. The genes for tRNA molecules as well as those for the histone proteins are repeated in the genomes of all eukaryotes. The copies of the tRNA genes are scattered, whereas those for the histones are usually clustered in tandem repeats. Each histone gene is repeated about 10 times per genome in chickens, 20 times in mammals, about 100 times in *Drosophila*, and as many as 600 times in some species of sea urchins.

Most of the dispersed middle-repetitive DNA in the *D. melanogaster* genome consists of transposable elements (Chapter 7). There are about 50 different families of such elements—with names such as *P, hobo, copia, jockey, Helena, roo, 412, 297,* and *mdg*—each of which is present in 10–50 copies that occur at widely scattered locations throughout the chromosomes.

8.8 Centromere and Telomere Structure

Eukaryotic chromosomes contain regions specialized for maneuvering the chromosomes in cell division and for capping the ends. These regions are discussed next.

Molecular Structure of the Centromere

The centromere is a specific region of the eukaryotic chromosome that becomes visible as a distinct morphological entity along the chromosome during condensation. It serves as a central component of the kinetochore, the complex of DNA and proteins to which the spindle fibers attach as they move the chromosomes in both mitosis and meiosis. The kinetochore is also the site at which the spindle fibers shorten, causing the chromosomes to move toward the poles. Electron microscopic analysis has shown that in some organisms—for example, the budding yeast *S. cerevisiae*—a single spindle-protein fiber is attached to centromeric chromatin. Most other organisms have multiple spindle fibers attached to each centromeric region.

Centromeres exhibit considerable structural variation among species. At one extreme are **holocentric chromosomes,** which appear to have centromeric sequences spread throughout their length (constituting what is called a *diffuse centromere*). The nematode *Caenorhabditis elegans* has holocentric chromosomes. In a holocentric chromosome, microtubules attach along the entire length of each chromatid. If a holocentric chromosome is broken into fragments by x rays, each fragment behaves as a separate, smaller holocentric chromosome that moves normally in cell division. Diffuse centromeres are very poorly understood and will not be discussed further.

The conventional type of centromere is the **localized centromere,** in which microtubules attach to a single region of the chromosome (the kinetochore). Localized centromeres fall into two categories, *point centromeres* and *regional centromeres.* Point centromeres are found in a number of different yeasts, including *S. cerevisiae,* and they are relatively small in terms of their DNA content. Other eukaryotes, including higher eukaryotes, have regional centromeres that may contain hundreds of kilobases of DNA.

The chromatin segment of the centromeres of *S. cerevisiae* has a unique structure that is exceedingly resistant to the action of various DNases; it has been isolated as a protein–DNA complex containing from 220 to 250 nucleotide pairs. The nucleosomal constitution and DNA sequences of all of the yeast centromeres are known. Several common features of the sequences are shown in **Figure 8.19**, part A. There are four regions, labeled CDE1, CDE2, CDE3, and CDE4. All yeast centromeres have sequences highly similar to those indicated for regions 1, 2, and 3, but the sequence of region CDE4 varies from one centromere to another. Region 2 is noteworthy in that

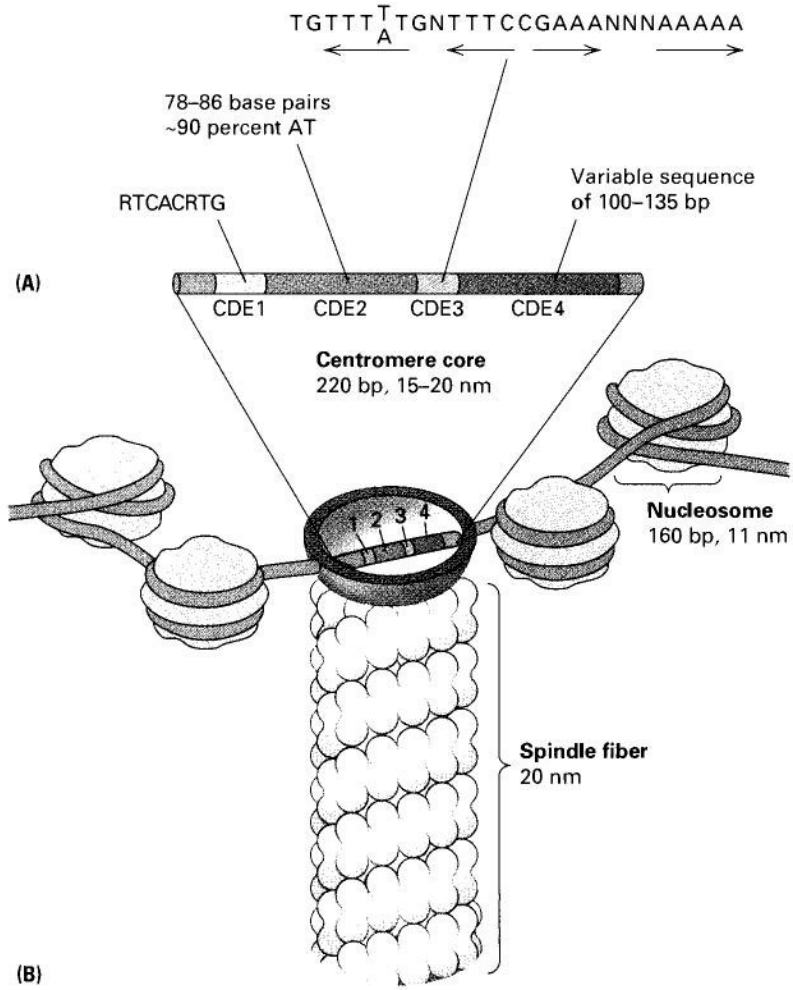


Figure 8.19 A centromere from the yeast *Saccharomyces cervisiae.* (A) Diagram of centromeric DNA, showing the major regions (CDE1 through CDE4) common to all yeast centromeres. The letter R stands for any purine (A or G), and the letter N indicates any nucleotide. Inverted-repeat segments in region 3 are indicated by arrows. The sequence of region CDE4 varies from one centromere to the next. (B) Positions of the centromere core and the nucleosomes on the DNA. The DNA is wrapped around histones in the nucleosomes, but the detailed organization and composition of the centromere core are unknown. [After K. S. Bloom, M. Fitzgerald-Hayes, and J. Carbon. 1982. *Cold Spring Harbor Symp. Quant. Biol.* 47: 1175.]

approximately 90 percent of the base pairs are A–T pairs. The centromeric DNA is contained in a structure designated *the centromeric core particle* that is larger and contains more DNA than the 160 nucleotide pairs in a typical yeast nucleosome core particle. This structure is responsible for the resistance of centromeric DNA to DNase. The spindle fiber is believed to be attached directly to this particle (Figure 8.19B).

The sequence organization of the yeast centromeres is not typical of other eukaryotic centromeres. In higher eukaryotes, the chromosomes are about 100 times larger than yeast chromosomes, and several spindle fibers are usually attached to each region of the centromere. Furthermore, the centromeric regions of the chromosomes of many higher eukaryotes contain large amounts of heterochromatin, consisting of repetitive satellite DNA, as described in Section 8.7. For example, the centromeric regions of human chromosomes contain a tandemly repeated DNA sequence of about 170 base pairs called the **alpha satellite** (**Figure 8.20**). The number of alpha-satellite copies in the centromeric region ranges from 5000 to 15,000, depending on the chromosome. The DNA sequences needed for spindle fiber attachment are interspersed among the alpha-satellite sequences, and the alpha-satellite sequences themselves contribute to centromere activity.

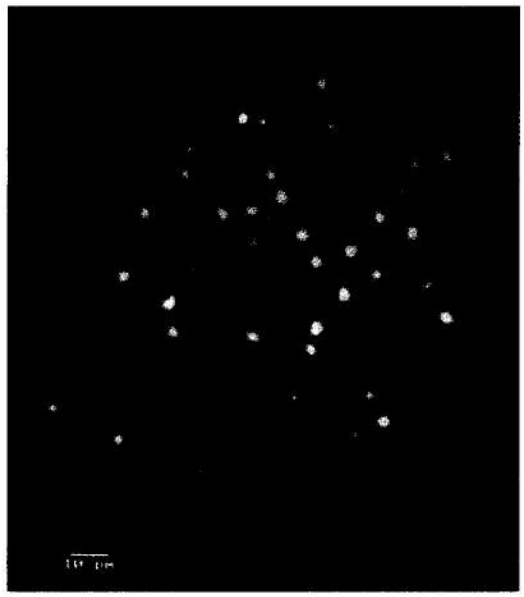


Figure 8.20 Hybridization of human metaphase chromosomes (red) with alpha-satellite DNA. The yellow areas result from hybridization with the labeled DNA. The sites of hybridization of the alpha satellite coincide with the centromeric regions of all 46 chromosomes. [Courtesy of Paula Coelho and Claudio E. Sunkel.]

Molecular Structure of the Telomere

Each end of a linear chromosome is composed of a special DNA–protein structure called a **telomere.** Genetic and microscopic observations first indicated that telomeres are essential for chromosome stability. In *Drosophila,* Hermann J. Muller found that chromosomes without ends could not be recovered after chromosomes were broken by treatment with x rays. In maize, Barbara McClintock observed that broken chromosomes frequently fuse with one another and form new chromosomes with abnormal structures often having two centromeres. As we saw in Chapter 6, DNA polymerase cannot initiate DNA synthesis by itself, but instead requires an RNA primase. At the end of a chromosome, the primer terminates the 5' end of one of the DNA strands, and when the primer is removed, the complementary strand has a short (8–12 nucleotides) single-stranded overhang at the 3' end. The single-stranded overhang is subject to degradation by exonucleases. Without some mechanism to restore the digested end, the DNA molecule in a chromosome would become slightly shorter with each replication. There is such a mechanism, and in mutant cells in which the mechanism is defective, each chromosome end does become shorter in each replication until, eventually, there is so much degradation that the cell dies.

The mechanism of restoring the ends of a DNA molecule in a chromosome relies on an enzyme called **telomerase.** This enzyme works by adding tandem repeats of a simple sequence to the 3' end of a DNA strand. In the ciliated protozoan *Tetrahymena,* in which the enzyme was first discovered, the simple repeating sequence is 5'-TTGGGG-3', and in humans and other vertebrate organisms it is 5'-TTAGGG-3'. (*Drosophila* is unusual even among insects

in having specialized non-LTR retrotransposable elements at the chromosome tips instead of simple repeats.) The tandem repeats of the simple sequence, along with associated proteins, constitute the telomere. As the repeating telomere sequence is being elongated, DNA replication occurs to synthesize a partner strand, and so, for example, the telomere sequence of the right-hand end of any *Tetrahymena* chromosome would be a DNA duplex of the form

```
5'...TTGGGGTTGGGGTTGGGGTTGGGGTTGGGGTTGGGG-3'
3'...AACCCCAACCCCAACCCC-5'
```

This DNA duplex has a single-stranded overhang at the 3' end that can be elongated further by the telomerase.The role of telomerase in the replication of chromosomal DNA is illustrated in **Figure 8.21**. Part A represents the duplex DNA in a chromosome; the telomere sequences are shown in red. Because DNA replication cannot start precisely at the 3' end, the 5' end of each daughter strand in part B is a little shorter than the template strand from which it was replicated. The unreplicated part of the telomere sequence is subject to degradation by nucleases. The 3' end of

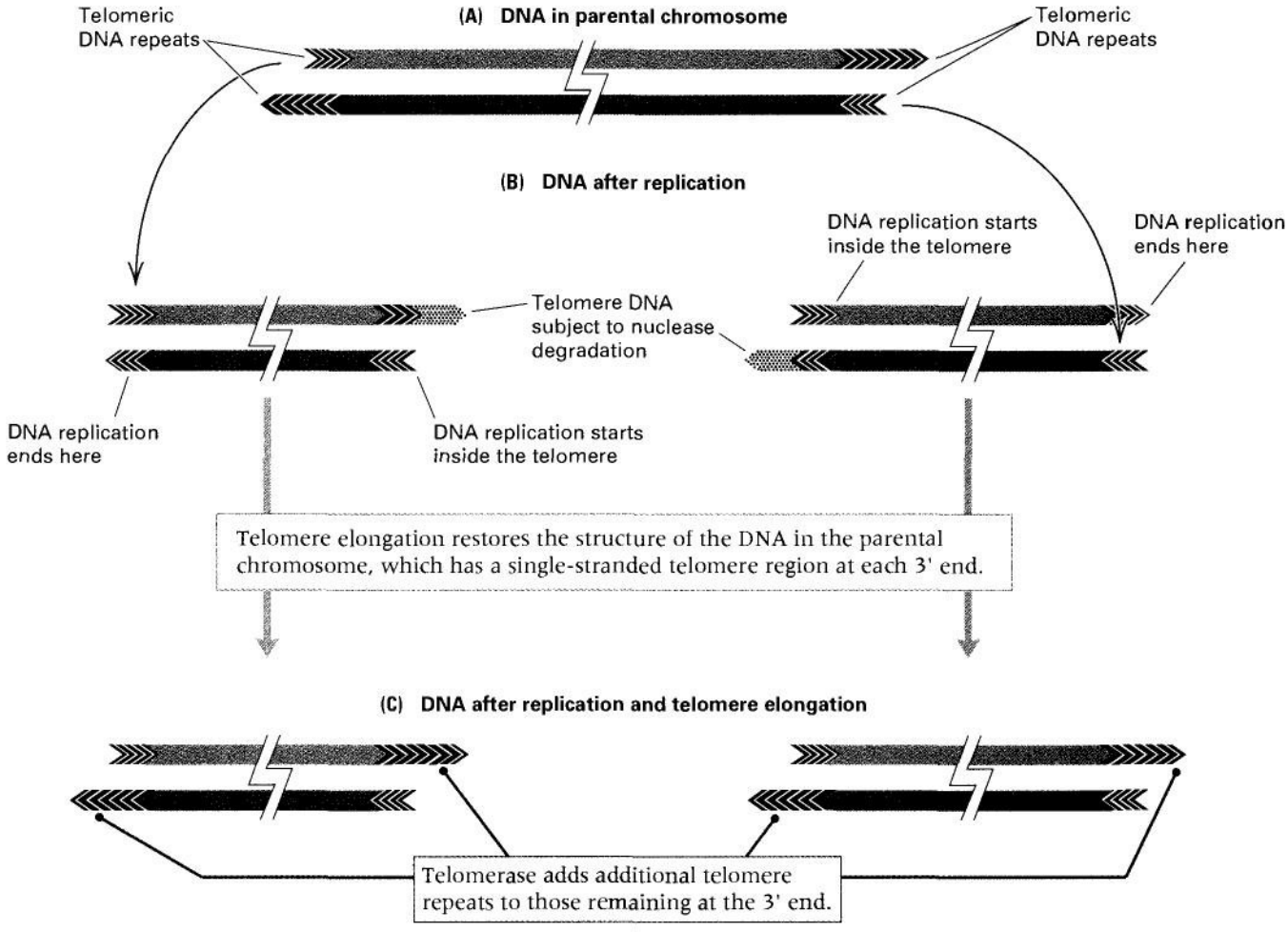


Figure 8.21 The function of telomerase. (A) Chromosomal DNA is double-stranded. Each end of each strand terminates in a set of telomere repeats, but the 3' end of each strand is longer as a result of telomerase action after the previous replication. (B) In the replication of each parental DNA strand, the new daughter DNA strand is initiated within the telomere repeat at the 3' end; the telomere at the 5' end in the new strand is shorter than that in the parental strand. The unreplicated 3' end of the parental strand is vulnerable to digestion by nucleases. (C) In the daughter DNA duplexes formed by replication, the 3' strand at each end is elongated by the addition of telomere repeats by the telomerase. The length and 3' overhang of the telomeres are restored to the state that was present in the original parental molecule.

CoNnection

Telomeres: The Beginning of the End

Carol W. Greider and Elizabeth H. Blackburn 1987
University of California,
Berkeley, California
The Telomere Terminal Transferase of Tetrahymena *Is a Ribonucleoprotein Enzyme with Two Kinds of Primer Specificity*

What a wonderful surprise that an RNA is a key ingredient in the formation of telomeres! Two limitations of DNA polymerase are that it requires a primer oligonucleotide and that it can elongate a DNA strand only at the 3' end. The limitations imply that the 3' end of a DNA strand should become progressively shorter with each round of replication, owing to the need for a primer at that extremity. (Chromosomes without proper telomeres do, in fact, become progressively shorter.) The organism used in this study, Tetrahymena, *is a ciliated protozoan. Each cell has a specialized type of nucleus called a macronucleus that contains many hundreds of small chromosomes. The level of telomerase activity is high because each of these tiny chromosomes needs a pair of telomeres. The convenience of using* Tetrahymena *for the study of telomere function illustrates a principle that runs through the history of genetics: Breakthroughs often come from choosing just the right organism to study. The authors speculate about the possible presence of a "guide" RNA in the telomerase. They were exactly right.*

Precise recognition of nucleic acids is often carried out by enzymes that contain both RNA and protein components. For some of these ribonucleoproteins (RNPs), the RNA components provide specificity to the reaction by base-pairing with the substrate. The recognition that RNA can act catalytically has led to the increase in the number of known RNP-catalyzed reactions. We report here that an RNP is involved in synthesizing the telomeric sequences found at the ends of *Tetrahymena* macronuclear chromosomes. . . . We have previously reported the identification of an activity in *Tetrahymena* cell extracts that adds telomeric repeats onto appropriate telomeric sequence primers in a nontemplated manner. . . . Repeats of the *Tetrahymena* telomeric sequence TTGGGG are added, 1 nucleotide at a time, onto the 3' end of the input primer. . . . We have begun characterizing and purifying the telomerase enzyme in order to investigate the mechanisms controlling the specificity of the reaction. . . . We propose that the RNA component(s) of telomerase may play a role in specifying the sequence of the added telomeric TTGGGG repeats, in recognizing the structure of the G-rich telomeric sequence primers, or both. . . . In the course of purifying the telomerase from crude extracts, we noted a marked sensitivity to salts [in which high] concentrations inactivated the telomeric elongation activity. The salt sensitivity and the large size of the enzyme suggested that the telomerase may be a complex containing a nucleic acid component. To test whether the telomerase contained an essential nucleic acid, we treated active fractions with either micrococcal nuclease or RNase A. The nuclease activity of each of these enzymes abolished the telomeric elongation activity. . . . These experiments suggest that the telomerase contains an essential RNA component. . . . The RNA of telomerase may simply provide a scaffold for the assembly of proteins in the active enzyme complex; however, . . . it is tempting to speculate that the RNA component of the telomerase might be involved in determining the sequence of the telomeric repeats that are synthesized and/or the specific primer recognition. If the RNA of telomerase contains the sequence CCCCAA, this sequence could act as an internal guide sequence.

. . . it is tempting to speculate that the RNA component of the telomerase might be involved in determining the sequence of the telomeric repeats that are synthesized . . .

Source: Cell 51: 887–898

each daughter molecule also has a shortened telomere, because this end is replicated from the underhanging 5' end of the telomere in the parental strand. The shortened telomere remaining at each 3' end is the substrate of the telomerase, which elongates each 3' end by the addition of more repeating telomere units—in the case of *Tetrahymena,* 5'-TTGGGG-3'. Telomere elongation restores the structure of the original parental chromosome in which each end has a larger number of telomere repeats at the 3' end and a smaller number of repeats at the 5' end.

Relatively few copies of the telomere repeat are necessary to prime the telomerase to add more copies and form a telomere. Remarkably, the telomerase enzyme incorporates an essential RNA molecule, called a **guide RNA,** that contains sequences complementary to the telomere repeat and that serves as a template for telomere synthesis and elongation. For example, the *Tetrahymena* guide RNA contains the sequence 3'-AACCCCAAC-5'. The guide RNA undergoes base pairing with the telomere repeat and serves as a template for telomere elongation by the addition of more repeating units (**Figure 8.22**). The complementary DNA strand of the telomere is synthesized exactly like the lagging strand in ordinary DNA replication (Chapter 6). In the telomeric regions of most eukaryotic chromosomes, there are also longer, moderately repetitive DNA sequences immediately preceding the terminal repeats. These sequences differ among organisms and even among different chromosomes in the same organism. Telomeric regions are also associated with proteins that inhibit gene expression in the region, which is known as *telomeric silencing.*

What limits the length of a telomere? In yeast, molecules of a protein called Rap1 bind to the telomere sequence as it is being elongated until about 17 Rap1 molecules have accumulated. At this point, telomere elongation stops, possibly because Rap1 itself inhibits telomerase activity, and then other proteins (including some proteins implicated in DNA repair) join the complex and bind to the extreme end of the telomere. Because each Rap1 molecule binds to approximately 18 base pairs of the telomere, the predicted length of a yeast telomere is 17 $\times$ 18 = 306 base pairs, which is very close to the value observed. Additional evidence for the role of Rap1 comes from mutations in the *RAP1* gene that produce a Rap1 protein that cannot bind to telomere sequences; in these mutant strains, massive telomere elongation is observed.

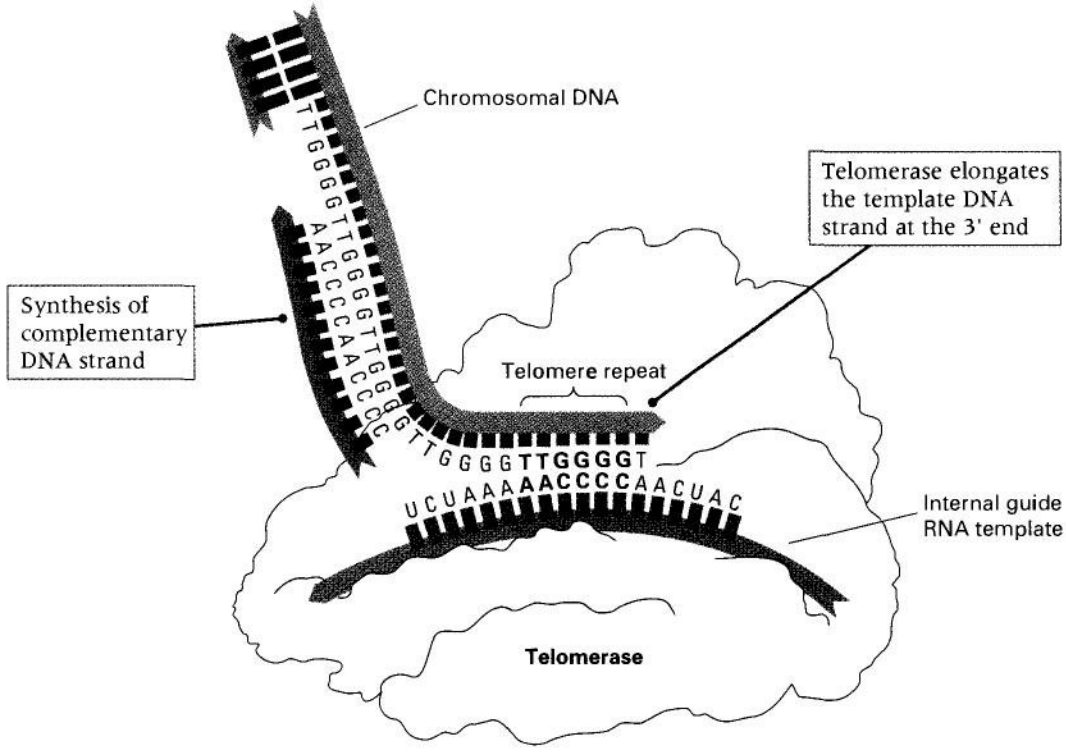


Figure 8.22 Telomere formation in *Tetrahymena*. The telomerase enzyme contains an internal RNA with a sequence complementary to the telomere repeat. The RNA undergoes base pairing with the telomere repeat and serves as a template for telomere elongation. The newly forming DNA strand is produced by DNA polymerase.

Telomeric sequences found in a variety of organisms are shown in **Table 8.2**. Each telomere sequence in Table 8.2 is written with the elongated 3' end at the far right. For example, the *Tetrahymena* formula reads C_4A_2/T_2G_4, which means the telomere sequence

3'-AACCCC-5'
5'-TTGGGG-3' →

where the arrow indicates elongation by telomerase (compare with Figure 8.22). Similarly, the formula for the vertebrate telomere is C_3TA_2/T_2AG_3, which means the telomere sequence

3'-AATCCC-5'
5'-TTAGGG-3' →

In many organisms the length of a run varies; for example, C_{1-8} means variation

Table 8.2 Sequences of telomeric DNAs

Organism	Sequence
Protozoa	
Tetrahymena	C_4A_2/T_2G_4
Paramecium	$C_3\overset{C}{A}A_2/T_2\overset{G}{T}G_3$
Oxytricha	C_4A_4/T_4G_4
Plasmodium	$C_3T\overset{A}{C}A_2/T_2\overset{T}{C}AG_3$
Trypanosoma	C_3TA_2/T_2AG_3
Giardia	C_3TA/TAG_3
Slime molds	
Physarum	C_3TA_2/T_2AG_3
Didymium	C_3TA_2/T_2AG_3
Dictyostelium	$C_{1-8}T/AG_{1-8}$
Fungi	
Saccharomyces	$C_{2-3}ACA_{1-6}/T_{1-6}GTG_{2-3}$
Kluyveromyces	$ACAC_2ACATAC_2TA_2TCA_3TC_2GA/TCG_2AT_3GAT_2AG_2TATGTG_2TGT$
Candida	$ACAC_2A_2GA_2GT_2AGACATC_2GT/ACG_2ATGTCTA_2CT_2CT_2G_2TGT$
Schizosaccharomyces	$C_{1-6}G_{0-1}T_{0-1}GTA_{1-2}/T_{1-2}ACA_{0-1}C_{0-1}G_{1-6}$
Neurospora	C_3TA_2/T_2AG_3
Podospora	C_3TA_2/T_2AG_3
Cryptococcus	$A_2C_{3-5}T/AG_{3-5}T_2$
Cladosporium	C_3TA_2/T_2AG_3
Invertebrates	
Caenorhabditis	GC_2TA_2/T_2AG_2C
Ascaris	GC_2TA_2/T_2AG_2C
Parascaris	$TGCA_2/T_2GCA$
Bombyx, other insects	C_2TA_2/T_2AG_2
Vertebrates	C_3TA_2/T_2AG_3
Plants	
Chlamydomonas	C_3TA_4/T_4AG_3
Chlorella	C_3TA_3/T_3AG_3
Arabidopsis	C_3TA_3/T_3AG_3
Tomato	$C_3\overset{T}{A}T_2/A_2\overset{A}{T}TG_3$

Source: Reprinted with permission from V.A. Zakian, *Science* 270: 1602. Copyright 1995 American Association for the Advancement of Science.

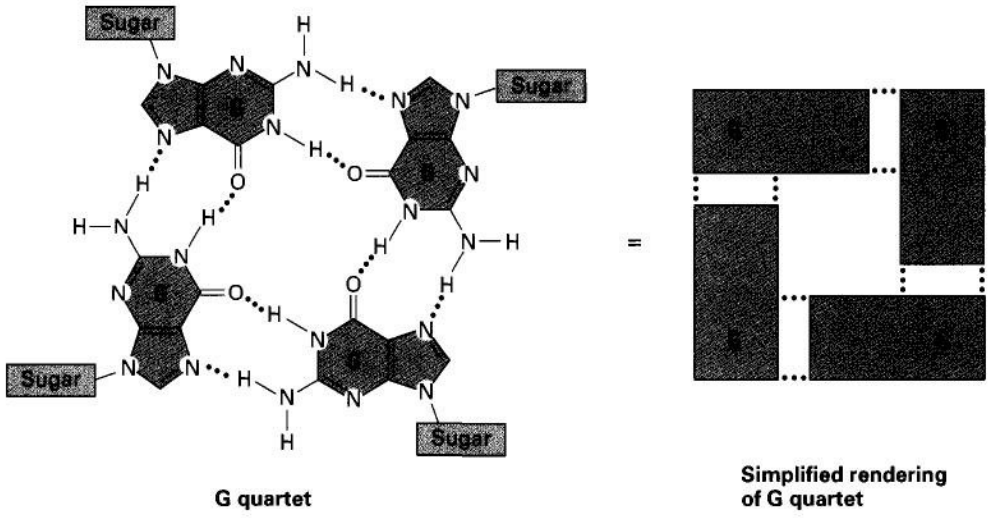


Figure 8.23 G-quartet structure formed by hydrogen bonding between four guanine bases present in a single DNA strand folded back upon itself. The rectangles on the right show how the guanines are oriented, with the deoxyribose sugars attached at the outside corners. On the left, the sugars are indicated by the small blue boxes. There is also a monovalent cation in the center of the quartet (not shown). [From J. R. Williamson, M. K. Raghuraman, and T. R. Cech. 1982. *Cell* 59: 871.]

ranging from as few as one C to as many as eight C's from one repeat to the next. There is also occasional variation in which nucleotide occupies a given position; this is indicated by small stacked letters.

There is a strong tendency for the elongated strand of the telomere repeats in Table 8.2 to be rich in guanine nucleotides. The guanine bases are special in that they have the capacity to hydrogen-bond to one another in a variety of ways. When the elongated 3' strand of the telomere folds back upon itself, the guanine bases can even pair to give the *G-quartet* structure illustrated in **Figure 8.23**. The hydrogen bonding in the G-quartet is called **Hoogstein base pairing** to distinguish it from ordinary Watson–Crick base pairing. A protein has been isolated from the ciliated protozoan *Oxytricha* that binds specifically to the telomeric DNA of linear chromosomes. The β subunit of this telomere-binding protein promotes G-quartet formation by the telomeric DNA of *Oxytricha:*

$$\cdots 5'\text{-TTTTGGGGTTTTGGGGT-}3' \cdots$$

or *Tetrahymena:*

$$\cdots 5'\text{-TTGGGGTTGGGGT-}3' \cdots$$

It is hypothesized that within telomeres, telomeric DNA may be organized in special three-dimensional conformations that include such G-quartets. Models for *Oxytricha* and *Tetrahymena* telomeric DNA showing the possible positions of G-quartets are presented in **Figure 8.24.**

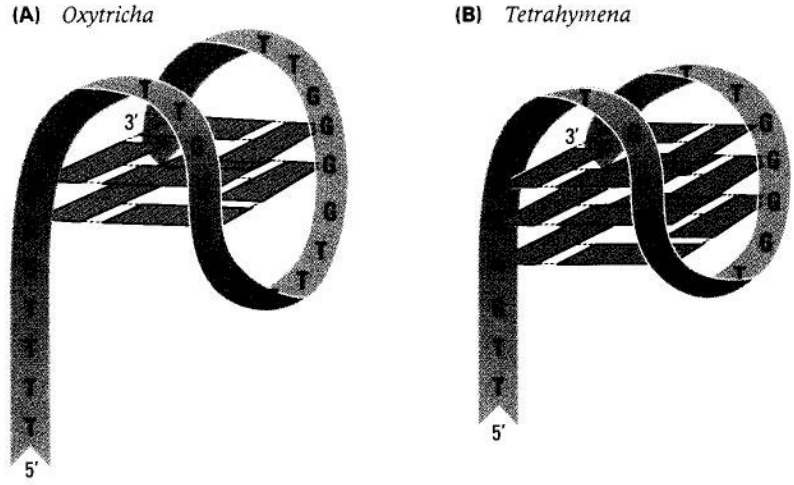


Figure 8.24 Models of telomere structure in *Oxytricha* (A) and *Tetrahymena* (B) that incorporate the G-quartet structure. The arrowhead indicates the 3' end of the DNA strand. The indicated G bases participate in the Hoogstein base pairing, and the G-quartets are drawn as green squares, as in Figure 8.23. There are two G-quartets postulated in the *Oxytricha* telomere, three in *Tetrahymena*. [From J. R. Williamson, M. K. Raghuraman, and T. R. Cech. 1982. *Cell* 59: 871.]

Chapter Summary

Most organisms have a genome composed of double-stranded DNA. (It is RNA or single-stranded DNA in some viruses.) Some chromosomes and plasmids in prokaryotic organisms are circular DNA molecules, whereas others are linear DNA molecules. Circular DNA molecules are invariably supercoiled. In bacteria the supercoiling is negative (right-handed twists) whereas in archaeans it is positive (left-handed twists). The chromosomal DNA of eukaryotes is nearly always linear and has little supercoiling except in regions where genes are actively being transcribed. Among metazoans, there is no correlation between the genome size of an organism and its evolutionary or metabolic complexity (the C-value paradox).

The DNA of both prokaryotic and eukaryotic cells and of viruses is never in a fully extended state but rather is folded in an intricate way, which reduces its effective volume. In viruses, the DNA is tightly folded but without bound protein molecules. In bacteria, the DNA is folded to form a multiply looped structure of independently supercoiled domains called a nucleoid, which includes several proteins that are essential for folding. In eukaryotes, the DNA is compacted into true chromosomes, which contain several proteins and which are large enough to be visible by light microscopy during the mitotic phase of the cell cycle. The DNA–protein complex of eukaryotic chromosomes is called chromatin. The protein component of chromatin consists primarily of five distinct proteins: histones H1, H2A, H2B, H3, and H4. The last four histones aggregate to form an octameric protein that contains two molecules of each. DNA is wrapped around the histone octamer, forming a structure called a nucleosome core particle. This wrapping is the first level of compaction of the DNA in chromosomes. Each nucleosome unit contains about 200 nucleotide pairs, of which about 145 are in contact with the core particle. The remaining 55 nucleotide pairs link adjacent nucleosomes. Histone H1 binds to the linker segment and draws the nucleosomes nearer to one another. The DNA in its nucleosome form is further compacted into a helical fiber, the 30-nm fiber. In forming a visible chromosome, this unit undergoes several additional levels of folding, producing a highly compact visible chromosome. The result is that a eukaryotic DNA molecule, whose length and width are about 50,000 μm and 0.002 μm, respectively, is folded to form a chromosome with a length of about 5 μm and a width of about 0.5 μm.

Polytene chromosomes are found in certain organs in insects. These gigantic chromosomes consist of about 1000 molecules of partly folded chromatin aligned side by side. Seen by microscopy, they have about 5000 transverse bands. Polytene chromosomes do not replicate further, and cells that contain them do not divide. They are useful to geneticists primarily as cytological markers for particular genes and chromosome segments.

The number of copies of individual nucleotide sequences in a DNA molecule can vary tremendously. In prokaryotic DNA, most sequences are unique. However, in eukaryotic DNA, only a fraction of the DNA consists of unique sequences present once per haploid genome. The sequence composition of complex genomes can be studied by DNA renaturation kinetics, or Cot analysis. In eukaryotes, many sequences are present in hundreds to millions of copies. Some highly repetitive sequences are primarily located in the centromeric regions of the chromosomes, whereas others are dispersed. A significant fraction of the DNA, the middle-repetitive DNA, consists of sequences of which from 10 to1000 copies are present per cell. Much of middle-repetitive DNA in higher eukaryotes consists of transposable elements.

Centromeres and telomeres are regions of eukaryotic chromosomes specialized for spindle fiber attachment and stabilization of the tips, respectively. The centromeres of most higher eukaryotes are associated with localized, highly repeated, satellite DNA sequences. Telomeres are formed by a telomerase enzyme that contains a guide RNA that serves as a template for the addition of nucleotides to the 3' end of a telomerase addition site. The complementary strand is synthesized in a manner analogous to the replication of the lagging strand at an ordinary replication fork. In mammals and other vertebrates, the 3' strand of the telomere terminates in tandem repeats of the simple sequence 5'-TTAGGG-3'. Relatively few copies of this sequence are needed to prime the telomerase. The special properties of the telomere may be related to the ability of guanine nucleotides to form complex hydrogen-bonded structures—in particular, the G-quartet.

Key Terms

alpha satellite
C-value paradox
chromatin
chromocenter
conserved sequence
core particle
Cot
Cot curve
cytological map
deoxyribonuclease (DNase)
euchromatin
guide RNA
heterochromatin
highly repetitive sequence
histone
holocentric chromosome
Hoogstein base pairing
localized centromere
middle-repetitive sequence
negative supercoiling
nucleoid

nucleosome
nucleosome core particle
polytene chromosome
relaxed DNA
repetitive sequence
satellite DNA
scaffold
single-copy sequence
supercoiled DNA
telomere
telomerase
30-nm fiber
topoisomerase
unique sequence

Review the Basics

- What is the "C-value paradox" and why is it regarded as a paradox?
- A circular, double-stranded DNA molecule is negatively supercoiled. What does this mean?
- What are the principal differences between prokaryotes and eukaryotes in the organization of the genetic material?
- Histone proteins are positively charged (basic) molecules that interact with negatively charged (acidic) molecules. What is the principal type of acid with which histones interact?
- The nucleosome, the fundamental unit of chromatin in eukaryotic cells, is sometimes described as including a "histone octamer." Is the term *octamer* literally applicable? Explain your answer.
- What is highly repetitive DNA? In DNA renaturation, highly repetitive DNA reanneals very rapidly. Explain why.
- In DNA renaturation, what types of sequences reanneal the most slowly, and why?
- Distinguish between euchromatin and heterochromatin. Is it correct to say that heterochromatin contains no functional genes? Explain your answer.
- For studies in genetics, what is special about the polytene chromosomes in salivary glands of dipteran fly larvae?
- What is a telomere? How are telomers elongated in each cell cycle? What specifies the nucleotide sequence of the telomere?

Guide to Problem Solving

Problem 1 Ignoring the effects of base stacking, which of the DNA molecules shown here would require a higher temperature for denaturation?

```
5'-ACTGTCATAGAT-3'
3'-TGACAGTATCTA-5'

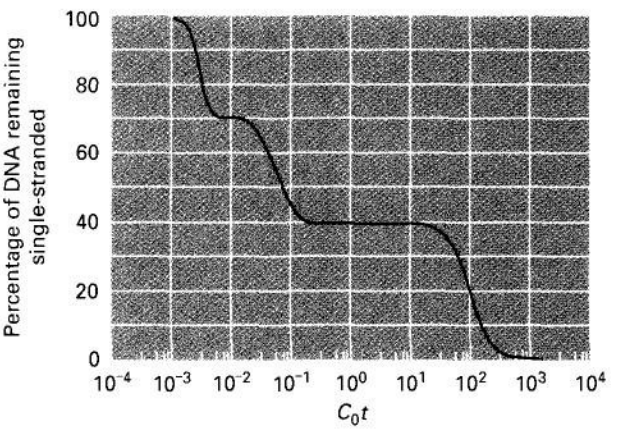
5'-GTCACGGCTAGC-3'
3'-CAGTGCCGAGCG-5'
```

Answer The molecules are of the same length, so the molecule with the greater proportion of G–C base pairs will be more stable, and so will require a higher temperature for denaturation, because each G–C pair has three hydrogen bonds whereas each A–T base pair has only two. Hence the duplex with 8 G–C base pairs will be more stable than the duplex with only 4.

Problem 2 Consider the Cot curve shown above at right.

(a) What fraction of the DNA is contained in components that are unique? in components that are repetititive?

(b) If the unique sequences are present once per haploid genome, how many copies per haploid genome are there of the other two classes?



Answer **(a)** The percentages are read off the y axis, subtracting the value for the percent DNA remaining single-stranded after each repetitive component of the DNA has renatured from the value prior to the renaturation of this component. The Cot values are read off the x-axis corresponding to the steepest part of the renaturation curve of each repetitive component. Hence $100 - 70 = 30$ percent consists of repetitive DNA with a Cot of 3×10^{-3}; $70 - 40 = 30$ percent consists of repetitive DNA with a

Cot of 3×10^{-2}; and $40 - 0 = 40$ percent consists of unique-sequence DNA with a Cot of 10^2. **(b)** The relative abundances of the sequence classes are inversely proportional to their Cot values, so in this example the repetitive-sequence classes have relative abundances of $1/(3 \times 10^{-3}) = 333$, $1/(3 \times 10^{-2}) = 33$, and $1/10^2 = 10^{-2}$, respectively.

Problem 3 Suppose that the DNA sequences repeated at the ends of a transposable element undergo pairing and recombination. What are the genetic consequences if:

(a) The sequences are direct repeats?

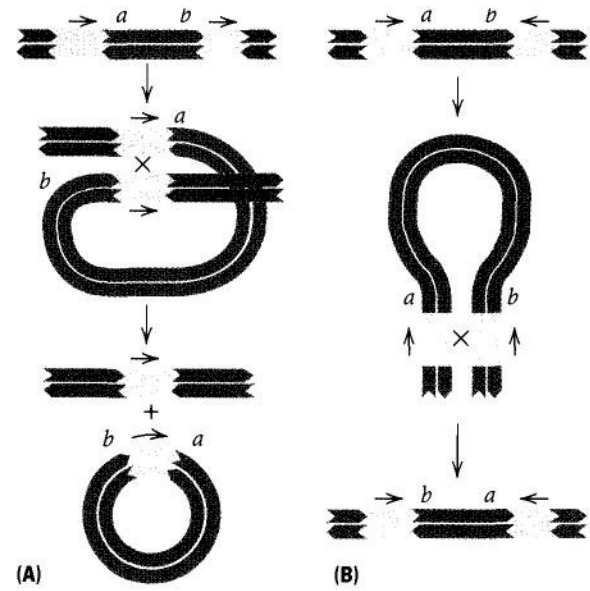
(b) The sequences are inverted repeats?

Draw diagrams to support your answers.

Answer The diagrams at right show the consequences.

(a) When direct repeats pair and undergo crossing-over, the result is the deletion of one of the repeats and the DNA segment between them (part A).

(b) When the repeats are inverted repeats, the result is that the transposable element is inverted in the chromosome (part B).



Analysis and Applications

8.1 If you had never seen a chromosome in a microscope but had seen nuclei, what feature of DNA structure would tell you that DNA must exist in a highly coiled state within cells?

8.2 Distinguish between a nucleosome and a nucleosome core particle. What protein molecules does each contain, and what length of DNA?

8.3 If the regulatory region at the 5′ end of a eukaryotic gene is 1 kb in length, how many nucleosomes need to be rearranged to give transcription factors and other regulatory proteins access to the DNA?

8.4 What causes the bands in a polytene chromosome? How do polytene chromosomes undergo division?

8.5 Do the banded regions of the *Drosophila* polytene chromosomes consist of euchromatin or heterochromatin? Explain your answer.

8.6 The polytene chromosomes of *Drosophila melanogaster* contain about 5000 bands. The size of the euchromatic genome is 120×10^6 base pairs, and there are about 13,600 genes. Assume that the amount of DNA between the bands is negligible.

(a) What is the average amount of DNA per band?

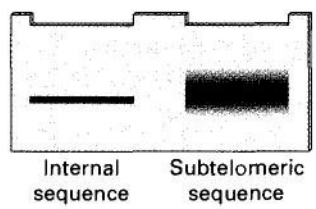
(a) What is the average number of genes per band?

8.7 Is there an analog of a telomere in the *E. coli* chromosome?

8.8 Why is it that telomeric DNA cannot be replicated by the normal mechanism of DNA replication?

8.9 What would be expected to happen to the telomeres of a cell through successive mitotic divisions if the telomerase were defective?

8.10 Genomic DNA is prepared from a sample of normal tissue and cleaved with a restriction enzyme that has a cleavage site in a single-copy sequence present in the subtelomeric DNA of a particular chromosome. A Southern blot is carried out using a probe for this single-copy subtelomeric sequence. As a control, a sample of the cleaved DNA is also hybridized with a probe homologous to a sequence known to be in the interior of the same chromosome. The result is shown in the accompanying gel diagram. Explain why the band produced by the probe for the internal sequence is very sharp and well defined, whereas that produced by the probe for the subtelomeric sequence is diffuse and spread over a range of sizes.



8.11 A *Drosophila* X chromosome is broken by x rays near the locus of the wildtype allele of the gene *white*. Although it lacks a telomere (*Drosophila* telomeres are unusual in being composed of specialized types of transposable elements rather than simple repeating sequences), the chromosome is stable enough to be transmitted through successive generations. However, many

males with white eyes are recovered; these result from a high rate of spontaneous mutation of $w^+ \rightarrow w$. Explain why this result might be expected, and specify the nature of the w mutations.

8.12 Many transposable elements contain a pair of repeated sequences. What is their position in the element? Are they in direct or inverted orientation?

8.13 Consider a long linear DNA molecule, one end of which is rotated four times with respect to the other end in the unwinding direction.

(a) If the two ends are joined to keep the molecule in the underwound state, how many base pairs will be broken?

(b) If the underwound molecule is allowed to form a supercoil, how many twists will be present?

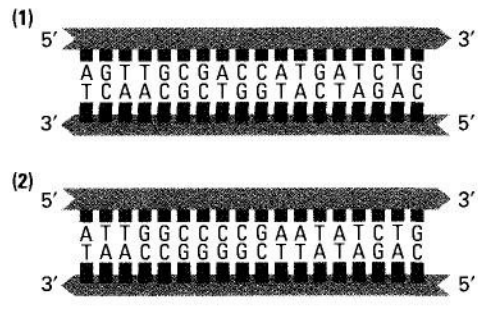
8.14 Endonuclease S1 can break single stranded DNA but does not break double-stranded linear DNA. However, S1 can cleave supercoiled DNA, usually making a single break. Why does this occur?

8.15 Circular DNA molecules of the same size migrate at different rates in electrophoretic gels, depending on whether they are supercoiled or relaxed circles. Why?

8.16 Denaturation of DNA refers to breakdown of the double helix and ultimate separation of the individual strands, and it can be induced by heat and other treatments. Because G−C base pairs have three hydrogen bonds and A−T base pairs have only two, the temperature required for denaturation increases with the G−C content and also with the length of continuous G−C tracts in the molecule. Renaturation of DNA refers to the formation of double-stranded DNA from complementary single strands.

(a) Which process—denaturation or renaturation—is dependent on the concentration of DNA?

(b) Which of the two illustrated DNA molecules (1 or 2) would have the lower temperature for strand separation? Why?



8.17 DNA from species A, labeled with ^{14}N and randomly fragmented, is renatured with an equal concentration of DNA from species B, labeled with ^{15}N and randomly fragmented, and then centrifuged to equilibrium in CsCl. Five percent of the total renatured DNA has a hybrid density. What fraction of the base sequences is common to the two species?

8.18 Mutations have not been observed that result in nonfunctional histones. Why should this be expected?

8.19 A sequence of middle-repetitive DNA from *Drosophila* is isolated and purified. It is used as a template in a polymerization reaction with DNA polymerase and radioactive substrates, and highly radioactive probe DNA is prepared. The probe DNA is then used in an *in situ* hybridization experiment with cells containing polytene chromosomes obtained from ten different flies of the same species. Autoradiography indicates that the radioactive material is localized to about 20 sites in the genome, but they are in different sites in each fly examined. What does this observation suggest about the DNA sequence being studied?

8.20 Suppose that a topoisomerase enzyme binds with a molecule of circular double-stranded DNA, nicks one strand, rotates it three complete revolutions around the other strand in the direction of overwinding, and then reseals the nick. The resulting molecule undergoes supercoiling to restore complete base pairing between the strands. Is the direction of supercoiling positive or negative? How many superhelical turns (complete rotations) are present?

8.21 A covalently closed circular DNA molecule with no supercoiling has 10 disrupted base pairs. If it is allowed to supercoil, how many turns of supercoiling are necessary to restore complete base pairing?

8.22 A solution contains supercoiled circular DNA molecules. If all the molecules were "frozen" at an instant in time and examined, would they all look the same?

8.23 Ten protein molecules are bound to a circular DNA molecule containing a nick. Each bound protein molecule breaks one base pair. The nick is then sealed with DNA ligase, and the protein molecules are removed by treatment with a protease. After the protein molecules are removed:

(a) Is the helix underwound or overwound?

(b) Will the molecule undergo supercoiling to relieve the stress?

(c) Is the supercoiling positive or negative?

8.24 Explain why DNA fragments containing the repeating sequence ATATAT . . . would renature much more rapidly than DNA fragments of the same length containing a random sequence of A's and T's.

8.25 The temperature at which half of the base pairs in a sample of double-stranded DNA are denatured is called the *melting temperature* and designated T_m. Over a wide range of base compositions, the T_m is approximately linear in the percentage of G−C base pairs, g. Deduce the approximate relationship between T_m and g, given that DNA with $g = 22$ percent G + C has $T_m = 78°C$ and that DNA with $g = 73$ percent G + C has $T_m = 98.9°C$.

GeNETics on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

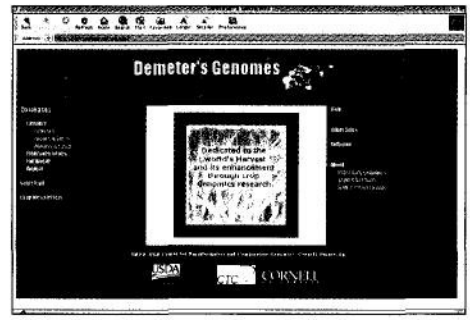
- Most bacterial cells must compact their DNA into a small volume because the cells are extremely small, yet the DNA must also be available for transcription. For example, the size of an *E. coli* cell is about 1 × 2 micrometers, yet it contains a circular DNA molecule of 4.6 million base pairs, which if opened and stretched to its full length would be about 70 times longer than the cell itself. The keyword **bacterial chromatin** will connect you to a site that focuses on the analysis of complete bacterial genome sequences with regard to the motifs involved in compaction of the DNA.

- The **centromere** sequences of the chromosomes of the yeast *Saccharomyces cerevisiae* include three conserved sequence elements that are sufficient for mitotic and meiotic segregation. The first is only 8 bp and is required for mitotic segregation; the second is a 78–86 bp, A–T-rich region required for chromosome disjunction in meiosis I; and the third is a 25-bp region essential for both mitotic and meiotic segregation. You can retrieve and compare the sequence of all the yeast centromeres via the search engine at this keyword site by entering the item name cen*.

- The **telomeres** of chromosomes are of great interest not only because they are specialized DNA–protein complexes that stabilize the ends of chromosomes but also because abnormal telomere maintenance is implicated in the progression of cancer cells (Chapter 15). In most organisms the telomeres consist of short repeating sequences that are synthesized by a telomerase enzyme containing an RNA template for the telomere repeat. You can learn more about telomere sequences and telomerases at this keyword site devoted to the structures.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 8 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 8.



8.26 Use the relationship in Problem 8.25 to find the approximate melting temperature for a bacterial species whose genome is:

(a) A + T rich (70 percent A + T).
(b) Composed of equal proportions of A–T and G–C base pairs.
(c) G + C rich (70 percent G + C).

8.27 Use the relationship in Problem 8.25 to find the G + content of:

(a) The human genome, with $T_m = 84.7°C$.
(b) The rice genome, with $T_m = 87.0°C$.

Challenge Problems

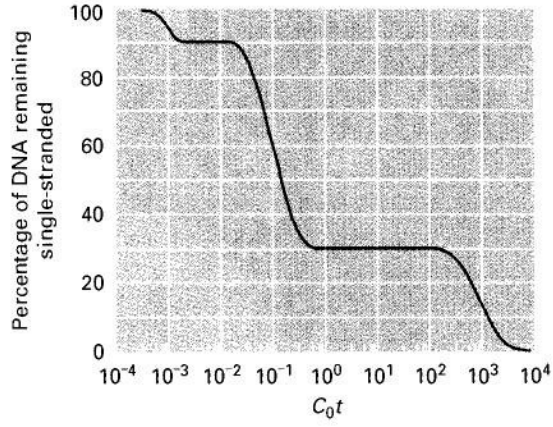
8.28 A sample of identical DNA molecules, each containing about 3000 base pairs per molecule, is mixed with histone octamers under conditions that allow formation of chromatin. The reconstituted chromatin is then treated with a nuclease, and enzymatic digestion is allowed to occur. The histones are removed, and the positions of the cuts in the DNA are identified by sequencing the fragments. It is found that the breaks have been made at random positions and, as expected, at about 200-base-pair intervals. The experiment is then repeated with a single variation. A protein known to bind to DNA is added to the DNA sample before addition of the histone octamers.

Again, reconstituted chromatin is formed and digested with nuclease. In this experiment, it is found that the breaks are again at intervals of about 200 base pairs, but they are localized at particular positions in the base sequence. At each position, the site of breakage can vary over only a 2- to 3-base range. However, examination of the base sequences in which the breaks have occurred does not indicate that breakage occurs in a particular sequence; in other words, each 2- to 3-base region in which cutting occurs has a different sequence. Explain the difference between the two experiments.

8.29 Consider the Cot curve shown in the figure.

(a) What fraction of the DNA is contained in components that are unique? in components that are repetitive?

(b) If the unique sequences are present once per haploid genome, how many copies per haploid genome are there of the other two classes?



Problem 8.29

Further Reading

Bauer, W. R., F. H. C. Crick, and J. H. White. 1980. Supercoiled DNA. *Scientific American*, July.

Bendich, A. J., and K. Drlic. 2000. Prokaryotic and eukaryotic chromosomes: What's the difference? *Bioessays* 22: 481.

Bickmore, W. A., and A. T. Sumner. 1989. Mammalian chromosome banding: An expression of genome organization. *Trends in Genetics* 5: 144.

Charlebois, R. L. 1999. *Organization of the Prokaryotic Genome*. Washington, DC: American Society for Microbiology.

Clarke, L. 1998. Centromeres: Proteins, protein complexes, and repeated domains at centromeres of simple eukaryotes. *Current Opinion in Genetics & Development* 8: 212.

Crane-Robinson, C. 1999. How do linker histones mediate differential gene expression? *Bioessays* 21: 367.

Elgin, S. C. R., ed. 1995. *Chromatin Structure and Gene Expression*. New York: IRL Press.

Greider, C. W., and E. H. Blackburn. 1996. Telomeres, telomerase and cancer. *Scientific American*, February.

Grunstein, M. 1992. Histones as regulators of genes. *Scientific American*, October.

Hsu, T. H. 1979. *Human and Mammalian Cytogenetics*. New York: Springer-Verlag.

Miller, O. J., and E. Therman. 2000. *Human Chromosomes*. 4th ed. New York: Springer.

Pardue, M. L., O. N. Danilevskaya, K. Lowenhaupt, F. Slot, and K. L. Traverse. 1996. *Drosophila* telomeres: New views on chromosome evolution. *Trends in Genetics* 12: 48

Prescott, J. C., and E. H. Blackburn. 1999. Telomerase: Dr. Jekyll or Mr. Hyde? *Current Opinion in Genetics & Development* 9: 368.

Razin, S. V. 1997. *The Nuclear Matrix and Spatial Organization of Chromosomal DNA Domains*. New York: Chapman & Hall.

Sachs, R. K., L. R. Hlatky, and B. J. Trask. 2000. Radiation-produced chromosome aberrations: Colourful clues. *Trends in Genetics* 16: 143.

Sharpe, M. E., and J. Errington. 1999. Upheaval in the bacterial nucleoid: An active chromosome segregation mechanism. *Trends in Genetics* 15: 70.

van Holde, K., and J. Zlatanova. 1999. The nucleosome core particle: Does it have structural and physiologic relevance? *Bioessays* 21: 776.

Wagner, R. P., M. P. Maguire, and R. L. Stallings. 1993. *Chromosomes*. New York: Wiley-Liss.

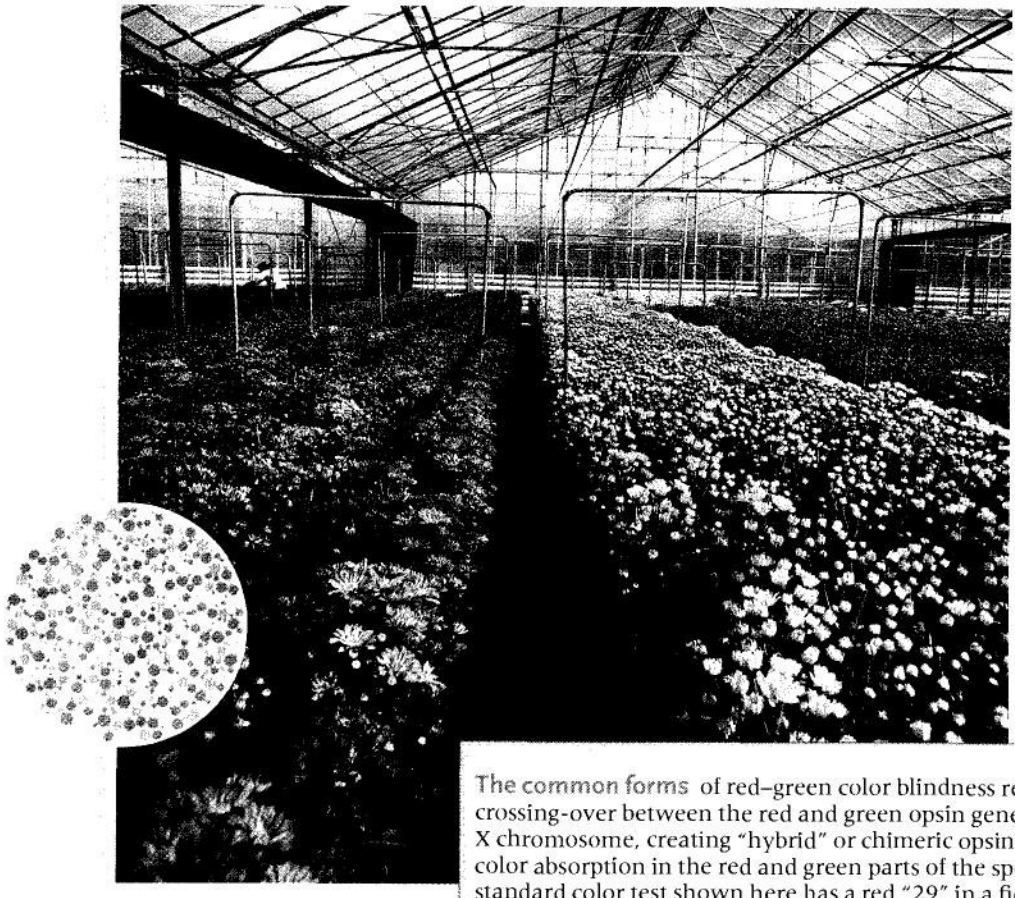
White, M. J. D. 1977. *Animal Cytology and Evolution*. London: Cambridge University Press.

Zakian, V. A. 1996. Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. *Annual Review of Genetics* 30: 141.

CHAPTER

9

Human Karyotypes and Chromosome Behavior



The common forms of red–green color blindness result from unequal crossing-over between the red and green opsin genes in the X chromosome, creating "hybrid" or chimeric opsins with atypical color absorption in the red and green parts of the spectrum. The standard color test shown here has a red "29" in a field of green. Males with red–green color blindness are unable to recognize the number. Nor are they able to distinguish red from green in this hothouse full of chrysanthemums. Some species of chrysanthemum differ in chromosome number: they can have 2, 4, 6, 8, or 10 copies of a basic set of 9 chromosomes. [Hothouse: Nigel Cattlin/Photo Researchers, Inc.]

CHAPTER OUTLINE

9.1 **The Human Karyotype**
Standard Karyotypes
The Centromere and Chromosome Stability
Dosage Compensation of X-linked Genes
The Calico Cat Is a Result of X-Chromosome Inactivation
Pseudoautosomal Inheritance
Gene Content and Evolution of the Y Chromosome

9.2 **Chromosome Abnormalities in Spontaneous Abortion**
Down Syndrome and Other Viable Trisomies
Trisomic Segregation
Sex-Chromosome Abnormalities

9.3 **Chromosomal Deletions and Duplications**
Deletions
Duplications
Unequal Crossing-Over in Red–Green Color Blindness

9.4 **Genetics of Chromosomal Inversions**
Paracentric Inversion (Not Including the Centromere)
Pericentric Inversion (Including the Centromere)

9.5 **Chromosomal Translocations**
Reciprocal Translocations
Translocation Complexes in *Oenothera*
Robertsonian Translocations
Translocations and Trisomy 21

9.6 **Genomic Position Effects on Gene Expression**

9.7 **Polyploidy in Plant Evolution**
Autopolyploids and Allopolyploids
Monoploid Oganisms

9.8 **Genome Evolution in the Grass Family (Gramineae)**

PRINCIPLES

- The standard human karyotype consists of 22 pairs of autosomes.
- Chromosome abnormalities are a major factor in human spontaneous abortions and an important cause of genetic disorders, such as trisomy 21 (Down syndrome).
- Dosage compensation in mammals results from genetic inactivation (silencing) of X chromosomes at an early stage in embryonic development, except that one X chromosome, chosen at random, remains active in each cell lineage (the single-active-X principle). Some genes in the "inactive" X chromosomes are not silenced.
- Genetically unbalanced (aneuploid) chromosomal complements due to an extra or missing chromosome often have more severe effects on phenotype than does addition of a genetically balanced (euploid) entire extra set of chromosomes.
- By a process of mispairing and unequal crossing-over, genes that are duplicated in tandem along the chromosome can give rise to chromosomes with even more copies.
- Gene-for-gene pairing between a wildtype chromosome and one that contains an inversion of a segment of genes results in the formation of a loop in one of the chromosomes; crossing-over within the "inversion loop" leads to chromosomal abnormalities.
- Reciprocal translocations result in abnormal gametes because they upset segregation.
- Duplication of the entire chromosome complement present in a species—or in a hybrid between species—is a major process in the evolution of higher plants.

CONNECTIONS

Lyonization of an X Chromosome
Mary F. Lyon 1961
Gene Action in the X Chromosome of the Mouse (Mus musculus L.)

The First Human Chromosomal Disorder Identified
Jerome Lejeune, Marthe Gautier, and Raymond Turpin 1959
Study of the Somatic Chromosomes of Nine Down Syndrome Children [original in French]

In most species of organisms, the chromosome complement is virtually identical from one individual to the next, with a few exceptions such as the XX—XY sex difference in many animals. As might be expected, organisms with an extra chromosome or a missing chromosome usually have developmental or other types of abnormalities, as a result of the increase or decrease in copy number (*dosage*) of the genes in this chromosome. Some organisms, usually rare, are found to have a variation in chromosome structure. The abnormal chromosome may have a particular segment missing, duplicated, reversed in orientation, or attached to a different chromosome. Each of these structural abnormalities has different genetic implications. In this chapter we consider the human chromosome complement and some of the major chromosomal abnormalities encountered in human populations. We also examine chromosome abnormalities in other organisms. Generally speaking, animals are much less tolerant of chromosomal changes than are plants. As we shall see, especially in plants, the acquisition of entire extra sets of chromosomes is not necessarily harmful. In some lineages of plants, the reduplication of entire chromosome sets has been a major process in genome evolution and the origin of species.

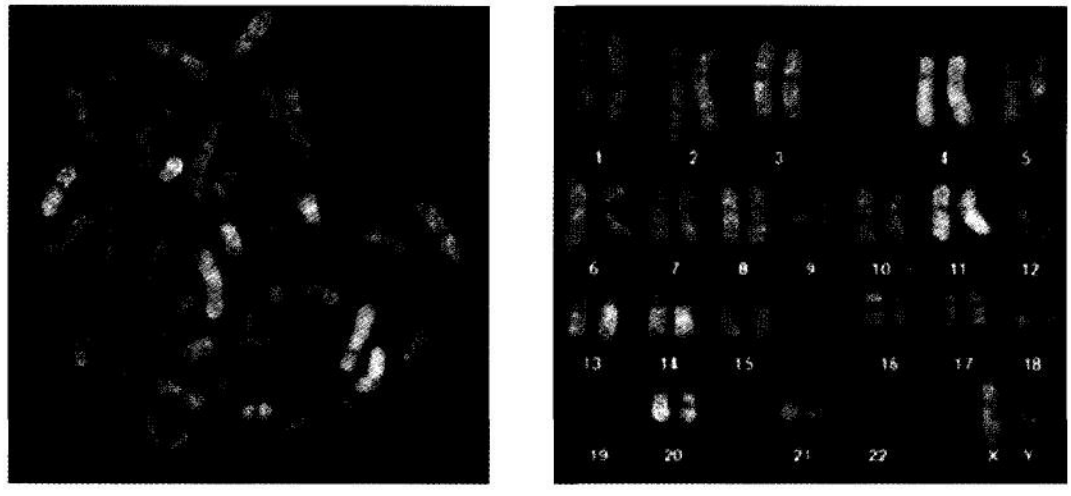


[Nigel Cattlin/Photo Researchers, Inc.]

Some varieties of cultivated banana are triploids. Abnormal chromosome numbers in the seeds arrest their development, resulting in the small edible seeds found in bananas in the fruit market.

9.1 The Human Karyotype

The normal chromosome complement of a cell in mitotic metaphase from a human male is illustrated in **Figure 9.1**. The chromosomes have been labeled via a technique called **chromosome painting,** in which



(A) (B)

Figure 9.1 Human chromosome painting, in which each pair of chromosomes is labeled by hybridization with a different fluorescent probe. (A) Metaphase spread showing the chromosomes in a random arrangement as they were squashed onto the slide. (B) A karyotype, in which the chomosomes have been grouped in pairs and arranged in conventional order. Chromosomes 1–20 are arranged in order of decreasing size, but for historical reasons, chromosome 21 precedes chromosome 22, even though chromosome 21 is smaller. [Courtesy of Johannes Wienberg and Thomas Ried.]

different colors are "painted" on each chromosome by hybridization with DNA strands labeled with different fluorescent dyes. Individual chromosomes are first isolated by any of a variety of techniques (for example, microdissection followed by DNA amplification with PCR); then the chromosome-specific DNA samples are labeled with fluorescence. A mixture of differently labeled strands from all the chromosomes is used in hybridization with metaphase chromosomes squashed onto a glass slide, allowing the fluorescent strands to hybridize with homologous strands present in the chromosomes. Unhybridized DNA is washed from the slide, and the preparation is examined through a confocal microscope to read the fluorescent signals for conversion into visible colors. (A confocal microscope produces images of a single region in a single plane, because it is able to reject scattered and extraneous light.)

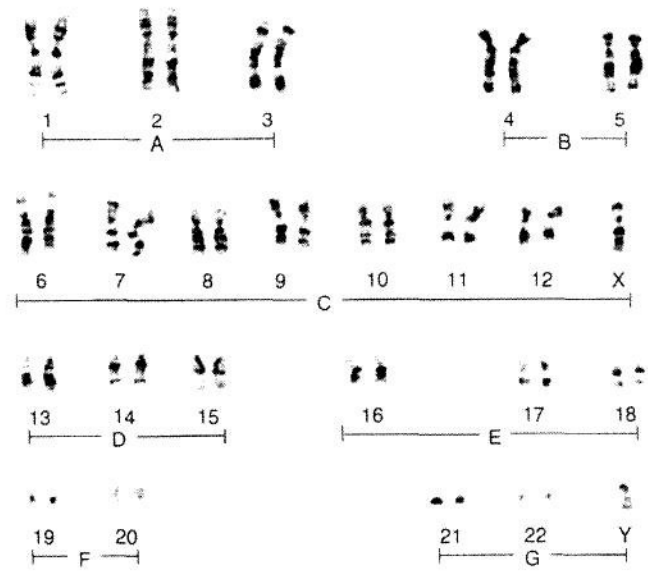
Standard Karyotypes

Chromosome painting dramatically identifies the pairs of homologous chromosomes. The presentation shown in Figure 9.1A is a *metaphase spread*, in which the chromosomes are arranged just as they appear in the cytological preparation. A more conventional representation, called a **karyotype,** is shown in Figure 9.1B. In a karyotype, the autosomes in the metaphase spread are rearranged systematically in pairs, from longest to shortest, and numbered from 1 (the longest) through 22. In this example the sex chromosomes are set off at the bottom right. The single X and Y chromosomes are evident. The karyotype of a normal human female has a pair of X chromosomes, instead of an X and a Y, in addition to the 22 pairs of autosomes. Chromosome painting is of considerable utility in human cytogenetics because even complex chromosome rearrangements can be detected rapidly and easily.

Another, less colorful metaphase spread and karyotype are shown in **Figure 9.2**. In this case the chromosomes have been treated with a staining reagent called Giemsa, which causes the chromosomes to exhibit transverse bands (**G-bands**) that are specific for each pair of homologs. These bands permit smaller segments of each chromosome arm to be identified. The



(A) Photograph of metaphase chromosomes



(B) Karyotype

Figure 9.2 A karyotype of a normal human male. Blood cells arrested in metaphase were stained with Giemsa and photographed with a microscope. (A) The chromosomes as seen in the cell by microscopy. (B) The chromosomes have been cut out of the photograph and paired with their homologs. [Courtesy of Patricia Jacobs.]

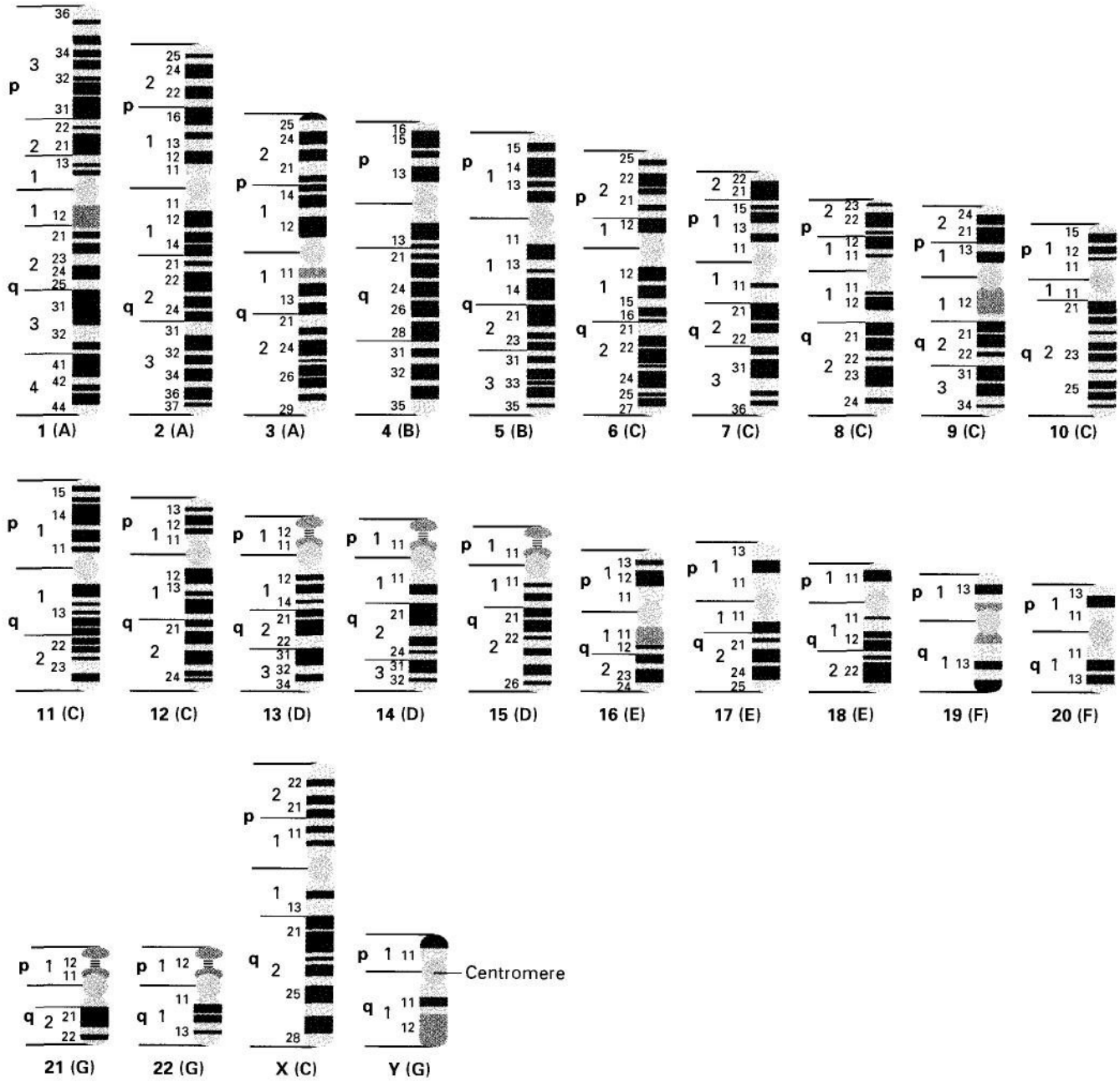


Figure 9.3 Designations of the bands and interbands in the human karyotype. Beneath each chromosome is the lettered group (A–G) to which it belongs.

chromosomes are grouped into seven sets denoted by the letters A through G. (The X chromosome is included in group C, the Y in group G.) These conventional groupings date from a time prior to G-banding and chromosome painting, when the chromosomes could be sorted only by size and centromere position.

The nomenclature of the banding patterns in human chromosomes is shown in Figure 9.3, where the red letter beneath each chromosome indicates its group. For each chromosome, the short arm is designated with the letter p, which stands for "petite," and the long arm by the letter q, which stands for "not-p." Within each arm the

Table 9.1 Conventional karyotype symbols used in human genetics

A–G	Chromosome groups
1–22	Autosome designations
X, Y	Sex-chromosome designations
p	Short arm of chromosome
q	Long arm of chromosome
ter	Terminal portion: pter refers to terminal portion of short arm, qter to terminal portion of long arm
+	Preceding a chromosome designation, indicates that the chromosome or arm is extra; following a designation, indicates that the chromosome or arm is larger than normal
−	Preceding a chromosome designation, indicates that the chromosome or arm is missing; following a designation, indicates that the chromosome or arm is smaller than normal
mos	Mosaic
/	Separates karyotypes of clones in mosaics—e.g., 47, XXX/45,X
dup	Duplication
dir dup	Direct duplication
inv dup	Inverted duplication
del	Deletion
inv	Inversion
t	translocation
rcp	Reciprocal translocation
rob	Robertsonian translocation
r	Ring chromosome
i	Isochromosome (two identical arms attached to a single centromere, like an attached-X chromosome in *Drosophila*)

regions are numbered. The first digit is the major region, numbered consecutively proceeding from the centromere toward the telomere; within each region, the second number indicates the next-smaller division, again numbered outward from the centromere. For example, the designation 1p34 indicates chromosome 1, short arm, division 34. Some divisions can be subdivided still further by the bands and interbands within them, which are numbered consecutively and indicated by a digit placed after a decimal point following the main division; for example, 1p36.2 means the second band in 1p36. Some familiar genetic landmarks in the human genome are the Rh (Rhesus) blood-group locus, which is somewhere in the region between 1p34 and 1p36.2; the ABO blood-group locus at 9q34; the red–green color blindness genes at Xq28, and the male-determining gene on the Y chromosome, called *SRY* (sex-determining region, Y) at Yp11.3.

For ease of reference, some of the principal conventions used in describing human chromosomes and human chromosome abnormalities are listed in **Table 9.1**. The suffix *ter* is used for "terminal"; hence pter and qter refer to the terminal parts of the short and long arms, respectively. The + and − symbols refer to extra or missing chromosomal material. Preceding a chromosome symbol they mean the entire chromosome is affected; following it they refer to only part of the chromosome. For example, +21 means the presence of an extra copy of chromosome 21 (characteristic of Down syndrome, also called Down's syndrome), whereas 21q22ter+ means the presence of an extra copy of part of chromosome 21 extending from band q22 through to the telomere. The designation

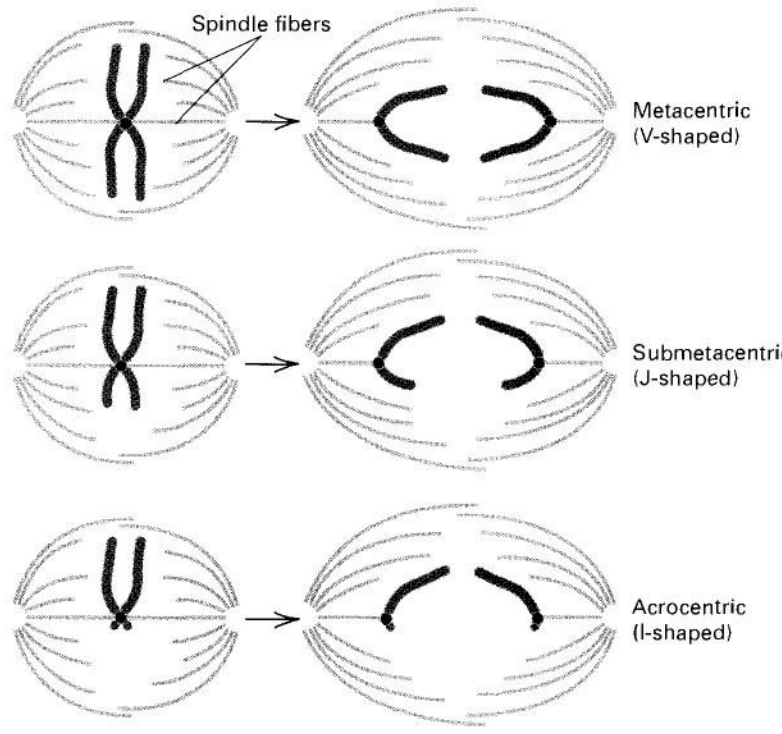


Figure 9.4 Three possible shapes of monocentric chromosomes in anaphase as determined by the position of the centromere (shown in dark blue).

mos means **mosaic,** an individual composed of two or more genetically distinct types of cells. The other terms in Table 9.1 are self-explanatory; many of these chromosome abnormalities are discussed in detail later in this chapter.

The Centromere and Chromosome Stability

As is true in nearly all eukaryotic organisms, each human chromosome is linear and has a single centromere. (Rarely, a *ring* chromosome is found, which results from breakage and loss of the telomere at each end of a linear chromosome and subsequent fusion of the broken ends.) Linear monocentric chromosomes are often classified according to the relative position of their centromeres, which determines the appearance of the daughter chromosomes as they separate from each other in anaphase (**Figure 9.4**). A chromosome with its centromere about in the middle is a **metacentric chromosome;** the arms are of approximately equal length, and each daughter chromosome forms a V shape at anaphase. When the centromere is somewhat off-center, the chromosome is a **submetacentric chromosome,** and each daughter chromosome forms a J shape at anaphase. A chromosome with the centromere very close to one end appears I-shaped at anaphase because the arms are grossly unequal in length; such a chromosome is **acrocentric.**

Chromosomes with a single centromere are usually the only ones that are reliably transmitted from parental cells to daughter cells or from parental organisms to their progeny. When a cell divides, spindle fibers attach to the centromere of each chromosome and pull the sister chromatids to opposite poles. A chromosome that lacks a centromere is an **acentric** chromosome. Acentric chromosomes are genetically unstable because they cannot be maneuvered properly during cell division and are lost. Occasionally, a chromosome arises that has two centromeres and is said to be **dicentric.** A dicentric chromosome is also genetically unstable because it is not transmitted in a predictable fashion. The dicentric chromosome is frequently lost from a cell when the two centromeres proceed to opposite poles in the course of cell division; in this

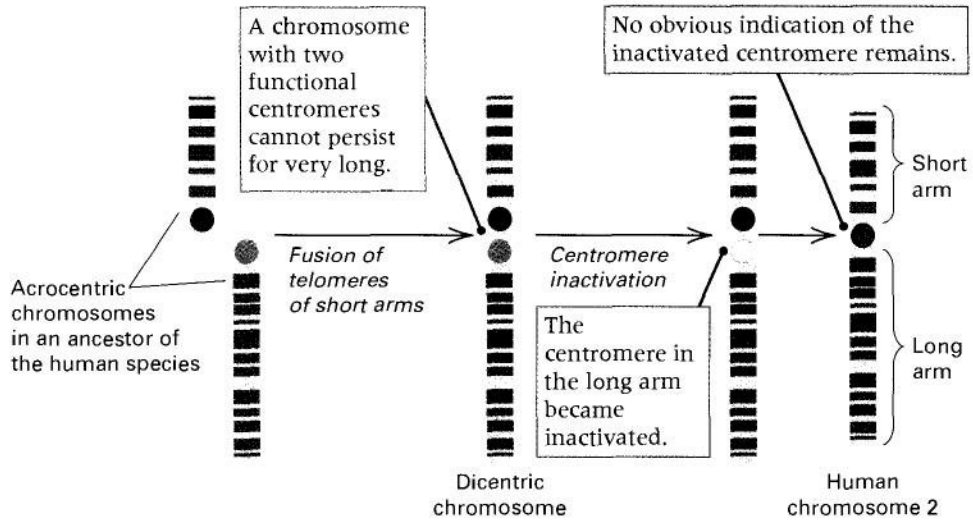


Figure 9.5 Human ancestors had 48 pairs of chromosomes rather than 46. In the evolution of the human genome, two acrocentric chromosomes fused to create human chromosome 2.

case, the chromosome is stretched and forms a *bridge* between the daughter cells. This bridge may not be included in either daughter nucleus, or it may break, with the result that each daughter nucleus receives a broken chromosome. We will consider one mechanism by which dicentric and acentric chromosomes are formed when we discuss inversions.

Although most dicentric chromosomes are genetically unstable, if the two centromeres are close enough together, they can frequently behave as a single unit and be transmitted normally. This principle was important in the evolution of human chromosome 2. Among higher primates, chimpanzees and human beings have 23 pairs of chromosomes that are similar in morphology and G-banding pattern, but chimpanzees have no obvious homolog of human chromosome 2, a large metacentric chromosome. Instead, chimpanzees have two medium-sized acrocentric chromosomes not found in the human genome. The cause of this situation is shown in **Figure 9.5**. The G-banding patterns indicate that human chromosome 2 was formed by fusion of the telomeres between the short arms of two acrocentric chromosomes that, in chimpanzees, remain acrocentrics. The fusion created a dicentric chromosome with its two centromeres close together. The new chromosome must have been sufficiently stable to be retained in the human lineage. The present-day human chromosome 2 has only a single functional centromere. The centromere on the slightly longer arm is inactive. Whether this inactivation occurred simultaneously with the chromosome fusion or evolved gradually is unknown. In any case, this chromosome fusion reduced the chromosome number in the human lineage from 48, which is characteristic of the great apes (chimpanzee, gorilla, and orangutan), to the number 46.

Dosage Compensation of X-linked Genes

For all organisms with XX–XY sex determination, there is a problem of the dosage of genes on the X and Y chromosome, because females have two copies of this chromosome whereas males have only one. (There is less of a problem with Y-linked genes, because the Y chromosome carries relatively few genes.) In most organisms, a mechanism of **dosage compensation** has evolved in which the unequal dosage in the sexes is corrected either by increasing the activity of genes in the X chromosome in males or by reducing the activity of genes in the X chromosome in females. The mechanism of dosage compensation in mammals is seemingly simple: In the early cleavage divisions of the embryo, all X chromosomes except one, chosen at random, are genetically inactivated. Different tissues undergo **X inactivation** at different times, and an X chromosome that is inactivated in

a particular somatic cell remains inactive in all the descendants of the cell (**Figure 9.6**).

The X-chromosome inactivation takes place in all embryos with two or more X chromosomes, including normal XX females. The inactivation process is one of chromosome condensation initiated at a site called *XIC* (for *X-inactivation center*) near the centromere on the long arm between Xq11.2 and Xq21.1. Cytologically, this site marks a "bend" in the chromosome in mitotic metaphase, which is thought to be a visible manifestation of the X inactivation. The finding that the Xq11.2–Xq21.1 region is present in all structurally abnormal X chromosomes is taken as one line of evidence for its essential function in X inactivation, because structurally abnormal X chromosomes lacking this region would not be inactivated and therefore cause lethality and not be observed. Another line of evidence is that the *XIC* includes a transcribed region in band Xq13 designated *Xist* (for *X-inactivation–specific transcript*). Transcription of *Xist* is the earliest event observed in X inactivation, and *Xist* is the only gene known to be transcribed only from the inactive X chromosome. As in many cases of transcriptionally silenced genes (Chapter 12), cytosine methylation targets are fully methylated in the 5' regulatory region of the untranscribed *Xist* gene in the otherwise active X chromosome, whereas those in the expressed allele in the otherwise inactive X are completely unmethylated. Remarkably, the spliced transcript of *Xist* does not contain an open reading frame encoding a protein. It appears to function as a structural RNA, and as transcription of *Xist* continues, the spliced transcript progressively coats the inactive X chromosome, spreading outward from the *XIC*. Thereafter, other molecular changes take place along the inactive X chromosome that are typically associated with gene silencing, including heavy cytosine methylation in 5' regulatory regions, acetylation of histones, and aggregation of heterochromatin-specific DNA-binding proteins. In mouse embryos in which the *Xist* homolog has been disrupted, X inactivation does not take place, which demonstrates that *Xist* is essential for inactivation.

In certain cell types, the inactive X chromosome in females can be observed microscopically as a densely staining heterochromatic body in the nucleus of in-

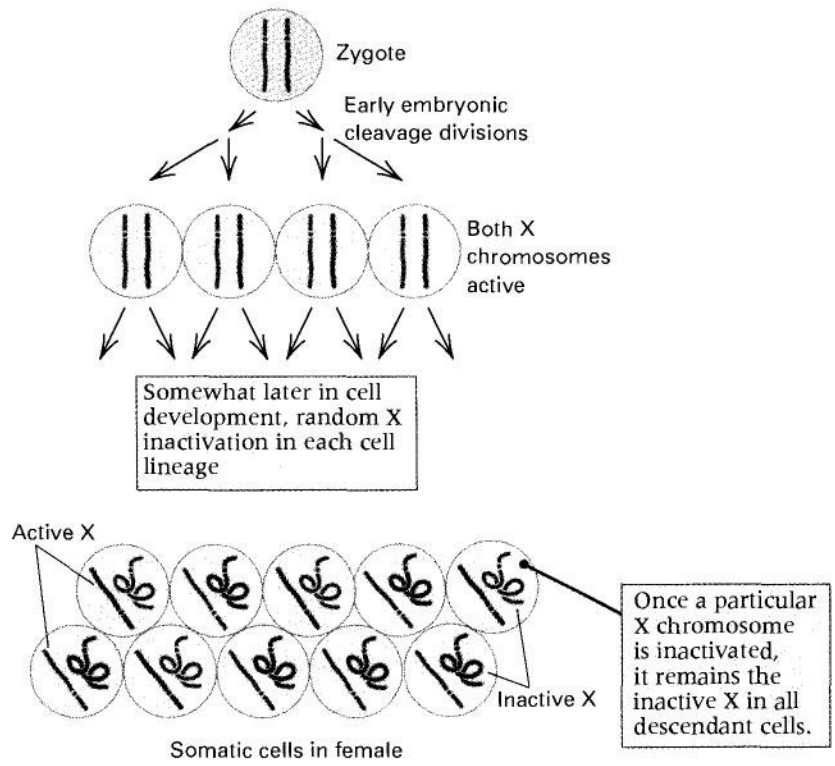


Figure 9.6 Schematic diagram of somatic cells of a normal female, showing that the female is a mosaic for X-linked genes. The two X chromosomes are shown in red and blue. An active X is depicted as a straight chromosome, an inactive X as a tangle. Each cell has just one active X, but the particular X that remains active is a matter of chance. In human beings, the inactivation includes all but a few genes in the tip of the short arm.

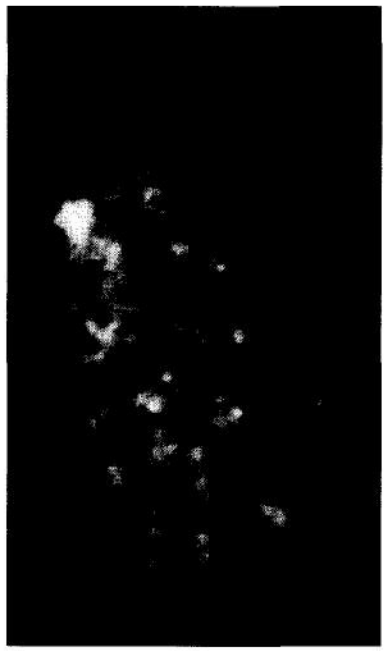


Figure 9.7 Fluorescence micrograph of a human cell, showing a Barr body (bright spot at the upper left indicated by arrow.) This cell is from a normal human female, and it has one Barr body. [Micrograph courtesy of A. J. R. de Jonge.]

terphase cells. This is called a **Barr body** (arrow in **Figure 9.7**). Although cells of normal females have one Barr body, cells of normal males have none. Persons with two or more X chromosomes have all but one X chromosome per cell inactivated, and the number of Barr bodies equals the number of inactivated X chromosomes. Hybridization *in situ* of labeled *Xist* RNA with chromosomes from a normal female demonstrates that the *Xist* RNA is localized within the nucleus at the same position as the Barr body.

X-chromosome inactivation has two consequences. First, it results in dosage compensation. It equalizes the number of active copies of X-linked genes in females and males. Although a female has two X chromosomes and a male has only one, because of inactivation of one X chromosome in each of the somatic cells of a female, both sexes have the same number of active X chromosomes. The mechanism of dosage compensation by means of X inactivation was originally proposed by Mary Lyon and is called the **single-active-X principle.**

The second consequence of X-chromosome inactivation is that a normal female is a sort of mosaic for X-linked genes. That is, each somatic cell expresses the genes in only one X chromosome, but the X chromosome that is active genetically differs from one cell to the next. This mosaicism can be observed directly in females that are heterozygous for X-linked alleles that determine different forms of an enzyme, A and B: When cells from a heterozygous female are individually cultured in the laboratory, half of the clones are found to produce only the A form of the enzyme and the other half to produce only the B form. Mosaicism can also be observed directly in women who are heterozygous for an X-linked recessive mutation that results in the absence of sweat glands; these women exhibit large patches of skin in which sweat glands are present (these patches are derived from embryonic cells in which the normal X chromosome remained active and the mutant X was inactivated) and other large patches of skin in which sweat glands are absent (these patches are derived from embryonic cells in which the normal X chromosome was inactivated and the mutant X remained active.)

Although all mammals use X-chromosome inactivation for dosage compensation in females, the choice of which chromosome to inactivate is not always random. In marsupial mammals, which include the kangaroo, the koala, and the wombat, the X chromosome that is inactivated is always the one contributed by the father. The result is that female marsupials are not genetic mosaics of paternal and maternal X-linked genes. A female marsupial expresses the X-linked genes that she inherited from her mother.

The Calico Cat Is a Result of X-Chromosome Inactivation

In nonmarsupial mammals, the result of random X inactivation in females can sometimes be observed in the external phenotype. One example is the "calico" pattern of coat coloration in female cats. Two alleles affecting coat color are present in the X chromosome in cats. One allele results in an orange coat color (sometimes referred to as yellow), the other in a black coat color. Because he has only one X chromosome, a normal male has either the orange or the black allele. A female can be heterozygous for orange and black, and in this case the coat color is "calico"—a mosaic of orange

CoNnection

Lyonization of an X Chromosome

Mary F. Lyon 1961
Medical Research Council,
Harwell, England
Gene Action in the X Chromosome of the Mouse (Mus musculus L.)

How do organisms solve the problem that females have two X chromosomes whereas males have only one? Unless there were some type of correction (called dosage compensation), the unequal number would mean that for all the genes in the X chromosome, cells in females would have twice as much gene product as cells in males. It would be difficult for the developing organism to cope with such a large difference in dosage for so many genes. The problem of dosage compensation has been solved by different organisms in different ways. The hypothesis put forward in this paper is that in the mouse (and, by inference, in other mammals), the mechanism is very simple: One of the X chromosomes, chosen at random in each cell lineage early in development, becomes inactivated and remains inactivated in all descendant cells in the lineage. In certain cells, the inactive X chromosome becomes visible in interphase as a deeply staining "sex-chromatin body." We now know that there are a few genes in the short arm of the X chromosome that are not inactivated. There is also good evidence that the inactivation of the rest of the X chromosome takes place sequentially from an "X-inactivation center." We also know that in marsupial mammals, such as the kangaroo, it is always the paternal X chromosome that is inactivated.

It has been suggested that the so-called sex-chromatin body is composed of one heteropyknotic [that is, deeply staining during interphase] X chromosome. . . . The present communication suggests that evidence of mouse genetics indicates: (1) that the heteropyknotic X chromosome can be either paternal or maternal in origin, in different cells of the same animal; (2) that it is genetically inactivated. The evidence has two main parts. First, the normal phenotype of XO females in the mouse shows that only one active X chromosome is necessary for normal development, including sexual development. The second piece of evidence concerns the mosaic phenotype of female mice heterozygous for some sex-linked [X-linked] mutants. All sex-linked mutants so far known affecting coat colour cause a "mottled" or "dappled" phenotype, with patches of normal and mutant colour. . . . It is here suggested that this mosaic phenotype is due to the inactivation of one or other X chromosome early in embryonic development. If this is true, pigment cells descended from the cells in which the chromosome carrying the mutant gene was inactivated will give rise to a normal-coloured patch and those in which the chromosome carrying the normal gene was inactivated will give rise to a mutant-coloured patch. . . . Thus this hypothesis predicts that for all sex-linked genes of the mouse in which the phenotype is due to localized gene action the heterozygote will have a mosaic appearance. . . . The genetic evidence does not indicate at what early stage of embryonic development the inactivation of the one X chromosome occurs. . . . The sex-chromatin body is thought to be formed from one X chromosome in the rat and in the opossum. If this should prove to be the case in all mammals, then all female mammals heterozygous for sex-linked mutant genes would be expected to show the same phenomena as those in the mouse. The coat of the tortoiseshell cat, being a mosaic of the black and yellow colours of the two homozygous genotypes, fulfills this expectation.

The coat of the tortoiseshell cat, being a mosaic of the black and yellow colours of the two homozygous genotypes, fulfills this expectation.

Source: Nature 190: 372



Figure 9.8 A female cat that is heterozygous for the orange and black coat color alleles and shows the classic "calico" pattern of patches of orange and black fur.

and black patches. The orange and black patches result from X-chromosome inactivation. In cell lineages in which the X chromosome bearing the orange allele is inactivated, the X chromosome with the black allele is active and so the fur is black. In cell lineages in which the X chromosome with the black allele is inactivated, the orange allele in the active X chromosome results in orange fur (**Figure 9.8**).

The white patches have a completely different explanation. The white patches are due to an autosomal gene *S* for white

spotting, which prevents pigment formation in the cell lineages in which it is expressed. Why the *S* gene is expressed in some cell lineages and not others is not known. Homozygous *SS* cats have more white than heterozygous *Ss* cats. The female in the photograph is homozygous.

Pseudoautosomal Inheritance

The silencing of genes in the inactive X chromosome evolved gradually as the X and Y chromosomes progressively diverged from their ancestral chromosomes and the Y chromosome began to lose the function of most of its genes. The gene inactivation in the inactive X chromosome therefore affects individual genes and blocks of genes, and some genes in the inactive X are not silenced. Some of the genes that escape X inactivation have functional homologs in the Y chromosome, whereas others do not.

Two continuous regions that escape X inactivation, each several megabases in length, are found at the tips of the long and short arms. These are regions in which the Y chromosome does retain functional homologous genes. These regions of homology enable the X and Y chromosomes to synapse in spermatogenesis, and (normally, at least) a crossover takes place that holds the chromosomes together to ensure their proper segregation during anaphase I. The regions of shared X–Y homology define the **pseudoautosomal regions:** *PARp* at the terminus of the short arms and *PARq* at the terminus of the long arms. Because crossing-over regularly takes place in the short pseudoautosomal regions, the rate of recombination per nucleotide pair is at least 20-fold greater in the pseudoautosomal regions than in the autosomes.

The pedigree patterns of inheritance of genes in the pseudoautosomal regions are indistinguishable from patterns characteristic of autosomal inheritance. The reason is that a mutant allele in a pseudoautosomal region is neither completely X-linked nor completely Y-linked but can move back and forth between the X and Y chromosomes because of recombination in the pseudoautosomal region. A gene that shows an autosome-like pattern of inheritance, but that is known from molecular studies to reside in a pseudoautosomal region, is said to show *pseudoautosomal inheritance*.

Gene Content and Evolution of the Y Chromosome

Comparative cytogenetic and molecular studies suggest that the X and Y chromosomes began their existence as a pair of ordinary autosomes in the common ancestor of modern mammals and birds. They started to diverge in DNA sequence and gene content at about the same time that the evolutionary lineage of mammals diverged from that of birds, some 300–350 million years ago (MY). One must assume that prior to this time, recombination took place at normal levels throughout the entire proto-X and proto-Y chromosomes and that their gene contents were identical.

In the short arm of the Y chromosome, near the pseudoautosomal region but not within it, the Y chromosome contains the master sex-determining gene *SRY* at position Yp11.3. *SRY* codes for a protein transcription factor, the **testis-determining factor (TDF),** that triggers male embryonic development by inducing the undifferentiated embryonic genital ridge, the precursor of the gonad, to develop as a testis. A transcription factor is a protein that stimulates transcription of its target genes.

Once *SRY* had evolved as a sex-determining mechanism, the Y chromosome began to diverge in DNA sequence from the X chromosome, and the region of possible X–Y recombination became progressively restricted to the telomeric regions. In regions with no X–Y recombination, there is a steady selection pressure for genes in the Y chromosome to undergo mutational degeneration into nonfunctional states. This results from the forced heterozygosity of the Y chromosome, which allows multiple deleterious mutations to accumulate through time because there is no opportunity for recombination to regenerate Y chromosomes that are free of deleterious mutations. Hence any Y-linked gene whose function is nonessential will tend to degenerate gradually because of the accumulation of mutations, and at the same time there will be selection pressure for dosage compensation of the homologous gene in the X chromosome. Eventually, only the dosage-compensated X-linked gene will remain functional.

Apparently blocks of genes were removed from the region of X–Y recombina-

tion in large chunks. Molecular evidence for this conclusion is summarized in **Figure 9.9**. Shown at the left are the locations of some protein-coding sequences in the short arm of the modern X chromosome from band Xp11 to the telomere. All of these genes have homologous sequences in the modern Y chromosome, as shown at the right. The amount of sequence divergence between the X and Y homologs shows a remarkable pattern. In the synonymous positions of the codons, where a nucleotide substitution can occur without changing the encoded amino acid, the sequence divergence (measured, roughly speaking, as the proportion of synonymous sites that differ between the X and Y homologs) for the genes *GYG2–AMELX* is 0.07–0.11, for *TB4X–UTX* it is 0.23–0.36, for *SMCX* it is 0.52, and for other genes outside the region shown it is > 0.94. Because the evolutionary rate of nucleotide substitutions at synonymous sites in mammalian genes is known, we can say that these levels of divergence correspond to divergence times of 30–50 million years (MY) for *GYG2–AMELX*, 80–130 MY for *TB4X–UTX*, 130–170 MY for *SMCX*, and 300–350 MY for other genes outside the region shown. The simplest explanation is that these were the times when successive blocks of genes were removed from the region of X–Y recombination by chromosome rearrangements—namely, inversions in which the Y chromosome was broken at two points and the middle segment inverted before rejoining of the broken ends. For each inversion, one breakpoint must have occurred outside the region shown, but the location of the other breakpoint

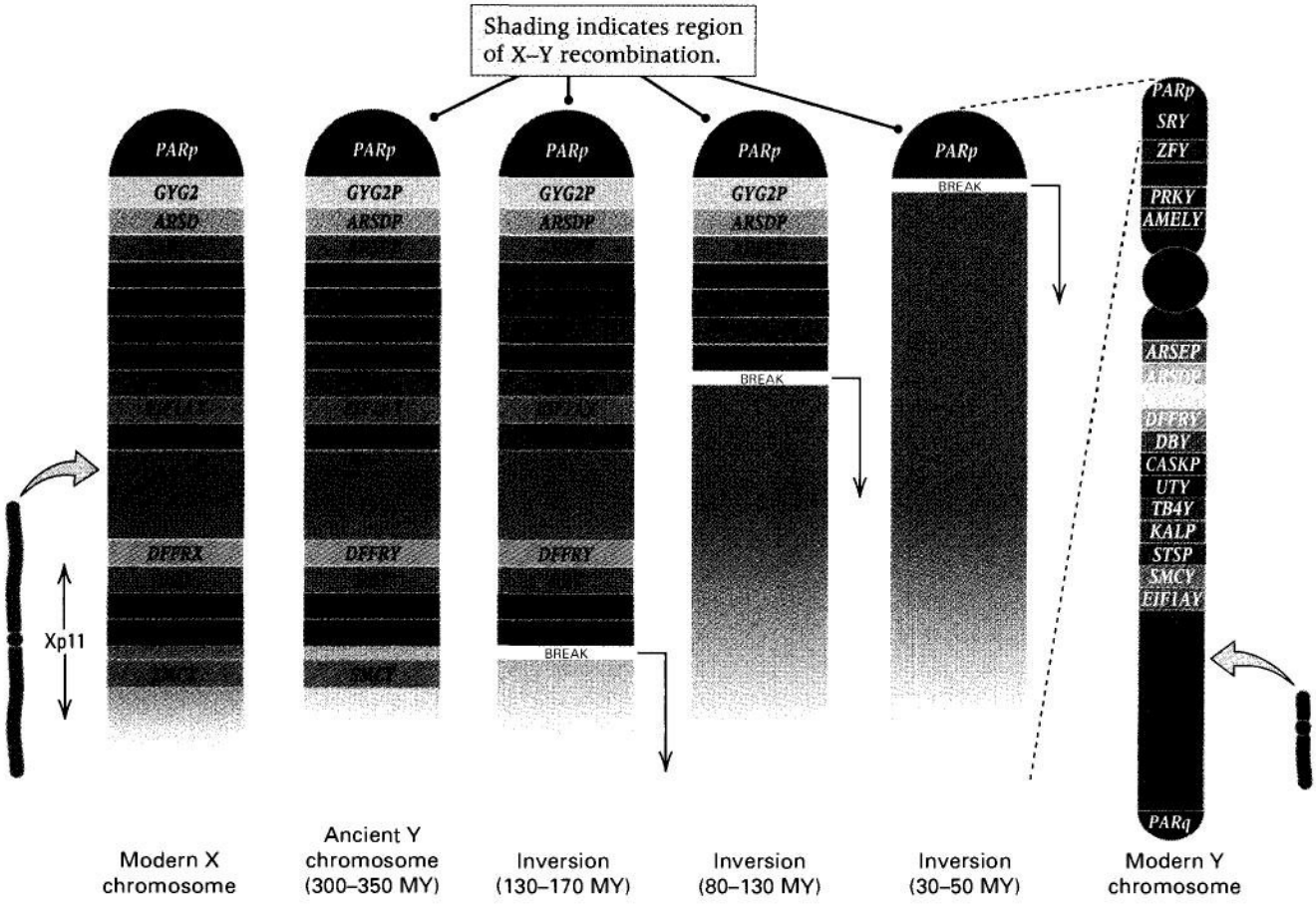


Figure 9.9 Progressive shortening of the mammalian X–Y pseudoautosomal region through time due to inversions in the Y chromosome, inferred from DNA sequence data. The arrows denote the distal (nearest the telemere) breakpoint of each successive inversion interrupting the pseudoautosomal region. [Data from B. T. Lahn and D. C. Page. 1999. *Science* 286:964.]

can be pinpointed as within the region between the blocks of genes. Hence, as shown in the evolutionary reconstruction in Figure 9.9, the 130–170 MY inversion breakpoint was adjacent to *UTY*, the 80–130 MY inversion breakpoint adjacent to *AMELY*, and the 30–50 MY year breakpoint adjacent to the present-day pseudoautosomal region on the short arm. As each of these inversions was fixed in the evolving Y chromosome, it removed the corresponding block of genes from the region of X–Y recombination, so that the rate of sequence divergence between the X and Y homologs accelerated. Other rearrangements in the Y chromosome, not able to be traced from these data, led to some additional scrambling of the gene order in the Y chromosome.

9.2 Chromosome Abnormalities in Spontaneous Abortion

Approximately 15 percent of all recognized pregnancies in human beings terminate in spontaneous abortion, and in about half of all spontaneous abortions, the fetus has a major chromosome abnormality. Table 9.2 summarizes the average rates of chromosome abnormality found per 100,000 recognized pregnancies in several studies. The term **trisomy** refers to an otherwise diploid organism that has an extra copy of an individual chromosome. Many of the spontaneously aborted fetuses have trisomy of one of the autosomes. Triploids, which have three sets of chromosomes (total count 69), and tetraploids, which have four sets of chromosomes (total count 92), are also common in spontaneous abortions. Triploids and tetraploids are examples of **euploid** conditions, because they have the same relative gene dosage as found in the diploid. In contrast, relative gene dosage is upset in a trisomic, because three copies of the genes located in the trisomic chromosome are present, whereas two copies of the genes in the other chromosomes are present. Such unbalanced chromosome complements are said to be **aneuploid.** Although it is not apparent in the data in Table 9.2, in most organisms, euploid abnormalities generally have less severe phenotypic effects than aneuploid abnormalities. For example, in *Drosophila*, triploid females are viable, fertile, and nearly normal in morphology, whereas trisomy for either of the two large autosomes is invariably lethal (the larvae die at an early stage). In Table 9.2 the term *balanced translocation* refers to a euploid condition in which nonhomologous chromosomes have an interchange of part, but all of the parts are present; the term *unbalanced translocation* refers to an aneuploid condition in

Table 9.2 Chromosome abnormalities per 100,000 recognized human pregnancies

	15,000 spontaneous abortions	85,000 live births
Trisomy		
A: 1	0	0
A: 2	159	0
A: 3	53	0
B: 4	95	0
B: 5	0	0
C: 6–12	561	0
D: 13	128	17
D: 14	275	0
D: 15	318	0
E: 16	1229	0
E: 17	10	0
E: 18	223	13
F: 19–20	52	0
G: 21	350	113
G: 22	424	0
Sex chromosomes		
XYY	4	46
XXY	4	44
XO	1350	8
XXX	21	44
Translocations		
Balanced	14	164
Unbalanced	225	52
Polyploid		
Triploid	1275	0
Tetraploid	450	0
Other (mosaics, etc.)	280	49
Total	**7500**	**550**

CoNnection

The First Human Chromosomal Disorder Identified

Jerome Lejeune, Marthe Gautier, and Raymond Turpin 1959
National Center for Scientific Research, Paris, France
Study of the Somatic Chromosomes of Nine Down Syndrome Children
[original in French]

Down syndrome had been one of the greatest mysteries in human genetics. One of the most common forms of mental retardation, the syndrome did not follow any pattern of Mendelian inheritance. Yet some families had two or more children with Down syndrome. (Many of these cases are now known to be due to a translocation involving chromosome 21.) This paper marked a turning point in human genetics by demonstrating that Down syndrome actually results from the presence of an extra chromosome. It was the first identified chromosomal disorder.

The excerpt uses the term telocentric, *which means a chromosome that has its centromere very near one end. In the human genome, the smallest chromosomes are three very small telocentric chromosomes. These are chromosomes 21, 22, and the Y. A normal male has five small telocentrics (21, 21, 22, 22, and Y); a normal female has four (21, 21, 22, and 22). (The X is a medium-sized chromosome with its centromere somewhat off center.) In the table that follows, note the variation in chromosome counts in the "doubtful" cells. The methods for counting chromosomes were then very difficult, and many errors were made either by counting two nearby chromosomes as one or by including in the count of one nucleus a chromosome that actually belonged to a nearby nucleus. Lejeune and collaborators wisely chose to ignore these doubtful counts and based their conclusion only on the "perfect" cells. Sometimes good science is a matter of knowing which data to ignore.*

Analysis of the chromosome set of the "perfect" cells reveals the presence in Down syndrome boys of 6 small telocentric chromosomes (instead of 5 in the normal man) and 5 small telocentric ones in Down syndrome girls (instead of 4 in the normal woman).

The culture of fibroblast cells from nine Down syndrome children reveals the presence of 47 chromosomes, the supernumerary chromosome being a small telocentric one. The hypothesis of the chromosomal determination of Down syndrome is consid-

which some part of the genome is missing. The much greater survivorship of the balanced translocation indicates that a euploid chromosomal abnormality is generally less harmful than an aneuploid chromosome abnormality.

When an otherwise diploid organism has a missing copy of an individual chromosome, the condition is known as *monosomy*. In most organisms, chromosome loss (resulting in monosomy) is a more frequent event than chromosome gain (resulting in trisomy). However, monosomies are conspicuously absent in the data on spontaneous abortions in Table 9.2. Their absence is undoubtedly due to another feature of monosomy:

> A missing copy of a chromosome (monosomy) usually results in more harmful effects than an extra copy of the same chromosome (trisomy).

In human fertilizations, monosomic zygotes are probably created in even greater numbers than the trisomic zygotes, but monosomy is not found among aborted fetuses because the abortions take place so early in development that the pregnancy goes unrecognized. The spontaneous abortions summarized in Table 9.2, although they represent a huge fetal wastage, serve the important biological function of eliminating many fetuses that are grossly abnormal in their development because of major chromosome abnormalities.

Down Syndrome and Other Viable Trisomies

Table 9.2 demonstrates that monosomy or trisomy of most human autosomes is incompatible with life. There are three exceptions: trisomy 13, trisomy 18, and trisomy 21. The first two are rare conditions associated with major developmental abnormalities, and the affected infants can survive for only a few days or weeks.

Trisomy 21 is **Down syndrome** (or *Down's syndrome*), which occurs in about 1 in 750 live-born children. Its major symptom is mental retardation, but there can

ered. . . . The observations made in these nine cases (five boys and four girls) are recorded in the table to the right.

The number of cells counted in each case may seem relatively small. This is due to the fact that only the pictures [of the spread chromosomes] that claim a minimum of interpretation have been retained in this table. The apparent variation in the chromosome number in the "doubtful" cells, that is to say, cells in which each chromosome cannot be noted individually with certainty, has been pointed out by several authors. It does not seem to us that this phenomenon represents a cytological reality, but merely reflects the difficulties of a delicate technique. It therefore seems logical to prefer a small number of absolutely certain counts ("perfect" cells in the table) to a mass of doubtful observations, the statistical variance of which rests solely on the lack of precision of the observations. Analysis of the chromosome set of the "perfect" cells reveals the presence in Down syndrome boys of 6 small telocentric chromosomes (instead of 5 in the normal man) and 5 small telocentric ones in Down syndrome girls (instead of 4 in the normal woman). . . . It therefore seems legitimate to conclude that there exists in Down syndrome children a small supernumerary telocentric chromosome, accounting for the abnormal figure of 47. To explain these observations, the hypothesis of nondisjunction of a pair of small telocentric chromosomes at the time of meiosis can be considered. . . . It is, however, not possible to say that the supernumerary small telocentric chromosome is indeed a normal chromosome and at the present time the possibility cannot be discarded that a fragment resulting from another type of aberration is involved.

		Number of chromosomes					
		"Doubtful" cells			"Perfect" cells		
		46	47	48	46	47	48
Boys	1	6	10	2	—	11	—
	2	—	2	1	—	9	—
	3	—	1	1	—	7	—
	4	—	3	—	—	1	—
	5	—	—	—	—	8	—
Girls	1	1	6	1	—	5	—
	2	1	2	—	—	8	—
	3	1	2	1	—	4	—
	4	1	1	2	—	4	—

Source: Comptes rendus des séances de l'Académie des Sciences 248: 1721–1722. Translation in S. H. Boyer, *Papers on Human Genetics* (Englewood Cliffs, NJ: Prentice-Hall, 1963)

also be multiple physical abnormalities, such as heart defects. Affected children are small in stature because of delayed maturation of the skeletal system; their muscle tone is poor, resulting in a characteristic facial appearance; and they have a shortened life span of usually less than 50 years. Nevertheless, for a major chromosomal abnormality, the symptoms are relatively mild, and most children with Down syndrome can relate well to other people.

> Children with Down syndrome usually take great pleasure in their surroundings, their families, their toys, their playmates. Happiness comes easily, and throughout life they usually maintain a childlike good humor. They are not burdened with the grown-up cares that come to most people with adolescence and adulthood. Life is simpler and less complex. The emotions that others feel seem to be less intense for them. They are sometimes sad, happy, angry, or irritable, like everyone else, but their moods are generally not so profound and they blow away more quickly. . . . Children with Down syndrome, though slow, are



Down syndrome child with his mother. [J. Cancalosi/DRK Photo.]

still very responsive to their environment, to those around them, and to the affection and encouragement they receive from others. [Quoted from D. W. Smith and A. A. Wilson. *The Child with Down's Syndrome.* (Philadelphia: Saunders, 1973.)]

Most cases of Down syndrome are caused by nondisjunction, which means the failure of homologous chromosomes to separate in meiosis, as explained in Chapter 4. The result of chromosome-21 nondisjunction is one gamete that contains two copies of chromosome 21 and one that contains none. If the gamete with two copies participates in fertilization, then a zygote with trisomy 21 is produced. The gamete with one copy may also participate in fertilization, but zygotes with monosomy 21 do not survive even through the first few days or weeks of pregnancy. About three-fourths of the trisomy-21 fetuses also undergo spontaneous abortion (Table 9.2). If this were not the case, and all trisomy-21 fetuses survived to birth, the incidence of Down syndrome would rise to 1 in 250, approximately a threefold increase over the incidence actually observed.

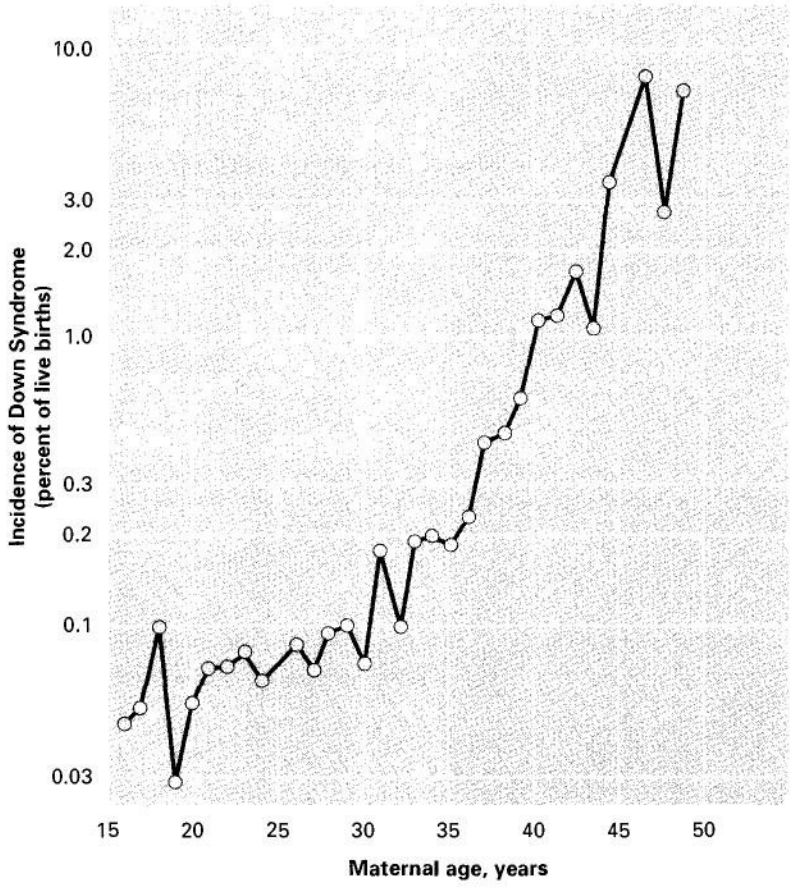


Figure 9.10 Frequency of Down syndrome (number of cases per 100 live births) related to age of mother. The graph is based on 438 Down syndrome births (among 330,859 total births) in Sweden in the period 1968 to 1970. [Data from E. B. Hook and A. Lindsjö. 1978. *Am. J. Human Genet.* 30:19.]

Chromosome 21 is a small chromosome and therefore is somewhat less likely to undergo meiotic crossing-over than a longer one. Noncrossover bivalents sometimes have difficulty aligning at the metaphase plate because they lack a chiasma to hold them together, so there is an increased risk of nondisjunction. Among the events of nondisjunction that result in Down syndrome, about 40 percent are derived from such nonexchange bivalents.

For unknown reasons, nondisjunction of chromosome 21 is more likely to happen in oogenesis than in spermatogenesis, and so the abnormal gamete in Down syndrome is usually the egg. Furthermore, the risk of nondisjunction of chromosome 21 increases dramatically with the age of the mother, resulting in a risk of Down syndrome that reaches 6 percent in mothers age 45 and older (**Figure 9.10**). For this reason, many physicians recommend that older women who are pregnant have cells from the fetus tested in order to detect Down syndrome prenatally. This can be done 15–16 weeks after fertilization by **amniocentesis,** in which cells of the developing fetus are obtained by insertion of a fine needle through the wall of the uterus and into the sac of fluid (the *amnion*) that contains the fetus, or it can be done at 10–11 weeks after fertilization by a procedure called *chorionic villus sampling* (*CVS*), which uses cells from a zygote-derived embryonic membrane (the *chorion*) associated with the placenta. Early diagnosis is desirable, but CVS has about a threefold higher risk of inducing a miscarriage than does amniocentesis.

About 3 percent of all cases of Down syndrome are due not to simple nondisjunction but to an abnormality in chromosome structure. In these cases the risk of recurrence of the syndrome in subsequent children is very high—up to 20 percent of births. The high risk is caused by a chromosomal translocation in one of the parents, in

which chromosome 21 has been broken and become attached to another chromosome. This situation is considered in Section 9.5.

Trisomic Segregation

In a trisomic organism, the segregation of chromosomes in meiosis is upset because the trisomic chromosome has two pairing partners instead of one. The behavior of the chromosomes in meiosis depends on the manner in which the homologous chromosome arms pair and on the chiasmata formed between them. In some cells, the three chromosomes form a **trivalent** in which distinct parts of one chromosome are paired with homologous parts of each of the others (**Figure 9.11**, part A). In metaphase, the trivalent is usually oriented with two centromeres pointing toward one pole and the other centromere pointing toward the other pole. The result is that at the end of both meiotic divisions, one pair of gametes contains two copies of the trisomic chromosome, and the other pair of gametes contains only a single copy. Alternatively, the trisomic chromosome can form one normal bivalent and one

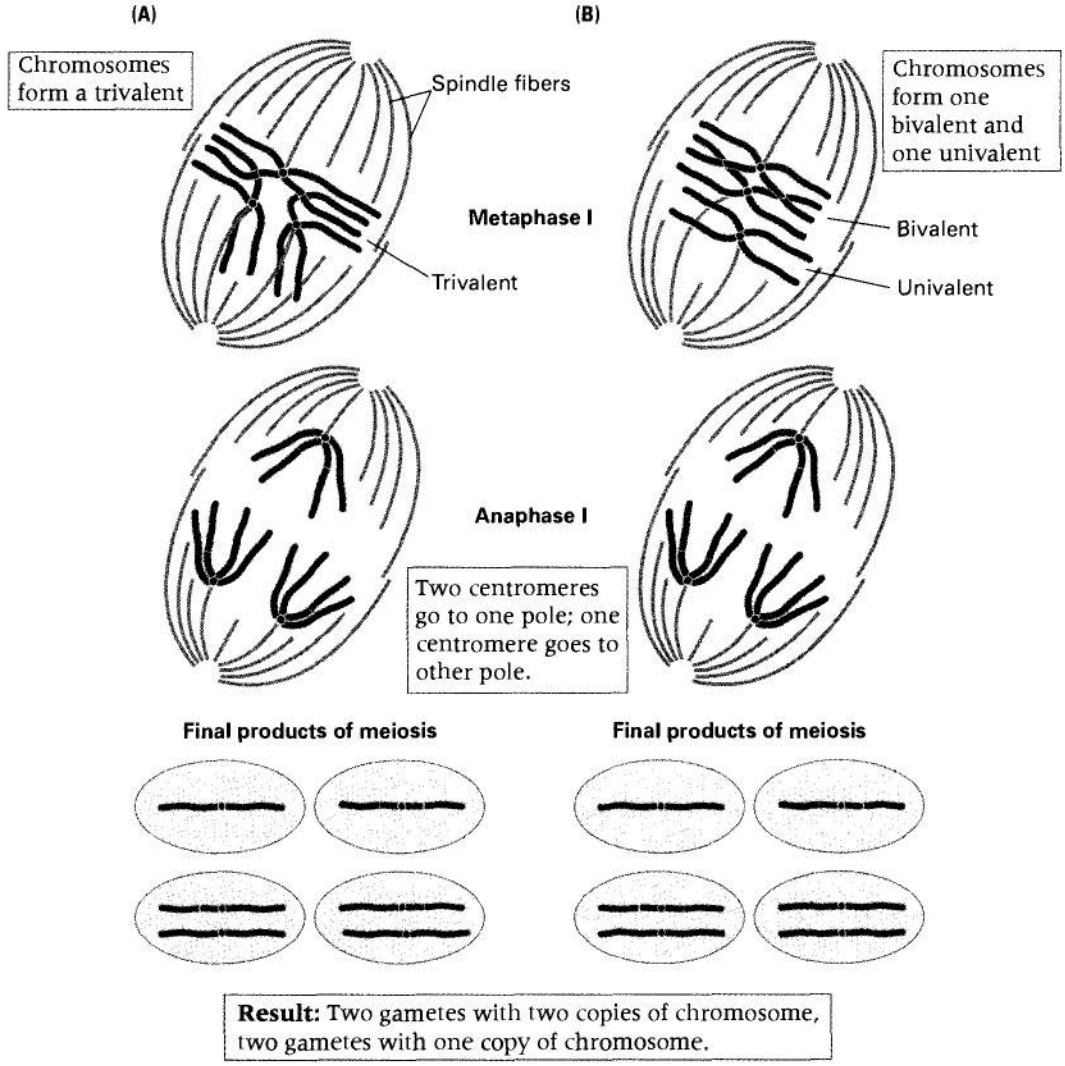


Figure 9.11 Meiotic pairing in a trisomic. (A) Formation of a trivalent. (B) Formation of a univalent and a bivalent. Both types of pairing result in one pair of gametes containing two copies of the trisomic chromosome and the other pair of gametes containing one copy of the trisomic chromosome.

univalent, or unpaired chromosome, as shown in Figure 9.11B. In anaphase I, the bivalent disjoins normally and the univalent usually proceeds randomly to one pole or the other. Again, the end result is the formation of two products of meiosis that contain two copies of the trisomic chromosome and two products of meiosis that contain one copy. To state the matter in another way, a trisomic organism with three copies of a chromosome (say, C C C) will produce gametes among which half contain two copies (C C) and half contain one (C). When mated with a chromosomally normal individual, a trisomic is therefore expected to produce trisomic or normal progeny in a ratio of 1 : 1. This theoretical expectation is borne out in experimental organisms by matings of trisomics and in human beings by the finding that, in the few known cases when a person with Down syndrome reproduces, the frequency of Down syndrome in the offspring is approximately 50 percent.

Sex-Chromosome Abnormalities

Sex-chromosome trisomies, as a group, are even more frequent among newborns than is trisomy 21 (Table 9.2). There are two reasons why extra sex chromosomes have phenotypic effects that are relatively mild compared with those of autosomal trisomies. First, the single-active-X principle results in the silencing of most X-linked genes in all but one X chromosome in each somatic cell. Second, the Y chromosome contains relatively few functional genes.

The four most common types of sex-chromosome abnormalities are described below. The karyotypes are given in the conventional fashion, with the total number of chromosomes listed first, followed by the sex chromosomes that are present. For example, in the designation 47,XXX the number 47 refers to the total number of chromosomes, and XXX indicates that the person has three X chromosomes.

- *47,XXX* This condition is often called simply **trisomy-X.** People with the karyotype 47,XXX are female. Most such women are mentally and phenotypically normal and give no physical or behavioral indication that an extra X chromosome is present. Mild mental impairment may be somewhat more frequent than among 46,XX females.
- *47,XYY* This condition is often called the **double-Y** karyotype. These people are male and tend to be taller than average, but they are otherwise phenotypically normal. At one time it was thought that 47,XYY males developed severe personality disorders and were at a high risk of committing crimes of violence, a belief based on the finding of 47,XYY among violent criminals. Further study indicated that the frequency of 47,XYY males in the general population was much higher than had been supposed. Some 47,XYY males have slightly impaired mental function. Although their rate of criminality is higher than that of normal males, the crimes are mainly nonviolent crimes such as petty theft. The majority of 47,XYY males are phenotypically and psychologically normal, have mental capabilities in the normal range, and have no criminal convictions.
- *47,XXY* This condition is called **Klinefelter syndrome.** Affected persons are male. They tend to be tall, do not undergo normal sexual maturation, are sterile, and in some cases have enlargement of the breasts. Mild mental impairment is common.
- *45,X* Monosomy of the X chromosome in females is called **Turner syndrome.** Affected persons are phenotypically female but short in stature and without sexual maturation. Mental abilities are typically within the normal range. Note in Table 9.2 that more than 99 percent of 45,X fetuses undergo spontaneous abortion. This indicates that the condition usually has profound deleterious effects on development. Table 9.2 also shows that monosomy of the X chromosome is much more frequent than trisomy, which is almost certainly true of autosomes as well, in spite of the absence of monosomics in recognized pregnancies.

9.3 Chromosomal Deletions and Duplications

We now turn to the genetic implications of aberrations in *chromosome structure.* In the history of genetics, chromosome abnormalities were originally discovered through their genetic effects, which, though confusing at first, eventually came to be understood and were later confirmed directly by microscopic observations. There are several principal types of structural aberrations—deletions, duplications, inversions, and translocations—each of which has characteristic genetic effects.

Deletions

A chromosome in which a segment is missing is said to have a **deletion** or a **deficiency.** Deletions are generally harmful to the organism, and the usual rule is the larger the deletion, the greater the harm. Very large deletions are usually lethal, even when heterozygous with a normal chromosome. Small deletions are often viable when they are heterozygous with a structurally normal homolog, because the normal homolog supplies gene products that are necessary for survival. However, even small deletions are usually homozygous-lethal (that is, when both members of a pair of homologous chromosomes carry the deletion).

Deletions can be detected genetically by making use of the fact that a chromosome with a deletion no longer carries the wildtype alleles of the genes that are missing. For example, in *Drosophila*, many deletions of a gene called *Notch*, which results in a notched wing margin, are large enough to remove the nearby wildtype allele of *white*. When these deleted chromosomes are heterozygous with a structurally normal chromosome carrying the recessive *w* allele, the fly has white eyes because the wildtype w^+ allele is no longer present in the deleted *Notch* chromosome. This **uncovering of the recessive allele** implies that the corresponding wildtype allele of *white* has also been deleted. Once a deletion has been identified, its size can be assessed genetically by determining which recessive mutations in the region are uncovered by the deletion. This method is illustrated in **Figure 9.12.**

With the banded polytene chromosomes in *Drosophila* salivary glands, it is possible to study deletions and other chromosome aberrations physically. For example, all the *Notch* deletions cause particular bands to be missing in the salivary chromosomes. Physical mapping of deletions also allows individual genes, otherwise known only from genetic studies, to be assigned to specific bands or regions in the salivary chromosomes.

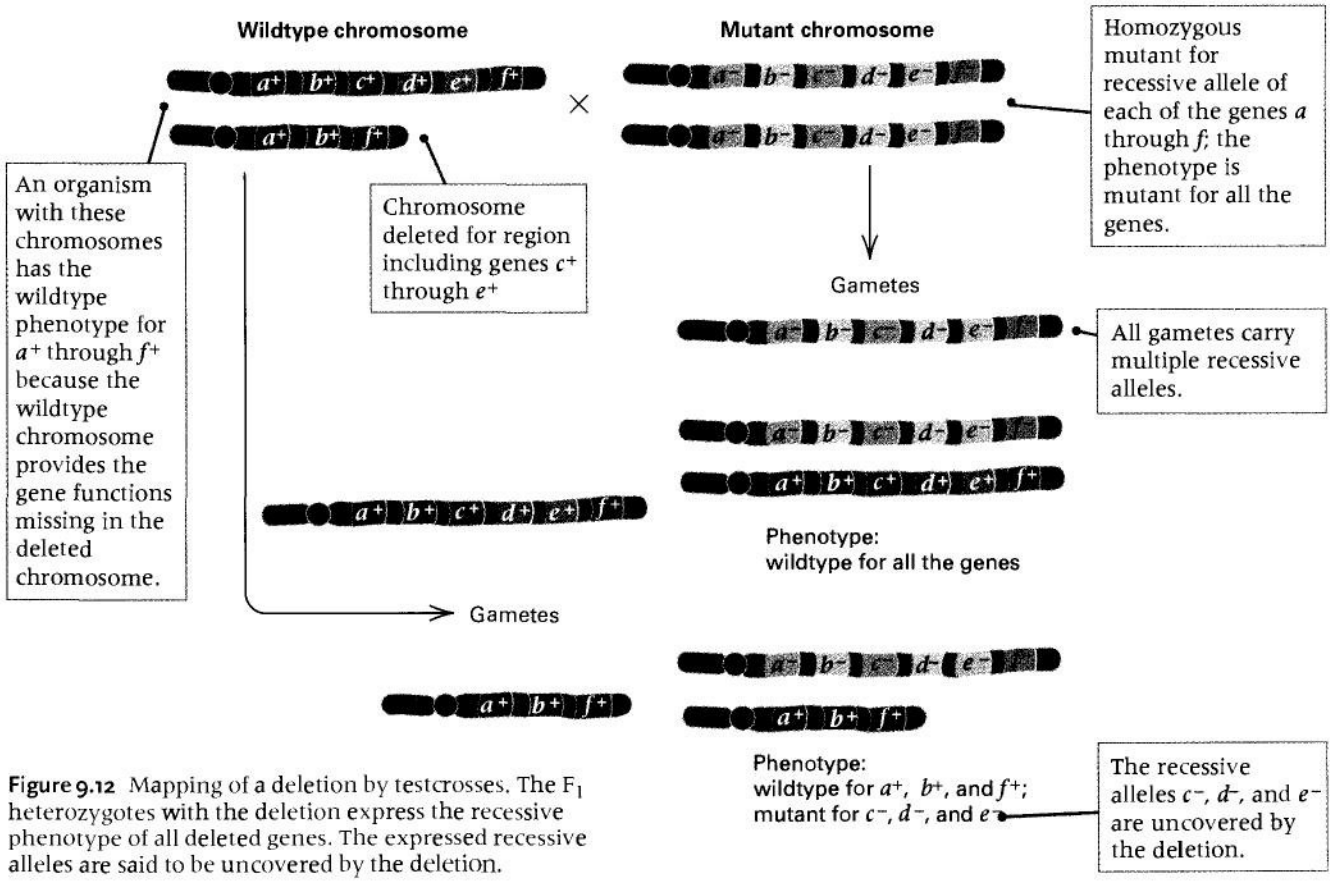


Figure 9.12 Mapping of a deletion by testcrosses. The F_1 heterozygotes with the deletion express the recessive phenotype of all deleted genes. The expressed recessive alleles are said to be uncovered by the deletion.

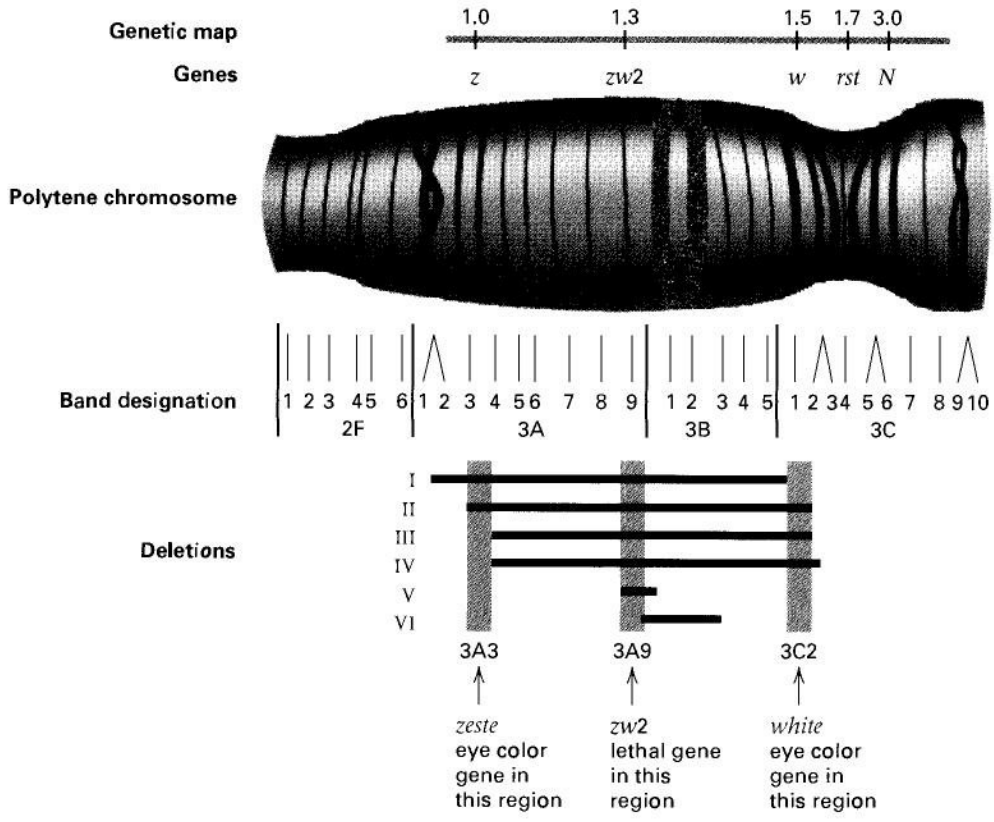


Figure 9.13 Part of the X chromosome in polytene salivary gland nuclei of *Drosophila melanogaster* and the extent of six deletions (I–VI) in a set of chromosomes. Any recessive allele that is uncovered by a deletion must be located inside the boundaries of the deletion. This principle can be used to assign genes to specific bands in the chromosome.

Physical mapping of genes in part of the *Drosophila* X chromosome is illustrated in **Figure 9.13**. The banded chromosome is shown near the top, along with the numbering system used to refer to specific bands. Each chromosome is divided into numbered sections (the X chromosome comprises sections 1 through 20), and each of the sections is divided into subdivisions designated A through F. Within each lettered subdivision, the bands are numbered in order, and so, for example, 3A6 is the sixth band in subdivision A of section 3 (Figure 9.13). On the average, each band contains about 20 kb of DNA, but there is considerable variation in DNA content from band to band.

In Figure 9.13, the mutant X chromosomes labeled I through VI have deletions. The deleted part of each chromosome is shown in red. These deletions define regions along the chromosome, some of which correspond to specific bands. For example, the deleted region in both chromosome I and chromosome II that is present in all the other chromosomes consists of band 3A3. In crosses, only deletions I and II uncover the mutation *zeste (z)*, so the *z* gene must be in band 3A3, as indicated at the top. Similarly, the recessive-lethal mutation *zw2* is uncovered by all deletions except VI; therefore, the *zw2* gene must be in band 3A9. As a final example, the *w* mutation is uncovered only by deletions II, III, and IV; thus the *w* gene must be in band 3C2. The *rst* (rough eye texture) and *N* (notched wing margin) genes are not uncovered by any of the deletions. These genes were localized by a similar analysis of overlapping deletions in regions 3C5 to 3C10.

Duplications

Some abnormal chromosomes have a region that is present twice, which is a **duplication.** Certain duplications have phenotypic effects of their own. An example is the *Bar* duplication in *Drosophila*, which is a tandem duplication of bands 16A1 through 16A7 in the X chromosome. A **tandem duplication** is one in which the duplicated segment is present in the same orientation immediately adjacent to the normal region in the chromosome. In the case of *Bar*, the tandem duplication produces a dominant phenotype of bar-shaped eyes.

Tandem duplications are able to produce even more copies of the duplicated region by means of a process called **unequal crossing-over.** Figure 9.14, part A, illustrates the chromosomes in meiosis of an organism that is homozygous for a tandem duplication (brown region). When they undergo synapsis, these chromosomes can mispair with each other, as illustrated in Figure 9.14B. A crossover within the mispaired part of the duplication (Figure 9.14C) will thereby produce a chromatid carrying a triplication and a reciprocal product (labeled "single copy" in Figure 9.14D) that has lost the duplication. For the *Bar* region, the triplication can be recognized because it produces an even greater reduction in eye size than the duplication.

The most frequent effect of a duplication is a reduction in the probability of survival (reduced viability). In general, survival decreases with increasing size of the duplication. However, deletions are usually more harmful than duplications of comparable size.

Unequal Crossing-Over in Human Red–Green Color Blindness

Human color vision is mediated by three light-sensitive protein pigments present in the cone cells of the retina. Each of the pigments is related to *rhodopsin*, the pigment that is found in the rod cells and mediates vision in dim light. The light sensitivities of the cone pigments are toward blue, red, and green. These are the pure colors that our visual pigments detect. We perceive all other colors as mixtures of blue, red, and green. The gene for the blue-sensitive pigment resides somewhere in 7q22-qter (which means 7q22 to the terminus), and the genes for the red and green pigments are on the X chromosome at Xq28 separated by less than 5 cM (approximately 5 Mb of DNA). Because the red and green pigments arose from the duplication of a single ancestral pigment gene and are still 96 percent identical in amino acid sequence, the genes are similar enough that they can pair and undergo unequal crossing-over. The process of unequal crossing-over is the genetic basis of red-green color blindness.

Almost everyone is familiar with **red–green color blindness;** it is one of the most common inherited conditions in human beings. Approximately 5 percent of males have some form of red–green color blindness. The preponderance of affected males immediately suggests X-linked inheritance, which is confirmed by pedigree studies (Section 4.4). Affected males have normal sons and carrier daughters, and the carrier daughters have 50 percent affected sons and 50 percent carrier daughters.

Actually, there are several distinct varieties of red–green color blindness. Defects in red vision include *protanopia*, an inability to perceive red, and *protanomaly*, an impaired

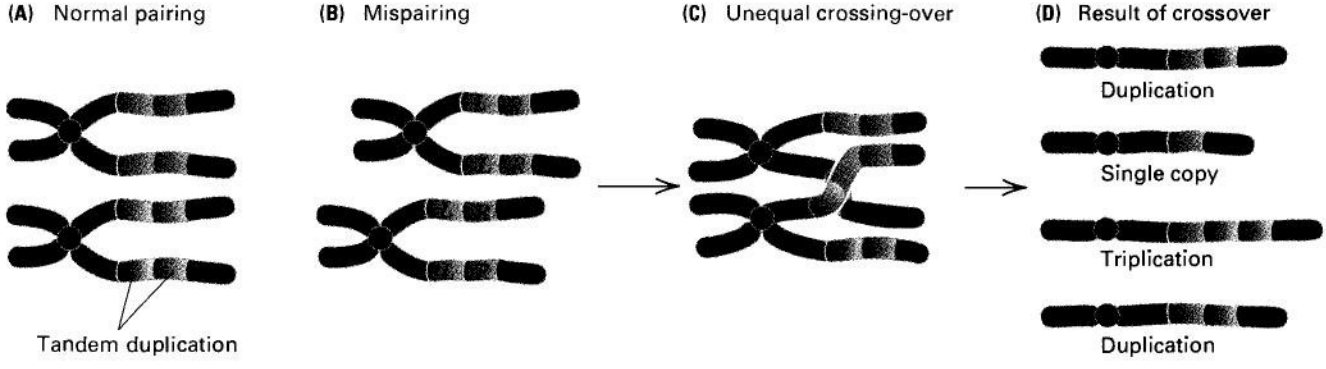


Figure 9.14 An increase in the number of copies of a chromosome segment resulting from unequal crossing-over of tandem duplications (brown). (A) Normal synapsis of chromosomes with a tandem duplication. (B) Mispairing. The right-hand element of the lower chromosome is paired with the left-hand element of the upper chromosome. (C) Crossing-over within the mispaired duplication, which is called unequal crossing-over. (D) The outcome of unequal crossing-over. One product contains a single copy of the duplicated region, another chromosome contains a triplication, and the two strands that do not participate in the crossover retain the duplication.

ability to perceive red. The comparable defects in green perception are called *deuteranopia* and *deuteranomaly,* respectively. Isolation of the red-pigment and green-pigment genes and study of their organization in people with normal and defective color vision have indicated quite clearly how the "-opias" and "-omalies" differ and have also explained why the frequency of color blindness is relatively so high.

The organization of the red-pigment and green-pigment genes in men with normal vision is illustrated in **Figure 9.15,** part A. Unexpectedly, a significant proportion of normal X chromosomes contain two or three green-pigment genes. How these arise by unequal crossing-over is shown in Figure 9.15B. The red-pigment and green-pigment genes pair, and a crossover takes place in the region of homology between the genes. The result is a duplication of the green-pigment gene in one chromosome and a deletion of the green-pigment gene in the other.

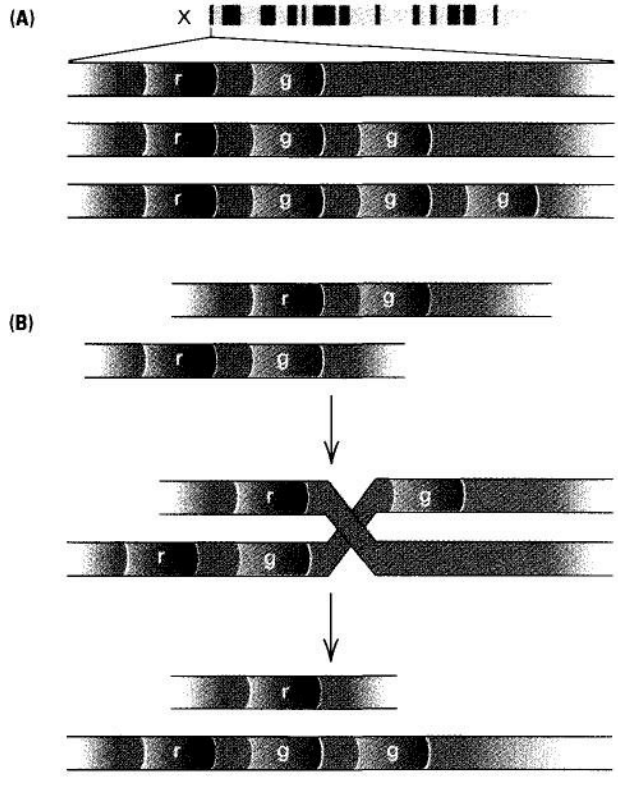


Figure 9.15 (A) Organization of red-pigment and green-pigment genes in normal X chromosomes. Some chromosomes contain one copy of the green-pigment gene, others two, still others three. (B) Origin of multiple green-pigment genes by unequal crossing-over in the region of DNA homology between the genes. Note that one product of unequal crossing-over is a chromosome that contains a red-pigment gene but no green-pigment gene.

The recombinational origins of the defects in color vision are illustrated in **Figure 9.16**. The top chromosome in Figure 9.16A is the result of deletion of the green-pigment gene shown earlier in Figure 9.15B. Males with such an X chromosome have deuteranopia, or "green-blindness." Other types of abnormal pigments result when crossing-over takes place within mispaired red-pigment and green-pigment genes. Crossing-over between the genes yields a **chimeric gene,** which is a composite gene: part of one joined with part of the other. The chimeric gene in Figure 9.16A joins the 5' end of the green-pigment gene with the 3' end of the red-pigment gene. If the

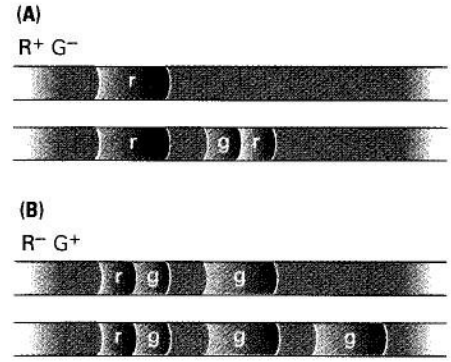


Figure 9.16 Genetic basis of absent or impaired red–green color vision. (A) Defects in green vision result from unequal crossing-over between mispaired red-pigment and green-pigment genes, yielding a green–red chimeric gene. If the green-pigment gene is missing or the chimeric gene is largely "red" in its sequence, deuteranopia is the result. If the chimeric gene is largely "green" in its sequence, deuteranomaly is the result. (B) Defects in red vision result from unequal crossing-over between mispaired red-pigment and green-pigment genes, again yielding a green–red chimeric gene. If the chimeric gene is largely "green," protanopia results; if it is largely "red," protanomaly results. Note that the red gene cannot be eliminated altogether (as the green gene can), because the red gene is at the end of the region of homology between the chromosomes.

crossover point is nearer the 5' end (toward the left in the figure), then the resulting chimeric gene is mostly "red" in sequence, so the chromosome causes deuteranopia, or "green-blindness." However, if the crossover point is nearer the 3' end (toward the right in the figure), then most of the green-pigment gene remains intact, and the chromosome causes deuteranomaly.

Chromosomes associated with defects in red vision are illustrated in Figure 9.16B. The chimeric genes are the reciprocal products of the unequal crossovers that yield defects in green vision. In this case, the chimeric gene consists of the red-pigment gene at the 5' end and the green-pigment gene at the 3' end. If the crossover point is nearer the 5' end, most of the red-pigment gene is replaced with the green-pigment gene. The result is protanopia, or "red-blindness." The same is true of the other chromosome indicated in Figure 9.16B. However, if the crossover point is nearer the 3' end, then most of the red-pigment gene remains intact, and the result is protanomaly.

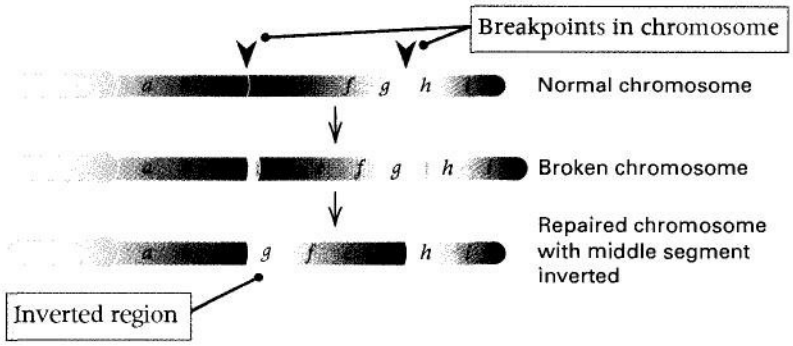


Figure 9.17 Origin of an inversion by reversal of the region between two chromosomal breakpoints.

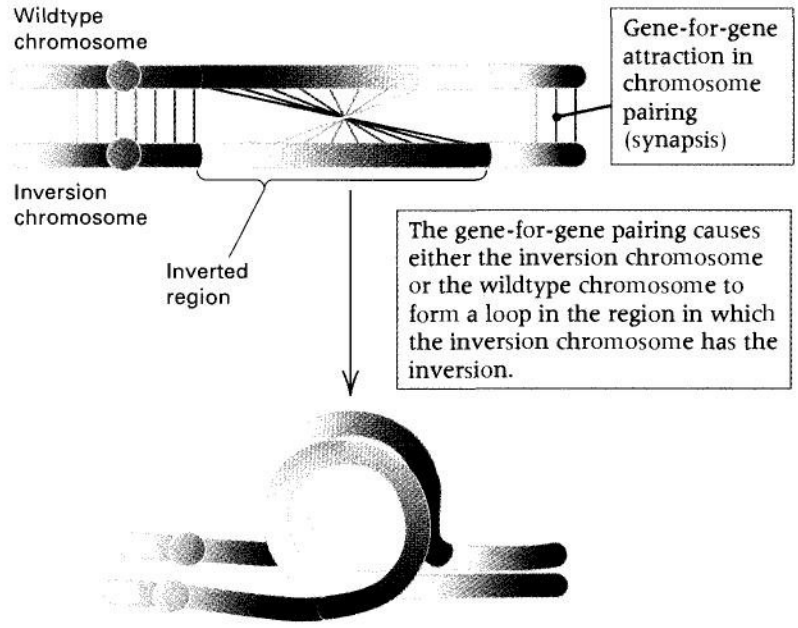


Figure 9.18 In an organism that carries a chromosome that is structurally normal along with a homologous chromosome bearing an inversion, the gene-for-gene attraction between the chromosomes during synapsis causes one of the chromosomes to form into a loop in the region in which the gene order is inverted. In this example, the structurally normal chromosome forms the loop. Only two of the four chromatids are shown.

9.4 Genetics of Chromosomal Inversions

Another important type of chromosome abnormality is an **inversion,** a chromosome in which the linear order of a group of genes is the reverse of the normal order. An example is shown in **Figure 9.17**.

In an organism that is heterozygous for an inversion, one chromosome is structurally normal (wildtype), and the other carries an inversion. These chromosomes pass through mitosis without difficulty, because each chromosome duplicates and its chromatids are separated into the daughter cells without regard to the other chromosome. There is a problem in meiosis, however. The problem is that the chromosomes are attracted gene for gene in the process of synapsis, as shown in **Figure 9.18**. In an inversion heterozygote, in order for gene-for-gene pairing to take place everywhere along the length of the chromosome, one or the other of the chromosomes must twist into a loop in the region in which the gene order is inverted. In Figure 9.18, it is the structurally normal chromosome that is shown as looped, but in other cells it may be the inverted chromosome that is looped. In either case, the loop is called an **inversion loop.**

Paracentric Inversion (Not Including the Centromere)

The loop itself does not create a problem. The looping apparently takes place without difficulty and can be observed through the microscope. As long as there is no crossing-over within the inversion, the homologous chromosomes can separate normally at anaphase I, as illustrated in **Figure 9.19**. On the other hand, when there is crossing-over within the inversion loop, the chromatids involved in the crossing-over become physically joined, and the result is the formation of chromosomes containing large duplications and deletions. The products of the crossing-over can be deduced from **Figure 9.20**, part A, by tracing along the chromatids. The outer chromatids are the ones not participating in the crossover. One of these contains the inverted sequence and the other the normal sequence, as shown in Figure 9.20B. Because of the crossover, the inner chromatids, which did participate in the crossover, are connected.

The inversion shown in Figure 9.20, in which the centromere is not included in the inverted region, is known as a **paracentric inversion,** which means inverted "beside" (*para-*) the centromere. As the figure shows, the products of crossing-over include a dicentric and an acentric chromosome. Neither the dicentric chromosome nor the acentric chromosome can be included in a normal gamete. The acentric chromosome is usually lost because it lacks a centromere and, in any case, has a deletion of the *a* region and a duplication of the *d* region. The dicentric chromosome is also often lost because it is held on the meiotic spindle by the chromatid bridging between the centromeres; in any case, this chromosome is deleted for the *d* region and duplicated for the *a* region. Hence, when there is a crossover in the inversion loop, the only chromatids that can be recovered in the gametes are the chromatids that did not participate in the crossover. One of these carries the inversion, and the other does not. It is for this reason that inversions prevent the recovery of crossover products. In the early years of genetics, before their identity as inversions was discovered, inversion-bearing chromosomes were known as "crossover suppressors."

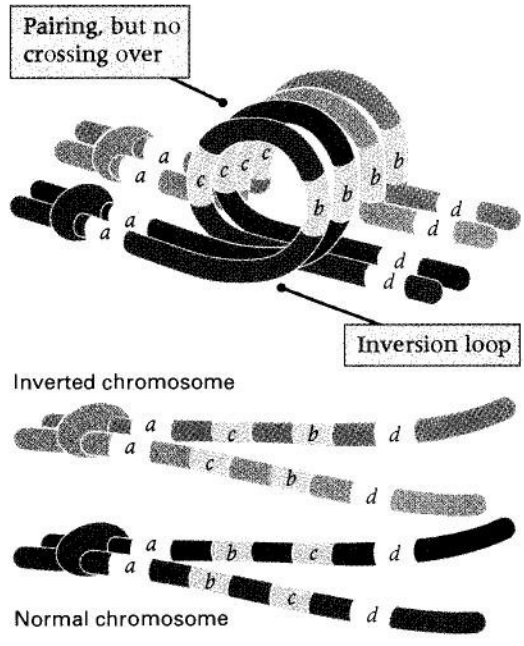


Figure 9.19 In an inversion heterozygote, if there is no crossing over within the inversion loop, then the homologous chromosomes disjoin without problems. Two of the resulting gametes carry the inverted chromosome, and two carry the structurally normal chromosome.

Pericentric Inversion (Including the Centromere)

When the inversion does include the centromere, it is called a **pericentric inversion,** which means "around" (*peri-*) the centromere. Chromatids with duplications and deficiencies are also created by crossing-over within the inversion loop of a pericentric inversion, but in this case the crossover products are monocentric. The situation is illustrated in **Figure 9.21**, part A. The diagram is identical to that in Figure 9.20 except for the position of the centromere. The products of crossing-over can again be deduced by tracing the chromatids. In this case, both products of the crossover are monocentric, but one chromatid carries a duplication of *a* and a deletion of *d*, and the other carries a duplication of *d* and a deletion of *a* (Figure 9.21B). Although either of these chromosomes could be included in a gamete, the duplication and deficiency usually cause inviability. Thus, as with the paracentric inversion, the products of recombination are not recovered, but for a different reason. Among the chromatids not participating in the crossover in Figure 9.21A, one carries the pericentric inversion, and the other has the normal sequence.

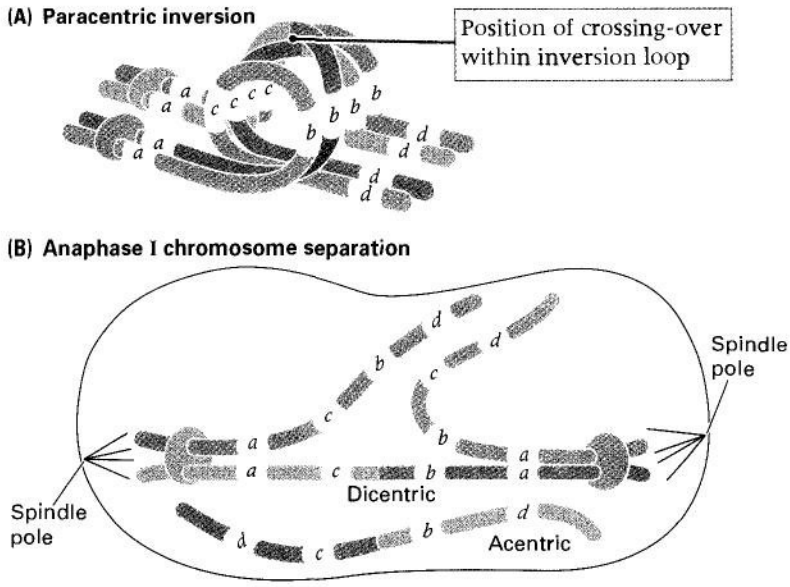


Figure 9.20 (A) Synapsis between homologous chromosomes, one of which contains an inversion. There is a crossover within the inversion loop. (B) Anaphase I configuration resulting from the crossover. Because the centromere is not included in the inverted region, one of the crossover products is a dicentric chromosome, and the reciprocal product is an acentric chromosome. Among the two chromatids not involved in the crossover, one carries the inversion and the other the normal gene sequence.

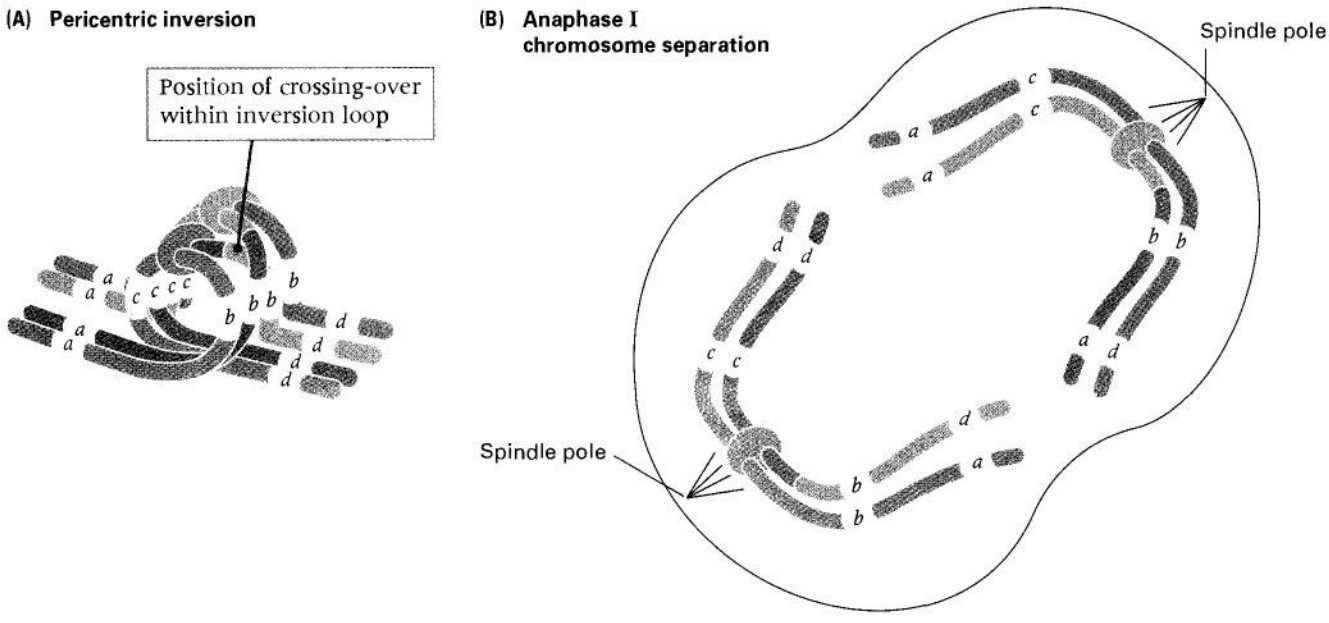


Figure 9.21 (A) Synapsis between homologous chromosomes, one of which carries a pericentric inversion. A crossover within the inversion loop is shown. (B) Anaphase I configuration resulting from the crossover. One of the crossover products is duplicated for *a* and deficient for *d;* the other is duplicated for *d* and deficient for *a*. Among the two chromatids not involved in the crossover, one carries the inversion and the other is normal.

9.5 Chromosomal Translocations

A chromosomal aberration resulting from the interchange of parts between nonhomologous chromosomes is called a **translocation.** In **Figure 9.22**, organism A is homozygous for two pairs of structurally normal chromosomes. Organism B contains one structurally normal pair of chromosomes and another pair of chromosomes that have undergone an interchange of terminal parts. The organism is said to be *heterozygous* for the translocation.

Reciprocal Translocations

The translocation in Figure 9.22 is properly called a **reciprocal translocation** because it consists of two reciprocally interchanged parts. As indicated in Figure 9.22C, an organism can also be homozygous for the translocation if both pairs of homologous chromosomes have undergone an interchange of parts.

An organism that is heterozygous for a reciprocal translocation usually produces only about half as many offspring as normal—a condition called **semisterility.** The reason for the semisterility is production of unbalanced gametes in meiosis. When meiosis takes place in a translocation heterozygote, the normal and translocated chromosomes must undergo synapsis, as shown in **Figure 9.23**. Ordinarily, there would also be chiasmata between nonsister chromatids in the arms of the homologous chromosomes, but these are not shown, as though the translocation were present in an organism with no crossing-over, such as a male *Drosophila*. Segregation from this configuration can take place in any of three ways. In the list that follows, the symbol $1 + 2 \longleftrightarrow 3 + 4$ means that at the first meiotic anaphase, the chromosomes in Figure 9.23 labeled 1 and 2 go to one pole and those labeled 3 and 4 go to the opposite pole. The red numbers indicate the two parts of the reciprocal translocation. The three types of segregation are as follows:

- $1 + 2 \longleftrightarrow 3 + 4$ This mode is called **adjacent-1** or *disjunctional* segregation. Homologous centromeres go to opposite poles, but each normal chromosome goes with one part of the reciprocal translocation. All gametes formed from adjacent-1 segregation have a large duplication and deficiency for the distal part of the translocated chromosomes. (The *distal* part of a chromosome is the part farthest from the centromere.) The pair of gametes originating from the 1 + 2 pole are duplicated for the distal part of the blue chromosome and deficient for the distal part of the red chromosome; the pair of gametes from the 3 + 4 pole have the reciprocal deficiency and duplication.
- $1 + 3 \longleftrightarrow 2 + 4$ This mode is **adjacent-2** or *nondisjunctional* segregation, in which homologous centromeres go to the same pole at anaphase I. In this case, all gametes have a large duplication and deficiency of the proximal part of the translocated chromosome. (The *proximal* part of a chromosome is the part closest to the centromere.) The pair of gametes from the 1 + 3 pole have a duplication of the proximal part of the red chromosome and a deficiency of the proximal part of the blue chromosome; the pair of gametes from the 2 + 4 pole have the reciprocal deficiency and duplication.
- $1 + 4 \longleftrightarrow 2 + 3$ In this type of segregation, which is called **alternate** segregation, the gametes are all balanced (euploid), which means that none has a duplication or deficiency. The gametes from the 1 + 4 pole have both parts of the reciprocal translocation; those from the 2 + 3 pole have both normal chromosomes.

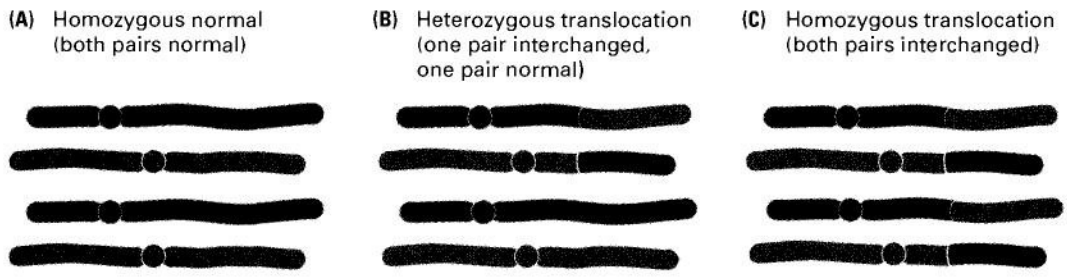


Figure 9.22 (A) Two pairs of nonhomologous chromosomes in a diploid organism. (B) Heterozygous reciprocal translocation, in which two nonhomologous chromosomes (the two at the top) have interchanged terminal segments. (C) Homozygous reciprocal translocation.

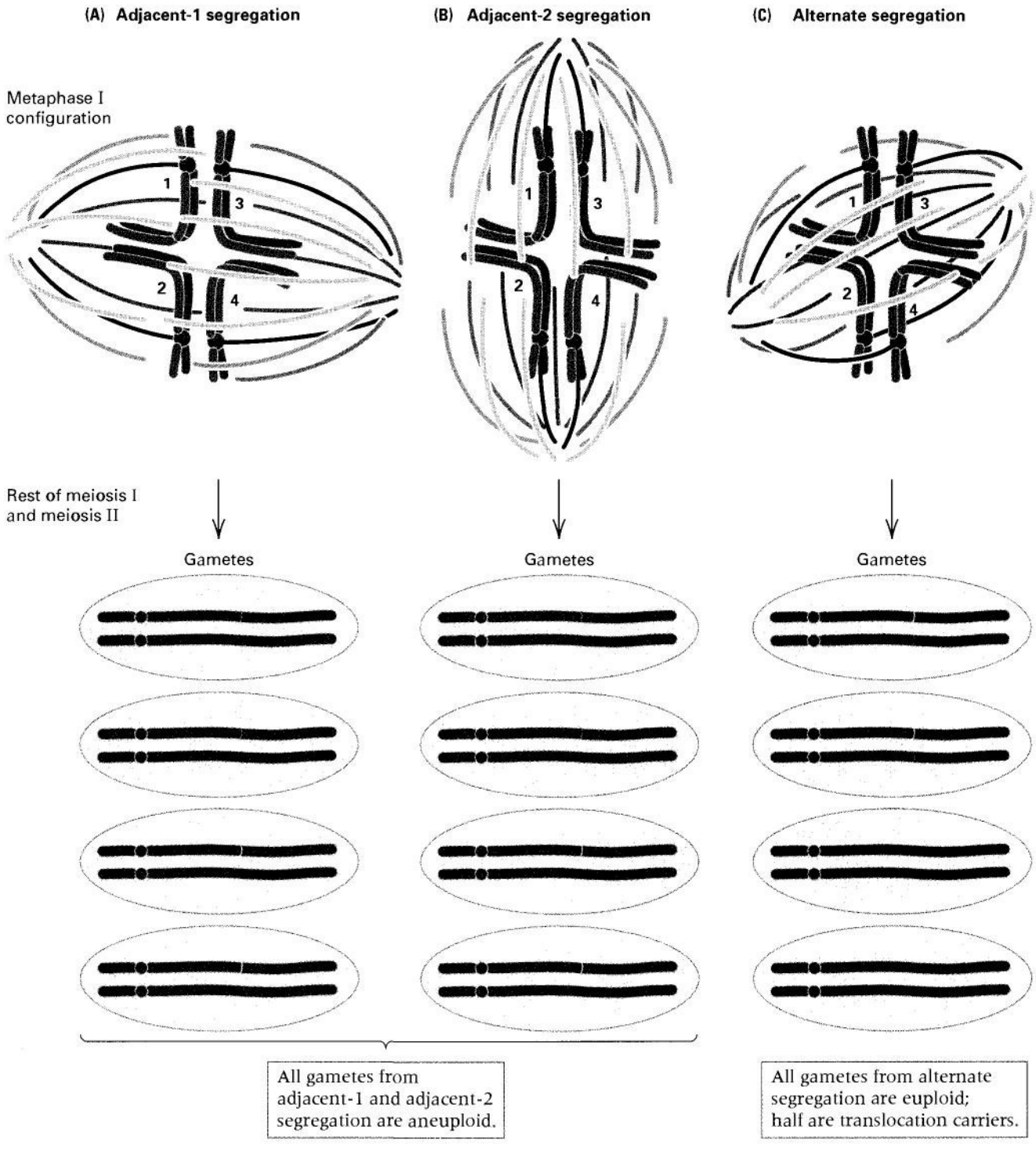


Figure 9.23 Quadrivalents formed in the synapsis of a heterozygous reciprocal translocation. The translocated chromosomes are numbered in red, their normal homologs in black. No chiasmata are shown. (A) Adjacent-1 segregation, in which homologous centromeres separate at anaphase I; all of the resulting gametes have a duplication of one terminal segment and a deficiency of the other. (B) Adjacent-2 segregation, in which homologous centromeres go together at anaphase I; all of the resulting gametes have a duplication of one proximal segment and a deficiency of the other. (C) Alternate segregation, in which half of the gametes receive both parts of the reciprocal translocation and the other half receive both normal chromosomes.

The semisterility of genotypes that are heterozygous for a reciprocal translocation results from lethality due to the duplication and deficiency gametes produced by adjacent-1 and adjacent-2 segregation. The frequency with which these types of segregation take place is strongly influenced by the position of the translocation breakpoints, by the number and distribution of chiasmata in the interstitial region between the centromere and each breakpoint, and by whether the quadrivalent tends to open out into a ring-shaped structure on the metaphase plate. A ring often forms if the breakpoints are in the middle region of the arms and there is no crossing-over in the interstitial regions; in this case, the frequencies of adjacent-1 : adjacent-2 : alternate segregation are approximately 1 : 1 : 2. If there is crossing-over in the interstitial regions, or if one breakpoint is near a telomere, then the orientation of the quadrivalent at metaphase tends to discourage adjacent-2 segregation. In some cases, adjacent-2 segregation is eliminated, and the ratio of adjacent-1 : alternate segregation is approximately 1 : 1. Adjacent-1 segregation is quite common in any event, which means that semisterility is to be expected from virtually all translocation heterozygotes.

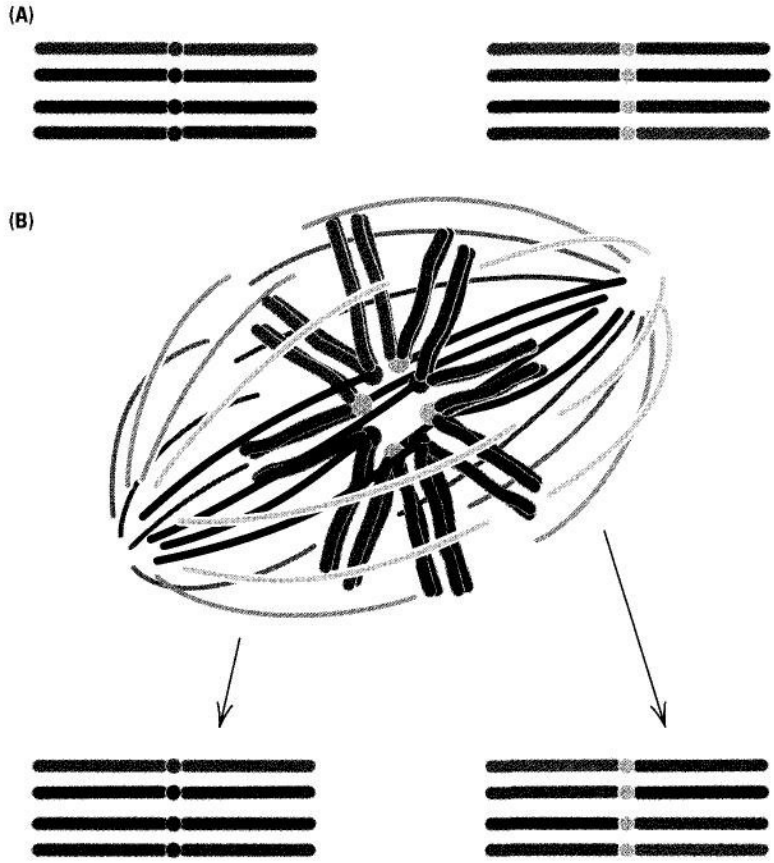


Figure 9.24 (A) Complex translocation heterozygote of the type found in some species of *Oenothera*. The chromosomes at the top left are not rearranged. Those at the top right are connected by a chain of translocations, each chromosome having exchanged an arm with the next chromosome in line. (B) At metaphase I in meiosis, the pairing configuration of the translocation heterozygote is a ring of chromosomes in which each arm is paired with its proper partner; note that each chromosome consists of two chromatids. Alternate segregation from the metaphase ring yields, after the second meiotic division, two types of gametes: those containing all normal chromosomes and those containing all translocated chromosomes.

Translocation semisterility is manifested in different stages in the life histories of plants and animals. Plants have an elaborate gametophyte phase of the life cycle, a haploid phase in which complex metabolic and developmental processes are necessary. In plants, large duplications and deficiencies are usually lethal in the gametophyte stage. Because the gametophyte produces the gametes, in higher plants the semisterility is manifested as pollen or seed lethality. In animals, by contrast, only minimal gene activity is necessary in the gametes, which function in spite of very large duplications and deficiencies. In animals, therefore, the semisterility is usually manifested as zygotic lethality.

Translocation Complexes in *Oenothera*

Certain groups of plants have reciprocal translocations present in natural populations without the semisterility usually expected. Among these are the evening primroses in the genus *Oenothera*, of which there are about 100 wild species native to North America and many cultivated varieties. *Oenothera* escapes the translocation semisterility because segregation is always in the alternate mode. In terms of Figure 9.23, the mode of segregation is always $1 + 4 \longleftrightarrow 2 + 3$. In some species of *Oenothera*, entire sets of chromosomes are interconnected through a chain of reciprocal translocations of the type illustrated in Figure 9.24A. The chromosomes at the top left are all normal; those at the top right are

translocated in such a way that each chromosome has exchanged an arm with the next in line. When the chromosomes in the complex translocation heterozygote undergo synapsis, the result is a ring of chromosomes (Figure 9.24B). The astonishing feature of meiosis in such *Oenothera* heterozygotes is that the segregation is exclusively of the alternate type, so the only gametes formed contain either the entire set of normal chromosomes or the entire set of translocated chromosomes. This is not the end of the surprises. In *Oenothera*, one of the gametic types is inviable in the pollen and the other is inviable in the ovule, so fertilization restores the karyotype of the complex translocation heterozygote!

In species in which translocation heterozygotes exhibit semisterility, the semisterility can be used as the phenotype to map the breakpoint of the translocation just as though it were a normal gene. The mapping procedure can be made clear by means of an example. Translocation *TB-10L1* is a translocation with one breakpoint in the long arm of chromosome 10 in maize, and it results in semisterility when heterozygous. A cross is made between a translocation heterozygote and a genotype homozygous for both *zn1* (*zebra necrotic 1*) and *tp2* (*teopod 2*), and semisterile progeny are testcrossed with *zn1 tp2* homozygotes. (The phenotypes of *zn1* and *tp2* are shown in Figure 5.8.) The parental genotype is therefore *Zn1 Tp2 TB-10L1/zn1 tp2* +, where the + denotes the position of the translocation breakpoint in the homologous chromosome. The progeny phenotypes are as follows:

nonzebrastripe	nonteopod	semisterile	392
nonzebrastripe	nonteopod	fertile	3
nonzebrastripe	teopod	semisterile	42
nonzebrastripe	teopod	fertile	73
zebrastripe	nonteopod	semisterile	83
zebrastripe	nonteopod	fertile	34
zebrastripe	teopod	semisterile	1
zebrastripe	teopod	fertile	372

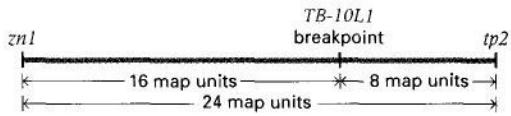
These data are analyzed exactly like the three-point crosses in Section 5.3. The double recombinants are present in the rarest classes of progeny and differ from the parental genotypes in exchanging the gene that is in the middle of the three. The double recombinants are *Zn1 Tp2* + and *zn1 tp2 TB-10L1*, which means that the translocation breakpoint lies between *zn1* and *tp2*. Hence the map distance between *zn1* and the breakpoint is

$$(3 + 73 + 83 + 1)/1000 = 16 \text{ cM}$$

and that between the translocation breakpoint and *tp2* is

$$(3 + 42 + 34 + 1)/1000 = 8 \text{ cM}$$

The genetic map of the region containing the *TB-10L1* translocation breakpoint is



Robertsonian Translocations

A special type of *non*reciprocal translocation is a **Robertsonian translocation,** in which the centromeric regions of two nonhomologous acrocentric chromosomes become fused to form a single centromere (**Figure 9.25**). Robertsonian translocations are important in human genetics, especially as a risk factor to be considered in Down syndrome. When chromosome 21 is one of the acrocentrics in a Robertsonian translocation, the rearrangement leads to a familial type of Down syndrome in which the risk of recurrence is very high. Approximately 3 percent of children with Down syndrome are found to have one parent with such a translocation.

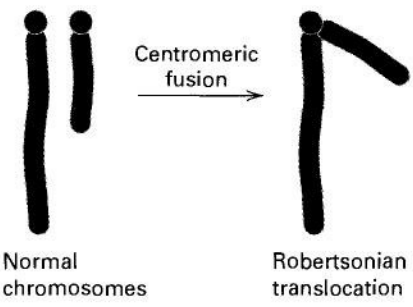


Figure 9.25 Formation of a Robertsonian translocation by fusion of two acrocentric chromosomes in the centromeric region.

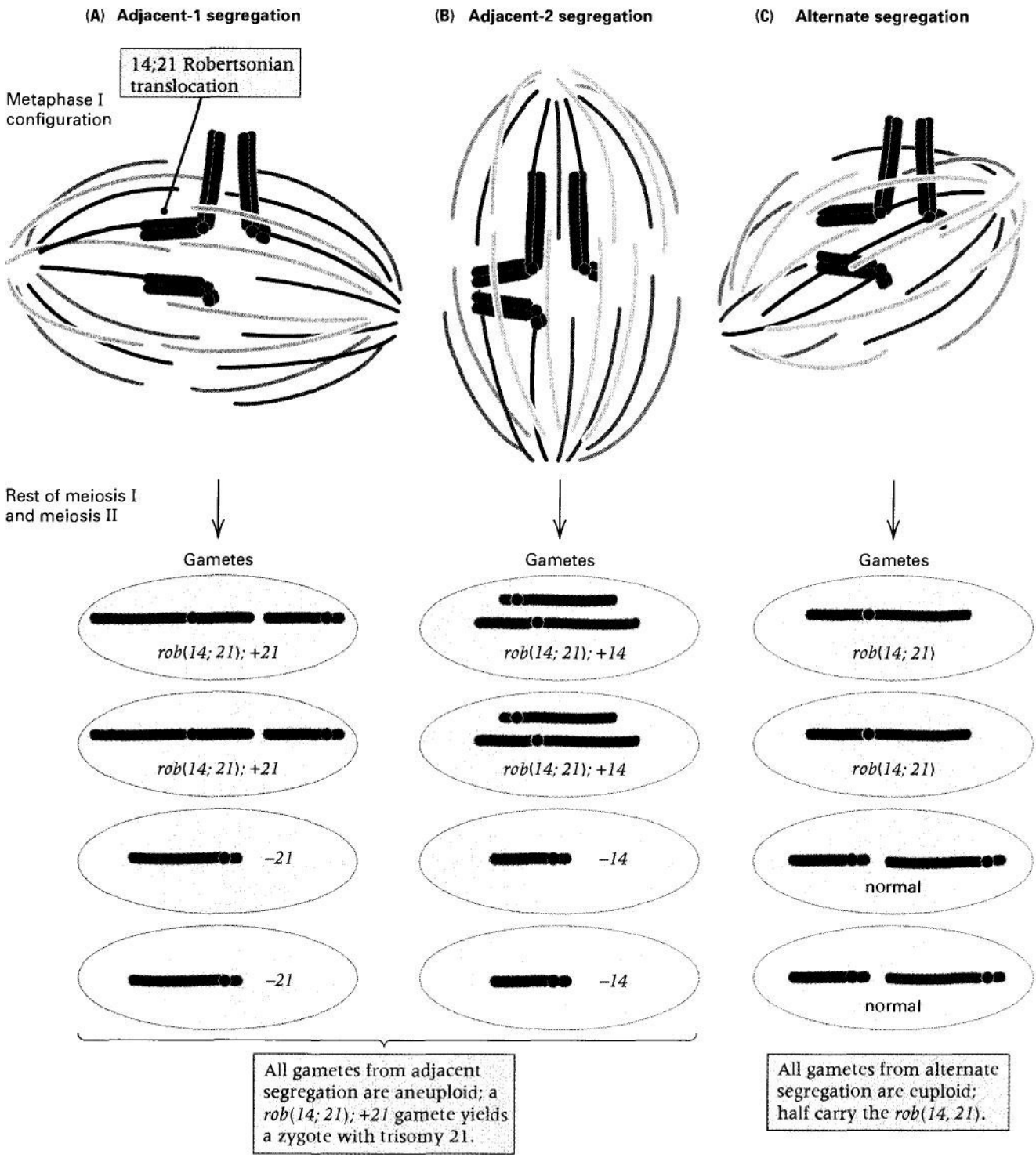


Figure 9.26 Segregation of Robertsonian translocation between chromosomes 14 and 21. (A) Adjacent-1 segregation, in which the gametes formed from the pole at the top have, effectively, an extra copy of chromosome 21. (B) Adjacent-2 segregation. The gametes are either duplicated or deficient for chromosome 14. (C) Alternate segregation. Half of the gametes give rise to phenotypically normal children who are carriers of the Robertsonian translocation.

Translocations and Trisomy 21

Heterozygous carriers of a Robertsonian translocation are phenotypically normal, but a high risk of Down syndrome results from aberrant segregation in meiosis. The possible modes of segregation are shown in **Figure 9.26**. A Robertsonian translocation can join chromosome 21 with any other acrocentric chromosome, but for concreteness we use chromosome 14. The symbol *rob* refers to the translocation, and the + sign or − sign preceding a chromosome number designates an extra copy, or a missing copy, of the entire chromosome.

Among the several possible types of gametes that can arise, one contains a normal chromosome 21 along with the 14/21 Robertsonian translocation (Figure 9.26A). If this aberrant gamete is used in fertilization, then the fetus will contain two copies of the normal chromosome 21 plus the 14/21 translocation. In effect, the fetus contains three copies of chromosome 21 and hence has Down syndrome. The other abnormal gametes that result from adjacent-1 or adjacent-2 segregation either are missing chromosome 21 or chromosome 14 or contain effectively two copies of chromosome 14 (Figure 9.26A and B). If these gametes participate in fertilization, the result is monosomy 21, monosomy 14, or trisomy 14, respectively. The monosomic embryos undergo very early spontaneous abortion; the trisomy-14 fetus undergoes spontaneous abortion later in pregnancy. Hence families with a high risk of translocation Down syndrome also have a high risk of spontaneous abortion due to other chromosome abnormalities. Alternate segregation of a Robertsonian translocation yields gametes that carry either the translocation or both normal chromosomes (Figure 9.26C). Because these gametes derive from reciprocal products of meiosis, the nonaffected children have a risk of 1/2 of carrying the translocation.

9.6 Genomic Position Effects on Gene Expression

Genes near the breakpoints of chromosomal rearrangement become repositioned in the genome and flanked by new neighboring genes. In many cases, the repositioning of a gene affects its level of expression or, in some cases, its ability to function; this is called **position effect.** Such effects have been studied extensively in *Drosophila* and also in yeast. In *Drosophila*, the most common type of position effect results in a mottled (mosaic) phenotype that is observed as interspersed patches of wildtype cells, in which the wildtype gene is expressed, and mutant cells, in which the wildtype gene is inactivated. The phenotype is said to show *variegation*, and the phenomenon is called **position-effect variegation (PEV).** In the older literature, PEV is often referred to as variegated or V-type position effect.

PEV usually results from a chromosome aberration that moves a wildtype gene from a position in euchromatin to a new position in or near heterochromatin (**Figure 9.27**).

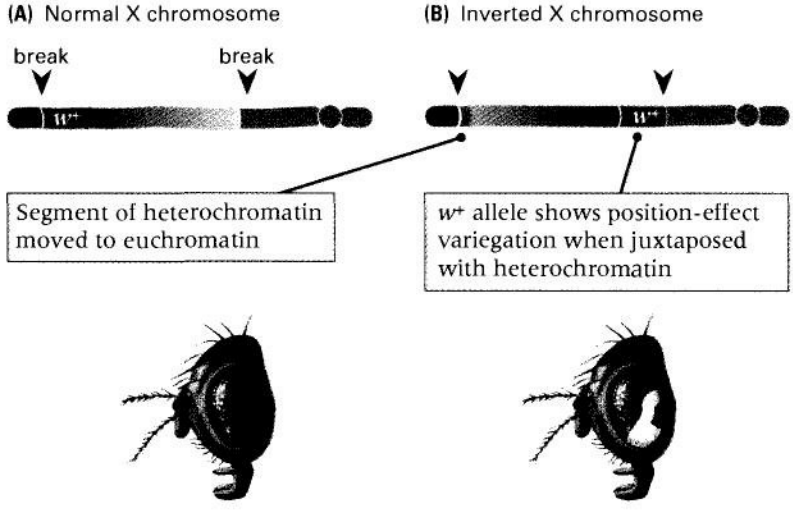


Figure 9.27 Position-effect variegation (PEV) is often observed when an inversion or other chromosome rearrangement repositions a gene normally in euchromatin to a new location in or near heterochromatin. In this example, an inversion in the X chromosome of *Drosophila melanogaster* repositions the wildtype allele of the *white* gene near heterochromatin. PEV of the w^+ allele is observed as mottled red and white eyes.

(Euchromatin and heterochromatin are discussed in Chapter 8.) **Figure 9.28** illustrates some of the patterns of wildtype (red) and mutant (white) facets that are observed in male flies that carry a rearranged X chromosome in which an inversion repositions the wildtype w^+ allele into heterochromatin. The same types of patterns are found in females heterozygous for the rearranged X chromosome and an X chromosome carrying the w allele. The patterns of w^+ expression coincide with the clonal lineages in the eye; that is, all of the red cells in a particular patch derive from a single ancestral cell in the embryo in which the w^+ allele was activated. In contrast with the pattern shown in Figure 9.28, other chromosome rearrangements with PEV yield a salt-and-pepper pattern of mosaicism, which consists of numerous very small patches of wildtype tissue, and still others yield a combination of many small and a few large patches. These patterns imply that gene activation can be very late in development as well as very early.

Although the mechanism of PEV is not understood in detail, it is thought to result from the unusual chromatin structure of heterochromatin interfering with gene activation. The determination of gene expression or nonexpression is thought to take place when the boundary between condensed heterochromatin and euchromatin is established. Where heterochromatin is juxtaposed with euchromatin, the chromatin condensation characteristic of heterochromatin may spread into the adjacent euchromatin, inactivating euchromatic genes in the cell and all of its descendants. A similar inactivation phenomenon takes place in cells of female mammals when euchromatic genes translocated to the X chromosome become heterochromatic and inactive. The length of the euchromatic region that is inactivated ranges from 1 to 50 bands in the *Drosophila* polytene chromosomes, depending on the particular chromosome abnormality. At the molecular level, this range is approximately 20–1000 kb. The term *spreading*, with its implication of smooth continuity, should not be taken literally, however, because in both *Drosophila* and mammals, chromosomal regions affected by PEV are known to contain active genes interspersed with inactive ones. It is likely that heterochromatic regions associate to form a relatively compact "compartment" in the nucleus, and the presence of heterochromatin near a euchromatic gene may silence the gene simply by attracting it into the heterochromatic compartment.

The heterochromatin of different chromosomes differs in its ability to induce variegation of specific genes and in its response to environmental or genetic modifiers. Environmental modifiers include tempera-

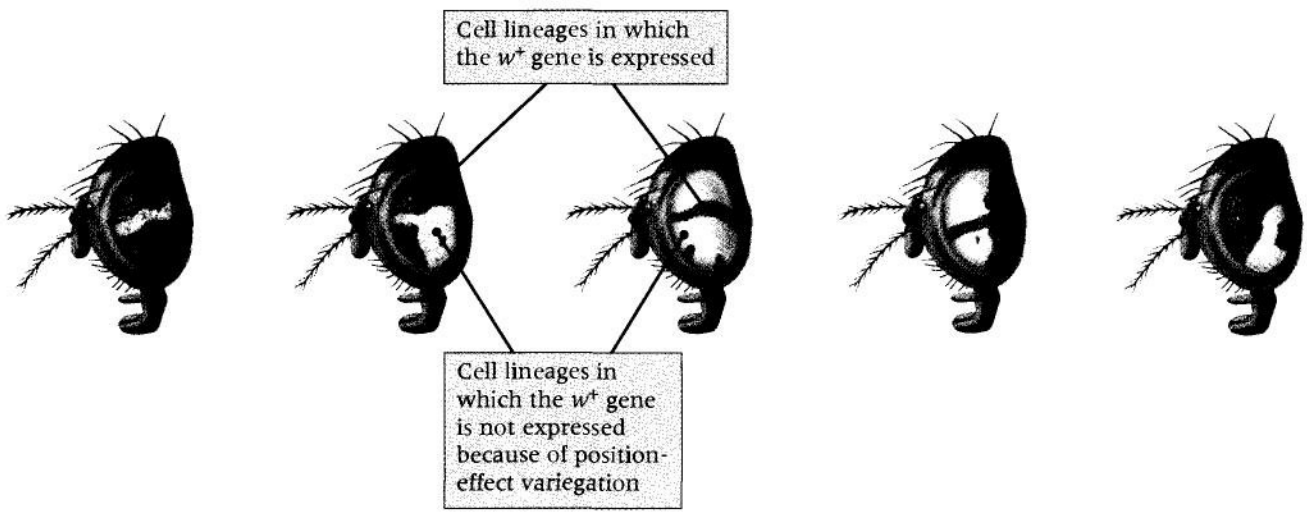


Figure 9.28 Patterns of red and white sectors in the eye of *Drosophila melanogaster* resulting from position-effect variegation. Each group of contiguous facets of the same color derives from a single cell in development. Such large patches of red are atypical. More often, one observes numerous very small patches of red or a mixture of many small and a few large patches.

ture; genetic modifiers include the presence of extra blocks of heterochromatin (such as a Y chromosome). Some mutations that affect chromatin structure also suppress PEV, and certain euchromatic DNA sequences, when repositioned into heterochromatin, act as "insulators" protecting nearby genes by binding with chromatin-associated proteins that prevent inactivation. Apart from position-effect variegation, genes are rarely shut down completely because of their position in the genome. Most genes can be expressed irrespective of neighboring genes, but the level of expression may vary substantially according to position.

9.7 Polyploidy in Plant Evolution

The genus *Chrysanthemum* illustrates **polyploidy,** an important phenomenon, found often in higher plants, in which a species has a genome composed of multiple complete sets of chromosomes. One *Chrysanthemum* species, a diploid species, has 18 chromosomes. A closely related species has 36 chromosomes. However, comparison of chromosome morphology indicates that the 36-chromosome species has two complete sets of the chromosomes found in the 18-chromosome species (**Figure 9.29**). The basic chromosome set in the group, from which all the other genomes are formed, is called the **monoploid** chromosome set. In *Chrysanthemum*, the monoploid chromosome number is 9. The diploid species has two complete copies of the monoploid set, or 18 chromosomes altogether. The 36-chromosome species has four copies of the monoploid set ($4 \times 9 = 36$) and is a **tetraploid.** Other species of *Chrysanthemum* have 54 chromosomes (6×9, constituting the *hexaploid*), 72 chromosomes (8×9, constituting the *octoploid*), and 90 chromosomes (10×9, constituting the *decaploid*).

In meiosis, the chromosomes of all *Chrysanthemum* species synapse normally in pairs to form bivalents (Section 4.3). The 18-chromosome species forms 9 bivalents, the 36-chromosome species forms 18 bivalents, the 54-chromosome species forms 27 bivalents, and so forth. Gametes receive one chromosome from each bivalent, so the number of chromosomes in the gametes of any species is exactly half the number of chromosomes in its somatic cells. The

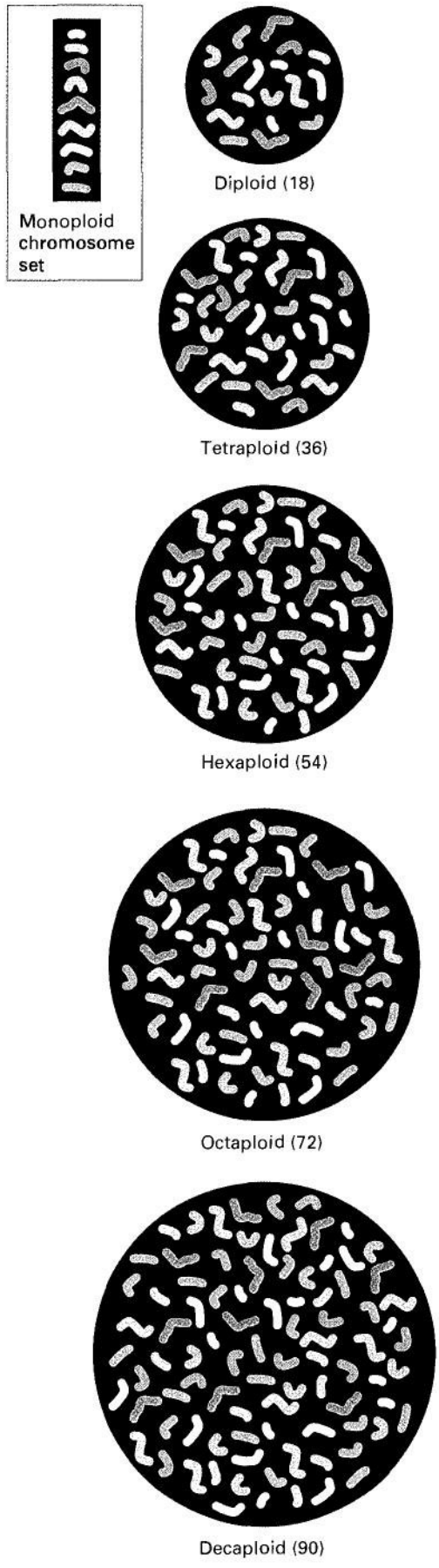


Figure 9.29
Chromosome numbers in diploid and polyploid species of *Chrysanthemum*. Each set of homologous chromosomes is depicted in a different color.

chromosomes present in the gametes of a species constitute the **haploid** set of chromosomes. In the species of *Chrysanthemum* with 90 chromosomes, for example, the haploid chromosome number is 45; in meiosis, 45 bivalents are formed, so each gamete contains 45 chromosomes. When two such gametes come together in fertilization, the complete set of 90 chromosomes in the species is restored. Thus the gametes of a polyploid organism are not always monoploid, as they are in a diploid organism; for example, a tetraploid organism has diploid gametes.

The distinction between the terms *monoploid* and *haploid* is subtle:

- The *monoploid* chromosome set is the basic set of chromosomes that is multiplied in a polyploid series of species, such as *Chrysanthemum*.
- The *haploid* chromosome set is the set of chromosomes present in a gamete, irrespective of the chromosome number in the species.

The potential confusion arises because of diploid organisms, in which the monoploid chromosome set and the haploid chromosome set are the same. Considering the tetraploid helps to clarify the difference: It contains four monoploid chromosome sets, and the haploid gametes are diploid.

Polyploidy is widespread in certain plant groups. Among angiosperms (flowering plants), at least 50 percent of existing species are thought to have originated as some form of polyploid. In some groups the frequency is even higher; for example, in approximately 95 percent of species of pteridophytes (a group that includes ferns, horsetails, and clubmosses), the frequency of polyploid species is approximately 95 percent. Valuable agricultural crops that are polyploid include wheat, oats, cotton, potatoes, bananas, coffee, and sugar cane. Polyploidy often leads to an increase in the size of individual cells, and polyploid plants are often larger and more vigorous than their diploid ancestors; however, there are many exceptions to these generalizations. Polyploidy is rare in vertebrate animals, but it is found in a few groups of invertebrates. One reason why polyploidy is rare in animals is the difficulty in regular segregation of the sex chromosomes. For example, a tetraploid animal with XXXX females and XXYY males would produce XX eggs and XY sperm (if all chromosomes paired to form bivalents), so the progeny would be exclusively XXXY and thus unlike either of the parents.

Polyploid plants found in nature nearly always have an even number of sets of chromosomes, because organisms that have an odd number have low fertility. Organisms with three monoploid sets of chromosomes are known as **triploids.** As far as growth is concerned, a triploid is quite normal because the triploid condition does not interfere with mitosis; in mitosis in triploids (or any other type of polyploid), each chromosome replicates and divides just as in a diploid. However, because each chromosome has more than one pairing partner, chromosome segregation is severely upset in meiosis, and most gametes are defective (Figure 9.11). Unless the organism can perpetuate itself by means of asexual reproduction, it will eventually become extinct. The infertility of triploids is sometimes of commercial benefit. For example, the seeds in commercial bananas are small and edible because the plant is triploid and most of the seeds fail to develop to full size. In Florida and in certain other states, weed control in waterways is aided by the release of weed-eating fish (the grass carp), which do not become overpopulated because the released fish are sterile triploids.

Tetraploid organisms can be produced in several ways. The simplest mechanism is a failure of chromosome separation in either mitosis or meiosis, which instantly doubles the chromosome number. Chromosome doubling through an abortive cell division is called **endoreduplication.** In a plant species that can undergo self-fertilization, endoreduplication creates a new, genetically stable species, because the chromosomes in the tetraploid can pair two by two in meiosis and therefore segregate regularly, each gamete receiving a full diploid set of chromosomes. Self-fertilization of the tetraploid restores the chromosome number, so the tetraploid condition can be perpetuated. The genetics of tetraploid species, and that of other polyploids, is more complex than that of diploid species because the organism carries more than two alleles of any gene. With two alleles in a diploid, only three genotypes are possible: *AA*, *Aa*, and *aa*. In a tetraploid, by contrast, five genotypes are possible: *AAAA*, *AAAa*, *AAaa*, *Aaaa*, and *aaaa*. Among these genotypes,

the middle three represent different types of tetraploid heterozygotes.

An octoploid species (eight sets of chromosomes) can be generated by failure of chromosome separation in mitosis in a tetraploid. If only bivalents form in meiosis, then an octoploid organism can be perpetuated sexually by self-fertilization or through crosses with other octoploids. Furthermore, cross-fertilization between an octoploid and a tetraploid results in a hexaploid (six sets of chromosomes). Repeated episodes of polyploidization and cross-fertilization may ultimately produce an entire polyploid series of closely related organisms that differ in chromosome number, as exemplified in *Chrysanthemum*.

Autopolyploids and Allopolyploids

Chrysanthemum represents a type of polyploidy, known as **autopolyploidy,** in which all chromosomes in the polyploid species derive from a single diploid ancestral species. In many cases of polyploidy, the polyploid species have complete sets of chromosomes from two or more *different* ancestral species. Such polyploids are known as **allopolyploids.** They derive from occasional hybridization between different diploid species when pollen from one species germinates on the stigma of another species and sexually fertilizes the ovule, followed by endoreduplication in the zygote to yield a hybrid plant in which each chromosome has a pairing partner in meiosis. The pollen may be carried to the wrong flower by wind, insects, or other pollinators. **Figure 9.30** illustrates hybridization between species A and B in which endoreduplication leads to the formation of an allopolyploid (in this case, an *allotetraploid*), which carries a complete diploid genome from each of its two ancestral species. The formation of allopolyploids through hybridization and endoreduplication is an

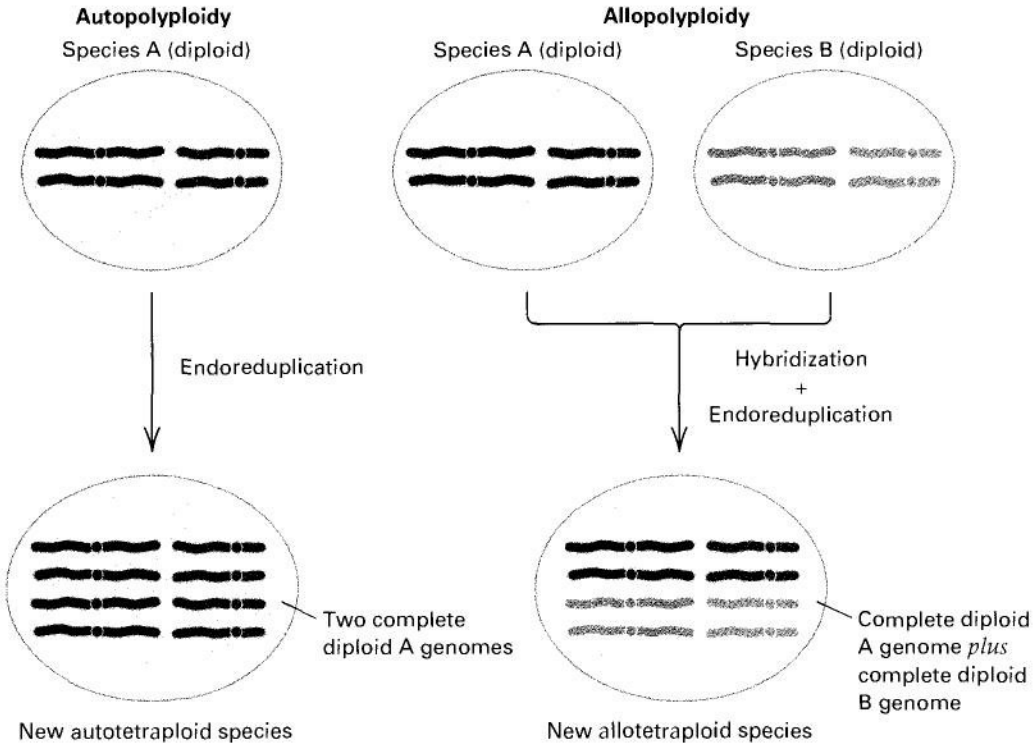


Figure 9.30 Endoreduplication (the doubling of the chromosome complement) plays a key role in the formation of polyploid species. When it takes place in a diploid species, endoreduplication results in the formation of an autotetraploid species. When it takes place in the hybrid formed by cross-fertilization between distinct species, endoreduplication results in the formation of an allotetraploid species that has a complete diploid set of chromosomes from each of the parental species.

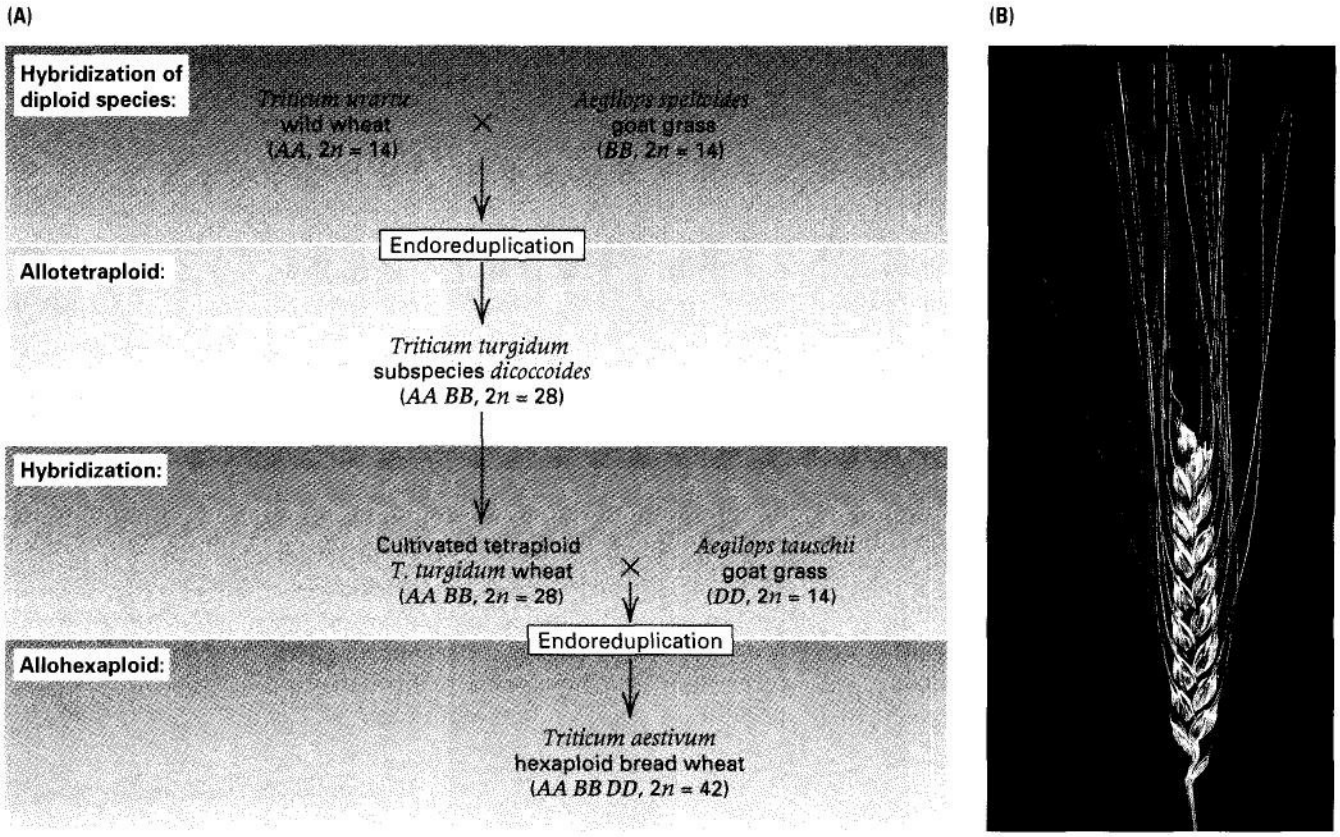


Figure 9.31 (A) Repeated hybridization and endoreduplication in the ancestry of cultivated bread wheat (*Triticum aestivum*), which is an allohexaploid containing complete diploid genomes (*AA, BB, DD*) from three ancestral species. (B) The spike of *T. turgidum* in the photograph is an allotetraploid. The large grains of this species made it attractive to early hunter–gatherer societies in the Middle East. One of the earliest cultivated wheats, *T. turgidum* is the progenitor of commercial macaroni wheat. [Photograph courtesy of Gordon Kimber.]

extremely important process in plant evolution and plant breeding. At least half of all naturally occurring polyploids are allopolyploids. Cultivated wheat provides an excellent example of allopolyploidy. Cultivated bread wheat is a hexaploid with 42 chromosomes constituting a complete diploid genome of 14 chromosomes from each of three ancestral species. The 42-chromosome allopolyploid is thought to have originated by the series of hybridizations and endoreduplications outlined in **Figure 9.31**.

An example of chromosome painting to reveal the ancestral origin of the chromosome sets in an allopolyploid is shown in **Figure 9.32**. The flower is from a variety of crocus called Golden Yellow. Its genome contains seven pairs of chromosomes, which are shown painted in yellow and green. Golden Yellow was thought to be an allopolyploid formed by hybridization of two closely related species followed by endoreduplication of the chromosomes in the hybrid. The putative ancestral species are *Crocus flavus*, which has four pairs of chromosomes, and *Crocus angustifolius*, which has three pairs of chromosomes. To paint the chromosomes of Golden Yellow, DNA from *C. flavus* was isolated and labeled with a fluorescent green dye, and that from *C. angustifolius* was isolated and labeled with a fluorescent yellow dye. The result of the chromosome painting is very clear: Three pairs of chromosomes hybridize with the green-labeled DNA from *C. flavus*, and four pairs of chromosomes hybridize with the yellow-labeled DNA from *C. angustifolius*. This pattern of hybridization strongly supports the hypothesis of the allopolyploid origin of Golden Yellow diagrammed in **Figure 9.33**.

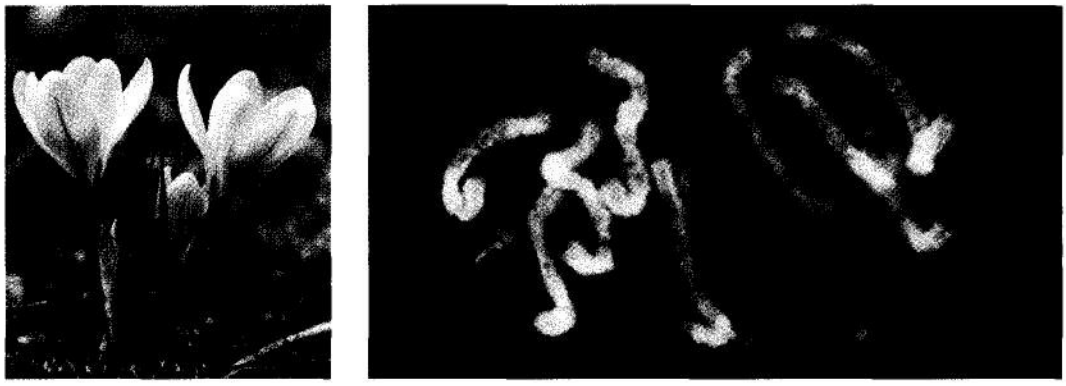


Figure 9.32 Flower of the crocus variety Golden Yellow and chromosome painting that reveals its origin as an allopolyploid. Its seven pairs of chromosomes are shown at the right. The chromosomes in green hybridized with DNA from *C. angustifolius*, which has three pairs of chromosomes, and those in yellow hybridized with DNA from *C. flavus*, which has four pairs of chromosomes. [Courtesy of J. S. Heslop-Harrison, John Innes Centre, Norwich, UK. With permission of the *Annals of Botany*.]

Monoploid Organisms

As noted, the monoploid chromosome set is the set of chromosomes multiplied in polyploid species. An organism is monoploid if it develops from a monoploid cell. Meiosis cannot take place normally in the germ cells of a monoploid, because each chromosome lacks a pairing partner, and hence monoploids are usually sterile. Monoploid organisms are quite rare, but they occur naturally in certain insect species (ants, bees) in which males are derived from unfertilized eggs. These monoploid males are fertile because the gametes are produced by a modified meiosis in which chromosomes do not separate in meiosis I.

Monoploids are important in plant breeding because, in the selection of diploid organisms with desired properties, favorable recessive alleles may be masked by heterozygosity. This problem can be avoided by studying monoploids, provided that their sterility can be overcome. In many plants, the production of monoploids capable of reproducing can be stimulated by conditions that yield aberrant cell divisions. Two techniques make this possible.

With some diploid plants, monoploids can be derived from cells in the anthers (the pollen-bearing structures). Extreme chilling of the anthers causes some of the haploid cells destined to become pollen grains to begin to divide. These cells are monoploid as well as haploid. If the cold-shocked cells are placed on an agar surface containing suitable nutrients and certain plant hormones, then a small dividing mass of cells called an *embryoid* forms. A subsequent change of plant hormones in the growth medium causes the embryoid to form a

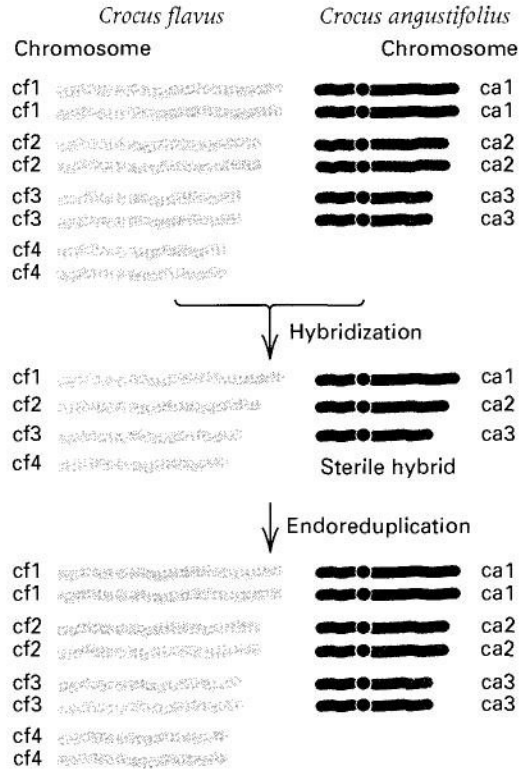


Figure 9.33 Allopolyploidy in the evolutionary origin of the Golden Yellow crocus. The original hybrid between *C. flavus* and *C. angustifolius* produced a sterile monoploid with one copy of each chromosome from each species. Chromosome endoreduplication in the monoploid results in a fertile allotetraploid with a complete diploid genome from each species.

small plant with roots and leaves that can be potted in soil and allowed to grow normally. Because monoploid cells have only a single set of chromosomes, their genotypes can be identified without regard to the dominance or recessiveness of individual alleles. A plant breeder can then select a monoploid plant with the desired traits. In some cases, the desired genes are present in the original diploid plant and are merely sorted out and selected in the monoploids. In other cases, the anthers are treated with mutagenic agents in the hope of producing the desired traits.

When a desired mutation is isolated in a monoploid, it is necessary to convert the monoploid into a homozygous diploid because the monoploid plant is sterile and does not produce seeds. Converting the monoploid into a diploid is possible by treatment of the meristematic tissue (the growing point of a stem or branch) with the substance **colchicine.** This chemical is an inhibitor of the formation of the mitotic spindle. When the treated cells in the monoploid meristem begin mitosis, the chromosomes replicate normally; however, the colchicine blocks metaphase and anaphase, so the result is endoreduplication (doubling of each chromosome in a given cell). Many of the cells are killed by colchicine, but, fortunately for the plant breeder, a few of the monoploid cells are converted into the diploid state (**Figure 9.34**). The colchicine is removed to allow continued cell multiplication, and many of the now-diploid cells multiply to form a small sector of tissue that can be recognized microscopically. If placed on a nutrient–agar surface, this tissue will develop into a complete plant. Such plants, which are completely homozygous, are fertile and produce normal seeds.

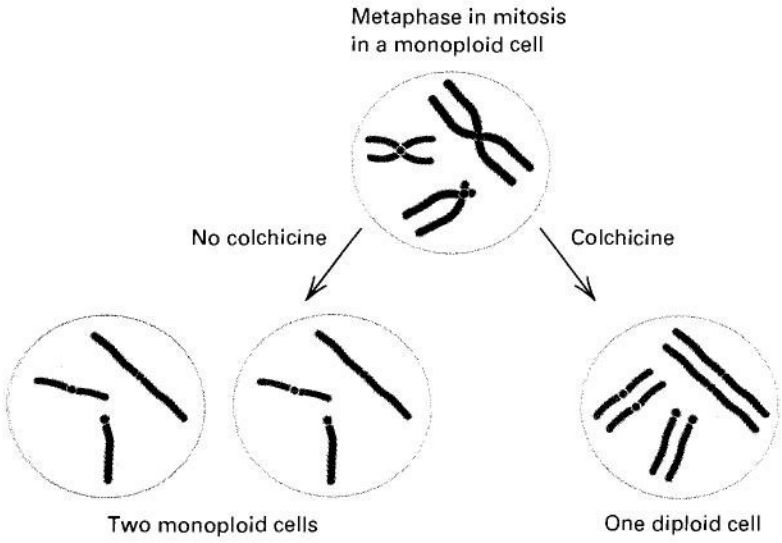


Figure 9.34 Production of a diploid from a monoploid by treatment with colchicine. The colchicine disrupts the spindle and thereby prevents separation of the chromatids after the centromeres (circles) divide.

9.8 Genome Evolution in the Grass Family (Gramineae)

The cereal grasses are our most important crop plants. They include rice, wheat, maize, millet, sugar cane, sorghum, and other cereals. The genomes of grass species vary enormously in size. The smallest, at 400 Mb, is found in rice; the largest, at 16,000 Mb, is found in wheat. Although some of the difference in genome size results from the fact that wheat is an allohexaploid whereas rice is a diploid, a far more important factor is the large variation from one species to the next in types and amount of repetitive DNA sequences present. Each chromosome in wheat contains approximately 25 times as much DNA as each chromosome in rice. For comparison, maize has a genome size of 2500 Mb; it is intermediate in size among the grasses and approximately the same size as the human genome.

In spite of the large variation in chromosome number and genome size in the grass family, there are a number of genetic and physical linkages between single-copy genes that are remarkably conserved amid a background of very rapidly evolving repetitive DNA sequences. In particular, each of the conserved regions can be identified in all the grasses and referred to a similar region in the rice genome. The situation is as depicted in **Figure 9.35**. The rice chromosome pairs are numbered R1 through R12, and the conserved regions within each chromosome are indicated by lowercase letters—for example, R1a and R1b. In each of the other species, each chromosome pair is diagrammed according to the arrangement of segments of the rice genome that contain single-copy DNA sequences homol-

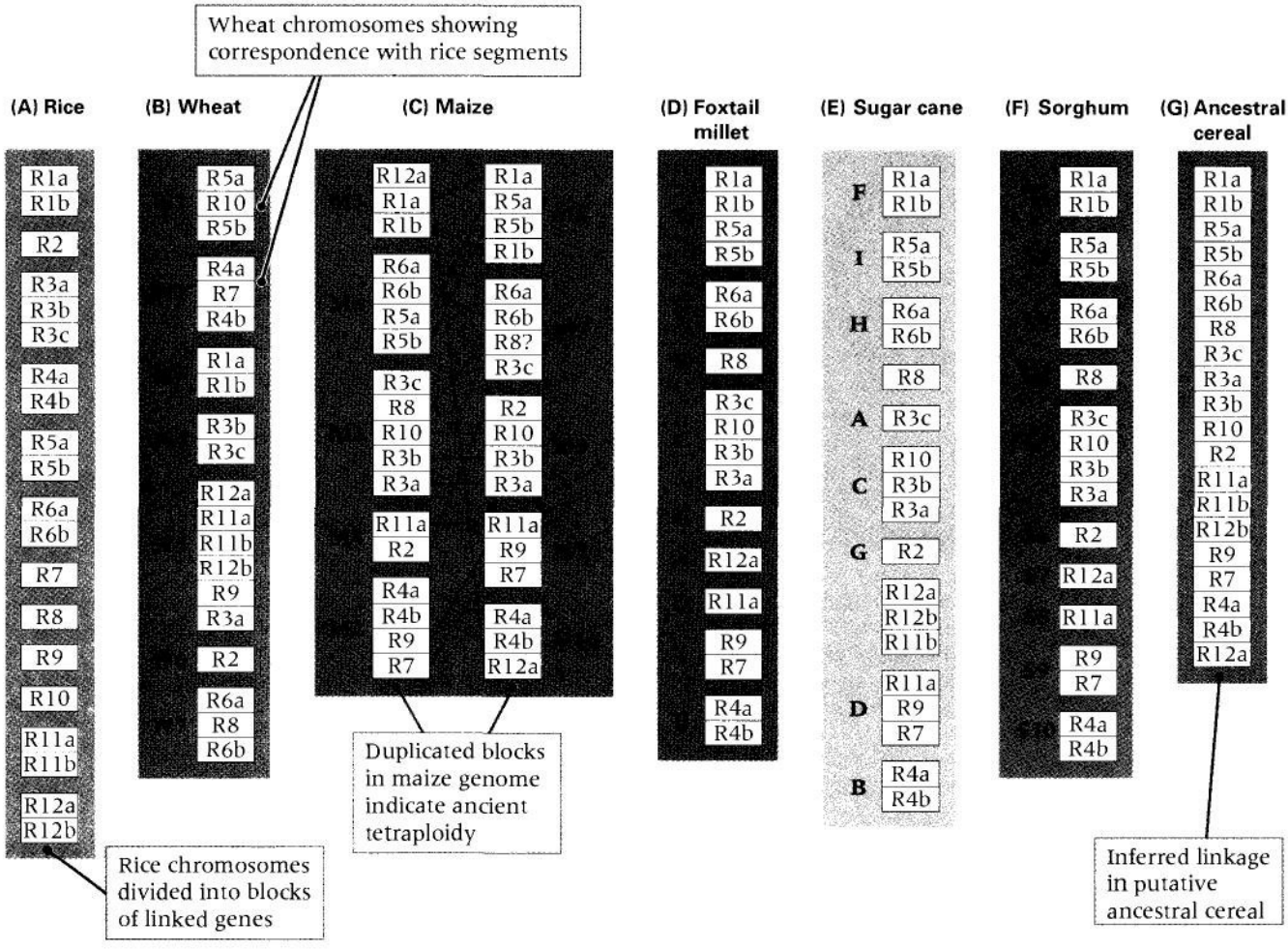
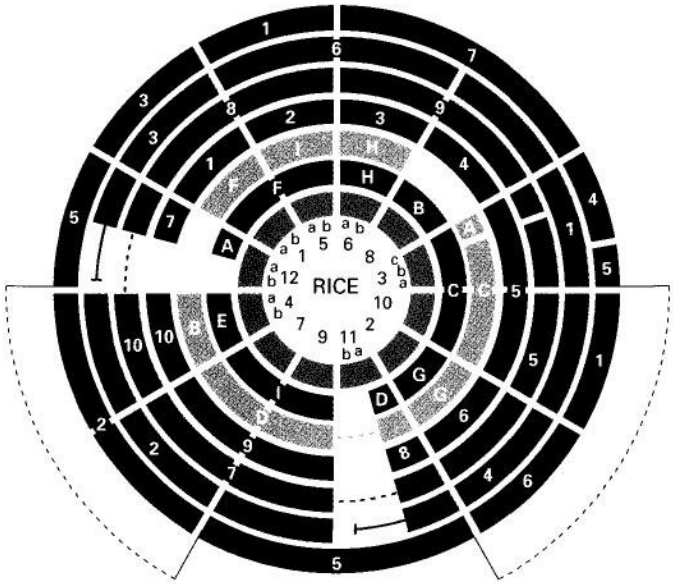


Figure 9.35 Conserved linkages (synteny groups) between the rice genome (A) and that of other grass species: wheat (B), maize (C), foxtail millet (D), sugar cane (E), and sorghum (F). Part (G) depicts the inferred order of segments in a hypothetical ancestral cereal genome consisting of a single chromosome pair. For each extant species, the hypothetical ancestral chromosome can be cleaved at different points to yield groups of blocks corresponding to the arrangement of the segments in the chromosomes of the species. [Courtesy of Graham Moore. From G. Moore, K. M. Devos, Z. Wang, and M. D. Gale. 1995. *Current Opinion Genet. Devel.* 5: 737.]

ogous to those in the corresponding region of the chromosome of the species in question. For example, the wheat monoploid chromosome set is designated W1 through W7. One region of W1 contains single-copy sequences that are homologous to those in rice segment R5a, another contains single-copy sequences that are homologous to those in rice segment R10, and still another contains single-copy sequences that are homologous to those in rice segment R5b. The genomes of the other grass species can be aligned with those of rice as shown. Each of such conserved genetic and physical linkages is called a **synteny group.**

Synteny groups are found in other species comparisons as well. For example, the human and mouse genomes share many synteny groups. These synteny groups are often useful in identifying the mouse homolog of a human gene.

Relative to the synteny groups in Figure 9.35, note that the maize genome has a repetition of segments, indicated by the connecting lines. The relationships confirm what some maize geneticists had long suspected: that maize is a complete, very ancient tetraploid with the complication that the two complete genomes are rearranged relative to each other. Furthermore, most of



Common name	Scientific name	Monoploid genome size (Mb)
Rice	*Oryza sativa*	400
Wheat	*Triticum aestivum*	16,000
Maize	*Zea mays*	2500
Foxtail millet	*Setaria itallica*	500
Sugar cane	*Saccharum officinarum*	3500
Sorghum	*Sorghum vulgare*	1000

Figure 9.36 Circular arrangement of synteny groups in the cereal grasses makes simultaneous comparisons possible. The thin dashed lines indicate connections between blocks of genes. A number of transpositions of genetic segments are not shown. In some cases, a region within a synteny group is inverted; these inversions are not shown. The circular diagram is for convenience only; there is no indication that the ancestral grass chromosome was actually circular. [Courtesy of Graham Moore. From G. Moore, K. M. Devos, Z. Wang, and M. D. Gale. 1995. *Current Opinion Genet. Devel.* 5: 737.]

the larger chromosomes (1, 2, 3, 4 and 6) constitute one of the genomes, and most of the short chromosomes (5, 7, 8, 9, and 10) make up the other.

The synteny groups among the grass genomes are shown in a different format in **Figure 9.36**. In this case, the segments are formed into a circle in the same order in which they are aligned in the hypothetical ancestral chromosome (Figure 9.35G). There is no evidence that the ancestral cereal chromosome was actually a circle. It seems highly unlikely that it was anything other than a normal linear chromosome. However, the value of the circular diagram is that it shows the arrangement of the synteny groups in all of the grass genomes simultaneously, which makes comparisons much easier. Because of the synteny groups in the genomes, homologous genes can often be identified by location alone. For example, both wheat and maize have dwarfing mutations in which the mutant plants are insensitive to the plant hormone gibberellin. A line through the positions of these mutations in the wheat and maize chromosomes passes through rice block 3b, which probably contains the homologous gene. Similarly, rice block 4 contains a gene for *liguless* that aligns with similar mutations in the chromosomes of barley and maize, which indicates that the mutations are almost certainly in homologous genes in all three species. Determining such relationships between genes on the basis of position in the physical maps affords an important method of positional cloning in all species in the grass family.

Chapter Summary

The normal human chromosome complement consists of 22 pairs of autosomes, assigned numbers 1 through 22 from longest to shortest, and one pair of sex chromosomes (XX in females and XY in males). Fetuses that contain an abnormal number of autosomes usually either fail to complete normal embryonic development or die shortly after birth, though people with Down syndrome (trisomy 21) sometimes survive for several decades. Persons with excess sex chromosomes survive, because the Y chromosome contains relatively few genes other than the

master sex-controller *SRY*, and because only one X chromosome is genetically active in the cells of females (dosage compensation through the single-active-X principle). The mosaic orange and black pattern of the female "calico" cat results from X inactivation, because these alternative coat color alleles are X-linked in cats.

Most structural abnormalities in chromosomes are duplications, deletions, inversions, or translocations. In a duplication, there are two copies of a chromosomal segment. In a deletion, a chromosomal segment is missing. An organism can often tolerate an imbalance of gene dosage resulting from small duplications or deletions, but large duplications or deletions are nearly always harmful. Chromosome rearrangements may affect gene expression through position effects. Although most genes can be expressed whatever their location in the genome, the level of expression may vary substantially according to position. A major effect of gene location on expression is observed in position-effect variegation (PEV), in which a wildtype allele repositioned in or near heterochromatin is unable to be expressed in a fraction of the cell lineages.

A chromosome that contains an inversion has a group of adjacent genes in reverse of the normal order. Expression of the genes is usually unaltered, so inversions rarely affect viability. However, crossing-over between an inverted chromosome and its noninverted homolog in meiosis yields abnormal chromatids. Crossing-over within a heterozygous paracentric inversion yields an acentric chromosome and a dicentric chromosome, both of which also have a duplication and a deficiency. Crossing-over within a heterozygous pericentric inversion yields monocentric chromosomes, but both have a duplication and a deficiency.

Two nonhomologous chromosomes that have undergone an exchange of parts constitute a reciprocal translocation. Organisms that contain a reciprocal translocation, as well as the normal homologous chromosomes of the translocation, produce fewer offspring (this is called semisterility) because of abnormal segregation of the chromosomes in meiosis. The semisterility is caused by the aneuploid gametes produced in adjacent-1 and adjacent-2 segregation. Alternate segregation yields equal numbers of normal and translocation-bearing gametes. In genetic crosses, the semisterility of a heterozygous translocation behaves like a dominant genetic marker that can be mapped like any other gene; however, what is actually mapped is the breakpoint of the translocation.

A translocation may also be nonreciprocal. A Robertsonian translocation is a type of nonreciprocal translocation in which the long arms of two acrocentric chromosomes are attached to a common centromere. In human beings, Robertsonian translocations that include chromosome 21 account for about 3 percent of all cases of Down syndrome, and the parents have a high risk of recurrence of Down syndrome in a subsequent child.

Polyploid organisms contain more than two complete sets of chromosomes. Polyploidy is widespread among higher plants and uncommon otherwise. Some 30–35 percent of all species of flowering plants are thought to have originated as some form of polyploid. An autopolyploid organism contains multiple sets of chromosomes from a single ancestral species; allopolyploid organisms contain complete sets of chromosomes from two or more ancestral species. The usual cytological mechanism that creates polyploidy is endoreduplication, in which a cell undergoes chromosome replication but not chromosome separation or cytoplasmic division.

Key Terms

acentric chromosome
acrocentric chromosome
adjacent-1 segregation
adjacent-2 segregation
allopolyploid
alternate segregation
amniocentesis
aneuploid
autopolyploidy
Barr body
chimeric gene
chromosome painting
colchicine
deficiency
deletion
dicentric chromosome
dosage compensation
double-Y
Down syndrome
duplication
endoreduplication
euploid
G-bands
haploid
inversion
inversion loop
karyotype
Klinefelter syndrome
metacentric chromosome
monoploid
mosaic
paracentric inversion
pericentric inversion
position effect
position-effect variegation (PEV)
polyploidy
pseudoautosomal region
reciprocal translocation
red–green color blindness
Robertsonian translocation
semisterility
single-active-X principle
submetacentric chromosome
synteny group
tandem duplication
testis-determining factor (TDF)
tetraploid
translocation
triploid
trisomy
trisomy-X
trivalent
Turner syndrome
uncovering of recessive allele
unequal crossing-over
univalent
X inactivation

Review the Basics

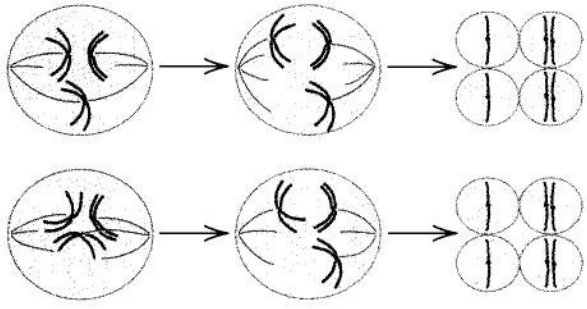
- Why does an acentric chromosome fail to align at the metaphase plate and fail to move to one of the poles in anaphase? What type of abnormal chromosome forms a chromosomal "bridge" between the daughter cells at anaphase?
- What does the term *dosage compensation* mean with reference to X-linked genes? In mammals, what is the mechanism of dosage compensation?
- What is meant by the term *pseudoautosomal inheritance?* Where in the genome are genes that show pseudoautosomal inheritance located?
- In human reproduction, what is the fate of most zygotes that are monosomic or trisomic for an autosome?
- What is Down syndrome, and how is the incidence of this condition related to the age of the mother?
- Explain how unequal crossing-over in the region of the red and green opsin genes is associated with red–green color blindness.
- What does it mean to say that a deletion "uncovers" a recessive mutation?
- Why do most chromosome rearrangements pass through mitosis without upsetting the process?
- Distinguish between a paracentric and a pericentric inversion. What types of chromosomal abnormalities in meiosis are associated with each of these? How are these abnormalities related to crossing-over?
- What types of chromosomally abnormal gametes are associated with a heterozygous reciprocal translocation?
- Distinguish between autopolyploidy and allopolyploidy. Which type of polyploidy would seem to favor the formation of bivalents in meiosis?
- Why do most naturally occurring polyploid species have an even-number multiple of the monoploid chromosome set?
- Why would the presence of sex chromosomes be a hindrance to the evolution of a series of related species with different levels of polyploidy?

Guide to Problem Solving

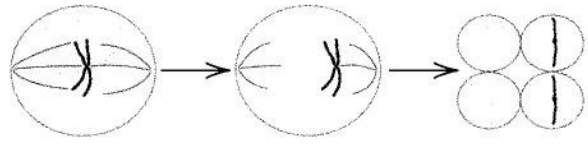
Problem 1 Indicate what types of gametes would be expected, and in what proportions, from an organism that is:
(a) Trisomic for a single chromosome.
(b) Monosomic for a single chromosome.

Answer

(a) In a trisomic organism, either two of the chromosomes will form a pair and the other remain unpaired, or all three chromosomes will come together, each chromosome pairing along part of its length with both of the others. These situations are shown in the illustration. The anaphase I segregation determines the chromosomes included in the gametes. In either type of pairing, two gametes receive two copies of the chromosome and two gametes receive one copy. Therefore, trisomic segregation leads to disomic or monosomic gametes in an expected ratio of 1 : 1.



(b) In a monosomic organism, the monosomic chromosome must remain unpaired. At anaphase I, it goes to one pole or the other. The end result is two gametes containing no copies of the chromosome and two gametes containing one copy. Therefore, monosomic segregation leads to nullisomic (missing the chromosome) gametes or monosomic gametes in an expected ratio of 1 : 1.

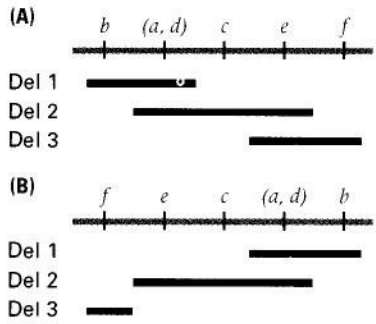


Problem 2 Genes *a*, *b*, *c*, *d*, *e*, and *f* are closely linked in a chromosome, but their order is unknown. Three deletions in the region are found to uncover recessive alleles of the genes as follows:

Deletion 1 uncovers *a*, *b*, and *d*.
Deletion 2 uncovers *a*, *d*, *c*, and *e*.
Deletion 3 uncovers *e* and *f*.

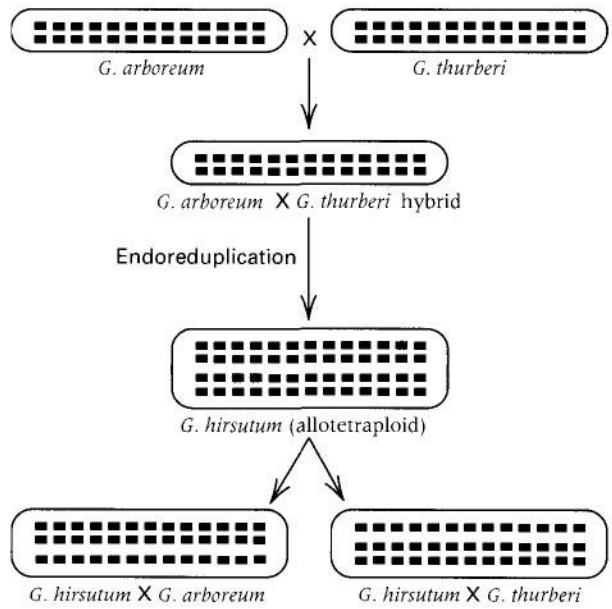
What is the order of the genes? In this problem, you will see that there is enough information to order most, but not all, of the genes. Suggest what experiments you might carry out to complete the ordering.

Answer Problems of this sort are worked by noting that genes uncovered by a single deletion must be contiguous. The gene order is deduced from the overlaps between the deletions. The overlaps are *a* and *d* between the first and second deletions and *e* between the second and third. The gene order (as far as can be determined from these data) is diagrammed in part A of the accompanying illustration. The deletions are shown in red. Gene *b* is at the far left, then *a* and *d* (in unknown order), then *c*, *e*, and *f*. (Gene *c* must be to the left of *e*, because otherwise *c* would be uncovered by deletion 3.) Part B is a completely equivalent map with gene *b* at the right. The ordering can be completed with a three-point cross between *b*, *a*, and *d* or between *a*, *d*, and *c* or by examining additional deletions. Any deletion that uncovers either *a* or *d* (but not both) plus at least one other marker on either side would provide the information to complete the ordering.



Problem 3 Asiatic wild cotton *(Gossypium arboreum)* has 13 pairs of chromosomes, and an American wild species *(G. thurberi)* also has 13 pairs of chromosomes. Interspecific crosses between the species are sterile because of highly irregular chromosome pairing in meiosis. The American cultivated cotton *(G. hirsutum)* has 26 pairs of chromosomes and is fully fertile. Crosses between *G. arboreum* and *G. hirsutum* and crosses between *G. thurberi* and *G. hirsutum* produce plants that, in meiosis, exhibit 13 pairs of chromosomes (bivalents) and 13 chromosomes with no pairing partner (univalents). What do these cytological data tell us about the genetic origin of present-day American cultivated cotton?

Answer It is helpful to draw a diagram showing the 13 paired chromosomes in *G. arboreum* and *G. thurberi*. The fact that the cross between these species yields a hybrid in which the chromosomes do not pair normally indicates that the chromosomes of the two parental species are quite different genetically. One possibility for the origin of *G. hirsutum* is that it arose by endoreduplication in such a hybrid, which would account for the 26 chromosome pairs in *G. hirsutum*. The backcross results are also consistent with this hypothesis, because each yields a hybrid with 13 bivalents and 13 univalents. These results are illustrated in the accompanying diagram. The data therefore suggest that present-day American cultivated cotton originally arose as an allotetraploid from a hybrid between a wild species related to *G. arboreum* and another wild species related to *G. thurberi*.



Analysis and Applications

9.1 A newborn is found to have 47 chromosomes. What are the most likely karyotypes?

9.2 A newborn is found to have 45 chromosomes. What are the most likely karyotypes?

9.3 The vast majority of progeny from either 47,XXX females or 47,XYY males are karyotypically normal 46,XX or 46,XY. Is this finding the one expected?

9.4 How many Barr bodies would be present in white blood cells of an individual of karyotype 49,XXXYY, and what do these represent?

9.5 How many Barr bodies would be present in each of the following human conditions?

(a) Klinefelter syndrome
(b) Turner syndrome
(c) Down syndrome
(d) XYY
(e) XXX

9.6 A spontaneously aborted human fetus was found to have the karyotype 92,XXYY. What might have happened to the chromosomes in the zygote to result in this karyotype?

9.7 A spontaneously aborted human fetus is found to have 45 chromosomes. What is the most probable karyotype? Had the fetus survived, what genetic disorder would it have had?

9.8 A recessive mutation in the human genome results in a condition called anhidrotic ectodermal dysplasia, which is associated with an absence of sweat glands. The condition can be detected by studies of the electrical conductivity of the skin, because skin without sweat glands has a lower electrical conductivity (higher resistance) than normal skin. In kinships in which the recessive allele is segregating, affected males are found to show low conductance uniformly across their skin surface, as do affected females. However, many females show a mosaic pattern with normal conductance in some patches of skin and low conductance in others. The pattern of tissue lacking sweat glands is different for each mosaic female examined. How could this pattern of gene expression be explained?

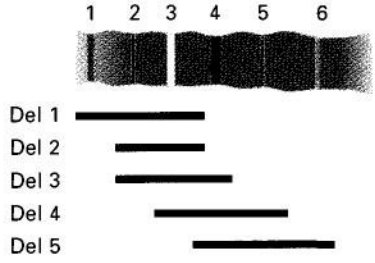
9.9 Color blindness in human beings is an X-linked trait. A man who is color blind has a 45,X (Turner syndrome) daughter who is also color blind. Did the nondisjunction that led to the 45,X child occur in the mother or the father? How can you tell?

9.10 A phenotypically normal woman has a child with Down syndrome. The woman is found to have 45 chromosomes. What kind of chromosome abnormality can account for these observations? How many chromosomes does the affected child have? How does this differ from the usual chromosome number and karyotype of a child with Down syndrome?

9.11 A chromosome has the gene sequence *A B C D E F G*. What is the sequence following a *C*-through-*E* inversion? Following a *C*-through-*E* deletion? Two chromosomes with the sequences *A B C D E F G* and *M N O P Q R S T U V* undergo a reciprocal translocation after breaks in *E*–*F* and *S*–*T*. What are the possible products? Which products are genetically stable?

9.12 Recessive genes *a*, *b*, *c*, *d*, *e*, and *f* are closely linked in a chromosome, but their order is unknown. Three deletions in the region are examined. One deletion uncovers *a*, *d*, and *e*; another uncovers *c*, *d*, and *f*; and the third uncovers *b* and *c*. What is the order of the genes?

9.13 Six bands in a salivary gland chromosome of *Drosophila* are shown in the accompanying figure, along with the extent of five deletions (Del1–Del5).

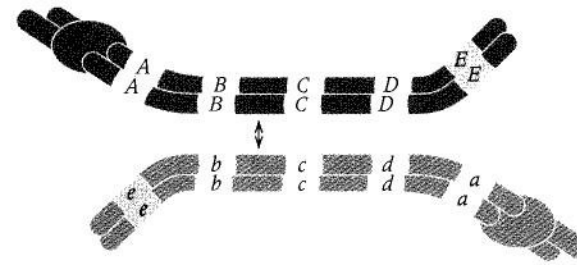


Recessive alleles *a*, *b*, *c*, *d*, *e*, and *f* are known to be in the region, but their order is unknown. When the deletions are heterozygous with each allele, the following results are obtained:

	a	*b*	*c*	*d*	*e*	*f*
Del 1	–	–	–	+	+	+
Del 2	–	+	–	+	+	+
Del 3	–	+	–	+	–	+
Del 4	+	+	–	–	–	+
Del 5	+	+	+	–	–	–

In this table, the – means that the deletion is missing the corresponding wildtype allele (the deletion uncovers the recessive allele) and + means that the corresponding wildtype allele is still present. Use these data to infer the position of each gene relative to the salivary gland chromosome bands.

9.14 The accompanying illustration is a simplified version of pairing between a chromosome with an inversion and its normal homolog, which is extremely useful when the question of interest concerns the consequences of crossing over within the inversion loop. In the simplified version, only the inverted region is shown paired, the normal regions unpaired. In the case of a paracentric inversion, this means that the centromeres are on opposite sides. (In real cells, of course, the inverted region forms a loop.) In this diagram it is easy to see that a crossover at the position of the arrow results in a dicentric chromosome and an acentric fragment (in addition to the normal and inverted chromosomes that do not participate in the crossover). Deduce the consequence of a two-strand double crossover when the first one takes place as shown and the second one takes place between the genes *C* and *D*.



9.15 A strain of corn that has been maintained by self-fertilization for many generations, when crossed with a normal strain, produces an F_1 in which many meiotic cells have dicentric chromatids and acentrics, and the amount of recombination in chromosome 6 is greatly reduced. However, in the original strain, no dicentrics or acentrics are found. What kind of abnormality in chromosome structure can explain these results?

9.16 Four strains of *Drosophila melanogaster* are isolated from different localities. The banding patterns of a particular region of salivary chromosome 2 have the following configurations (each letter denotes a band).

(a) a b f e d c g h i j
(b) a b c d e f g h i j
(c) a b f e h g i d c j
(d) a b f e h g c d i j

Assuming that part **(c)** is the ancestral sequence, deduce the evolutionary ancestry of the other chromosomes.

9.17 Two species of Australian grasshoppers coexist side by side. In meiosis, each has 8 pairs of chromosomes. When the species are crossed, the chromosomes in the hybrid form 6 pairs of chromosomes and one group of 4. What alteration in chromosome structure could account for these results?

9.18 Why are translocation heterozygotes semisterile? Why are translocation homozygotes fully fertile? If a translocation homozygote is crossed with an individual that has normal chromosomes, what fraction of the F_1 is expected to be semisterile?

9.19 From how many possible pathways of hybridization and endoreduplication could the diploid species with genomes *AA*, *BB*, and *CC* yield an allohexaploid with the genome *AABBCC*? Each of *A*, *B*, and *C* stands for one monoploid set of chromosomes. (One such pathway, for example, is $(AA \times BB) \times CC$, which means that *A* and *B* underwent hybridization and reduplication to yield the allotetraploid *AABB*, which then underwent hybridization and endoreduplication with *CC* to yield the allohexaploid *AABBCC*.)

9.20 From how many possible pathways of hybridization and endoreduplication could the diploid species with genomes *AA*, *BB*, *CC*, and *DD* yield an allooctaploid with the genome *AABBCCDD*? Each of *A*, *B*, *C*, and *D* stands for one monoploid set of chromosomes.

9.21 What types of hybridization and endoreduplication could account for a polyploid species with the genome composition *AABBBBCC*, where the diploid ancestors have genome compositions *AA*, *BB*, and *CC*, respectively? (This type of polyploidy is known as *segmental allopolyploidy*.)

9.22 An autopolyploid series similar to *Chrysanthemum* consists of five species. The basic monoploid chromosome number in the group is 5. What chromosome numbers would be expected among the species?

9.23 A plant species S coexists with two related species A and B. All species are interfertile. In meiosis, S forms 26 bivalents, and A and B form 14 and 12 bivalents, respectively. Hybrids between S and A form 14 bivalents and 12 univalents, and hybrids between S and B form 12 bivalents and 14 univalents. Suggest a probable evolutionary origin of species S.

9.24 Curly wings (*Cy*) is a dominant mutation in the second chromosome of *Drosophila*. A *Cy*/+ male was irradiated with x rays and crossed with +/+ females, and the *Cy*/+ sons were mated individually with +/+ females. From one cross, the progeny were

curly males	146	curly females	0
wildtype males	0	wildtype females	163

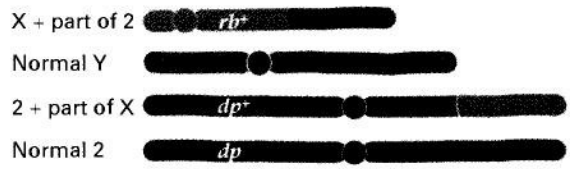
What abnormality in chromosome structure is the most likely explanation for these results? (Remember that crossing-over does not take place in male *Drosophila*.)

9.25 Yellow body (*y*) is a recessive mutation near the tip of the X chromosome of *Drosophila*. A wildtype male was irradiated with x rays and crossed with *yy* females, and one y^+ son was observed. This male was mated with *yy* females, and the offspring were

yellow females	256
yellow males	0
wildtype females	0
wildtype males	231

The yellow females were found to be chromosomally normal, and the y^+ males were found to breed in the same manner as their father. What type of chromosome abnormality could account for these results?

9.26 In *Drosophila*, ruby eye color (*rb*) results from a recessive allele in the X chromosome, and dumpy wings (*dp*) results from a recessive allele in chromosome 2. Among the progeny of flies exposed to x rays, a phenotypically normal male carrying a reciprocal translocation between the X chromosome and chromosome 2 was obtained. The genetic and chromosomal constitution of this male is shown in the illustration. What phenotypes and what relative frequencies would be expected among the progeny from the mating of this male to a female with ruby eyes and dumpy wings? (Remember that there is no crossing over in *Drosophila* males.)



9.27 In the homologous chromosomes shown here, the red region represents an inverted segment of chromosome.



An individual of this genotype was crossed with an *a b c d e* homozygote. Most of the offspring were either *A B C D E* or *a b c d e*, but a few rare offspring were obtained that were *A B c D E*. What events occurring in meiosis in the inversion heterozygote can explain these rare progeny? Is the gene sequence in the rare progeny normal or inverted?

9.28 A strain of semisterile maize heterozygous for a reciprocal translocation between chromosomes 1 and 2 was crossed with chromosomally normal plants homozygous for the recessive mutations *brachytic* and *fine-stripe* on

chromosome 1. When semisterile F_1 plants were crossed with plants of the *brachytic, fine-stripe* parental strain, the following phenotypes were found in a total of 682 F_2 progeny.

	Semisterile	Fertile
wildtype	333	19
brachytic	17	6
fine-stripe	1	8
brachytic, fine-stripe	25	273

What are the recombination frequencies between *brachytic* and the translocation breakpoint and between *fine-stripe* and the translocation breakpoint?

9.29 Semisterile maize plants heterozygous for a reciprocal translocation between chromosomes 1 and 2 were crossed with chromosomally normal plants homozygous for a recessive allele in chromosome 2 that results in virescent leaves. When semisterile F_1 plants were testcrossed, the following data were obtained:

semisterile virescent	39
fertile virescent	231
semisterile nonvirescent	287
fertile nonveriscent	31

(a) What is the frequency of recombination between the gene for virescent leaves and the translocation breakpoint in chromosome 2?

(b) What phenotypic ratio in the testcross progeny would be expected if the gene for virescent leaves had not been in one of the chromosomes involved in the translocation?

9.30 A wildtype strain of yeast, thought to be a normal haploid, was crossed with a different haploid strain carrying the mutation *his7*. This mutation is located in chromosome 2 and is an allele of a gene normally required for synthesis of the amino acid histidine. Among 9 tetrads analyzed from this cross, the following types of segregation were observed.

4 wildtype : 0 *his7*	4 tetrads
2 wildtype : 2 *his7*	1 tetrad
3 wildtype : 1 *his7*	4 tetrads

However, when the same wildtype strain was crossed with haploid strains with recessive markers on other chromosomes, segregation in the tetrads was always 2 : 2. What type of chromosome abnormality in the wildtype strain might account for the unusual segregation when the strain is mated with *his7*?

Challenge Problems

9.31 Use the "trick" given in Problem 9.14 for drawing the paired configuration between an inversion-bearing chromosome and its homolog to deduce the consequences of the following types of double crossovers, where the first is at the position shown and the second is between the genes *C* and *D*.

(a) Two-strand double crossover.

(b) Three-strand double crossover (one normal homolog not included).

(c) Three-strand double crossover (one inverted homolog not included).

(d) Four-strand double crossover.

9.32 Females heterozygous for the reciprocal translocation in Problem 9.26 and carrying normal X and 2 chromosomes containing *rb* and *dp*, respectively, are crossed with ruby, dumpy males. What genotypes and phenotypes are expected among 1000 progeny if the recombination frequency between *rb* and the translocation breakpoint is 10 percent and that between *dp* and the translocation breakpoint is 30 percent. Assume there is no interference.

9.33 An otherwise diploid strain of yeast is trisomic for one chromosome. The strain carries a recessive allele *a* in one copy of the trisomic chroomosome and the dominant allele *A* in each of the other copies, yielding the genotype *AAa*. The strain is sporulated and tetrads are classified for *A*− and *a*. Three types of tetrads are found: (1) 4 *A*− : 0 *a*, (2) 2 *A*− : 2 *a*, and (3) 3 *A*− : 1 *a*. Assume that the trisomic chromosome forms a bivalent and a univalent at random. Show that:

(a) if *A* and *a* are completely linked to the centromere, then the expected ratio of 4 : 0 to 2 : 0 to 3 : 1 tetrads is 1 : 2 : 0.

(b) if *A* and *a* are unlinked to the centromere ($r = 0.5$), then the expected ratio of 4 : 0 to 2 : 0 to 3 : 1 tetrads is 4 : 1 : 4.

Further Reading

Bickmore, W. A., and A. T. Sumner. 1989. Mammalian chromosome banding: An expression of genome organization. *Trends in Genetics* 5: 144.

Cavenee, W. K., and R. L. White. 1995. The genetic basis of cancer. *Scientific American*, March.

Cicchetti, D., and M. Beeghly, eds. 1990. *Children with Down Syndrome: A Developmental Perspective*. New York: Cambridge University Press.

Epstein, C. J. 1988. Mechanisms of the effects of aneuploidy in mammals. *Annual Review of Genetics* 22: 51.

Hernandez, D., and E. M. C. Fisher. 1999. Mouse autosomal trisomy: Two's company, three's a crowd. *Trends in Genetics* 15: 241.

Hsu, T. H. 1979. *Human and Mammalian Cytogenetics*. New York: Springer-Verlag.

GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web.** You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- Many people think that **birth defects** are rare and happen only to other people. But about 4–5 percent of all newborns have some type of birth defect, amounting to 150,000 cases each year in the United States alone, and birth defects are the leading cause of infant death and disability. Some 3000–5000 different birth defects have been described. The most common are listed at this keyword site. The list includes only those conditions that can be diagnosed immediately at birth and hence does not include such conditions as cystic fibrosis, Tay–Sachs disease, and sickle-cell anemia, which become apparent only in the first weeks or months after birth. Note that defects of the heart and circulatory system affect more infants than any other type of birth defect.

- About half a million human **karyotypes** are analyzed every year in North America in the course of genetic diagnosis and genetic counseling. At this keyword site, you can arrange and analyze digital images of human chromosomes from actual patients and then draw your own conclusions much as you would if were you carrying out genetic analysis in a hospital or clinical setting.

- Molecular studies of the red and green opsin genes indicate abnormalities in 15.7 percent of Caucasian men. This frequency is substantially greater than the incidence of color blindness, as shown by color-vision testing, which suggests that some chimeric genes created by unequal crossing-over may be compatible with normal color vision. More information about the molecular genetics and physiology of color vision can be found by searching the two keyword sites **protanopia** and **deauteranopia.** You will also find a description of the "Nagel anomaloscope" used in testing color vision.

- **Down syndrome** affects more than 350,000 people in the United States, and there are approximately 5000 new cases every year. There are many myths about the condition. One is that people with Down syndrome are severely retarded; another is that their parents are usually old. At this keyword site you can find these and other myths and the truths that contradict them.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 9 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 9.



Hurst, L. D., and Randerson, J. P. 1999. An exceptional chromosome. *Trends in Genetics* 15: 383.

Kimber, G., and M. Feldman. 1987. *Wild Wheat: An Introduction.* Columbia, MO: University of Missouri Press.

Manning, C. H., and H. O. Goodman. 1981. Parental origin of chromosomes in Down's syndrome. *Human Genetics* 59: 101.

Miller, O. J., and E. Therman. 2000. *Human Chromosomes* 4th ed. New York: Springer-Verlag.

Nathans, J., 1989. The genes for color vision. *Scientific American,* February.

Rattmer, A., H. Sun, and J. Nathans. 1999. Molecular genetics of human retinal disease. *Annual Review of Genetics* 33: 89.

Richards, R. I., and G. R. Sutherland. 1992. Fragile X syndrome: The molecular picture comes into focus. *Trends in Genetics* 8: 249.

Stebbins, G. L. 1971. *Chromosome Evolution in Higher Plants.* Reading, MA: Addison-Wesley.

Stewart, G. D., T. J. Hassold, and D. M. Kurnit. 1988. Trisomy 21: Molecular and cytogenetic studies of nondisjunction. *Advances in Human Genetics* 17: 99.

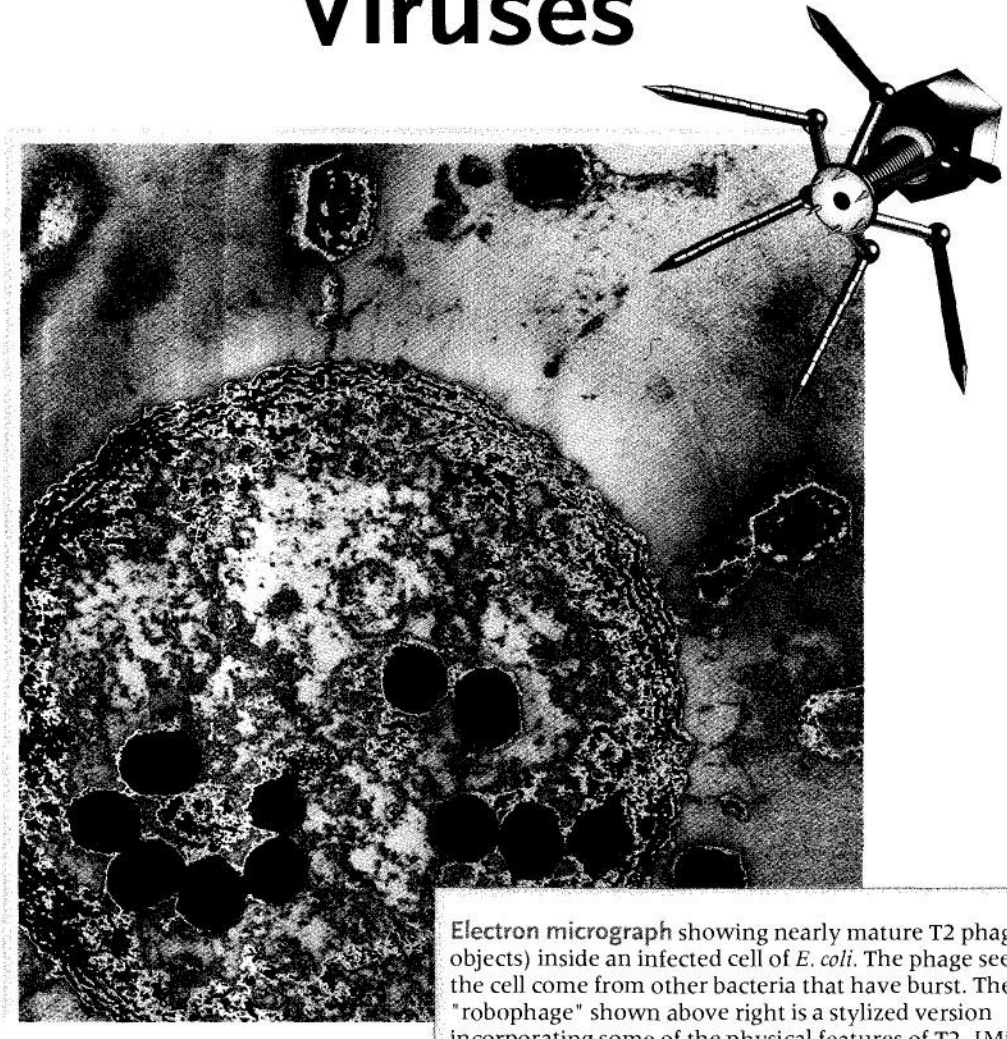
Vogt, P. K., ed. 1997. *Chromosomal Translocations and Oncogenic Transcription Factors.* New York: Springer-Verlag.

Wagner, R. P., M. P. Maguire, and R. L. Stallings. 1993. *Chromosomes.* New York: Wiley-Liss.

White, M. J. D. 1977. *Animal Cytology and Evolution.* London: Cambridge University Press.

CHAPTER TO

Genetics of Bacteria and Their Viruses



Electron micrograph showing nearly mature T2 phage (dark objects) inside an infected cell of *E. coli*. The phage seen outside the cell come from other bacteria that have burst. The "robophage" shown above right is a stylized version incorporating some of the physical features of T2. [Micrograph: Lee Simon/Photo Researchers, Inc.; robophage: courtesy Paul Dowrick, © Phage *et al* Ltd.]

CHAPTER OUTLINE

10.1 Mobile DNA
Plasmids
The F plasmid: A Conjugative Plasmid
Insertion Sequences and Transposons
Mobilization of Nonconjugative Plasmids
Integrons and Antibiotic-Resistance Cassettes
Multiple-Antibiotic-Resistant Bacteria

10.2 Bacterial Genetics
Mutant Phenotypes
Mechanisms of Genetic Exchange

10.3 DNA-Mediated Transformation

10.4 Conjugation
Cointegrate Formation and Hfr Cells
Time-of-Entry Mapping
F' Plasmids

10.5 Transduction
The Phage Lytic Cycle
Generalized Transduction

10.6 Bacteriophage Genetics
Plaque Formation and Phage Mutants
Genetic Recombination in the Lytic Cycle
Fine Structure of the *rII* Gene in Bacteriophage T4

10.7 Lysogeny and Specialized Transduction
Site-Specific Recombination and Lysogeny
Specialized Transduction

PRINCIPLES

- Bacteria take advantage of several mechanisms by which DNA sequences can move from one DNA molecule to another, from one cell to another, or even from one bacterial species to another; these mechanisms have led to the evolution of multiple-antibiotic-resistant bacteria.
- In transformation, a bacterial cell takes up DNA from the surrounding medium and incorporates the DNA into its genome by homologous recombination, replacing some host genes in the process.
- In *E. coli*, the F (fertility) plasmid can mobilize the chromosome for transfer to another cell in the process of conjugation.
- Some types of bacteriophages can incorporate bacterial genes and transfer them into new host cells in the process of transduction.
- DNA molecules from related bacteriophages that are present in the same host cell can undergo genetic recombination.
- Some viruses are able to integrate their DNA into that of the host cell, where it replicates along with the host DNA and is transmitted to progeny cells.

CONNECTIONS

The Sex Life of Bacteria
Joshua Lederberg and Edward L. Tatum 1946
Gene Recombination in Escherichia coli

Origin of Phage Genetics
Alfred D. Hershey and Raquel Rotman 1948
Genetic Recombination Between Host-range and Plaque-type Mutants of Bacteriophage in Single Bacterial Cells

Artoo
Seymour Benzer 1955
Fine Structure of a Genetic Region in Bacteriophage

Bacteria and their viruses (bacteriophage) have unique reproductive systems with multiple and novel mechanisms of genetic exchange. Some bacterial DNA sequences can become mobile by any of a variety of mechanisms. This feature enables them to become widely disseminated within a bacterial population and even to spread between species. In this chapter we discuss the genetic systems of bacteria and bacteriophage. We begin by examining **mobile DNA:** sequences that can be transferred between DNA molecules and from one cell to another. The ability to share genes in this manner, even among different bacterial species, is a unique feature of bacterial genetic systems.

10.1 Mobile DNA

A high percentage of bacteria isolated from clinical infections are resistant to one or more antibiotics. Most of them are resistant to multiple antibiotics. Some are resistant to all antibiotics in routine use. The problem has become so severe that many of the antibiotics that were at one time most effective and had the fewest side effects are now virtually useless. The widespread antibiotic-resistance genes almost never originate from new mutations in the bacterial genome. They are acquired, usually several at a time, in various forms of mobile DNA.

Plasmids

Plasmids are nonessential DNA molecules that exist inside bacterial cells. They replicate independently of the bacterial genome and segregate to the progeny when a bacterial cell divides, so they can be maintained indefinitely in a bacterial lineage. Many plasmids are circular DNA molecules, but others are linear. The number of copies of a particular plasmid in a cell varies from one plasmid to the next, depending on the mechanism by which replication is regulated. High-copy-number plasmids are found in as many as 50 copies per host cell, whereas low-copy-number plasmids are present in 1–2 copies per cell. Plasmids range in size from a few kilobases to a few hundred kilobases (**Figure 10.1**) and are found in most bacterial species that have been studied. In *E. coli*, most plasmids are either quite small (up to about 10 kb) or quite large (greater than 40 kb). A typical *E. coli* isolate contains three different small plasmids, each

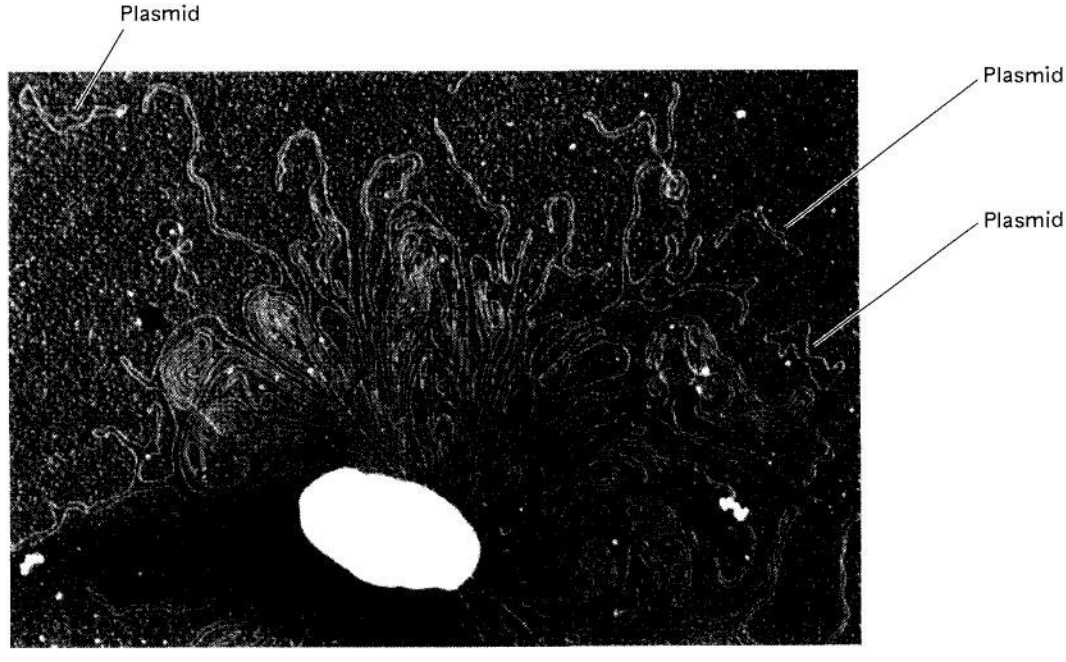


Figure 10.1 Electron micrograph of a ruptured *E. coli* cell, showing released chromosomal DNA and several plasmid molecules. [Courtesy of David Dressler and Huntington Potter.]

present in multiple copies per cell, and one large plasmid present in a single copy per cell. The presence of plasmids can be detected physically by electron microscopy, as in Figure 10.1, or by gel electrophoresis of DNA samples. Some plasmids can be detected because of phenotypic characteristics that they confer on the host cell. The phenotype most commonly studied is antibiotic resistance. For example, a plasmid containing a tetracycline-resistance gene (*tet-r*) will enable the host bacterial cell to form colonies on medium containing tetracycline.

Plasmids rely on the DNA-replication enzymes of the host cell for their reproduction, but the initiation of replication is controlled by plasmid genes. In high-copy-number plasmids, replication is initiated multiple times during replication of the host genome, but in low-copy-number plasmids, replication is initiated only once, or little more than once, per round of replication of the host genome. All types of plasmids contain sequences that promote their segregation into both daughter cells produced by fission of the host cell, so spontaneous loss of plasmids is uncommon.

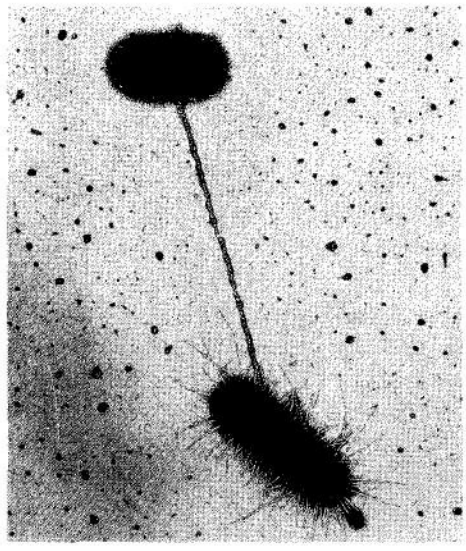


Figure 10.2 Pilus connecting two cells of *E. coli*. The cell with the multiple appendages (used for colonizing the intestine) contains an F plasmid that encodes the proteins necessary to produce the pilus. Prior to transfer of the plasmid DNA, the pilus shortens somewhat and draws the cells closer together. For ease of visualization, this pilus is coated with a bacteriophage that targets the F pilus. [Courtesy of C. C. Brinton, Jr., and J. Carnahan.]

The F Plasmid: A Conjugative Plasmid

Many large plasmids contain genes that enable the plasmid DNA to be transferred between cells. The transfer is mediated by a tube-like structure called a *pilus* (plural *pili*), formed between the cells, through which the plasmid DNA passes (**Figure 10.2**). The joining of bacterial cells in the transfer process is called **conjugation,** and the plasmids that can be transferred in this manner are called **conjugative plasmids.** Not all plasmids are conjugative plasmids. Most small plasmids are nonconjugative: They can be maintained in a bacterial lineage as the cells divide, but they do not contain the approximately 20 genes necessary for pilus assembly or those for DNA transfer. Hence they are unable to be transferred on their own. As we shall see later, however, they are able to employ the genetic trickery of recombination in order to tag along with conjugative plasmids, and in this way nonconjugative plasmids can be mobilized for cell-to-cell transfer.

The pilus between the *E. coli* cells in Figure 10.2 is an *F pilus* whose synthesis results from the presence of a conjugative plasmid called the **F factor** (the F stands for *fertility*). Cells that contain the F plasmid are donors and are designated F^+ ("F plus"); those lacking F are recipients and are designated F^- ("F minus"). The F plasmid is a low-copy-number plasmid, present in 1–2 copies per cell. It replicates once per cell cycle and segregates to both daughter cells in cell division. The F factor is approximately 100 kb in length and contains many genes that govern its maintenance in the cell and its transmission between cells.

Conjugation begins with physical contact between a donor cell and a recipient cell, as in Figure 10.2. The tube-like F pilus projects from the donor cell and, when it meets with the recipient cell, forms a hollow passageway between the donor and recipient cells. Through this passageway, the donor DNA moves from the donor to the recipient and takes up residence in the recipient. The transfer is always accompanied by replication of the plasmid. Contact between an F^+ and an F^- cell initiates rolling-circle replication (explained in Section 6.2), which results in the transfer of a single-stranded linear branch of the rolling circle

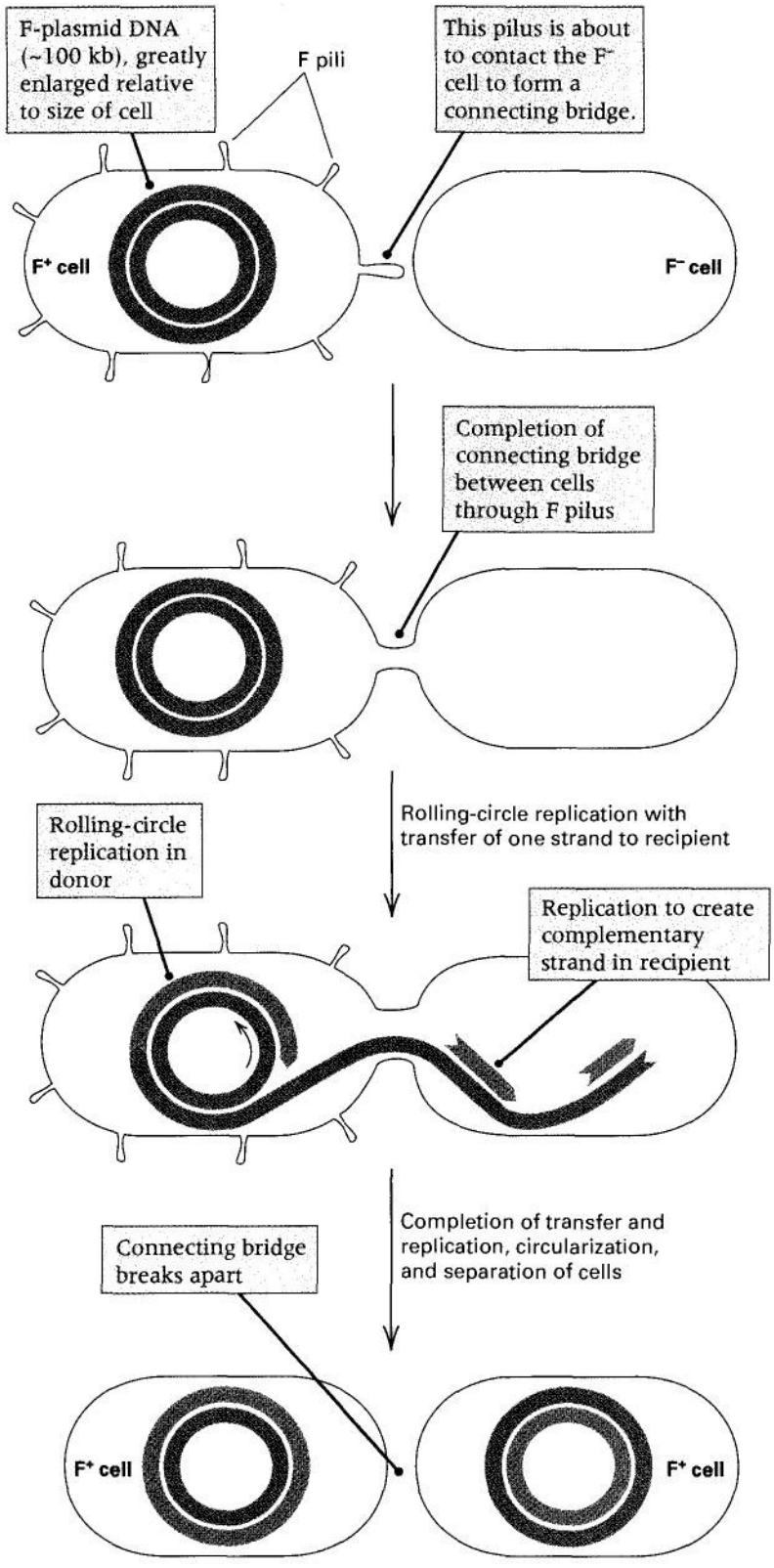


Figure 10.3 Transfer of F from an F^+ to an F^- cell. Pairing of the cells triggers rolling-circle replication. Pink represents DNA synthesized during pairing. For clarity, the bacterial chromosome is not shown, and the plasmid is drawn overly large; the plasmid is in fact much smaller than a bacterial chromosome.

to the recipient cell. During transfer, DNA is synthesized in both donor and recipient (**Figure 10.3**). Leading-strand synthesis in the donor replaces the transferred single strand, and lagging-strand synthesis in the recipient converts the transferred single strand into double-stranded DNA. When transfer is complete, the linear F strand becomes circular again in the recipient cell. Note that because one replica remains in the donor while the other is transferred to the recipient, after transfer both cells contain F and can function as donors. The F^- cell has been converted into an F^+ cell.

The transfer of the F plasmid requires only a few minutes. In laboratory cultures, if a small number of donor cells are mixed with an excess of recipient cells, F spreads throughout the population in a few hours, and all cells ultimately become F^+. Transfer is not so efficient under natural conditions, and only about 10 percent of naturally occurring *E. coli* cells contain the F factor. Conjugation normally takes place only between F^+ and F^- cells, because the F plasmid contains two genes for *surface exclusion*, which prevents an F^+ cell from conjugating with any other cell containing the same or a closely related plasmid. Most conjugative plasmids have analogous exclusion mechanisms.

Insertion Sequences and Transposons

Transposable elements are the key agents of gene mobilization. These DNA sequences were discussed in Chapter 7 in terms of their molecular mechanisms of transposition and their role as mutagenic agents either through insertional inactivation or misexpression of genes or through ectopic recombination. Here we focus on their role in bacterial genetics, particularly as agents through which antibiotic-resistance genes can be mobilized.

Bacteria contain a wide variety of transposable elements. The smallest and simplest are **insertion sequences,** or **IS elements,** which are typically 1–3 kb in length and usually encode only the transposase protein required for transposition and one or more additional proteins that regulate the rate of transposition. Like many transposable elements in eukaryotes, they possess inverted-repeat sequences at their termini, which

are used by the transposase for recognizing and mobilizing the IS element. Upon insertion, they create a short, direct duplication of the target sequence at each end of the inserted element. The DNA organization of the insertion sequence IS*50* is diagrammed in **Figure 10.4**, part A.

Other transposable elements in bacteria contain one or more genes unrelated to transposition that can be mobilized along with the transposable element; this type of element is called a **transposon.** The length of a typical transposon is several kilobases, but a few are much longer. Much of the widespread antibiotic resistance among bacteria is due to the spread of transposons that include one or more (usually multiple) antibiotic-resistance genes. When a transposon mobilizes and inserts into a conjugative plasmid, it can be widely disseminated among different bacterial hosts by means of conjugation.

Some transposons have composite structures with antibiotic resistance sandwiched between insertion sequences, as is the case with the Tn*5* element illustrated in Figure 10.4B, which terminates in two IS*50* elements in inverted orientation. Transposons are usually designated by the abbreviation Tn followed by an italicized number (for example, Tn*5*). When it is necessary to refer to genes carried in such an element, the usual designations for the genes are used. For example, Tn*5* (*neo-r ble-r str-r*) contains genes for resistance to three different antibiotics: neomycin, bleomycin, and streptomycin.

Mobilization of Nonconjugative Plasmids

Nonconjugative and conjugative plasmids typically coexist in the same cell along with host genomic DNA, and when a transposable element is mobilized, all of the DNA molecules present are potential targets for insertion. In time, most or all of the plasmids in a bacterial lineage will acquire copies of any transposable element present in the host DNA, and the host DNA will acquire copies of any transposable elements present in the plasmids. In this manner, transposable elements become disseminated among independently replicating DNA molecules. The result is that most bacteria contain multiple copies of different types of transposable elements, some in the host genome, some in plasmids, and some

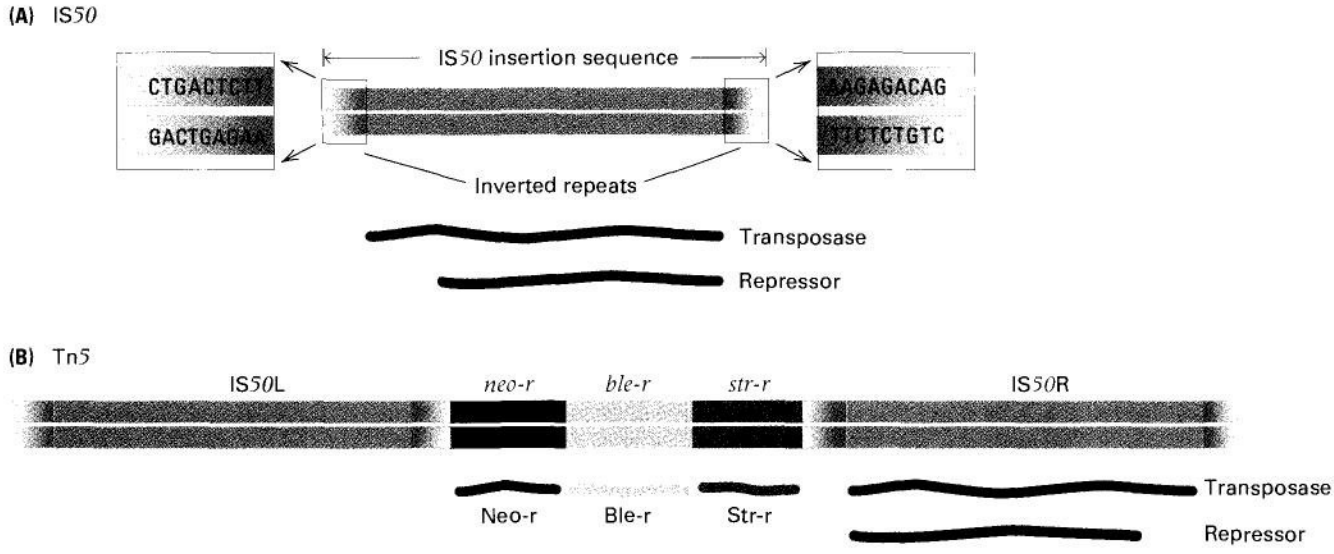


Figure 10.4 Transposable elements in bacteria. (A) Insertion sequence IS*50*. The element is terminated by short, nearly perfect inverted-repeat sequences, the terminal nine base pairs of which are indicated. IS*50* contains a region that codes for the transposase and for a repressor of transposition. The coding regions are identical in the region of overlap, but the repressor is somewhat shorter because it begins at a different place. (B) Composite transposon Tn*5*. The central sequence contains genes for resistance to neomycin, *neo-r;* bleomycin, *ble-r;* and streptomycin, *str-r.* It is flanked by two copies of IS*50* in inverted orientation. The left-hand element (IS*50*L) contains mutations and is nonfunctional, so the transposase and repressor are made by the right-hand element (IS*50*R).

in both. In *E. coli*, for example, natural isolates contain an average of 1–6 genomic copies of each of six naturally occurring IS elements, and among the cells that contain a particular IS element, 20–60 percent also contain copies in one or more plasmids.

Thus it happens that many nonconjugative and conjugative plasmids present in a bacterial cell come to carry one or more copies of the same transposable element. Because these copies are homologous DNA sequences, they can serve as substrates for recombination by the processes of strand exchange and Holliday-junction resolution discussed in Section 6.8. What happens when two plasmids undergo recombination in a region of homology? The result is shown in **Figure 10.5**. The recombination forms a composite plasmid called a **cointegrate.** If one of the participating plasmids is nonconjugative and the other is conjugative, then the cointegrate is also a conjugative plasmid and so can be transferred in conjugation. After conjugation, the nonconjugative plasmid can become free of the cointegrate by recombination between the same sequences that created it. By the mechanism of cointegrate formation, therefore, nonconjugative plasmids can temporarily ride along with conjugative plasmids and be transferred from cell to cell.

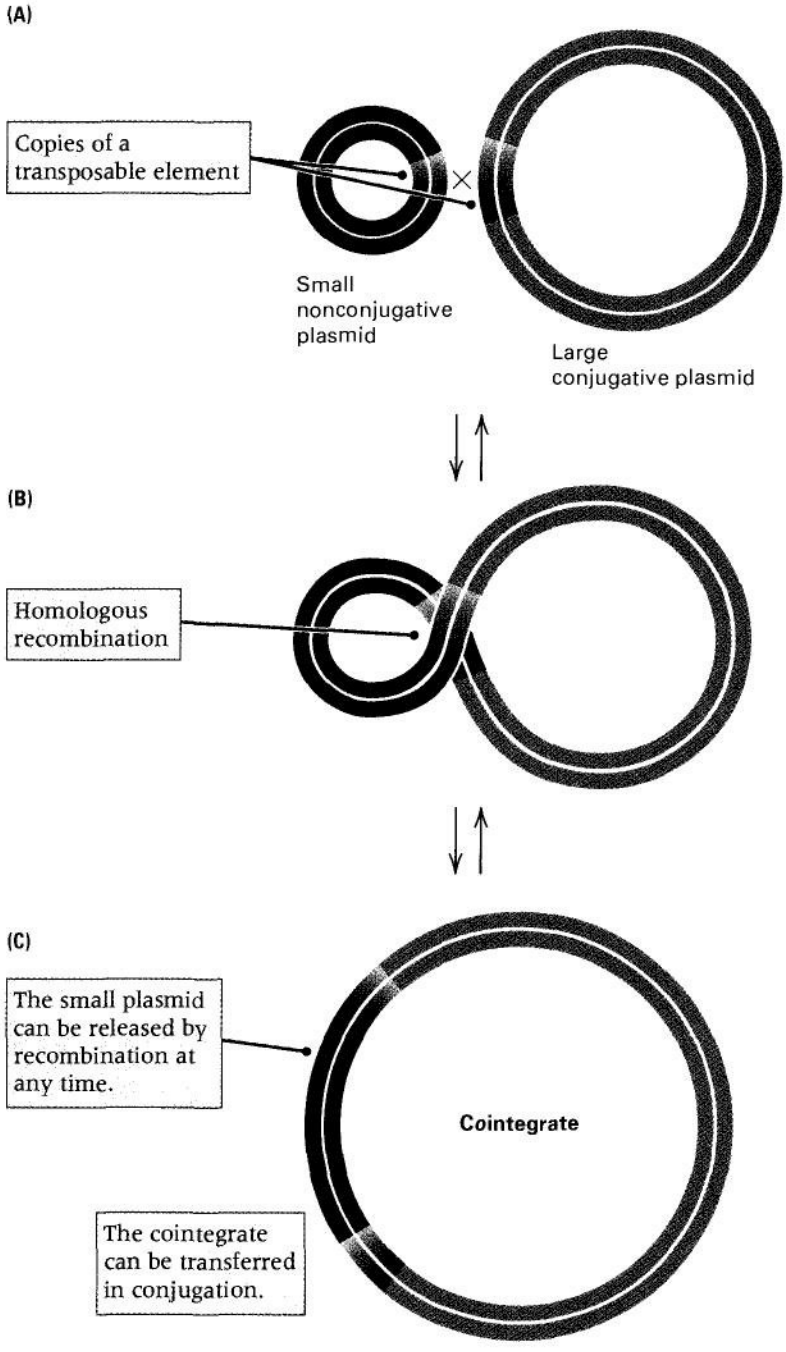


Figure 10.5 Cointegrate formed between two plasmids by recombination between homologous sequences (for example, copies of a transposable element) present in both plasmids.

Integrons and Antibiotic-Resistance Cassettes

In the evolution of multiple antibiotic resistance, bacteria have also made liberal use of a set of enzymes known as *site-specific recombinases*, which were present in bacterial populations long before the antibiotic era and functioned in the evolution of other traits. Each type of **site-specific recombinase** binds with a specific nucleotide sequence in duplex DNA. When the site is present in each of two duplex DNA molecules, the recombinase brings the sites together and catalyzes a reciprocal exchange between the duplexes. An example is shown in **Figure 10.6**, where the site-specific recombinase joins a circular DNA molecule with a linear DNA molecule. Note that the reaction can proceed in the reverse direction, too, and free the circle from the cointegrate.

An example of a site-specific recombinase is an enzyme called the *Cre recombinase*, which is encoded in a gene in the *E. coli* bacteriophage P1. The Cre recognition sequence is called *loxP;* it is the 34-bp sequence shown in Figure 10.6. Although the terminal 13 base pairs at each end of the *loxP* sequence form a perfect inverted repeat, the *loxP* sequence nevertheless has a

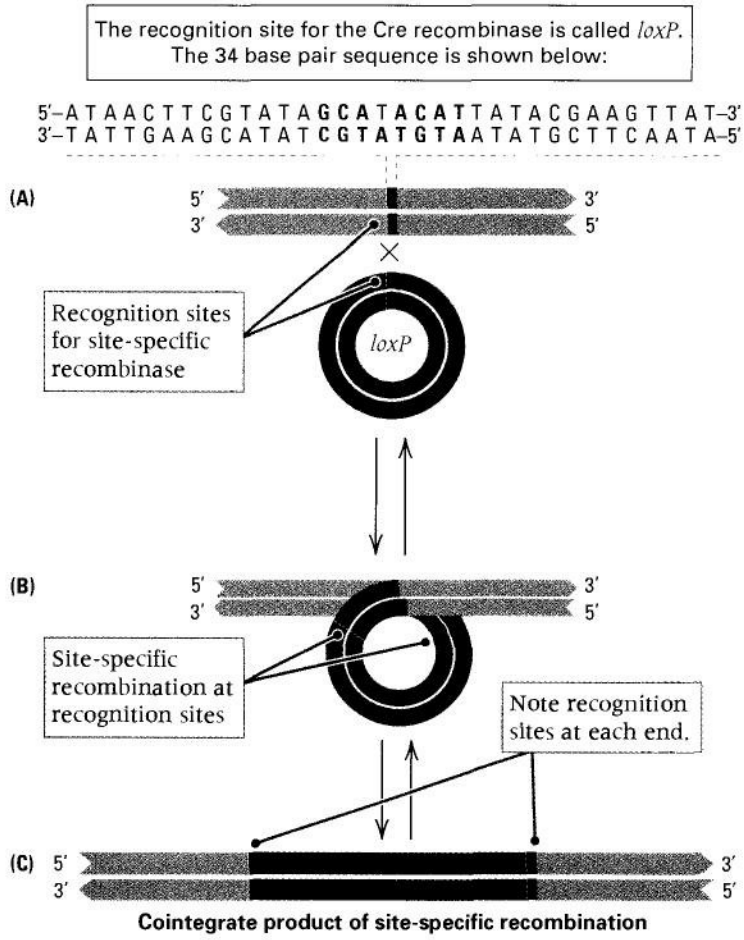


Figure 10.6 A site-specific recombinase catalyzes a reciprocal exchange between two specific sequences. No other sequences can serve as substrates. The recognition site for the Cre recombinase (*loxP*) is shown at the top.

directionality because the central region is asymmetrical. This kind of partial symmetry is typical of site-specific recombinases. Recombination between two *loxP* sites preserves the *loxP* sequences because the participating sites are identical, and hence the recombination reaction is reversible. Some site-specific recombinases favor the reaction that brings two molecules together into a cointegrate. Others (including Cre) favor the reaction that splits a cointegrate into two separate molecules. Some site-specific recombinases bring together and recombine sites that are similar but not identical; in these cases the recombination does not preserve the recognition sites, and so the reaction is not necessarily reversible.

Site-specific recombinases are used in the assembly of multiple-antibiotic-resistance units called *integrons*. An **integron** is a DNA element that encodes a site-specific recombinase as well as a recognition region that allows other sequences with similar recognition regions to be incorporated into the integron by recombination. The elements that integrons acquire are known as *cassettes*. In the context of integrons, a **cassette** is a circular antibiotic-resistance-coding region flanked by a recognition region for an integron. Because the site-specific recombinase integrates cassettes, the integron recombinase is usually called an **integrase.**

Several different types of integrons have been characterized. The best known of these are the Class 1 integrons, which include a site-specific recombinase denoted Int1 and invariably a coding region (*sul1*) that confers resistance to sulphonamide antibiotics. The molecular structure of a Class

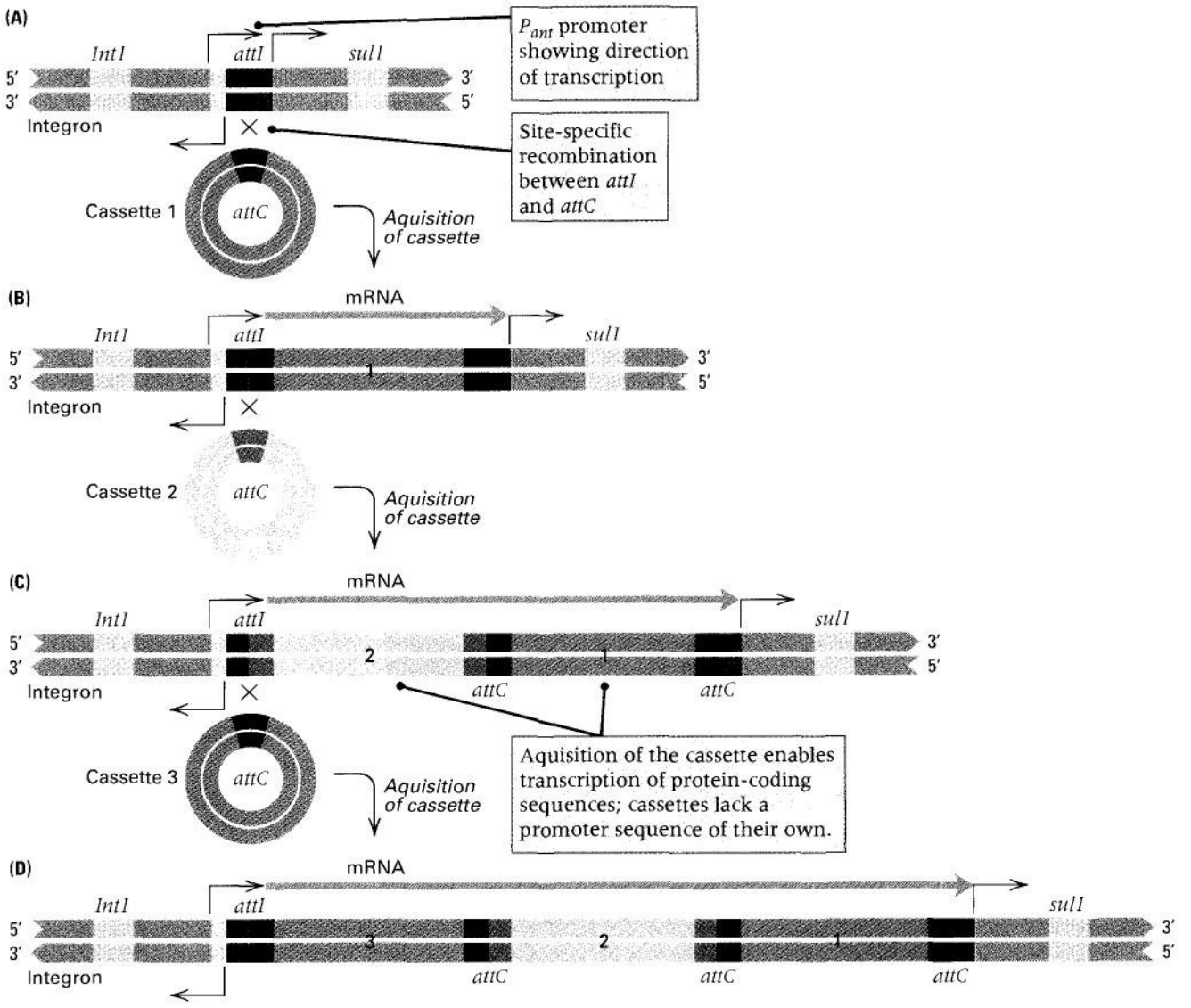


Figure 10.7 Mechanism by which an integron sequentially captures cassettes by site-specific recombination between the *attI* site in the integron and the *attC* site in the cassette.

1 integron is shown in **Figure 10.7**, part A. Also shown is the mechanism by which antibiotic-resistance cassettes are sequentially acquired. The Int1 integrase catalyzes a site-specific recombination between a sequence denoted *attI* present in the integron and a similar sequence denoted *attC* (also called the *59-base element*) in the cassette. All *attC* regions are similar, but no two are identical.

Figure 10.7A shows how a cassette is captured by site-specific recombination between *attI* and *attC*. In general, antibiotic-resistance cassettes contain protein-coding regions but lack promoter sequences of their own. They can be transcribed only by read-through transcription from an adjacent promoter. The integron provides the needed promoter, called P_{ant}, at a position upstream from the *attI* site, so that when a cassette is captured, the coding sequence can be expressed. More than 40 different promoter-less cassettes have been identified that encode proteins for resistance to antibiotics including β-lactams, aminoglycosides, chloramphenicol, trimethoprim, and streptothricin.

Once one cassette is in place, as shown in Figure 10.7B, a second can be captured using the same *attI* site and the *attC* present in the new cassette. Note that the new cassette is integrated immediately adjacent to

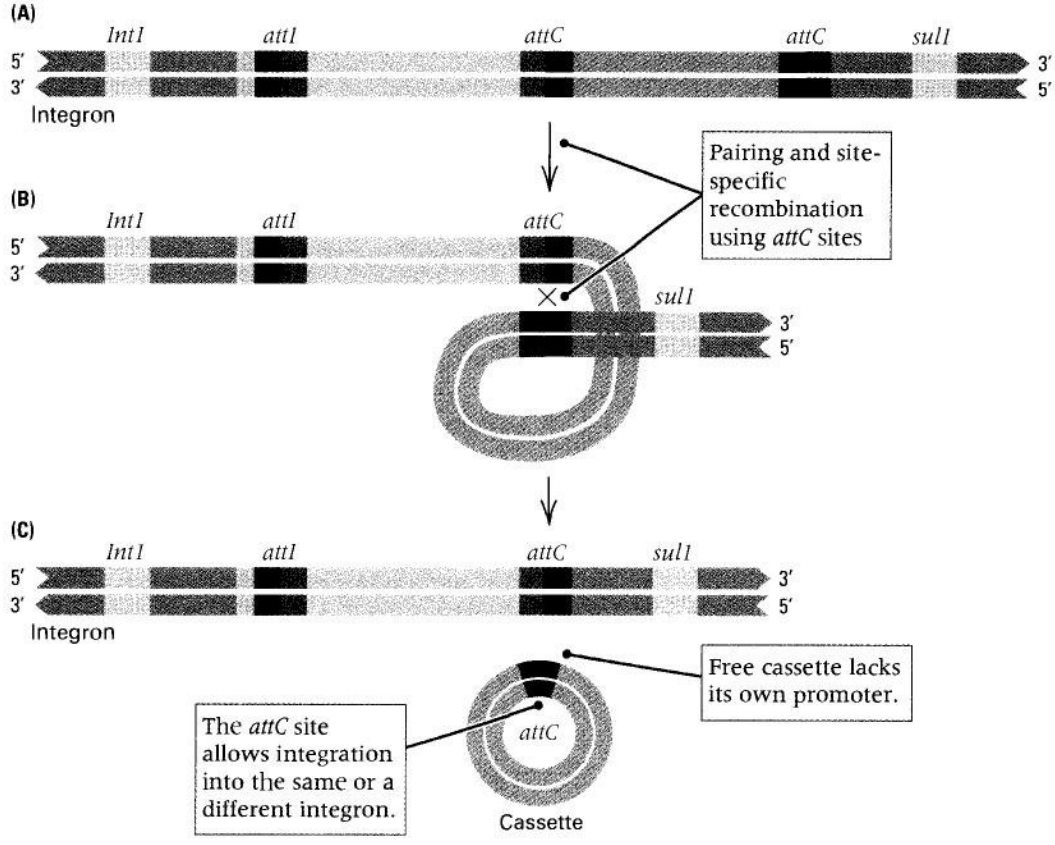


Figure 10.8
Mechanism of cassette excision from an integron by site-specific recombination between two *attC* sites.

the *attI* site and that the mRNA produced from the P_{ant} promoter includes the coding sequences for both cassettes. In Figure 10.7C a third cassette is added to the integron, and the mRNA from P_{ant} becomes even longer. When there are multiple cassettes, as shown here, all of them are cotranscribed from P_{ant}, but the downstream coding sequences are transcribed less frequently because there is a greater chance that transcription will terminate before reaching them. This constraint sets a practical limit on the number of cassettes that can be transcribed efficiently, but integrons with up to 10 antibiotic-resistance cassettes have been found.

The Int1 integrase can also catalyze the reverse of the cassette-capture reaction, though at a much lower level. This reaction generally uses two *attC* sites. An example is shown in **Figure 10.8**. Site-specific recombination between adjacent *attC* sites releases a circular cassette containing a promoter-less coding sequence. The cassette cannot replicate because it lacks an origin of replication, but the capture reaction is efficient enough that the cassette will often be recaptured by the same integron (which repositions the cassette immediately adjacent to *attI*) or by a different integron in the same cell (adding to the repertoire of cassettes it already contains).

Although integrons cannot mobilize themselves, they are present in transposons, conjugative plasmids, and nonconjugative plasmids, as well as in bacterial chromosomes. The integrons that are parts of mobile DNA elements are particularly important in the evolution of antibiotic resistance, because they can capture antibiotic-resistance cassettes and thereby make possible not only the transcription of the antibiotic-resistance coding sequences but also their mobilization.

Multiple-Antibiotic-Resistant Bacteria

In nature, a conjugative plasmid can, through time, accumulate different transposons containing multiple independent antibiotic-resistance genes, or transposons containing integrons that have acquired multiple-antibiotic-resistance cassettes, with

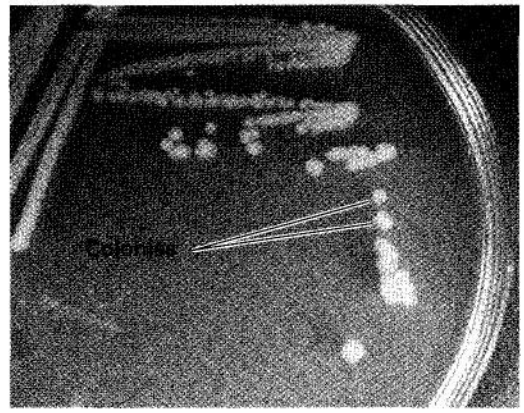


Figure 10.9 A petri dish with bacterial colonies that have formed on a solid medium. The heavy streaks of growth result from colonies so densely packed that there is no space between them.

the result that the plasmid confers resistance to a large number of completely unrelated antibiotics. These multiple-resistance plasmids are called **R plasmids.** Some R plasmids are closely related to the F plasmid and clearly evolved from the F factor. The evolution of R plasmids is promoted by the use (and overuse) of antibiotics, which selects for resistant cells because, in the presence of antibiotics, resistant cells have a growth advantage over sensitive cells. The presence of multiple antibiotics in the environment selects for multiple-drug resistance. Serious clinical complications result when plasmids resistant to multiple drugs are transferred to bacterial pathogens, or agents of disease. Infections with some pathogens that contain R factors are extremely difficult to treat, because the pathogen may be resistant to most or all antibiotics currently in use.

10.2 Bacterial Genetics

Bacteria and bacteriophage bring to traditional types of genetic analysis four important advantages over multicellular plants and animals. First, they are haploid, so dominance or recessiveness of alleles is not a complication in identifying genotype. Second, a new generation is produced in minutes rather than in weeks or months, which vastly increases the rate of accumulation of genetic data. Third, they are easy to grow in enormous numbers under controlled laboratory conditions, which facilitates molecular studies and the analysis of rare genetic events. Fourth, the individual members of these large populations are genetically identical; that is, each laboratory population is a *clone* of genetically identical cells.

Mutant Phenotypes

Many species of bacteria can be grown in liquid medium or on the surface of a semisolid growth medium gelled with agar. Bacteria used in genetic analysis are usually grown on agar. A single bacterial cell placed on an agar medium will grow and divide many times, forming a visible cluster of cells called a *colony* (**Figure 10.9**). The number of bacterial cells present in a liquid culture can be determined by spreading a known volume of the culture on a solid medium and counting the number of colonies that form. Typical *E. coli* cultures contain up to 10^9 cells/ml. The appearance of colonies, or the ability or inability to form colonies on particular media, can in some cases be used to identify the genotypes of bacterial cells.

As we have seen in earlier chapters, genetic analysis requires mutants. With bacteria, three types of mutants are particularly useful:

- ***Antibiotic-resistant mutants*** These mutants are able to grow in the presence of an antibiotic, such as streptomycin (Str) or tetracycline (Tet). For example, streptomycin-sensitive (Str-s) cells have the wildtype phenotype and fail to form colonies on medium that contains streptomycin, but streptomycin-resistant (Str-r) mutants can form colonies on such a medium.
- ***Nutritional mutants*** Wildtype bacteria can synthesize most of the complex nutrients they need from simple molecules present in the growth medium. The wildtype cells are said to be **prototrophs.** The ability to grow in simple medium can be lost by mutations that disable the enzymes used in synthesizing the complex nutrients. Mutant cells are unable to synthesize an essential nutrient and cannot grow unless the required nutrient is supplied in the medium. Such a mutant bacterium is said to be an **auxotroph** for the particular nutrient. For example, a methionine auxotroph cannot grow on a **minimal medium** that contains only inorganic salts and a source of energy and carbon atoms (such as glucose), but the methionine auxotroph *is* able to grow if the minimal medium is supplemented with methionine.

- ***Carbon-source mutants*** Such mutant cells cannot utilize particular substances as sources of carbon atoms or of energy. For example, Lac$^-$ mutants cannot utilize the sugar lactose for growth and are unable to form colonies on minimal medium that contains only lactose as the carbon source.

A medium on which all cells can form colonies is called a **nonselective medium.** Mutants and wildtype cells may or may not be distinguishable by growth on a nonselective medium. If the medium allows growth of only one type of cell (either wildtype or mutant), then it is said to be **selective.** For example, a medium containing streptomycin is selective for the Str-r phenotype and selective against the Str-s phenotype; similarly, minimal medium containing lactose as the sole carbon source is selective for Lac$^+$ cells and against Lac$^-$ cells.

In bacterial genetics, phenotype and genotype are designated in the following way. A phenotype is designated by three letters, the first of which is capitalized, with a superscript + or − to denote presence or absence of the designated phenotype, and with s or r for sensitivity or resistance. A genotype is designated by lowercase italicized letters. Thus a cell unable to grow without a supplement of leucine (a leucine auxotroph) has a Leu$^-$ phenotype; this would usually result from a *leu*$^-$ mutation. Often the − superscript is omitted, but using it prevents ambiguity.

Mechanisms of Genetic Exchange

A bacterial chromosomal DNA molecule almost never encounters another complete DNA molecule from a different bacterial cell. Instead, genetic exchange is usually between a chromosomal fragment from one cell and intact chromosomal DNA from another cell. Furthermore, a clear donor–recipient relationship exists: The donor cell is the source of a DNA fragment, which is transferred to the recipient cell by one of several mechanisms, and exchange of genetic material takes place in the recipient by means of reciprocal recombination between homologous DNA sequences. When the donor DNA is a linear molecule, incorporation of donor DNA into the recipient chromosome requires at least two exchange events, one at each end. Because the donor molecule contains only a fraction of a complete genome, only an even number of exchanges results in a viable product. The usual outcome of these events is the recovery of only one of the crossover products. However, in some situations, the transferred DNA is circular, and a single exchange results in total incorporation of the circular donor DNA into the chromosome of the recipient.

Three major types of genetic transfer are found in bacteria: *transformation*, in which a DNA molecule is taken up from the external environment and incorporated into the genome; *conjugation*, in which donor DNA is transferred from one bacterial cell to another by direct contact; and *transduction*, in which DNA is transferred from one bacterial cell to another by a bacterial virus. Recombination frequencies that result from these processes are used to produce genetic maps of bacteria. Although the maps are exceedingly useful, they differ in major respects from the types of maps obtained from crosses in eukaryotes, because genetic maps in eukaryotes are based on frequencies of exchange events between chromatids in meiosis.

10.3 DNA-Mediated Transformation

Transformation is a process by which recipient cells acquire genes from free DNA molecules in the surrounding medium. Transformation with purified DNA was the first experimental proof that DNA is the genetic material (Chapter 1). In these experiments, a rough-colony phenotype of *Streptococcus pneumoniae* was changed to a smooth-colony phenotype by exposure of the cells to DNA from a smooth-colony strain. In the laboratory, donor DNA is usually isolated from donor cells and then added to a suspension of recipient cells. In natural settings, such as soil, free DNA can become available by spontaneous breakage (lysis) of donor cells.

Transformation begins with a recipient cell's uptake of a DNA fragment from the surrounding medium and terminates with *one strand* of donor DNA replacing the homologous segment in the recipient DNA. Most bacterial species are probably capable

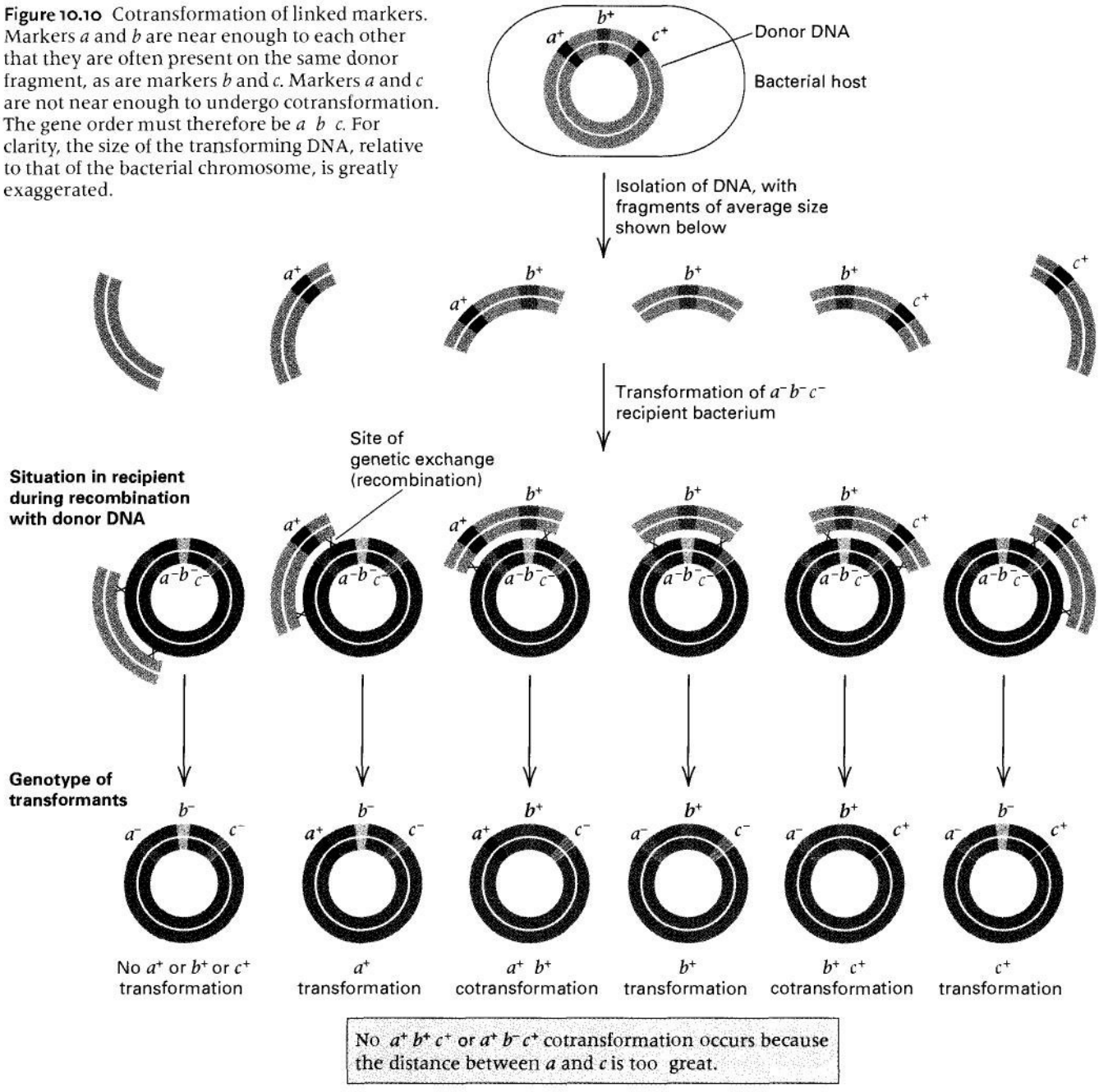


Figure 10.10 Cotransformation of linked markers. Markers *a* and *b* are near enough to each other that they are often present on the same donor fragment, as are markers *b* and *c*. Markers *a* and *c* are not near enough to undergo cotransformation. The gene order must therefore be *a b c*. For clarity, the size of the transforming DNA, relative to that of the bacterial chromosome, is greatly exaggerated.

of the recombination step, but many species have only a very limited ability to take up free DNA efficiently. Even in a species capable of transformation, DNA is able to penetrate only some of the cells in a growing population. However, many bacterial species can be made competent to take up DNA, provided that the cells are subjected to an appropriate chemical treatment (for example, treatment with $CaCl_2$).

Transformation is a convenient technique for gene mapping in some species. When DNA is isolated from a donor bacterium (**Figure 10.10**), it is invariably broken into small fragments. In most species, with suitable recipient cells and excess external DNA, transformation takes place at a frequency of about 1 transformed cell per 10^3 cells. If two genes, *a* and *b*, used as genetic markers, are so widely separated in the donor chromosome that they are always contained in two different DNA fragments, then the probability of simultaneous transformation (**cotransformation**) of an a^-b^-

recipient into wildtype is the product of the probabilities of transformation of each genetic marker, or roughly $10^{-3} \times 10^{-3}$, which equals one a^+b^+ transformant per 10^6 recipient cells. However, if the two genes are so near one another that they are often present in a single donor fragment, then the frequency of cotransformation is nearly the same as the frequency of single-gene transformation, or one wildtype transformant per 10^3 recipients. The general principle is as follows:

> Cotransformation of two genes at a frequency substantially greater than the product of the single-gene transformations implies that the two genes are close together in the bacterial chromosome.

Studies of the ability of various pairs of genes to be cotransformed also yield gene order. For example, if genes *a* and *b* can be cotransformed, and genes *b* and c can be cotransformed, but genes *a* and *c* cannot, then the gene order must be *a b c* (Figure 10.10). Note that cotransformation frequencies are not equivalent to the recombination frequencies used in mapping in eukaryotes, because they are determined by the size distribution of donor fragments and the likelihood of recombination between bacterial DNA molecules rather than by the occurrence of chiasmata in synapsed homologous chromosomes (Chapter 5).

10.4 Conjugation

Conjugation is a process in which DNA is transferred from a bacterial donor cell to a recipient cell by cell-to-cell contact. We have already examined this process in the context of conjugative plasmids. In this section we shall see how the same process can transfer genes present in the bacterial chromosome.

Cointegrate Formation and Hfr Cells

Transfer of chromosomal genes between *E. coli* cells was first observed by Joshua Lederberg in 1951. Although it was not known at the time, the exchange took place because the donor cells were F^+, and in a few cells the F factor had become integrated into the bacterial chromosome (Figure 10.11). These are known as **Hfr cells.** Hfr stands for **high frequency of recombination,** which refers to the relatively high frequency with which donor genes are transferred to the recipient. The integration process is essentially the same as the formation of a cointegrate between two plasmids illustrated in Figure 10.5. Insertion sequences (Section 8.1) are instrumental in the origin of Hfr bacteria from F^+ cells, because the F plasmid normally integrates through genetic exchange between an IS element present in F and a homologous copy that has transposed to an essentially random site in the bacterial chromosome. Because the F factor can exist either separate from the chromosome or incorporated into it, it qualifies as an **episome:** a genetic element that can exist free in the cell or as a segment of DNA integrated into the chromosome.

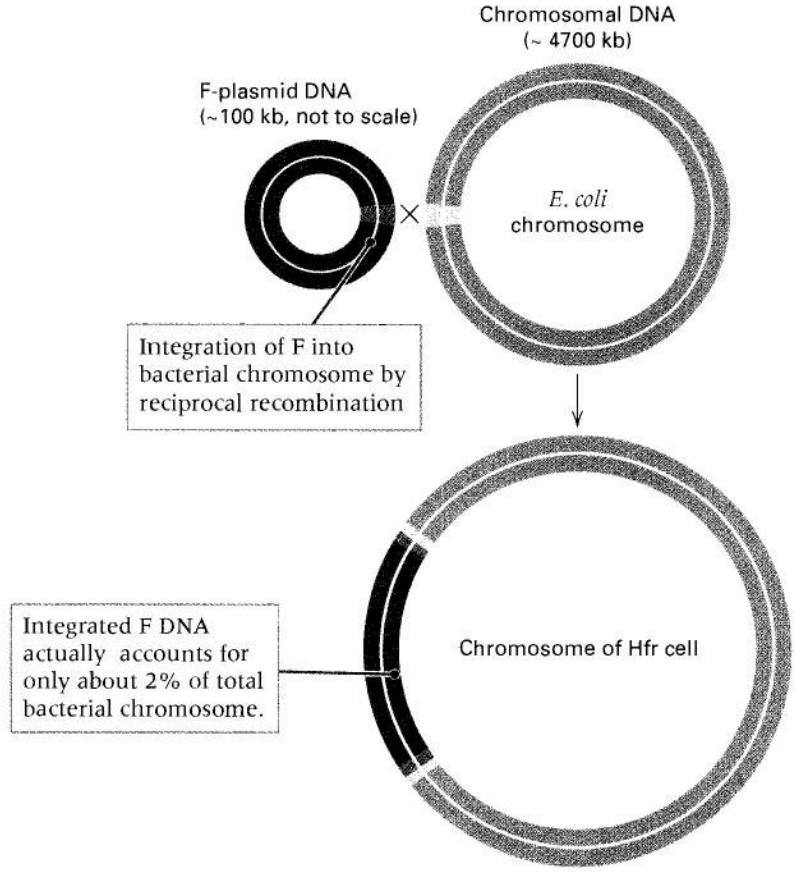


Figure 10.11 Integration of F (blue circle) by recombination between a nucleotide sequence in F and a homologous sequence (usually an insertion sequence) in the bacterial chromosome. The F plasmid DNA is shown greatly enlarged relative to the size of the bacterial chromosome.

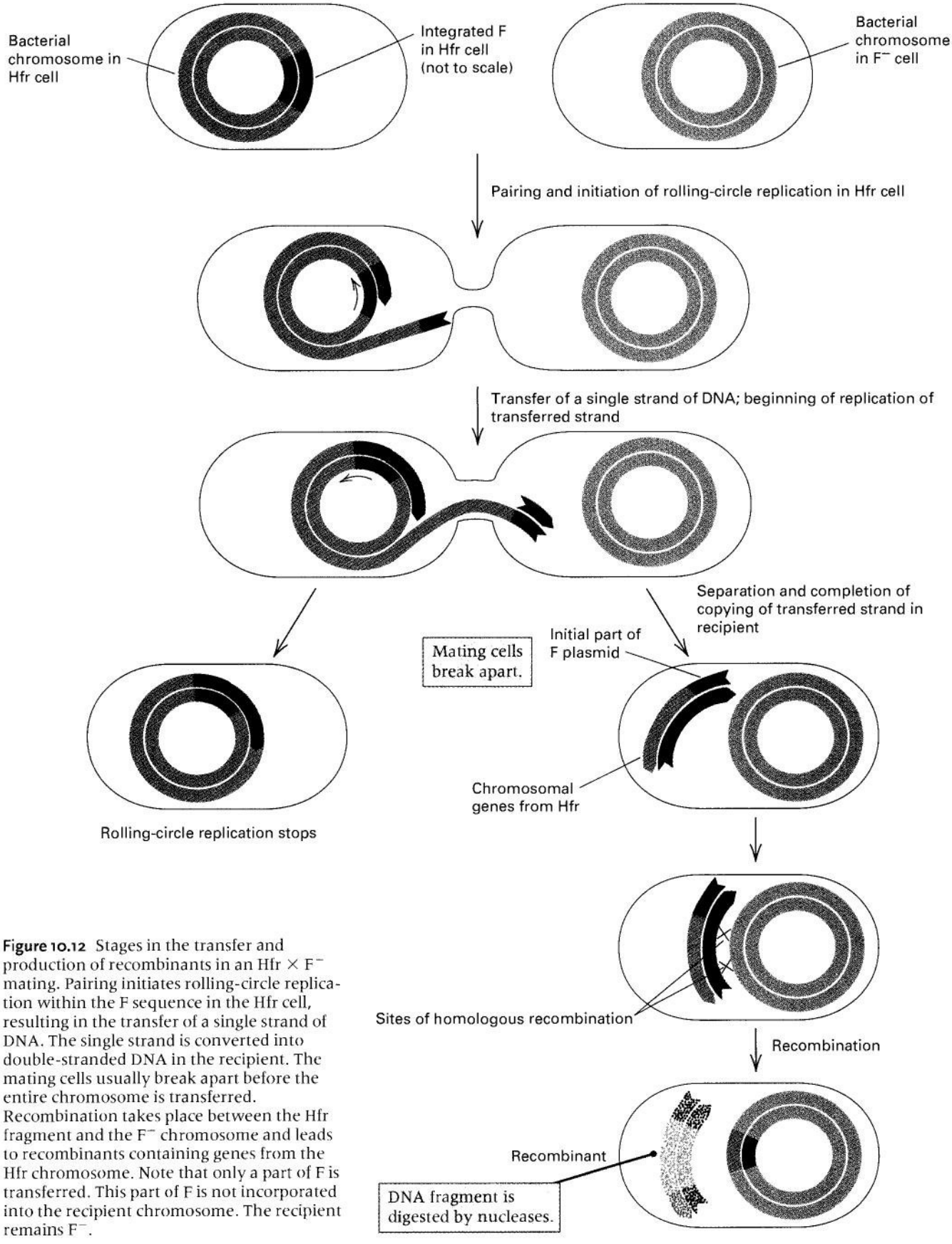


Figure 10.12 Stages in the transfer and production of recombinants in an Hfr × F⁻ mating. Pairing initiates rolling-circle replication within the F sequence in the Hfr cell, resulting in the transfer of a single strand of DNA. The single strand is converted into double-stranded DNA in the recipient. The mating cells usually break apart before the entire chromosome is transferred. Recombination takes place between the Hfr fragment and the F⁻ chromosome and leads to recombinants containing genes from the Hfr chromosome. Note that only a part of F is transferred. This part of F is not incorporated into the recipient chromosome. The recipient remains F⁻.

In an Hfr cell (Figure 10.11), the bacterial chromosome remains circular, though enlarged about 2 percent by the integrated F-factor DNA. Integration of F is an infrequent event, but single cells containing integrated F can be isolated and cultured. When an Hfr cell undergoes conjugation, the process of transfer of the F factor is initiated in the same manner as in an F^+ cell. However, because the F factor is part of the bacterial chromosome, transfer from an Hfr cell also includes DNA from the bacterial chromosome.

The Hfr $\times$ F^- conjugation process is illustrated in **Figure 10.12**. The stages of transfer are much like those by which F is transferred to F^- cells: coming together of donor and recipient cells, rolling-circle replication in the donor cell, and conversion of the transferred single-stranded DNA into double-stranded DNA by lagging-strand synthesis in the recipient. However, in the case of Hfr matings, the transferred DNA does not become circular and is not capable of further replication in the recipient, because the transferred F factor is not complete. The replication and associated transfer of the chromosomal DNA are controlled by the F factor and are initiated in the Hfr chromosome at the origin of replication in F, the same position at which replication and transfer begin in a free F plasmid. A part of F is the first DNA transferred, chromosomal genes are transferred next, and the remaining part of F is the last DNA to enter the recipient. Because the conjugating cells usually break apart long before the entire bacterial chromosome is transferred, the final part of F is almost never transferred into the recipient. Incorporation of donor DNA into the recipient chromosome requires homologous recombination in the recipient. Two recombination events result in the replacement of a segment of recipient DNA with a homologous segment of donor DNA, and the segment of recipient DNA that has been replaced is degraded (Figure 10.12).

Several differences between F transfer and Hfr transfer are notable.

- It takes 100 minutes under the usual conditions for an entire bacterial chromosome to be transferred, in contrast with about 2 minutes for the transfer of F. The difference in time is a result of the relative sizes of F and the chromosome (100 kb versus 4600 kb).
- During transfer of Hfr DNA into a recipient cell, the mating pair usually breaks apart before the entire chromosome is transferred. Under usual conditions, several hundred genes are transferred before the cells separate.
- In a mating between Hfr and F^- cells, the F^- recipient remains F^- because cell separation usually takes place before the final segment of F is transferred.
- In Hfr transfer, some regions in the transferred DNA fragment become incorporated into the recipient chromosome. The incorporated regions replace homologous regions in the recipient chromosome. The result is that some F^- cells become recombinants, containing one or more genes from the Hfr donor cell. For example, in a mating between Hfr *leu*$^+$ and F^- *leu*$^-$, some F^- *leu*$^+$ cells arise. However, *the genotype of the donor Hfr cell remains unchanged.*

Genetic analysis requires that recombinant recipients be identified. Because the recombinants derive from recipient cells, a method is needed to eliminate the donor cells. The usual procedure is to employ an F^- recipient containing an allele that can be selected. Genes that confer antibiotic resistance are especially useful for this purpose. For instance, after a mating between Hfr *leu*$^+$*str-s* and F^- *leu*$^-$*str-r* cells, the Hfr Str-s cells can be selectively killed by plating the mating mixture on medium containing streptomycin. A selective medium that lacks leucine can then be used to distinguish between the nonrecombinant and the recombinant recipients. The F^- *leu*$^-$ parent cannot grow in medium that lacks leucine, but recombinant F^- *leu*$^+$ cells can grow because they possess a *leu*$^+$ gene. Only recombinant recipients—that is, cells having the genotype *leu*$^+$*str-r*—form colonies on a selective medium containing streptomycin and lacking leucine. The selected allele should be located at such a place in the chromosome that most mating cells will have broken apart before the selected gene is transferred, and the selected allele must not be present in the Hfr cell. The selective agent can then be used to select the F^- cells and eliminate the Hfr donors.

When a mating is done in this way, the transferred marker that is selected by the growth conditions (*leu*$^+$ in this case) is called a **selected marker,** and the marker

used to prevent growth of the donor (*str-s* in this case) is called the **counterselected marker.** Selection and counterselection are necessary in bacterial matings because recombinants constitute only a small proportion of the entire population of cells (in spite of the name "high frequency of recombination").

Time-of-Entry Mapping

Genes can be mapped by Hfr × F⁻ matings. However, the genetic map is quite different from all maps that we have seen so far in that it is not a linkage map but a transfer-order map. It is obtained by deliberate interruption of DNA transfer in the course of mating—for example, by violent agitation of the suspension of mating cells in a kitchen blender. The time at which a particular gene is transferred can be determined by breaking the mating cells apart at various times and noting the earliest time at which breakage no longer prevents recombinants from appearing. This procedure is called the **interrupted-mating technique.** When this is done with Hfr × F⁻ matings, the number of recombinants of any particular allele increases with the time during which the cells are in contact. This phenomenon is illustrated in **Table 10.1**. The reason for the increase is that different Hfr × F⁻ pairs initiate conjugation and chromosome transfer at slightly different times.

Table 10.1 Data showing the production of Leu⁺ Str-r recombinants in a cross between Hfr *leu*⁺ *str-s* and F⁻ *leu*⁻ *str-r* cells when mating is interrupted at various times

Minutes after mating	Number of Leu⁺ Str-r recombinants per 100 Hfr cells
0	0
3	0
6	6
9	15
12	24
15	33
18	42
21	43
24	43
27	43

Note: Minutes after mating means minutes after the Hfr and F⁻ cell suspensions are mixed. Extrapolation of the recombination data to a value of zero recombinants indicates that the earliest time of entry of the *leu*⁺ marker is 4 minutes.

A greater understanding of the transfer process can be obtained by observing the results of a mating with several genetic markers. For example, consider the mating

$$\text{Hfr } a^+ b^+ c^+ d^+ e^+ str\text{-}s \times \text{F}^- a^- b^- c^- d^- e^- str\text{-}r$$

in which a^- cells require nutrient A, b^- cells require nutrient B, and so forth. At various times after mixing of the cells, samples are agitated violently and then plated on a series of media that contain streptomycin and different combinations of the five substances A through E (in each medium, one of the five is left out). Colonies that form on the medium lacking A are $a^+ str\text{-}r$, those growing without B are $b^+ str\text{-}r$, and so forth. All of these data can be plotted on a single graph to give a set of curves, as shown in part A of **Figure 10.13**. Four features of this set of curves are notable.

1. The number of recombinants in each curve increases with length of time of mating.
2. For each marker, there is a time (the **time of entry**) before which no recombinants are detected.
3. Each curve has a linear region that can be extrapolated back to the time axis, defining the time of entry of each gene $a^+, b^+, \ldots, e^+$.
4. The number of recombinants of each type reaches a maximum, the value of which decreases with successive times of entry.

The explanation for the time-of-entry phenomenon is the following: Not all donor cells start transferring DNA at the same time, so the number of recombinants increases with time. Transfer begins at the replication origin of F in the Hfr chromosome. Genes are transferred in linear order to the recipient, and the time of entry of a gene is the time at which that gene first enters a recipient in the population. Separation of a mating pair prevents further transfer and limits the number of recombinants seen at a particular time.

The times of entry of the genes used in the mating just described can be placed on a map, as shown in Figure 10.13B. The num-

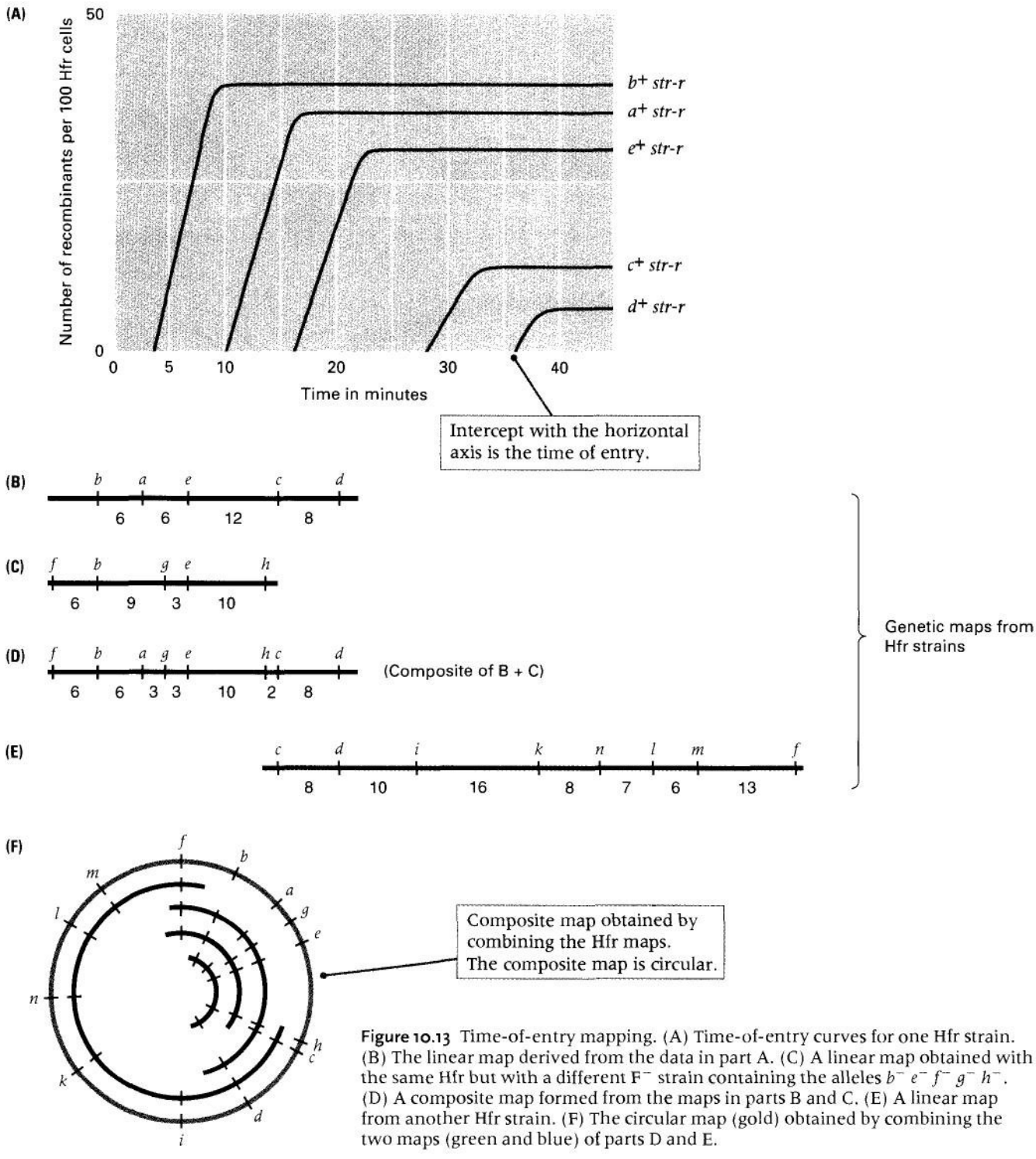


Figure 10.13 Time-of-entry mapping. (A) Time-of-entry curves for one Hfr strain. (B) The linear map derived from the data in part A. (C) A linear map obtained with the same Hfr but with a different F^- strain containing the alleles b^- e^- f^- g^- h^-. (D) A composite map formed from the maps in parts B and C. (E) A linear map from another Hfr strain. (F) The circular map (gold) obtained by combining the two maps (green and blue) of parts D and E.

bers on this map and the others are genetic distances between the markers. In this context, "genetic distance" is measured as minutes between the times of entry. Mating with another F^- with genotype b^- e^- f^- g^- h^- *str-r* could be used to locate the three genes *f*, *g*, and *h*. Data for the second recipient would yield a map such as that shown in Figure 10.13C. Because genes *b* and *e* are common to both maps, the two maps can be combined to form a more comprehensive map, as shown in Figure 10.13D.

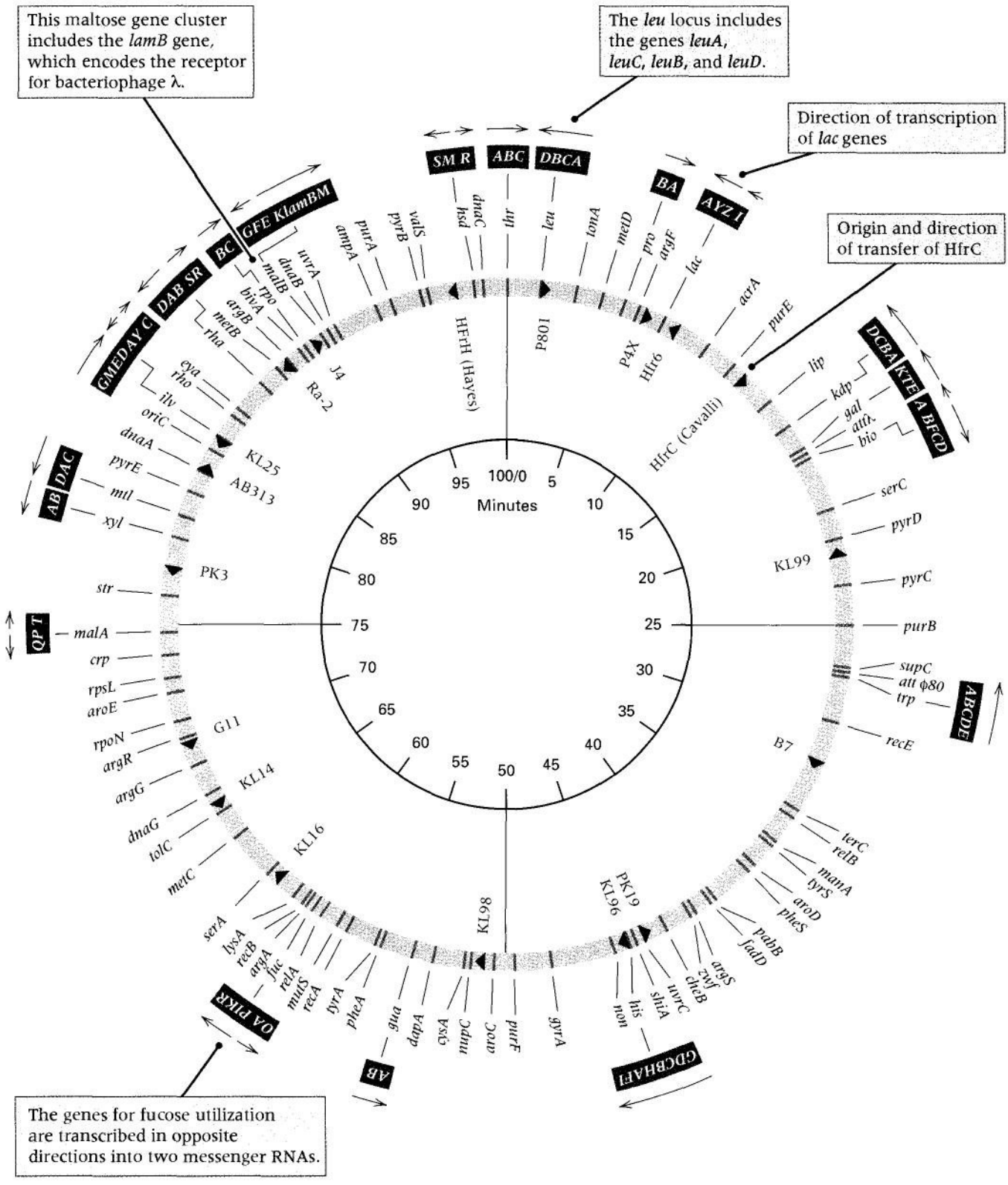


Figure 10.14 Circular genetic map of *E. coli*. Map distances are given in minutes; the total map length is 100 minutes. For some of the loci that encode functionally related gene products, the map order of the clustered genes is shown, along with the direction of transcription and length of transcript (black arrows). The purple arrowheads show the origin and direction of transfer of a number of Hfr strains. For example, HfrH transfers *thr* very early, followed by *leu* and other genes in a clockwise direction.

Studies with different Hfr strains (Figure 10.13E) also are informative. It is usually found that different Hfr strains are distinguishable by their origins and directions of transfer, which indicates that F can integrate at numerous sites in the chromosome and in two different orientations. Combining the maps obtained with different Hfr strains yields a composite map that is *circular*, as illustrated in Figure 10.13F. The circularity of the map is a result of the circularity of the *E. coli* chromosome in F^- cells and the multiple points of integration of the F plasmid; if F could integrate at only one site and in one orientation, the map would be linear.

A great many such mapping experiments have been carried out, and the data have been combined to provide an accurate genetic map of the entire *E. coli* chromosome (**Figure 10.14**). Both the DNA molecule and the genetic map are circular. The entire chromosome requires 100 minutes to be transferred (it usually breaks first), so the total map length is 100 minutes. In the outer circle in Figure 10.14, the arrows indicate the direction of transcription and the coding regions included in each transcript. The purple arrowheads show the origin and direction of transfer of a number of Hfr strains. Transfer from HfrC, for example, goes counterclockwise, starting with *purE acrA lac*.

The region between minutes 84 and 86 of the *E. coli* map is shown in greater detail in **Figure 10.15**. Genes and gene clusters transcribed as a unit are indicated by the arrows pointing in the direction of transcription, shown to the right of the blue line. To the left of this line are either regions that are not transcribed or genes in which the direction of transcription is unknown. The symbol *oriC* represents the origin of DNA replication of the *E. coli* chromosome.

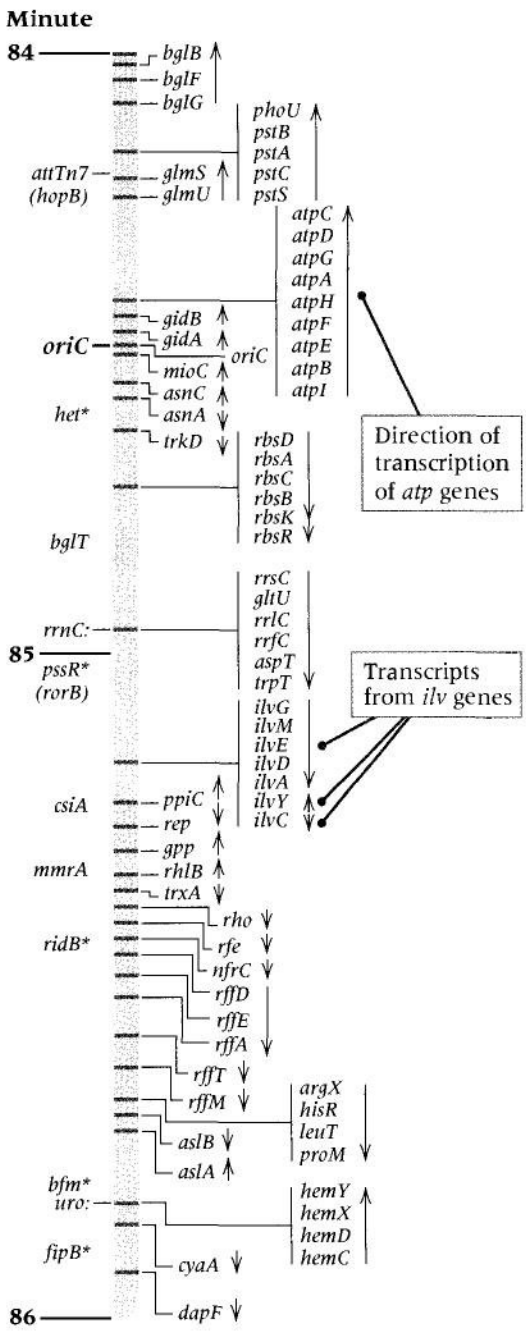


Figure 10.15 A more dense genetic map, showing genes in the region between 84 and 86 minutes of the *E. coli* chromosome. Arrows on the right indicate the direction of transcription and the length of the messenger RNA for each gene or gene cluster. On the left, *oriC* is the origin of replication of the chromosome. [From M. K. B. Berlyn, K. B. Low, and K. E. Rudd. 1996. In *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed. (F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, eds.) Washington, DC: American Society for Microbiology.]

F' Plasmids

Occasionally, F is excised from Hfr DNA by an exchange between the same sequences used in the integration event. However, in some cases, the excision process is not a precise reversal of integration. Instead, breakage and reunion take place between nonhomologous sequences at the boundary of F and nearby chromosomal DNA

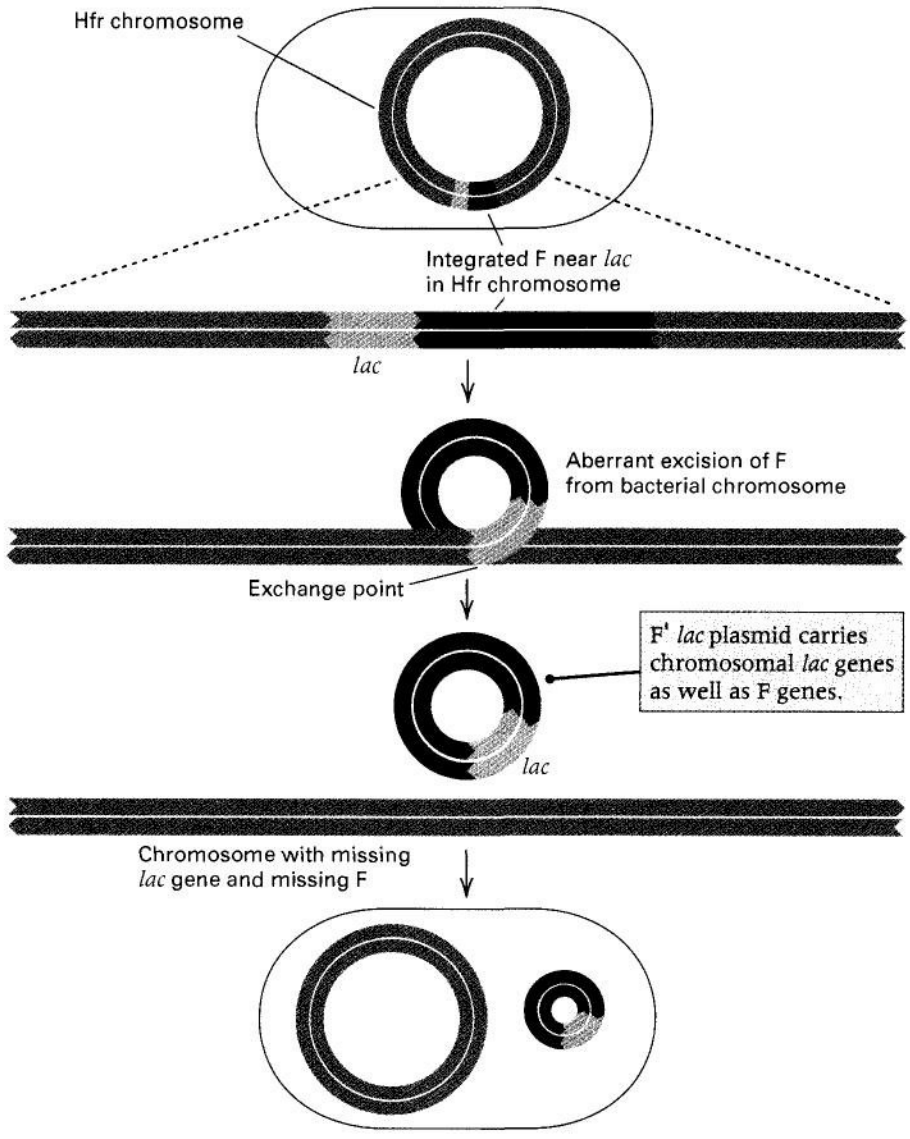


Figure 10.16 Formation of an F' *lac* plasmid by aberrant excision of F from an Hfr chromosome. Breakage and reunion are between nonhomologous regions.

(Figure 10.16). Aberrant excision creates a plasmid containing a fragment of chromosomal DNA, which is called an **F' plasmid** ("F prime"). By the use of Hfr strains having different origins of transfer, F' plasmids with chromosomal segments from many regions of the chromosome have been isolated. These elements are extremely useful because they render any recipient cell diploid for the region of the chromosome carried by the plasmid. These diploid regions make possible dominance tests and gene-dosage tests (studies of the effects on gene expression of increasing the number of copies of a gene). Because only a part of the genome is diploid, cells containing an F' plasmid are **partial diploids,** also called **merodiploids.** We will examine genetic analysis using F' plasmids in Chapter 12 when we discuss the *E. coli lac* genes.

10.5 Transduction

In the process of **transduction,** a bacterial DNA fragment is transferred from one bacterial cell to another by a phage particle containing the bacterial DNA. Such a particle is called a **transducing phage.**

CoNnection

The Sex Life of Bacteria

Joshua Lederberg and Edward L. Tatum
1946
Yale University,
New Haven, Connecticut
Gene Recombination in Escherichia coli

After their discovery in the nineteenth century, bacteria were considered "things apart"—unlike other organisms in fundamental ways. Lederberg and Tatum's discovery of what at first appeared to be a conventional sexual cycle was a sensation, completely unexpected. It meant that bacteria could be considered "genetic organisms" along with yeast, Neurospora, Drosophila, *and other genetic favorites. For this and related discoveries, Lederberg and Tatum were awarded the 1958 Nobel Prize along with George W. Beadle. In this excerpt, you will note that the authors discuss bacterial recombination as requiring a cell fusion that would bring both parental genomes together. This interpretation shows that it is possible to make exactly the right observation, and realize its significance, but not quite grasp what is really going on. The conclusion that bacterial recombination involved unidirectional transfer was reached much later, after the discovery of Hfr strains and the development of the interrupted-mating technique.*

Analysis of mixed cultures of nutritional mutants has revealed the presence of new types which strongly suggest the occurrence of a sexual process in the bacterium *Escherichia coli*. The mutants consist of strains which differ from their parent wildtype strain K-12 in lacking the ability to synthesize growth factors. As a result of these deficiencies they will only grow in media supplemented with their specific nutritional requirements. In these mutants single nutritional requirements are established as single mutational steps under the influence of x-rays or ultraviolet light. By successive treatments, strains with several requirements have been obtained. In the recombination studies here reported, two triple mutants have been used, one requiring threonine, leucine and thiamin, the other requiring biotin, phenylalanine and cystine. The strains were grown in mixed culture in complete medium. The cells were washed with sterile water and inoculated heavily into synthetic agar medium, to which various supplements had been added to allow the growth of colonies of various nutritional types. This procedure readily allows the detection of very small numbers of cell types different from the parental forms. The only new types found in "pure" cultures of the individual mutants were occasional forms which had reverted for a single factor, giving strains which required only two of the original three substances. In mixed cultures, however, a variety of types has been found. These include wildtype strains with no growth-factor deficiencies and single mutant types requiring only thiamin or phenylalanine. . . . These types can most reasonably be interpreted as instances of the assortment of genes in new combinations. In order that various genes may have the opportunity to recombine, a cell fusion would be required. . . . The fusion presumably occurs only rarely, since in the cultures investigated only one cell in a million can be classified as a recombinant type. . . . These experiments imply the occurrence of a sexual process in the bacterium *Escherichia coli*.

These types can most reasonably be interpreted as instances of the assortment of genes in new combinations

Source: Nature 158: 558

The Phage Lytic Cycle

In the **lytic cycle** of a bacteriophage, a phage particle attaches to a host bacterium and injects its nucleic acid into the cell. The phage nucleic acid is transcribed and translated into phage proteins, and it replicates many times. Finally, newly synthesized nucleic acid molecules are packaged into protein shells (forming progeny phage), and then the particles are released from the cell. We have already examined the lytic cycle of phage T2 (see Figure 1.4 on page 6). All types of bacteriophage can undergo a lytic cycle; a phage capable *only* of lytic growth is called **virulent.** T2 is a virulent phage.

Generalized Transduction

Generalized transduction is a process in which a bacteriophage particle can carry a small piece of any part of the host bacterial genome. Types of phage able to carry out this process are known as *generalized transducing phage.* In the lytic cycle, some particles are produced that contain only DNA obtained from the host bacterium, rather

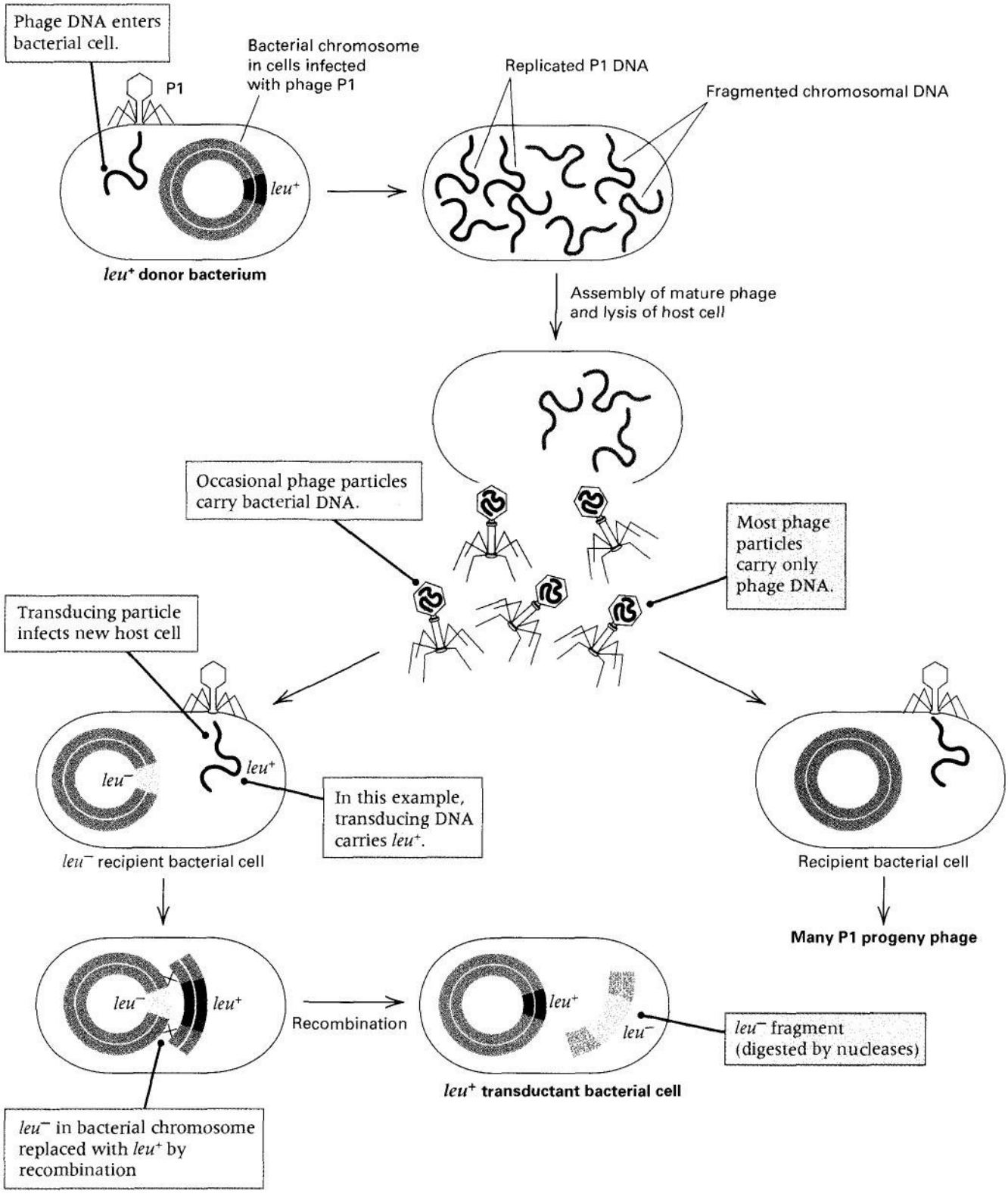


Figure 10.17 Generalized transduction. Phage P1 infects a *leu*$^+$ donor, yielding predominantly normal P1 progeny, but an occasional particle carries *leu*$^+$ (or any other small segment of bacterial DNA) instead of phage DNA. When the progeny population infects a bacterial culture, the transducing particle can yield a transductant. Note that the recombination step requires two exchanges. For clarity, double-stranded phage DNA is drawn as a single line.

than phage DNA. The term *general* is used to emphasize that the bacterial DNA fragment can be derived from *any* part of the bacterial chromosome.

An example of a generalized transducing phage is the *E. coli* phage P1. During the lytic cycle of phage P1, illustrated in **Figure 10.17**, the overwhelming majority of progeny particles contain only phage DNA. However, as part of the process of infection, the phage makes a nuclease that cleaves the bacterial DNA into fragments. Occasionally, single fragments of bacterial DNA comparable in size to P1 DNA are packaged into phage particles in place of P1 DNA (Figure 10.17). The positions of the nuclease cuts in the host chromosome are random, so a transducing particle may contain a fragment derived from any region of the host DNA. A large population of P1 phages will contain a few particles carrying any bacterial gene. On the average, any particular gene is present in roughly one transducing particle per 10^6 progeny phages. When a transducing particle adsorbs to a bacterium, the bacterial DNA contained in the phage head is injected into the cell and becomes available for recombination with the homologous region of the host chromosome. A typical P1 transducing particle contains 100–115 kb of bacterial DNA.

Let us now examine the events that follow infection of a bacterium by a generalized transducing particle obtained by growth of P1 on wildtype *E. coli* containing a leu^+ gene. This is shown at the bottom in Figure 10.17. When such a particle adsorbs to a bacterial cell of leu^- genotype and injects the DNA that it contains into the cell, the cell survives because the phage head contained only bacterial genes and no phage genes. A recombination event exchanging the leu^+ allele carried by the phage for the leu^- allele carried by the host converts the genotype of the host cell from leu^- into leu^+. In such an experiment, typically about 1 leu^- cell in 10^6 becomes leu^+. Such frequencies are easily detected on selective growth medium. For example, if the infected cell is placed on solid medium that lacks leucine, it is able to multiply and a leu^+ colony forms. A colony does not form unless recombination inserted the leu^+ allele.

The fragment of bacterial DNA contained in a transducing particle is large enough to include about 50 genes, so transduction provides a valuable tool for genetic linkage studies of short regions of the bacterial genome. Consider a population of P1 prepared from a leu^+ gal^+ bio^+ bacterium. This sample contains particles able to transfer any of these alleles to another cell; that is, a leu^+ particle can transduce a leu^- cell to leu^+, or a gal^+ particle can transduce a gal^- cell to gal^+. Furthermore, if a leu^- gal^- culture is infected with phage, then both leu^+ gal^- and leu^- gal^+ bacteria are produced. However, leu^+ gal^+ colonies do not arise because the *leu* and *gal* genes are too far apart to be included in the same DNA fragment (**Figure 10.18**, part A).

The situation is quite different with a recipient cell with genotype gal^- bio^-, because the *gal* and *bio* genes are so closely linked that both genes are sometimes present in a single DNA fragment carried in a transducing particle—namely, a gal^+–bio^+ particle (Figure 10.18B). However, not all gal^+ transducing particles also include bio^+, nor do all bio^+ particles include gal^+. The probability of both markers being in a single particle—and hence the probability of simultaneous transduction of both markers (**cotransduction**)—depends on how close to each other the genes are. The closer they are, the greater the frequency of cotransduction.

Cotransduction of the gal^+–bio^+ pair can be detected by plating infected cells on the appropriate growth medium. If bio^+ transductants are selected by spreading the infected cells on a glucose-containing medium that lacks biotin, then both gal^+ bio^+ and gal^- bio^+ colonies will grow. If these colonies are tested for the *gal* marker, then 42 percent are found to be gal^+ bio^+ and the rest gal^- bio^+; similarly, if gal^+ transductants are selected, then about 42 percent are found to be gal^+ bio^+. In other words, the *frequency of cotransduction* of *gal* and *bio* is 42 percent, which means that 42 percent of all transducing particles that contain one gene also include the other.

Studies of cotransduction can be used to map closely linked genetic markers by means of three-factor crosses analogous to those described in Chapter 5. That is, P1 is grown on wildtype bacteria and used to transduce cells that carry a mutation of each of three closely linked genes. Cotransductants containing various pairs of wildtype alleles are examined. The gene

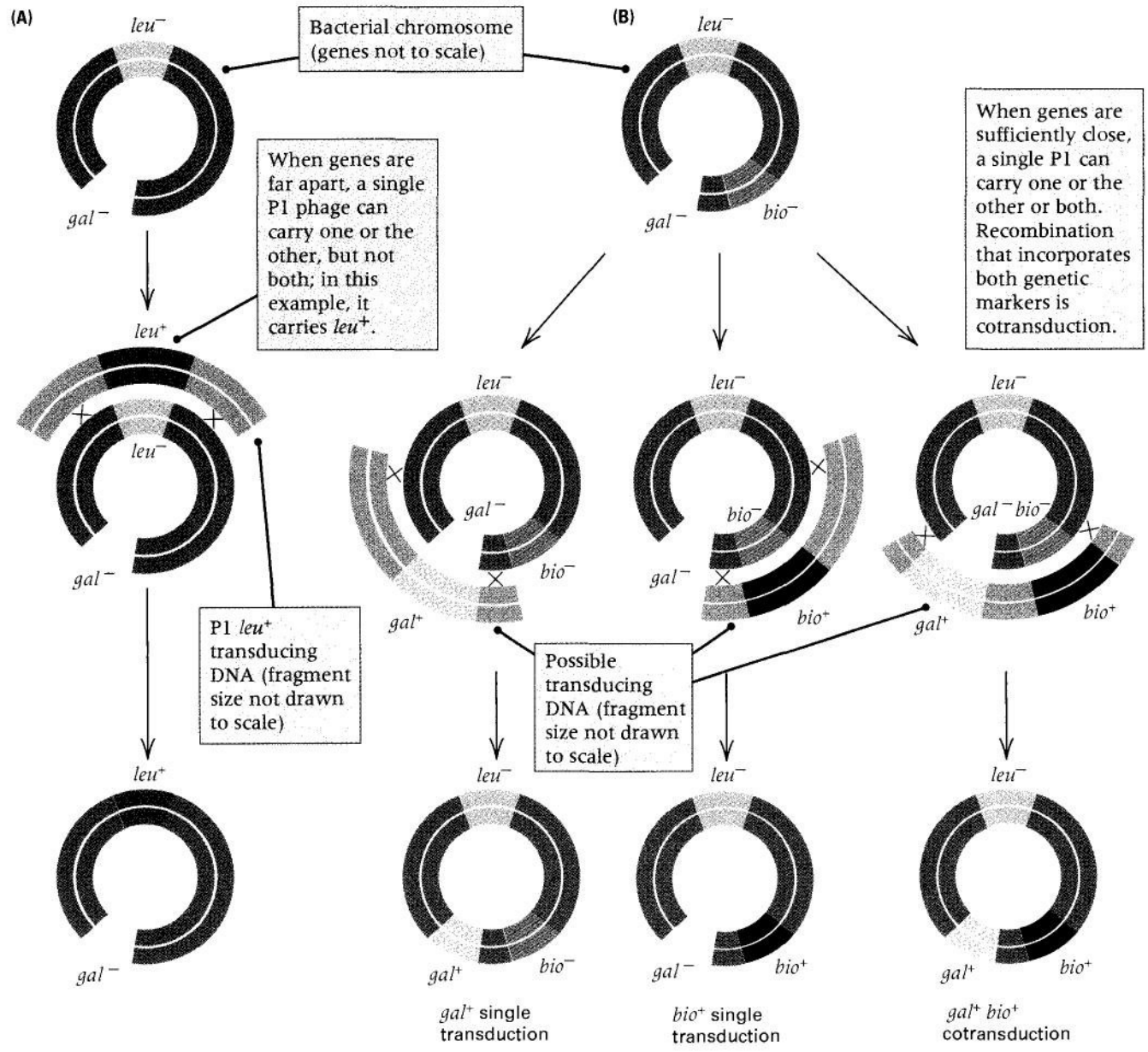


Figure 10.18 Demonstration of linkage of the *gal* and *bio* genes by cotransduction. (A) A P1 transducing particle carrying the leu^+ allele can convert a leu^- gal^- cell into a leu^+ gal^- genotype (but cannot produce a leu^+ gal^+ genotype). (B) The transductants that could be formed by three possible types of transducing particles—one carrying gal^+, one carrying bio^+, and one carrying the linked alleles gal^+ bio^+. The third type results in cotransduction. For clarity, the size of the DNA fragment in a transducing particle, relative to the size of the bacterial chromosome, is not drawn to scale.

located in the middle can be identified because its wildtype allele is nearly always cotransduced with the wildtype alleles of the genes that flank it. For example, in Figure 10.18B, a genetic marker located between gal^+ and bio^+ will almost always be present in gal^+ bio^+ transductants.

How is the frequency of cotransduction between genes related to their map distance in minutes in the standard *E. coli* map? The theoretical relationship depends on the size of a molecule in a transducing phage relative to the size of the entire chromosome. For bacteriophage P1, the distance in min-

CoNnection

Origin of Phage Genetics

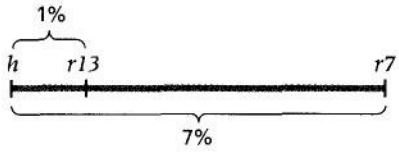
Alfred D. Hershey and Raquel Rotman 1948
Washington University,
St. Louis, Missouri

Genetic Recombination Between Host-range and Plaque-type Mutants of Bacteriophage in Single Bacterial Cells

Are bacteriophages "organisms"? Well . . . yes and no. If by the term organism *you mean a cell or group of cells that function together as a whole to maintain life and its activities, then bacteriophages are not organisms. By* organism *you might also mean "anything resembling a living thing in its complexity of structure and function." In this sense bacteriophages are organisms. Even though they require a bacterial host to survive and reproduce, their reproductive cycle conforms in all respects to that of cellular organisms, including DNA as their genetic material, semiconservative replication of DNA, heritable mutations, and, as shown first in this report, a process of recombination. Hershey and Rotman were quite unsure whether phage recombination resulted from physical exchange of DNA molecules and whether, when two molecules recombined, both reciprocal products of the exchange were produced. Later research showed that the exchanges are reciprocal and that they are physical exchanges.*

We have shown that any two of several mutants of the bacterial virus T2 interact with each other, in bacterial cells infected with both, to give rise to wildtype and double mutant genetic recombinants. . . . In principle, the experimental technique we have to describe is very similar to genetic crossing. One starts with a pair of mutants, each corresponding to a mutant haploid germ cell differing from wildtype by a different unit change. Bacterial cells are infected with both members of the pair, and during viral growth the pair interact to produce viral progeny corresponding to germ cells of a new generation, but now including some individuals differing from wildtype by both unit changes, and other individuals differing from wildtype not at all. The analogy to other genetic recombination is obvious, and it is natural to look for a common mechanism. . . . The crosses between *h* and *r* mutants yield the linkage system shown below.

The analogy to other genetic recombination is obvious, and it is natural to look for a common mechanism.



In the diagram, *h* refers to bacterial host range and *r13* and *r7* to two different, rapidly lysing mutants. The data show that the *h* locus is very closely linked to *r13* (less than 1 percent of wildtype), and that the linkage relation between *r13* and *r7* is 6 percent. . . . The results further show that the two recombinants [wildtype and double mutant] appear in equal numbers in any one cross, and that pairs of reverse crosses yield equal numbers of recombinants. It is these relations that increase the resemblance to simple types of Mendelian segregation. . . . A hypothesis is proposed according to which one visualizes genetic interaction not between two viral particles, but between two sets of independently multiplying chromosome-like structures. Genetic exchange occurs either by reassortment of these structures, or by something like crossing-over between homologous pairs, depending on the structural relation between the genetic factors concerned. The interpretation made brings the linkage relations into superficial agreement with the requirements of linear structure, but there is little evidence that the genetic exchanges are reciprocal, and accordingly little evidence that they are material exchanges.

Source: Genetics 34: 44–71

utes between markers on a conjugative map is related to the frequency of cotransduction approximately as follows:

$$\text{Map distance in minutes} = 2 - 2 \times (\text{Cotransduction frequency})^{1/3} \quad (1)$$

This equation implies that the frequency of cotransduction falls off very rapidly as the distance between the markers increases. In the formula, the shortest distance for which the cotransduction frequency equals 0 is 2 minutes; hence any markers separated by more than 2 minutes will not be cotransduced by phage P1. For *gal* and *bio*, which have a cotransduction frequency of about 42 percent, the distance equals 0.5 minute. (The *gal* and *bio* genes are located at 17.0 and 17.5 minutes, respectively, in the genetic map in Figure 10.14.)

10.6 Bacteriophage Genetics

Except for newly arising mutations, phage progeny from a bacterium infected by a single phage have the same genotype as the parental phage. However, if two phage particles with *different* genotypes infect a single bacterial cell, new genotypes can arise by genetic recombination. This process differs significantly from genetic recombination in eukaryotes in two ways:

1. The number of participating DNA molecules differs from one cell to the next.
2. Reciprocal recombinants are not always recovered in equal frequencies from a single infected cell.

Several phage particles can infect a single bacterium, and the DNA of each particle can replicate. In a multiply infected cell in which a lytic cycle is under way, there are genetic exchanges between phage DNA molecules. If the infecting particles carry different mutations as genetic markers, then recombinant phages result. Recombination can take place anywhere along two homologous DNA molecules.

Plaque Formation and Phage Mutants

Phages are easily detected because, in a lytic cycle, an infected cell breaks open (a process called **lysis**) and releases phage particles to the growth medium. The formation of plaques can be observed as outlined in part A of **Figure 10.19**. A large number of bacteria (about 10^8) are placed on a solid medium. After a period of growth, a continuous turbid layer of bacteria results. If a phage is present at the time the bacteria are placed on the medium, it adsorbs to a cell, and shortly afterward the infected cell lyses and releases many phage. Each of these progeny adsorbs to a nearby bacterium, and after another lytic cycle, these bacteria in turn release phage that can infect still other bacteria in the vicinity. (Progeny phage remain close to their site of origin because their size prevents diffusion in agar medium.) These cycles of infection continue, and after several hours the phage destroy all of the bacteria in a localized area, giving rise to a clear, transparent region—a **plaque**—in the otherwise turbid layer of confluent bacterial growth. Phage can multiply only in growing bacterial cells, so exhaustion of nutrients in the growth medium limits phage multiplication and the size of the plaque. Because a plaque is a result of an initial infection by one phage particle, the number of individual phage originally present on the medium can be counted.

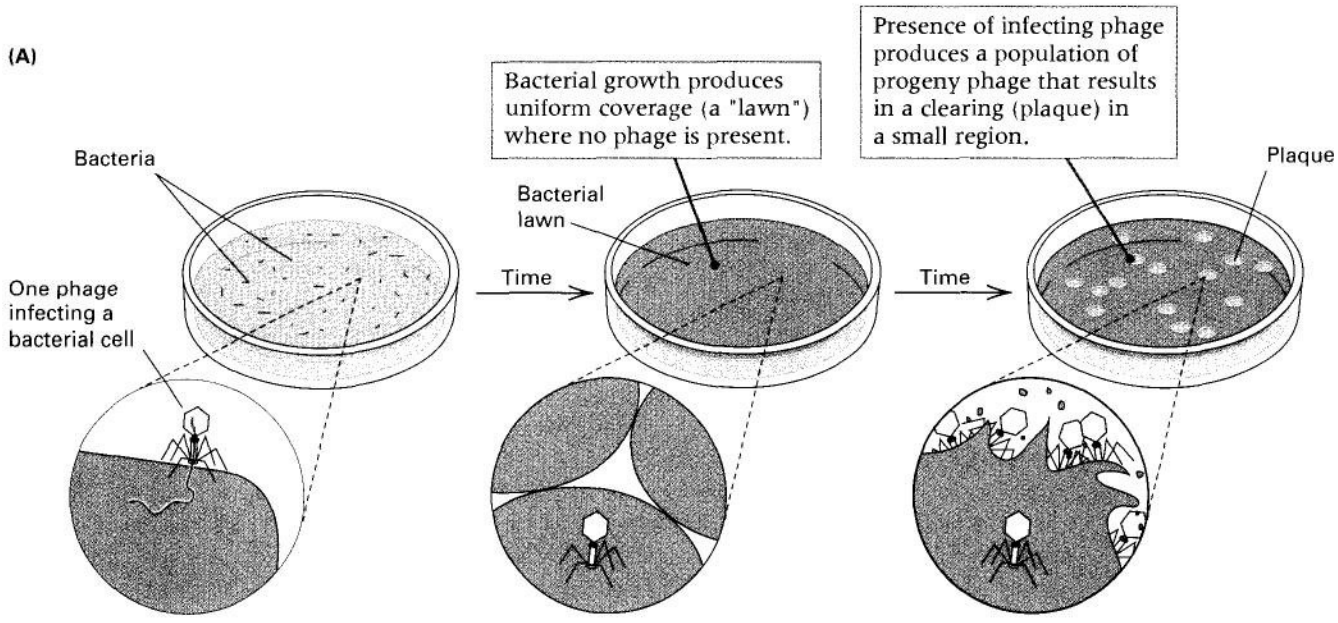
The genotypes of phage mutants can be determined by studying the plaques. In some cases, the appearance of the plaque is sufficient. For example, phage mutations that decrease the number of phage progeny from infected cells often yield smaller plaques. Large plaques can be produced by mutants that cause premature lysis of infected cells, so that each round of infection proceeds more quickly (Figure 10.19B). Another type of phage mutation can be identified by the ability or inability of the phage to form plaques on a particular bacterial strain.

Genetic Recombination in the Lytic Cycle

If two phage particles with different genotypes infect a single bacterium, then some phage progeny are genetically recombinant. **Figure 10.20** shows plaques from progeny of a mixed infection with *E. coli* phage T4 mutants. The r^- (*rapid lysis*) allele results in large plaques, and the h^- (*host range*) allele results in clear plaques. The cross is written as

$$r^- h^+ \text{ (large turbid plaque)} \times r^+ h^- \text{ (small clear plaque)}$$

Four plaque types can be seen in Figure 10.20. Two—the large turbid plaque and the small clear plaque—correspond to the phenotypes of the parental phages. The other two phenotypes—the large clear plaque and the small turbid plaque—are recombinants that correspond to the genotypes $r^- h^-$ and $r^+ h^+$, respectively. When many bacteria are infected, approximately equal numbers of reciprocal recombinant types are usually found among the progeny phage. In an experiment like that shown in Figure 10.20, in which each of the four genotypes yields a different phenotype of plaque morphology, the number of each of



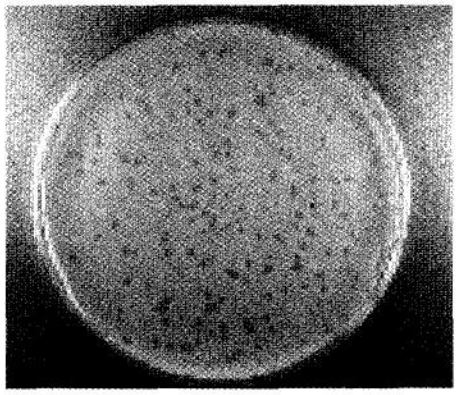


Figure 10.19 Plaque formation. (A) In the absence of a phage, bacterial cells grow and form a translucent lawn. Bacterial cells deposited in the vicinity of a phage are infected and lyse. Progeny phage diffusing outward from the original site infect other cells and cause their lysis, so no bacteria can grow in a small region (the transparent plaque) around the site of each phage particle originally present in the medium. (B) Large plaques in a lawn of *E. coli* formed by infection with a mutant of bacteriophage λ. Each plaque results from an initial infection by a single bacteriophage.

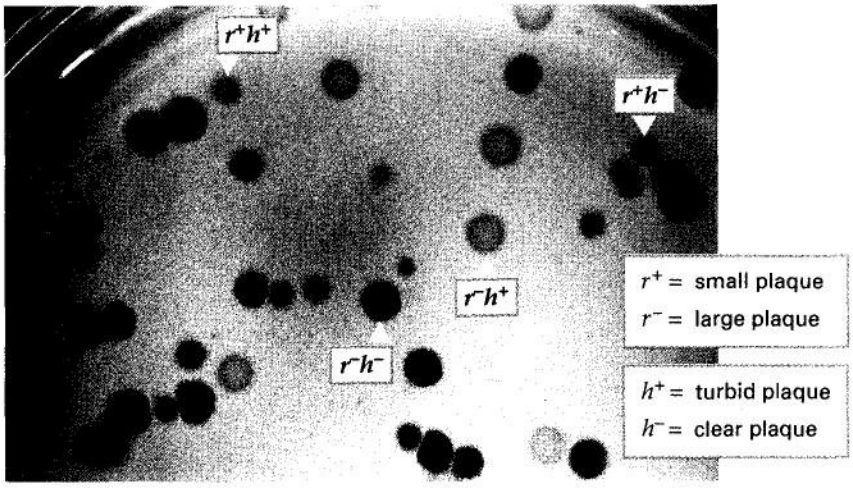


Figure 10.20 A phage cross is performed by infecting host cells with both parental types of phage simultaneously. This example shows the progeny of a cross between T4 phage of genotypes $r^- h^+$ and $r^+ h^-$ when both parental phage infect cells of *E. coli*. The r^+h^+ and r^-h^- genotypes are recombinants. [Courtesy of Leslie Smith and John W. Drake.]

the genotypes can be counted by examining each of the plaques that is formed. The recombination frequency, expressed as a percentage, is defined as

$$\text{Recombination frequency} = \frac{\text{Number of recombinant phage}}{\text{Total number of phage}} \times 100$$

Recombination frequencies can be used to estimate map distances, just as they are in eukaryotes. Early mapping experiments indicated that mutations in T4 mapped in three separate clusters. However, all three clusters showed linkage to one another. In elegant experiments with three-point crosses, George Streisinger and colleagues demonstrated in 1964 that the genetic map for T4 phage is actually circular.

In each cross, they mapped three or four genetic markers with respect to one another and proceeded systematically through the entire T4 genome, eventually demonstrating all of the linkages shown in **Figure 10.21**. Many additional genes were identified and mapped later by other researchers (**Figure 10.22**), and the results were fully consistent with the circular map. In Figure 10.22, the regions indicated in the innermost circle are the three clusters of T4 markers that had initially been identified and mapped. The outermost circle in Figure 10.22 presents a much larger set of markers establishing the overall circularity of the genetic map beyond doubt. The T4 genetic map in Figure 10.22 also indicates that genes in T4 show extensive clustering according to the function for which they are required. For example, there is a large

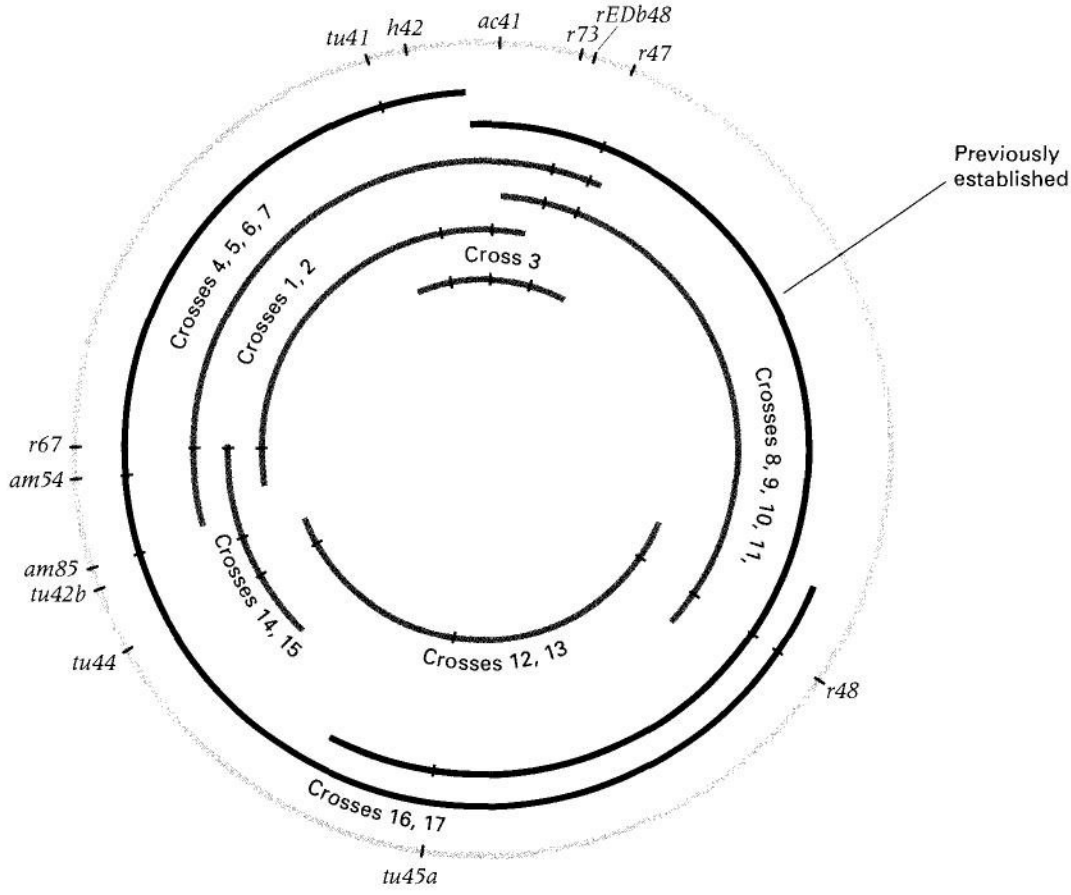


Figure 10.21 Circular genetic map of phage T4. Each arc connects three or four markers that were mapped in the crosses indicated. The circular map results when all data are considered together. [After G. Streisinger, R. S. Edgar, and G. H. Denhardt. 1964. *Proc. Natl. Acad. Sci. USA* 51: 775.]

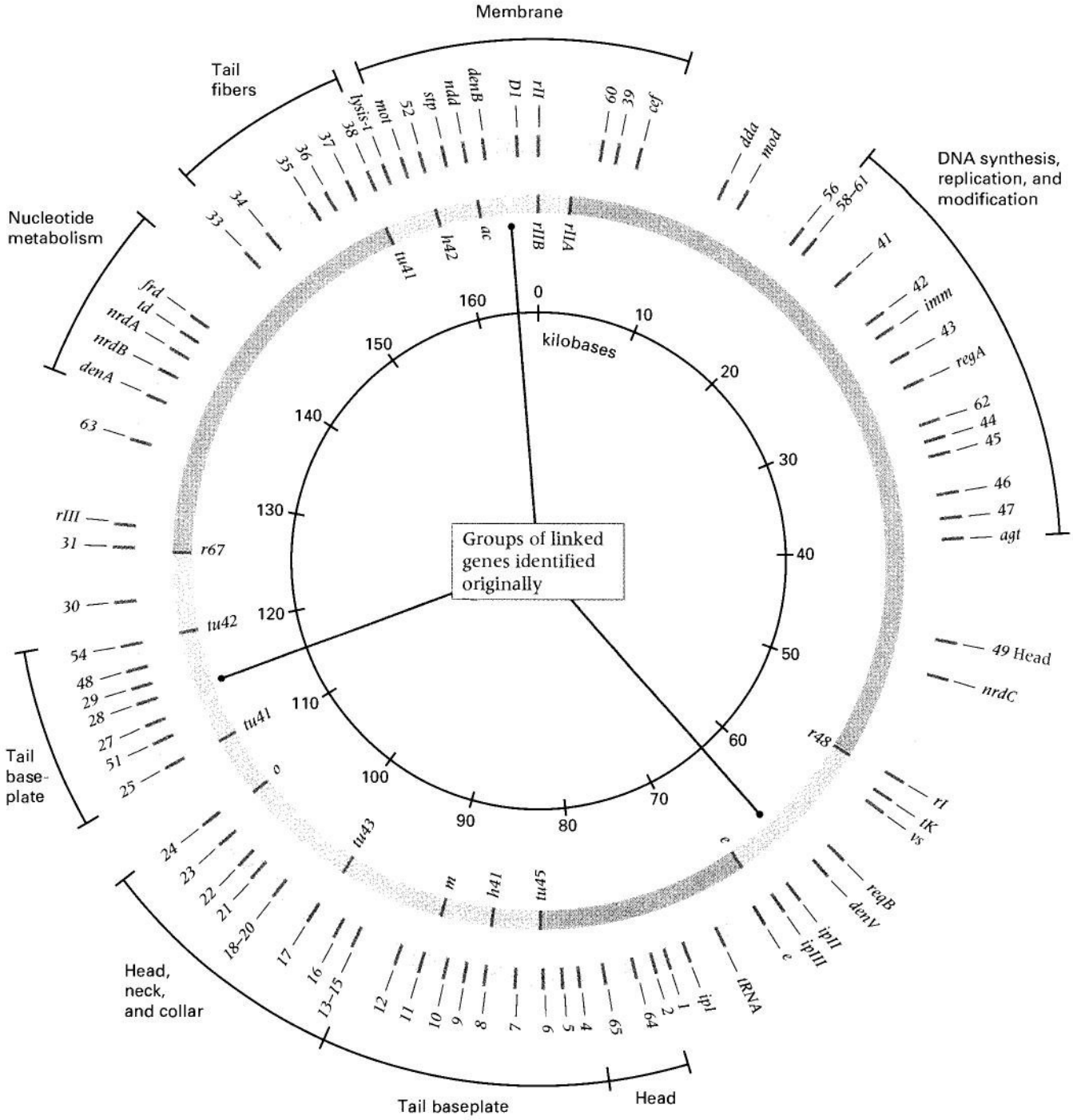


Figure 10.22 More extensive genetic map of bacteriophage T4. The units are kilobases of DNA. The innermost map contains the markers used in the earliest T4 experiments. The outermost map contains many genes (numbered) first identified as temperature-sensitive lethals. Many of the genes show clustering according to function. [After W. B. Wood and H. R. Revel. 1976. *Bacteriol. Rev.* 40:847.]

cluster of genes for DNA replication in the upper right quadrant, and there is a cluster of genes for phage head components near the bottom.

In view of the circular nature of the T4 genetic map, it came as quite a surprise to find that the DNA molecule in a T4 phage particle is a single *linear* molecule. This discovery was completely unexpected and at first seemed inconsistent with the genetic data. However, the discrepancy was resolved by the finding that the very ends of the phage T4 DNA are duplicated, or have **terminal redundancy.** Because of this redundancy, each molecule is about 2 percent longer than would be expected. As the DNA is replicating inside the cell, recombination between each of the duplicated ends of one T4 genome with homologous sequences in other T4 genomes results in products much longer than can be contained in a T4 head (Figure 10.23). Such *concatemeric* molecules are formed because recombination in the T4 genome is very frequent, averaging about 20 recombination events per chromosome. When the DNA is packaged, it is enzymatically cleaved into "headful" packages consisting of about 102 percent of the minimum length of the T4 genome; hence the duplication at the ends. Because of the headful packaging mechanism, each T4 DNA molecule is terminally redundant. Moreover, in a whole population of DNA molecules present in a phage sample, except for the short terminal redundancy, each of the molecules is also a **circular permutation** of the others. (A molecule is a circular permutation of another if the other sequence can be created by joining its ends and cleaving at an internal position.) Because each T4 molecule is a circular permutation, different molecules begin at different points in the DNA sequence, but they always incorporate a little more than one complete genome. The properties of terminal redundancy and circular permutation account for the circular genetic map.

Fine Structure of the *rII* Gene in Bacteriophage T4

In Section 5.6, we described experiments with *Drosophila* that showed that recombination within a gene could take place. A

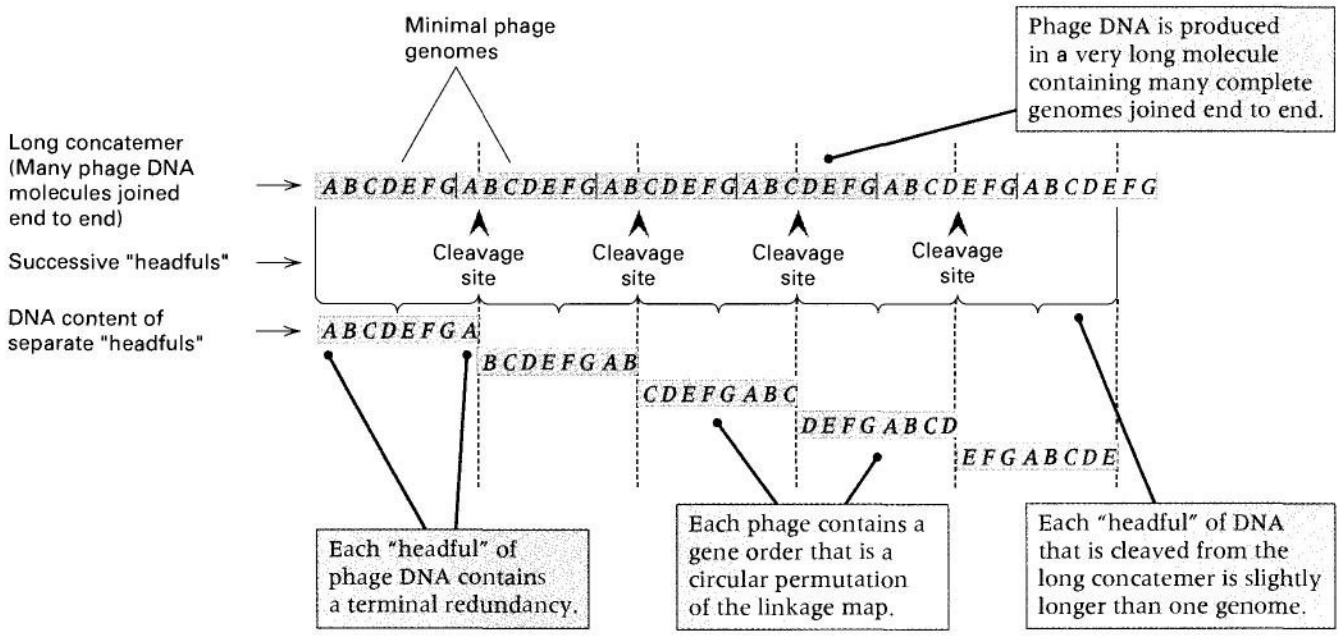
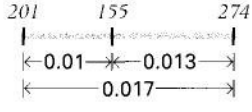


Figure 10.23 Terminal redundancy and circular permutation of the phage T4 linear DNA molecule. A long *concatemer,* or series of molecules linked together, is formed inside the cell by genetic recombination. Each successive headful of DNA is slightly longer than unit size and contains terminal redundancy. Except for the terminal redundancy, the molecules are also circular permutations of each other.

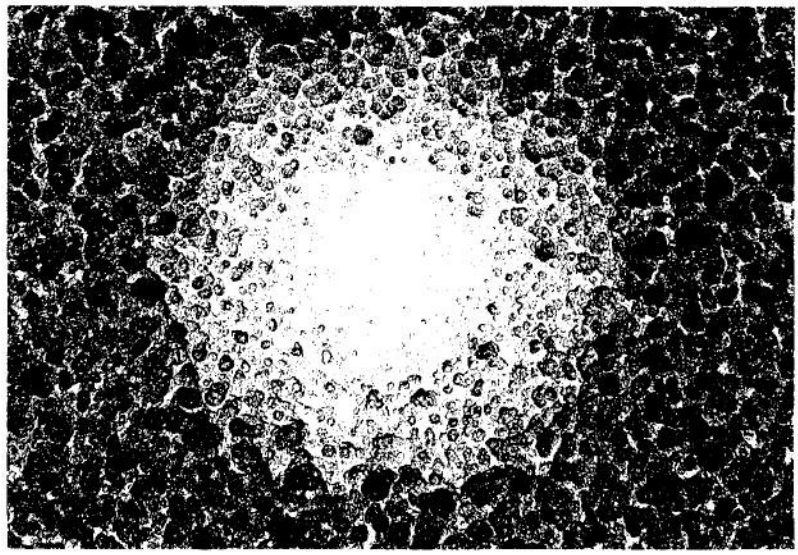
genotype heterozygous for two different mutant alleles of the same gene could, at low frequency, generate recombinant chromosomes carrying either both mutations or neither mutation. These experiments gave the first indication of intragenic recombination and the first evidence that genes have fine structure. Other studies of genes were performed in the years that followed this experiment, but none could equal the fine-structure mapping of the *rII* gene in bacteriophage T4 carried out by Seymour Benzer. Using novel genetic mapping techniques that reduced the number of required crosses from more than half a million to several thousand, Benzer succeeded in mapping 2400 independent mutations in the *rII* locus of phage T4.

Wildtype T4 bacteriophage is able to multiply in *E. coli* strains B and K12(λ) and gives small ragged plaques. Mutations in the *rII* gene of T4 result in large round plaques on strain B but completely prevent T4 from propagating in strain K12(λ). If *E. coli* cells of strain B are all infected with two different *rII* mutants, then recombination between the mutations can be detected, even if the frequency is very low, by taking advantage of the inability of *rII* mutants to grow on K12(λ). Plating the progeny phages on K12(λ) selects for growth of the rII^+ recombinant progeny, because only these recombinants can grow. Furthermore, because very large numbers of progeny phage can be examined (numbers of 10^{10} bacteriophage/ml are not unusual), even very low frequencies of recombination can be detected. Typical results for three different *rII* mutations might yield a map like the following:



where the numbers above the line designate *rII* alleles and the numbers below are map distances given as two times the percent of rII^+, namely

$$\text{Map distance} = \frac{2 \times \left[\begin{array}{c}\text{Number of}\\ \text{plaques per phage}\\ \text{on strain K12(λ)}\end{array}\right]}{\begin{array}{c}\text{Number of plaques}\\ \text{per phage on strain B}\end{array}} \times 100$$



High-resolution image of a plaque formed by a rapid-lysis (*r*) mutant of T4 bacteriophage. Infection occurs only in rapidly dividing bacterial cells. As the lawn of bacteria grows around the plaque, the growth rate of the cells decreases, and the slowly growing cells begin to encroach on the cleared region in the center of the plaque. [Ken Wagner/Visuals Unlimited.]

(In this case the ratio is multiplied by 2 because only the rII^+ recombinants are detected, rather than both reciprocal products of recombination.)

Some mutations failed to recombine with several mutations, each of which recombined with the others. These were interpreted to be deletion mutations, because they prevented recombination with two or more "point" mutations known to be at different sites in the gene. Each deletion eliminated a part of the bacteriophage genome, including a region of the *rII* gene. The use of deletions greatly simplified the ordering and mapping of thousands of mutations.

Figure 10.24 depicts the array of deletion mutations used for mapping. Deletion mapping is based on the presence or absence of recombinants; each cross yields a yes-or-no answer and so avoids many of the ambiguities of genetic maps based on frequencies of recombination. In any cross between an unknown "point" mutation (for example, a simple nucleotide substitution) and one of the deletions, the presence of wildtype progeny means that the point mutation is

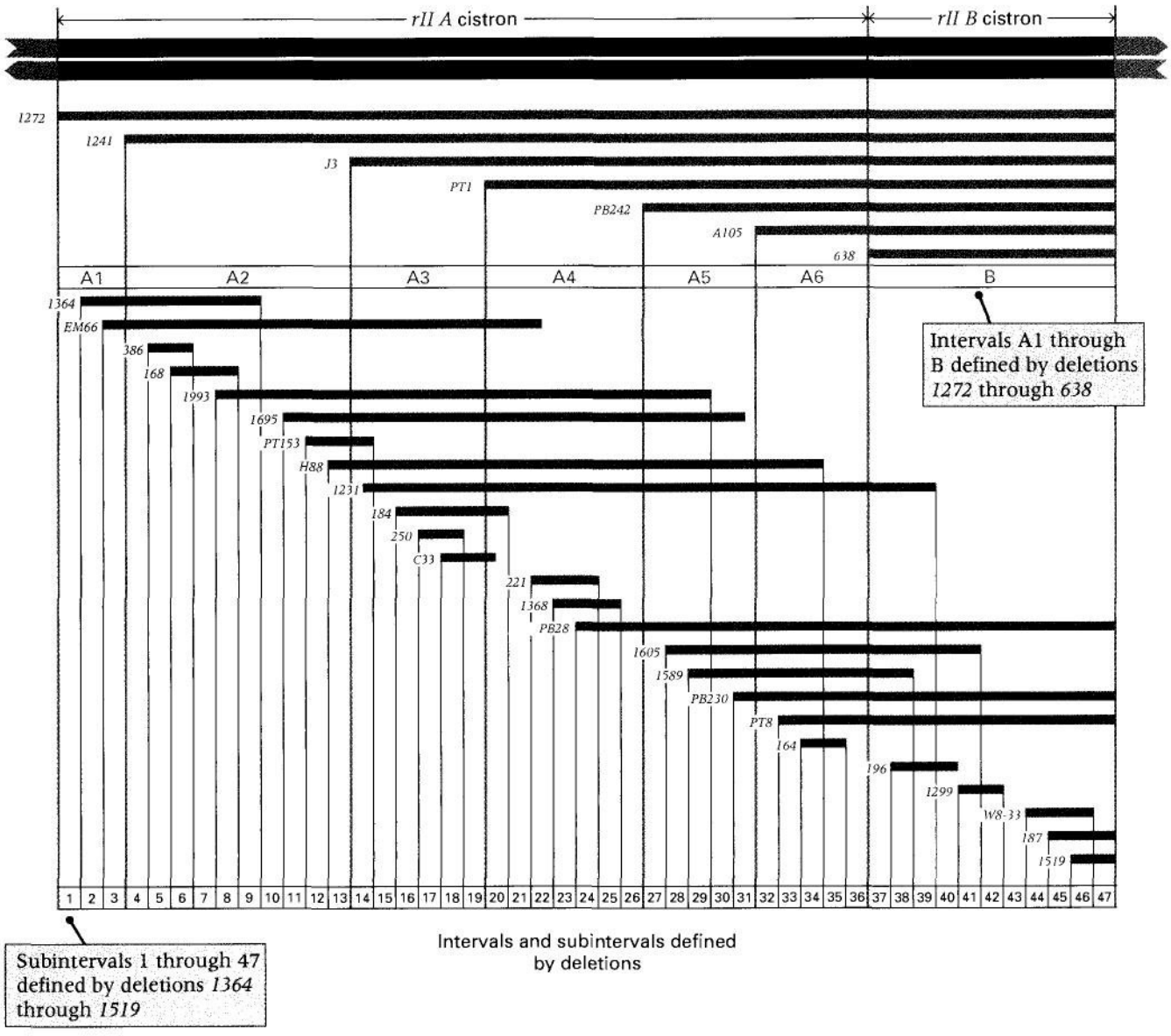


Figure 10.24 The array of deletion mutations used to divide the *rII* locus of bacteriophage T4 into 7 regions and 47 smaller subregions. The extent of each deletion is indicated by a horizontal orange or dark grey bar. Any deletion endpoint used to establish a boundary between regions or subregions is indicated with a vertical line. [After S. Benzer. 1961. *Proc. Natl. Acad. Sci. USA* 47:403.]

outside of the region missing in the deletion. (The reciprocal product of recombination, which carries the deletion plus the point mutation, is not detected in these experiments.) On the other hand, if the point mutation is present in the region missing in the deletion, then wildtype recombinant progeny cannot be produced. Because each cross clearly reveals whether a particular mutation is within the region missing in the deletion, deletion mapping also substantially reduces the amount of work needed to map a large number of mutations.

The series of crosses that would be made in order to map a particular *rII* mutation is presented in **Figure 10.25**, in which the large intervals A1 through A6 plus B are the same as those in Figure 10.24. To illustrate

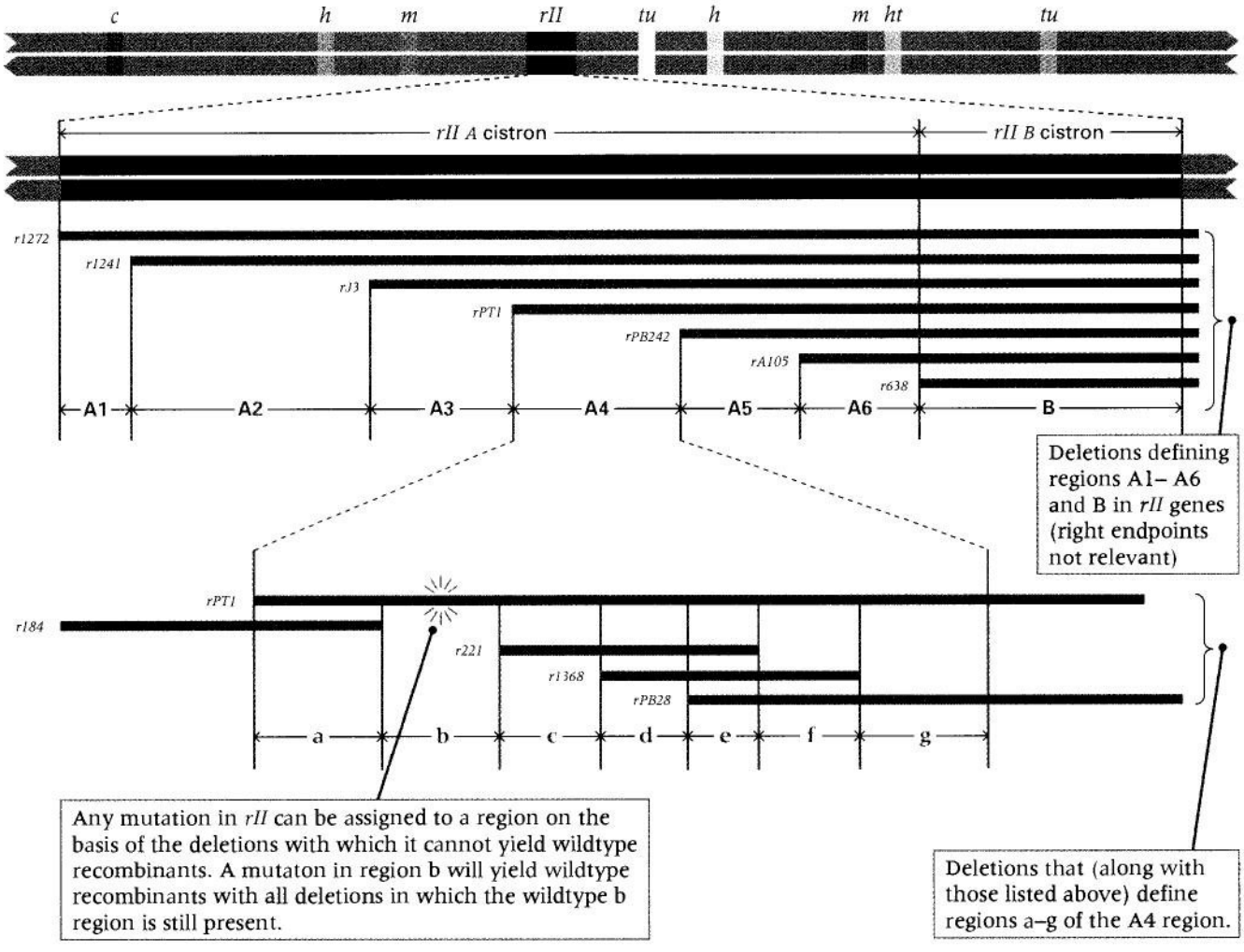


Figure 10.25 The location of the *rII* locus (green) in relation to other genetic markers. Major subdivisions of the *rII* region are defined by the left ends of seven deletion mutations (deleted DNA indicated in orange). All seven deletions extend through the right end of the *rII* region. Further subdivision of the A4 region is made possible by the additional deletions shown at the bottom of the figure (deleted DNA indicated in dark grey). For simplicity, only the deleted regions are shown.

the method, suppose that a particular mutation being examined is located in the region denoted A4. This mutation would fail to yield wildtype recombinants in crosses with the large deletions *r1272*, *r1241*, *rJ3*, and *rPT1*, but it would yield wildtype recombinant progeny in crosses with the large deletions *rPB242*, *rA105*, and *r638*. Conversely, any mutation that yielded the same pattern of outcomes in crosses with the large deletions would be assigned to region A4. Still finer resolution of the genetic map within region A4 is made possible by the set of deletions shown at the bottom of Figure 10.25, the endpoints of which define seven subregions (a through g) within A4. For example, a mutation in region A4 that yields wildtype recombinants with the deletion *r1368* but not with *r221* would be assigned to the c subregion. At the finest level of resolution, mutations within a subregion are ordered via crosses among themselves. In phage T4, mutant sites that are very close can be separated by recombination, because on the average, 1 percent recombination corresponds to a distance of about 100 base pairs. Hence, any two mutations that fail to recombine can be assigned to the same site

within the gene. The genetic map generated for a large number of independent *rII* mutations is given in **Figure 10.26**.

The *rII* mutation and mapping studies were important because they gave experimental support to these conclusions:

- Genetic exchange can take place within a gene and probably between any adjacent base pairs.
- Mutations are not produced at equal frequencies at all sites within a gene. For example, the 2400 *rII* mutations were located at only 304 sites. One of these sites accounted for 474 mutations (Figure 10.26); a site that shows such a high frequency of mutation is called a **hotspot** of mutation. At sites other than hotspots, mutations were recovered only once or a few times.

The *rII* analysis was also important because it helped to distinguish experimentally between three distinct meanings of the word *gene*. Most commonly, the word *gene* refers to a unit of function. Physically, this

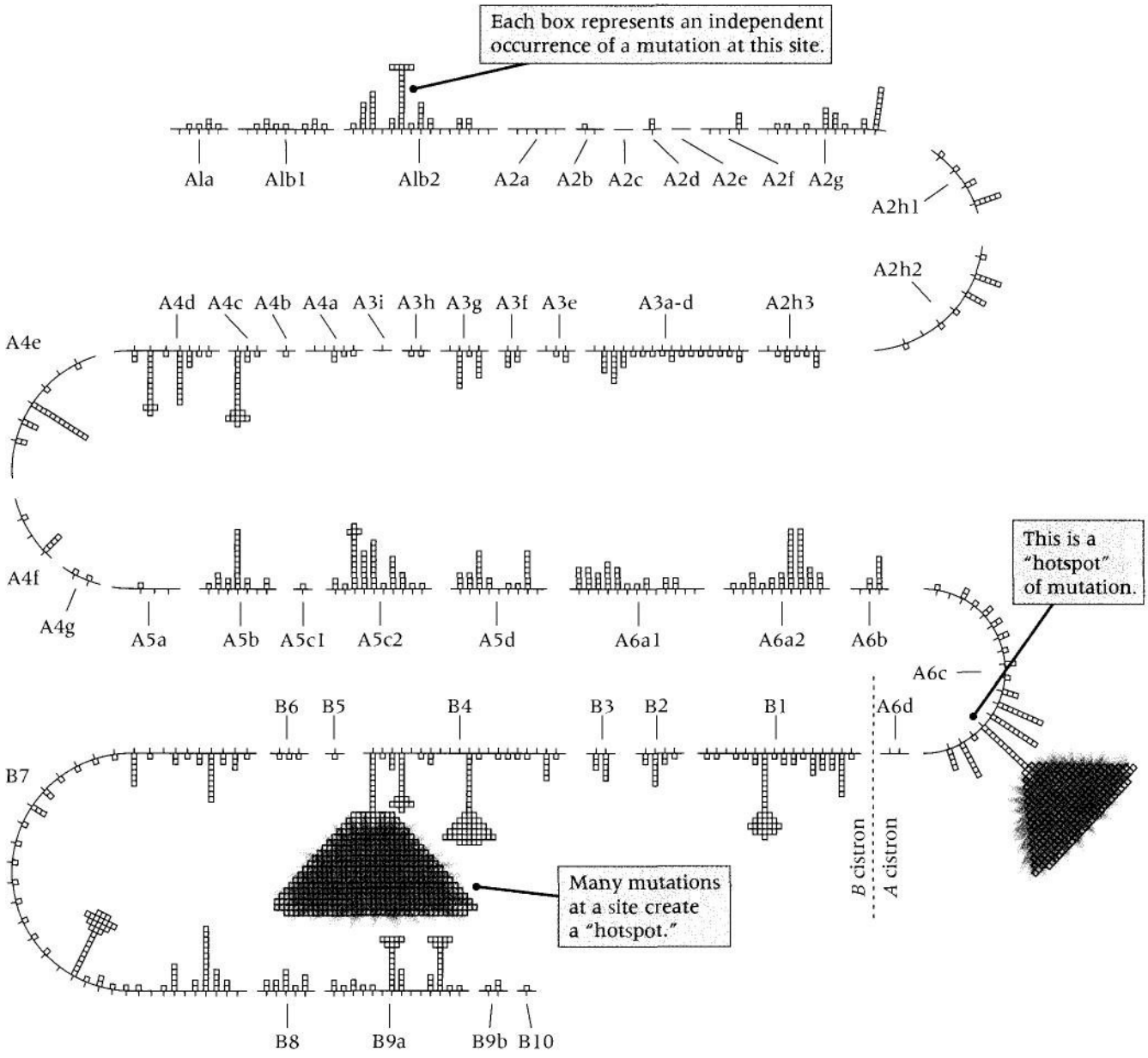


Figure 10.26 Genetic map of the *rII* locus of phage T4. Each small square indicates a separate, independent occurrence of a mutation at the indicated site. The arrangement of sites within each A or B segment is arbitrary. [After S. Benzer. 1961. *Proc. Natl. Acad. Sci. USA* 47:403.]

corresponds to a protein-coding segment of DNA. Benzer assigned the term **cistron** to this unit of function, and the term is still occasionally used. The unit of function is normally defined experimentally by a complementation test (see Section 3.7), and indeed, two units of function, the *rIIA* cistron and the *rIIB* cistron, were defined by complementation. The limits of *rIIA* and *rIIB* are shown in Figures 10.24 and 10.26. The complementation between *rIIA* and *rIIB* is observed when two types of T4 phage, one with a mutation in *rIIA* and the other with a mutation in *rIIB* , are used simultaneously to infect *E. coli* strain K12(λ). The mutations complement: Multiply infected cells produce normal numbers of phage progeny. In contrast, when K12(λ) is simultaneously infected with two phage having different *rIIA* mutations or with two phage having different *rIIB* mutations, no progeny phage are produced.

Besides the meaning of function, clarified by use of the term *cistron*, the term *gene* has two other distinct meanings: (1) the unit of genetic transmission that participates in recombination and (2) the unit of genetic change or mutation. Physically, both the recombinational and the mutational units correspond to the individual base pairs in a gene. Despite potential ambiguity, the term *gene* is still the most important word in genetics, and in most cases, the shade of meaning intended is clear from the context.

10.7 Lysogeny and Specialized Transduction

A genetic map of the *E. coli* phage λ based on recombination in the lytic cycle is shown in **Figure 10.27**. The map is linear, and the DNA molecule in the λ phage particle is also linear.

The complete DNA sequence of the genome of λ phage is known, and many of

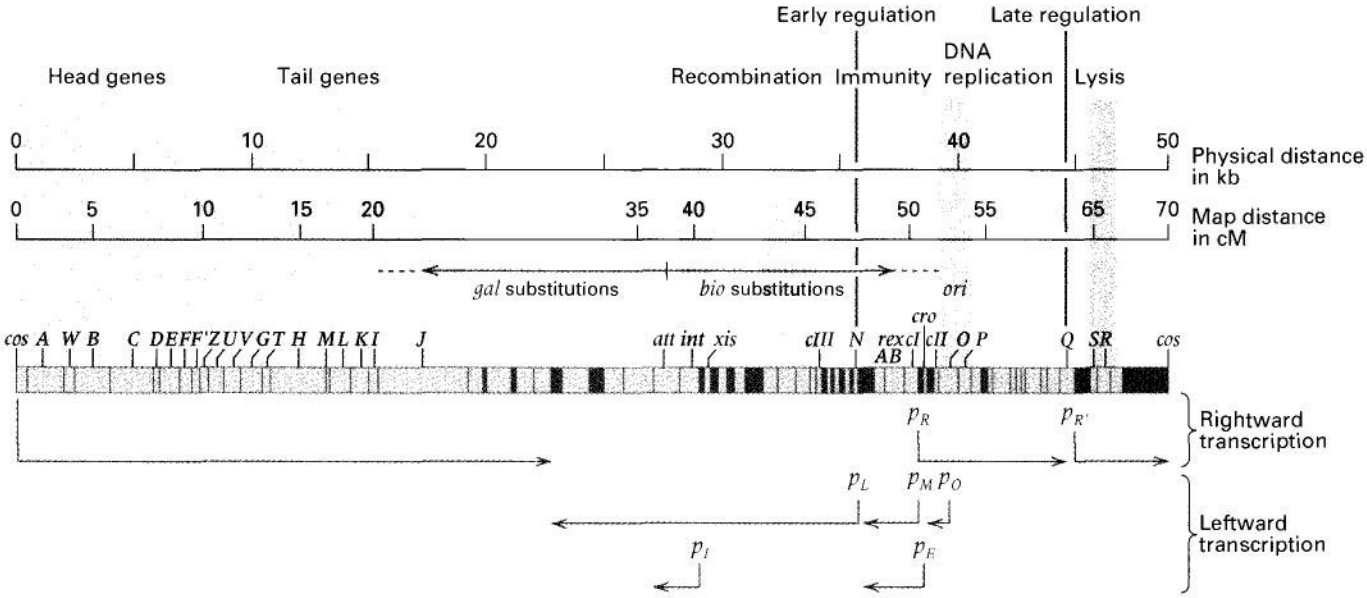


Figure 10.27 Molecular and lytic-cycle genetic map of bacteriophage λ. The molecular map is scaled in kilobase pairs (kb), the genetic map in centimorgans (cM), a unit equivalent in this case to percent recombination. Clusters of genes with related functions are indicated by the pastel-colored boxes. The positions of the regulatory genes *N* and *Q* are indicated by vertical lines at about 35 and 45 kb, respectively. Promoters are indicated by the letter *p*. The direction of transcription and length of transcript are indicated by arrows. Light brown regions are coding sequences; dark brown regions are unlikely to code for proteins. The *att* (attachment site) is the site of recombination for integration of the λ prophage. The origin of replication (*ori*) is the region denoted *O*. The extents of possible substitutions of λ DNA with *E. coli* DNA in the specialized transducing phages λ*dgal* and λ*dbio* are indicated by arrows just above the gene map; such transducing phages are discussed in the section on specialized transduction.

CoNnection Artoo

Seymour Benzer 1955
Purdue University,
West Lafayette, Indiana
Fine Structure of a Genetic Region in Bacteriophage

Just as every fan of "Star Wars" can identify the lovable robot Artoo-Detoo (a.k.a. R2-D2), every geneticist can identify the gene rII. *The* rII *gene in bacteriophage T4 was the first experimental example of genetic fine structure. Benzer used the special property that* rII *mutants cannot grow on* E. coli *strain K but can grow on strain B to examine recombination between different nucleotides within the* rII *gene. He demonstrated that the* rII *gene was divisible by recombination. (It is now known that, in principle, recombination can take place between any adjacent nucleotides.) But if the gene can be subdivided by recombination, then what is a gene, anyway? If two different mutations can undergo recombination whether or not they are in the same gene, then how can one decide, experimentally, whether two different mutations are or are not alleles? Benzer realized that the key experimental operation in the definition of allelism was not recombination but rather the complementation test. This is a rare paper with two great ideas in it: recombination within a gene, and the use of the complementation test to determine experimentally whether two different mutations are or are not alleles of the same gene.*

The phenomenon of genetic recombination provides a powerful tool for separating mutations and discerning their positions along a chromosome. When it comes to very close neighboring mutations, a difficulty arises, since the closer two mutations lie to one another, the smaller is the probability that recombination between them will occur. Therefore, failure to observe recombinant types ordinarily does not justify the conclusion that the two mutations are inseparable. . . . A high degree of resolution can best be achieved if there is available a selective feature for the detection of small proportions of recombinants. Such a feature is offered by the case of the *rII* mutants of T4 bacteriophage described in this paper. The wild-type phage produces plaques on either of two bacterial hosts, *Escherichia coli* strain B or strain K, while a mutant of the *rII* group produces plaques only on B. Therefore, if a cross is made between two different *rII* mutants any wildtype recombinants which arise, even in proportions as low as 10^{-8}, can be detected by plating on strain K. . . . In this way, a series of eight *rII* mutants of T4 have been crossed with each other. The results of these crosses are given in the figure on the facing page.

The functional relatedness of two closely linked mutations causing similar defects may be tested by constructing diploid heterozygotes containing the two mutations.

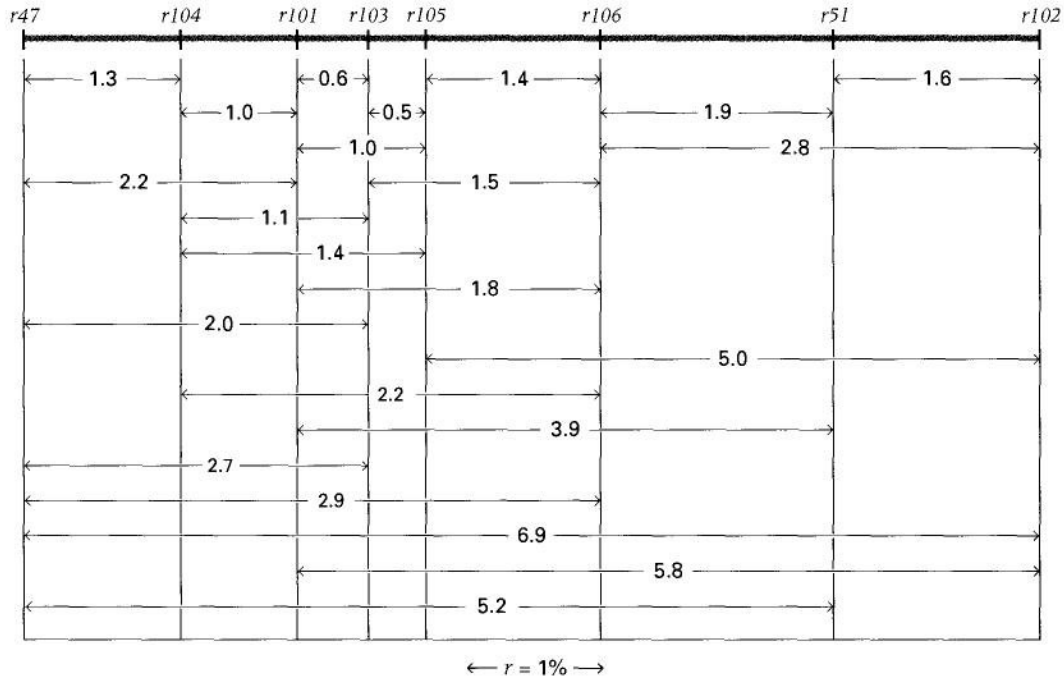
The distances are only roughly additive; there is some systematic deviation in the sense that a long distance tends to be smaller than the sum of its component shorter ones, [which is accounted for by multiple recombination events]. . . . Thus, while all *rII* mutants in this set fall into a small portion of the phage linkage map, it is possible to seriate [order] them unambiguously, and their positions *within* the region are well scattered. . . . *Test for allelism.* The functional relatedness of two closely linked mutations causing similar defects may be tested by constructing diploid heterozygotes containing the two mutations. . . .

the genes and gene products and their functions have been identified. The map of genes in Figure 10.27 shows where each gene is located along the DNA molecule, scaled in kilobase pairs (kb), rather than in terms of its position in the genetic map. However, a genetic map with distances in centimorgans (cM), or percent recombination, has been placed directly below the molecular scale. Comparing the scales reveals that the frequency of recombination is not uniform along the molecule. For example, the 5 map units between genes *H* and *I* span about 3.1 kb, whereas the 5 map units between the genes *int* and *cIII* span about 4.8 kb.

The λ map is also interesting for what it implies about the evolution of the phage genome. The genes in λ show extensive clustering by function. The left half of the map consists entirely of genes whose products (head and tail proteins) are required for assembly of the phage structure, and within this region the head genes and the tail genes themselves form subclusters. The right half of the λ genome also shows sev-

The *trans* form, containing one of the mutations in each chromosome, may or may not produce the wild phenotype. If it does [complementation], it is concluded that the two mutations in question are located in separate functional units. [They are alleles of different genes.] . . . In order to characterize a unit of genetic "function," it is necessary to define what function is meant. . . . On the basis of phenotype tests of *trans* configuration heterozygotes [complementation tests], the *rII* region can be subdivided into two functionally separable segments. Each segment may have the "function" of specifying the sequence of amino acids in a polypeptide chain.

Source: Proceedings of the National Academy of Sciences of the USA 41: 344–354



eral gene clusters, which include genes for DNA replication, recombination, and lysis. The genes are clustered not only by function but also according to the time at which their products are synthesized. For example, the *N* gene acts early; genes *O* and *P* are active later; and genes *Q*, *S*, and *R* and the head-tail cluster are expressed last. The transcription patterns for mRNA synthesis are thus very simple and efficient. There are only two rightward transcripts, and all late genes except for *Q* are transcribed into the same mRNA.

Site-Specific Recombination and Lysogeny

Infection with the phage λ can have an outcome other than the lytic cycle, however. The phage λ is an example of a **temperate phage,** which means that it can undergo a process called *lysogeny.* In a lysogenic infection, the phage does not multiply, the bacterial cell does not lyse, and no progeny phage are produced. The bacterial cell is seemingly unaffected, and no phage are apparent within it. In fact, in **lysogeny,** what

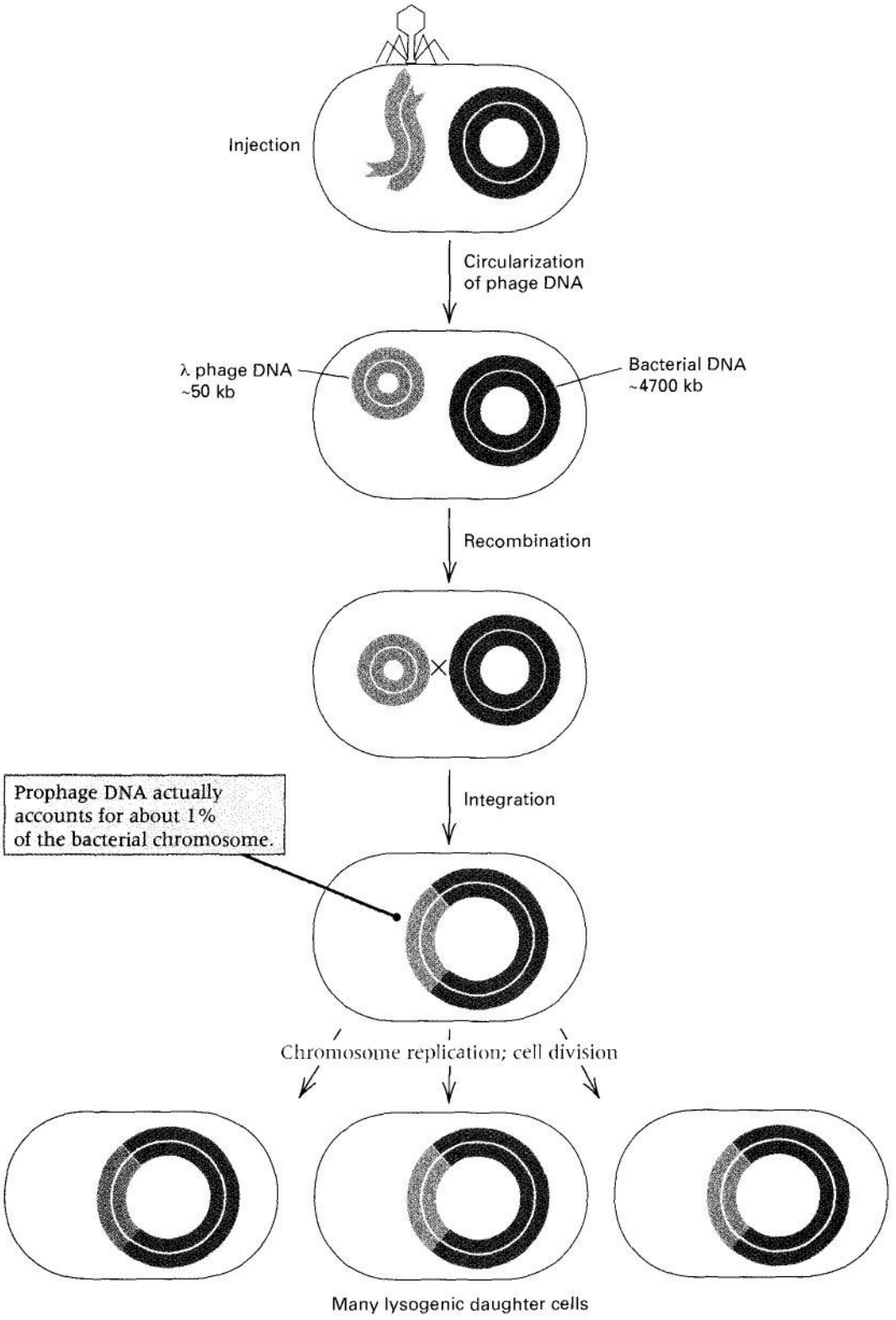


Figure 10.28 Integration of phage DNA into the bacterial chromosome during lysogeny. Genes needed to establish lysogeny are expressed early in infection and then repressed. The integrated region in brown represents the prophage DNA. For clarity, the phage DNA is drawn much larger than to scale; the size of phage λ DNA is actually about 1 percent of the size of the *E. coli* genome.

happens is that the phage DNA becomes integrated into the bacterial chromosome and is replicated passively along with it. As a result, a phage DNA molecule is transmitted to each bacterial daughter cell.

Integration of the phage DNA into the bacterial chromosome is a function of a site-specific recombinase (integrase) encoded in the phage. Each type of temperate phage has its own integration site in the bacterial chromosome. For phage λ the integration site (called *attλ*) is at 17.3 minutes on the genetic map, between *gal* and *bio*, and for phage φ80 (phi 80) the integration site (*attφ80*) is at 28.2 minutes; both sites are shown in the map in Figure 10.14. The integrated phage DNA is called a **prophage,** and the bacterial cell carrying the prophage is called a **lysogen.** A strain lysogenic for λ has the symbol (λ) appended to its name. For example, the strain *E. coli* K12(λ) that made possible the fine-structure analysis of the *rII* region of phage T4 was a K12 strain that had become lysogenic for λ.

Lysogeny comes about when the λ integrase catalyzes a site-specific recombination between *attλ* in the bacterial chromosome and a site called *att* in the middle of the lytic phage map (Figure 10.27). For this mechanism to work, the phage DNA must be in the form of a circle. But in the mature phage, the DNA is a linear molecule, so how is the circular molecule formed? As shown in **Figure 10.28**, the circular molecule is formed inside the bacterial cell immediately after infection. The extreme ends of phage λ DNA are single-stranded, each with 12 unpaired bases, which are complementary in sequence so that they can undergo base pairing with each other to form the circular molecule. The single-stranded ends are called **cohesive ends** to indicate their ability to undergo base pairing. The packaging of DNA in phage λ does not follow a headful mechanism, like T4. Rather, the λ packaging process recognizes specific sequences that are cleaved to produce the cohesive ends. Upon entering the cell, the complementary ends anneal to form a nicked circle, and ligation seals the nicks (**Figure 10.29**). Circularization, which takes place early in both the lytic and lysogenic cycles, is a necessary event in both cycles: for DNA replication in the lytic mode, for prophage integration in lysogeny. In about 75 percent of infected cells, the circular molecule replicates, and the lytic cycle ensues. However, in about 25 percent of infected cells, the circular λ molecule and the circular *E. coli* DNA molecule undergo site-specific recombination catalyzed by the λ integrase, and the phage DNA becomes incorporated into the bacterial chromosome. Because λ can exist either as an autonomous genetic element (in the lytic cycle) or as an integrated element in the chromosome (in lysogeny), λ, like the F factor, is classified as an episome.

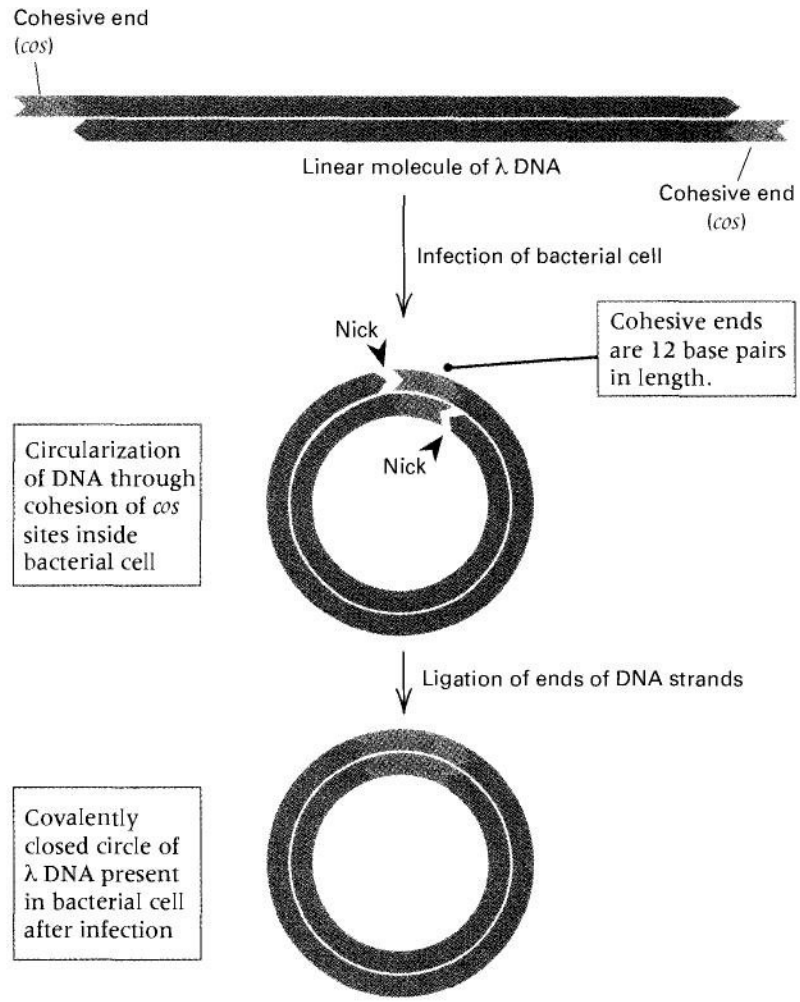


Figure 10.29 A diagram of a linear λ DNA molecule, showing the cohesive ends (complementary single-stranded ends). Circularization by means of base pairing between the cohesive ends forms an open (nicked) circle, which is converted into a covalently closed (uninterrupted) circle by sealing (ligation) of the single-strand breaks. The length of the cohesive ends is 12 base pairs in a total molecule of approximately 50 kb.

The bacterial and phage attachment sites consist of three segments each. The central segment has the same nucleotide sequence in both attachment sites and is the region in which the recombination actually takes place. The phage attachment site is denoted *POP'* (*P* for *phage*), and the bacterial attachment site is denoted *BOB'* (*B* for *bacteria*). The geometry of phage integration is shown in **Figure 10.30**. As a result of the recombination event, the genetic map of the prophage is not the same as the map of the phage but is, rather, a circular permutation of it that arises from the central location of *att* (the *POP'* site) in the phage lytic map and the circularization of the phage DNA.

The correct model for prophage integration was first suggested by Allan Campbell in 1962. The model was confirmed via bacterial crosses of lysogens with nonlysogens, as well as transduction by phage P1. Campbell found that the order of genes in the integrated prophage was

N — *mi R A m6* — *J*

where *m6* is a mutation affecting head formation and *mi* is one affecting lysis. In contrast, the order of genes in the free phage as determined by general recombination is

A m6 — *J att N* — *mi R*

The prophage map is thus a circular permutation of the map of the lytic phage. The prophage is inserted into the *E. coli* chromosome between the genes *gal* and *bio*, as indicated in **Figure 10.31**. An additional finding that confirmed the physical insertion of λ was that the integrated prophage increased the distance between *gal* and *bio*. In a

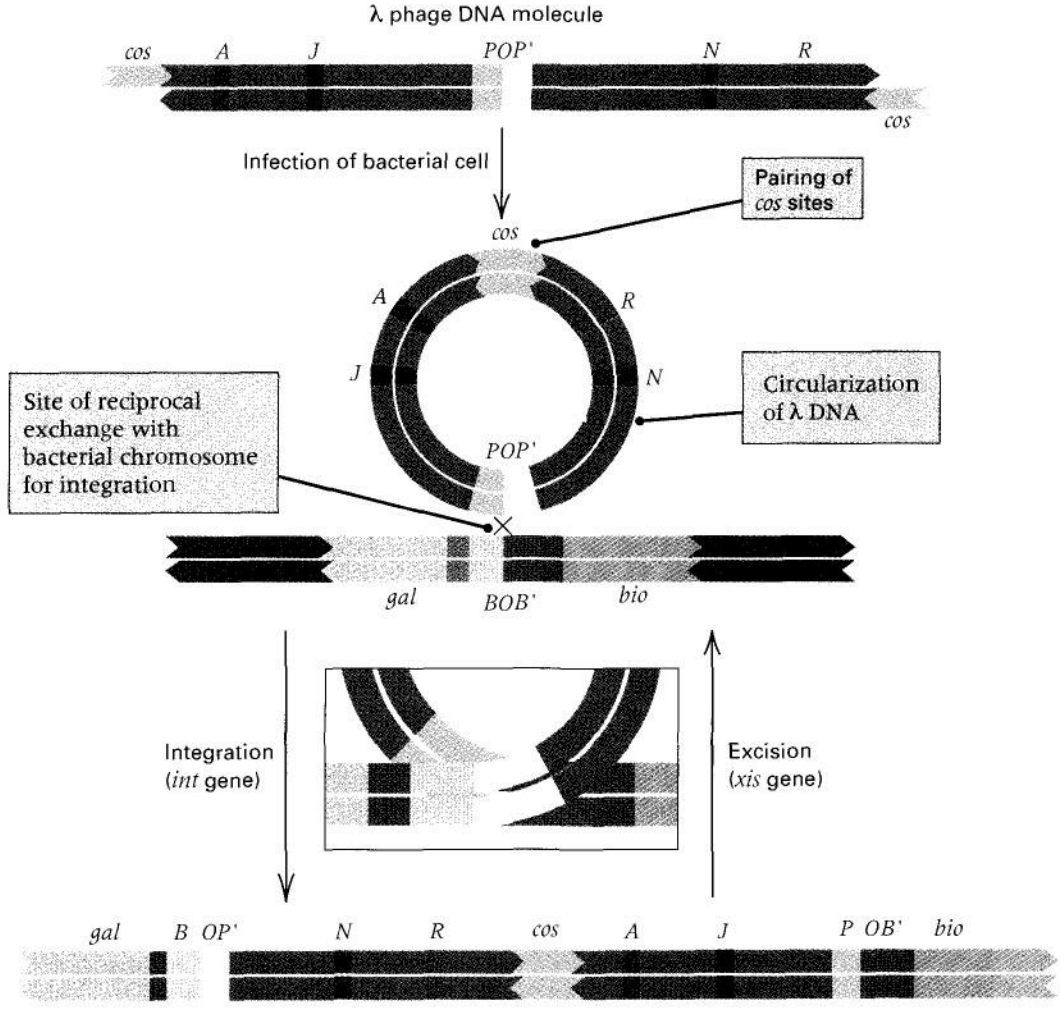


Figure 10.30 The geometry of integration and excision of phage λ. The phage attachment site is *POP'*. The bacterial attachment site is *BOB'*. The prophage is flanked by two hybrid attachment sites denoted *BOP'* and *POB'*.

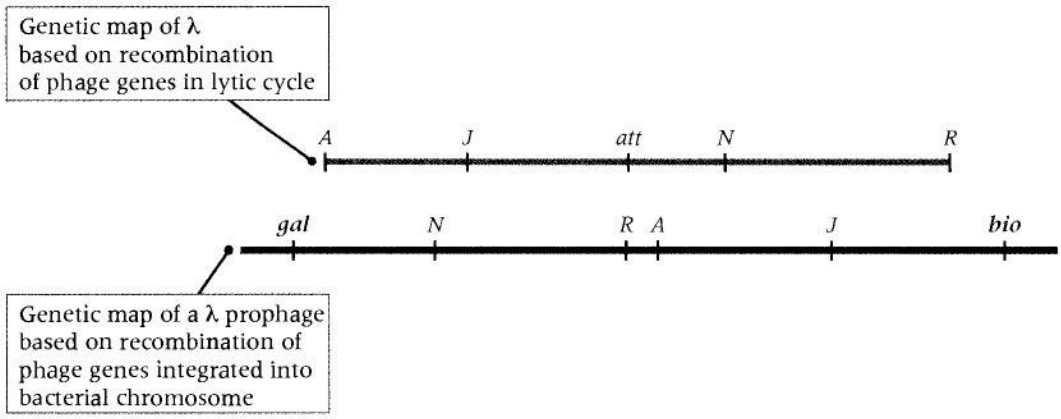


Figure 10.31 The lytic genetic map of phage λ as determined by recombination in the lytic cycle (gold) compared with the prophage genetic map. Only a few genes are shown in order to provide reference markers. The *gal* and *bio* genes (red) are bacterial genes flanking the *attλ* in the *E. coli* chromosome.

nonlysogen, the cotransduction frequency between *gal* and *bio* is about 42 percent, which from Equation (1) corresponds to a map distance of 0.5 minute. In a λ lysogen, the *gal–bio* cotransduction frequency is 1.6 percent, corresponding to a map distance of 1.5 minutes. The 1-minute difference in map distance corresponds to 46 kb, because the complete chromosome of 4600 kb is transferred in 100 minutes; this value is within experimental error of the physical size of λ (50 kb).

When a temperate phage undergoes lysogeny, the phage genes become part of the bacterial chromosome, so it might be expected that the phenotype of the bacterium would change. But most phage genes in a prophage are kept in an inactive state by a **phage repressor** protein encoded in one of the phage genes. The repressor protein is synthesized initially by the infecting phage and then continually by the prophage. The gene that codes for the repressor is frequently the only prophage gene that is expressed in lysogens. If a lysogen is infected with a phage of the same type as the prophage—for example, λ infecting a λ lysogen—then the repressor present within the cell from the prophage prevents expression of the genes of the infecting phage. This resistance to infection by a phage identical with the prophage, which is called **immunity,** is the usual criterion for determining whether a bacterial cell contains a particular prophage. For example, λ will not form plaques on bacteria that contain a λ prophage.

A lysogenic cell can replicate nearly indefinitely without the release of phage progeny. However, the prophage can sometimes become activated to undergo a lytic cycle in which the usual number of phage progeny are produced. This phenomenon is called **prophage induction,** and it is initiated by damage to the bacterial DNA. The damage sometimes happens spontaneously but is more often caused by some environmental agent, such as chemicals or radiation. The ability to be induced is advantageous for the phage because the phage DNA can escape from a damaged cell. The biochemical mechanism of induction is complex and will not be discussed, but the excision of the phage is straightforward.

Excision is another site-specific recombination event that reverses the integration process. Excision requires the phage enzyme integrase plus an additional phage protein called *excisionase*. Genetic evidence and studies of physical binding of purified excisionase, integrase, and λ DNA indicate that excisionase binds to integrase and thereby enables the latter to recognize the prophage attachment sites *BOP'* and *POB'*; once bound to these sites, integrase makes cuts in the *O* sequence and recreates the *BOB'* and *POP'* sites. This reverses the integration reaction, causing excision of the prophage (Figure 10.30).

Specialized Transduction

When a bacterium lysogenic for phage λ is subjected to DNA damage that leads to induction, the prophage is usually excised from the chromosome precisely. However, in about 1 cell per 10^6 to 10^7 cells, an

excision error is made (**Figure 10.32**), and a chance breakage in two nonhomologous sequences takes place—one break within the prophage and the other in the bacterial DNA. The free ends of the excised DNA are then joined to produce a DNA circle capable of replication. The sites of breakage are not always located so as to produce a length of DNA that can fit in a λ phage head, and the DNA may be too large or too small. Sometimes, however, a molecule forms that can replicate and be packaged. In λ lysogens, the prophage lies between the *gal* and *bio* genes, and because the aberrant cut in the host DNA can be either to the right or to the left of the prophage, particles can arise that carry either the *bio* genes (cut at the right) or the *gal* genes (cut at the left). The resulting phage are called λ*bio* and λ*gal* transducing particles. These are *specialized transducing phages* because they can transduce only certain bacterial genes (*gal* or *bio*), in contrast with the P1-type of generalized transducing phage, which can transduce any gene. **Specialized transduction** is therefore a process in which the DNA of an excised prophage can carry a small piece of the host bacterial genome adjacent to the phage attachment site.

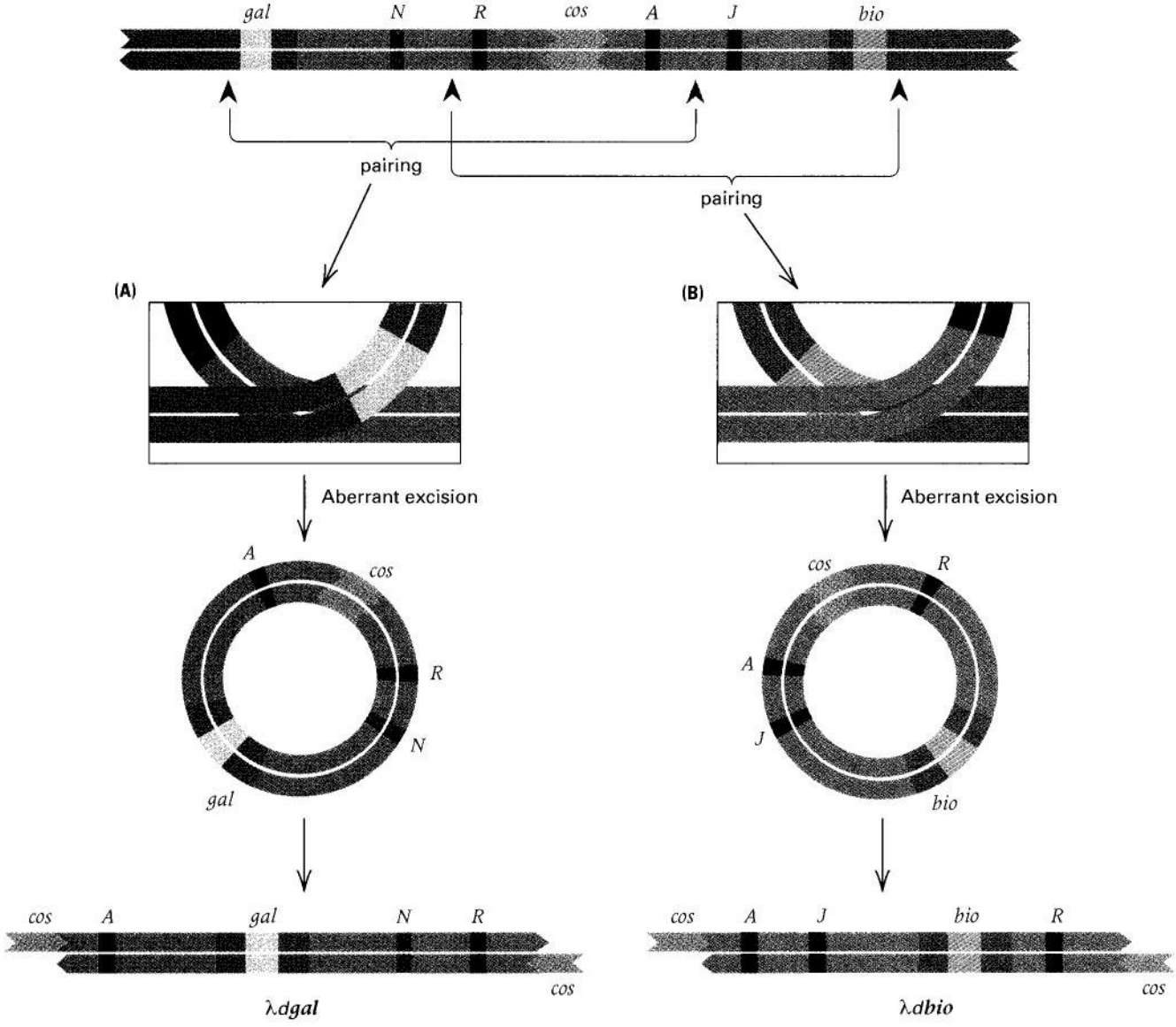


Figure 10.32 Aberrant excision leading to the production of specialized λ transducing phages. (A) Formation of a *gal*-transducing phage (λ*dgal*). (B) Formation of a *bio*-transducing phage (λ*dbio*).

Often the specialized transducing phage are defective, because essential λ genes are replaced by bacterial genes during formation of the *gal*$^+$-bearing or *bio*$^+$-bearing λ molecules. The phage are thus called λ*dgal* (defective, *gal*$^+$-transducing) and λ*dbio* (defective, *bio*$^+$-transducing), respectively. They are unable by themselves to infect *E. coli* productively. Addition of a wildtype helper phage both allows production of a lysate rich in λ*dgal* or λ*dbio* transducing phage and formation of a double lysogen. Presumably, integration of the wildtype λ provides hybrid attachment sites *POB'* and *BOP'* that allow integration of λ*dbio* or λ*dgal* by homologous recombination between identical hybrid attachment sites.

Chapter Summary

Bacterial cells often contain plasmids—nonessential DNA molecules that replicate and are transmitted to progeny cells. Some plasmids, like the F (fertility) factor, are conjugative: They contain genes that make possible their transfer between cells through a tube-like projection. Other plasmids are nonconjugative, but through homologous recombination they can form a cointegrate with a conjugative plasmid and be transferred by a sort of hitchhiking. Most bacteria also possess transposable elements, which are capable of moving from one part of the DNA to another. Transposition does not require sequence homology. Bacterial insertion sequences (IS) are small transposable elements that code for their own transposase and that have inverted-repeat sequences at the ends. Some larger transposons have a composite structure consisting of a central region, often containing one or more antibiotic-resistance genes, flanked by two IS sequences in inverted orientation. DNA sequences called integrons encode a site-specific recombinase (integrase) and can capture circular cassettes of DNA by site-specific recombination between an *attI* site in the integron and an *attC* site in the cassette. Accumulation of antibiotic-resistance transposons and integrons in plasmids has given rise to multiple-drug-resistant R plasmids, which confer resistance to many chemically unrelated antibiotics.

DNA can be transferred between bacteria in three ways: transformation, conjugation, and transduction. In transformation, free DNA molecules, obtained from donor cells, are taken up by recipient cells; by a recombinational mechanism, a single-stranded segment becomes integrated into the recipient chromosome, replacing a homologous segment.

In conjugation, donor and recipient cells pair, and a single strand of DNA is transferred by rolling-circle replication from the donor cell to the recipient. Transfer is mediated by cointegrate formation between the F plasmid and the bacterial chromosome, creating an Hfr (high frequency of recombination) donor strain, and transfer begins at the replication origin of F and proceeds linearly. Only part of the donor chromosome is usually transferred, and it can be maintained in the recipient only after an exchange event. About 100 minutes are required to transfer the entire *E. coli* chromosome, but the mating cells usually break apart before transfer is complete. The times at which donor markers first enter recipient cells—the times of entry—can be arranged in order, yielding a genetic map that is circular because of the multiple sites at which an F plasmid integrates into the bacterial chromosome. Occasionally, F is excised from the chromosome in an Hfr cell; aberrant excision, in which one cut is made at the end of F and the other cut is made in the chromosome, gives rise to an F' plasmid, which carries and can transfer bacterial genes. The term *episome* applies to DNA elements, such as the F factor, that can be maintained within the cell either as an autonomously replicating molecule or by integration into the chromosome.

In generalized transduction, a transducing phage infects a donor cell, fragments the host DNA, and packages fragments of host DNA into phage particles. The transducing particles contain no phage DNA but can inject donor DNA into a recipient bacterium. By genetic exchange, the transferred DNA can replace homologous DNA of the recipient, producing a recombinant bacterium called a transductant.

When bacteria are infected with several phage, genetic exchange can take place between phage DNA molecules, generating recombinant phage progeny. Measurement of recombination frequency yields a genetic map of the phage. A common feature of these maps is clustering of phage genes with related functions. The genetic map of bacteriophage T4, which infects cells of *E. coli*, is circular, but the genome consists of a linear DNA molecule. The circularity of the genetic map results from a packaging mechanism in which successive "headfuls" of DNA, each slightly longer than the minimum size required to contain all genes, are cleaved from a long concatemer of phage genomes. In a population of phage, the DNA molecules are related as circular permutations, and every DNA molecule has a short duplication at each end. Bacteriophage T4 undergoes efficient recombination. Fine-structure genetic analysis of the *rII* region has led to the discovery of hotspots of mutation and to clarification of the term *gene* as meaning (1) the unit of function (cistron) defined by a complementation test, (2) the unit of genetic transmission that participates in recombination, or (3) the unit of genetic change or mutation.

In contrast to virulent phage, such as T4, temperate phage (including phage λ) possess mechanisms for recombining phage and bacterial DNA. A bacterium that contains integrated phage DNA is called a lysogen, the integrated phage DNA is a prophage, and the overall phenomenon is lysogeny. Integration results from

site-specific recombination between attachment sites in the bacterial and phage DNA. The bacterial and phage attachment sites are not identical but have a short common sequence in which the recombination takes place. Temperate phage circularize their DNA after infection. Because the phage has a single attachment site, which is not terminally located, the order of the prophage genes is a circular permutation of the order of the genes in the phage particle. The integrated prophage DNA is stable, but if the bacterial DNA is damaged, then prophage induction is initiated, the phage DNA is excised, and the lytic cycle of the phage ensues. Defective excision of a prophage yields a specialized transducing phage capable of transducing bacterial genes on either side of the phage attachment site.

Key Terms

auxotroph
cassette
circular permutation
cistron
cohesive end
cointegrate
conjugation
conjugative plasmid
cotransduction
cotransformation
counterselected marker
episome
F factor
F' plasmid
generalized transduction
Hfr cell
high-frequency of recombination
hotspot of mutation
immunity
insertion sequence (IS)
integrase
integron
interrupted-mating technique
IS element
lysis
lysogen
lysogeny
lytic cycle
merodiploid
minimal medium
mobile DNA
nonselective medium
partial diploid
phage repressor
plaque
plasmid
prophage
prophage induction
prototroph
R plasmid
selected marker
selective medium
site-specific recombinase
specialized transduction
temperate phage
terminal redundancy
time of entry
transducing phage
transduction
transformation
transposon
virulent phage

Review the Basics

- What is a plasmid? a conjugative plasmid? a nonconjugative plasmid? a cointegrate?
- What is an integron? a cassette in an integron? an integrase?
- Name three phenotypes of mutations that are particularly useful for genetic analysis in bacteria.
- Distinguish between a selective medium and a nonselective medium, and give an example of each.
- What growth medium would you use to distinguish a Lac^+ Amp-r bacterial strain from one that is Lac^- Amp-s? (Amp-r and Amp-s denote resistance or sensitivity to the antibiotic ampicillin.)
- What is the physical basis of cotransformation? If two genes can be cotransformed, what does this observation imply about the ability of each gene to be cotransformed with a genetic marker located between them?
- How does the physical state of the F factor differ between an F^+ bacterial cell and an Hfr bacterial cell?
- In a mating between Hfr and F^- cells, why does the recipient cell usually remain F^-?
- Transfer of DNA in conjugation is always accompanied by DNA replication. What is the mode of replication and in which cell does it take place?
- What is the difference between a selected marker and a counterselected marker? Why are both necessary in analyzing the progeny of an Hfr $\times$ F^- mating?
- When an Hfr bacterial cell transfers its chromosomal DNA into an F^- recipient cell, where in the chromosome does the transfer process begin? How long does it take to transfer the entire chromosome? Why does complete chromosome transfer happen only rarely?
- How does the process of transduction differ from that of transformation?
- Distinguish between generalized and specialized transduction.
- What prevents a bacterial cell that is lysogenic for a temperate bacteriophage from being reinfected with a phage of the same type?

Guide to Problem Solving

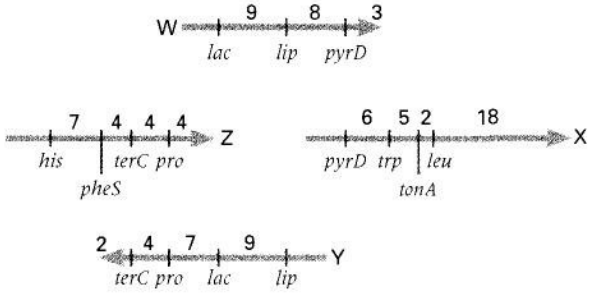
Problem 1 Four independent integrations of the F factor into the chromosome of an unusual strain of *E. coli* yielded four different Hfr derivatives of the strain (HfrW, HfrX, HfrY, and HfrZ), each with a different origin and possibly a different direction of transfer of markers. These were examined in interrupted-mating experiments and were found to transfer chromosomal genes at the times shown in the accompanying table.

(a) Draw a circular genetic map, with position 0 (also 100) minutes at the top, showing the order of the chromosomal genes and the distance (in minutes) between adjacent genes. (The marker *leu* is near 2 minutes on the standard map.) Annotate the genetic map with arrows indicating the origin and direction of transfer of each Hfr and the distance (in minutes) from the origin of transfer to the first marker transferred.

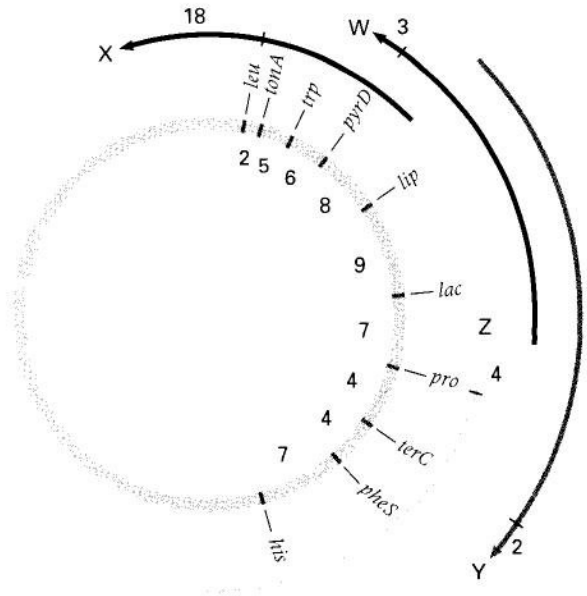
(b) How does the genetic map of this particular strain compare with that of the standard *E. coli* strain in Figure 8.9? Suggest an explanation of any discrepancy between the genetic maps.

	Genetic marker									
Hfr	***his***	***lac***	***leu***	***lip***	***pheS***	***pro***	***pyrD***	***terC***	***tonA***	***trp***
W	–	20	–	11	–	–	3	–	–	–
X	–	–	18	–	–	–	31	–	20	25
Y	–	13	–	22	–	6	–	2	–	–
Z	19	–	–	–	12	4	–	8	–	–

Answer **(a)** Consider each Hfr in turn and, starting with the earliest-entering gene, write down the name of each gene as it enters. The difference in time of entry between adjacent genes is the distance in minutes between the genes. When a partial genetic map of gene transfer from each Hfr has been made, arrange the maps so that any shared markers between two or more Hfr strains coincide. This process yields the genetic maps shown below, where the arrows indicate direction of transfer and the numbers are times of transfer in minutes.



Now arrange the composite genetic map in the manner required, in the form of a circle of 100 minutes, with 0 minutes at the top and *leu* at about 2 minutes. The map is as shown in the next comumn. **(b)** Comparison of this map with that in Figure 10.14 indicates that a group of markers, present in a contiguous segment of chromosome, is in the wrong order as compared with the standard map. The simplest explanation is that the strain in question has undergone an inversion of part of the chromosome, with breakpoints in the *tonA–pro* and *trp–terC* intervals in the standard map.



Problem 2 An Hfr *str-s* strain that transfers genes in alphabetical order

$$a^+ b^+ c^+ \ldots x^+ y^+ z^+$$

with *a* transferred early and *z* a terminal marker, is mated with $F^- z^-$ *str-r* cells. The mating mixture is agitated violently 15 minutes after mixing to break apart conjugating cells and then plated on a medium that lacks nutrient Z and contains streptomycin. The *z* gene is far from the *str* gene. The yield of z^+ *str-r* colonies is about one per 10^7 Hfr cells. What are two possible modes of origin and genotypes of such a colony? How could the two possibilities be distinguished?

Answer The time of 15 minutes is too short to allow the transfer of a terminal gene. Thus, one explanation for a z^+ *str-r* colony is that aberrant excision of the F factor occurred in an Hfr cell, yielding an F' plasmid carrying terminal markers. When the F' z^+ plasmid is transferred into an F^- recipient, the genotype of the resulting cell would be F' z^+/z^- *str-r*. Alternatively, a z^+ *str-r* colony can originate by reverse mutation of $z^- \rightarrow z^+$ in the F^- recipient. The possibilities can be distinguished because the F' z^+/z^- *str-r* cell is able to transfer the z^+ marker to other cells.

Problem 3 Cotransduction experiments were carried out to determine the order of the closely linked genes *tolC*, *metC*, and *ebg* in the chromosome of *E. coli*. P1 phage grown on a strain of genotype

$$tolC^+\ metC^+\ ebg^+$$

were used to transduce a recipient strain of genotype

$$tolC^-\ metC^-\ ebg^-$$

The results were as shown in the accompanying table. What order of the genes is consistent with these results?

Selected marker	Genotypes of unselected markers	Observed percent
tolC$^+$	*metC*$^+$ *ebg*$^+$	2
	metC$^+$ *ebg*$^-$	12
	metC$^-$ *ebg*$^+$	30
	metC$^-$ *ebg*$^-$	56
metC$^+$	*tolC*$^+$ *ebg*$^+$	1
	tolC$^-$ *ebg*$^+$	0
	tolC$^+$ *ebg*$^-$	34
	tolC$^-$ *ebg*$^-$	65

Answer The gene order is specified if we can deduce the gene in the middle. One approach is to note that the wild-type allele of the gene in the middle will be cotransduced at high frequency (usually greater than 90 percent) when the flanking markers are both transduced. Consider the results when *tolC*$^+$ is the selected gene: Among the *tolC*$^+$ *metC*$^+$ transductants, 2/14 are *ebg*$^+$, which suggests that *ebg* is not in the middle; among the *tolC*$^+$ *ebg*$^+$ transductants, 2/32 are *metC*$^+$, which suggests that *metC* is not in the middle. These comparisons would suggest that *tolC* is the gene in the middle, but it is important to see whether this hypothesis is consistent with the other results. Among the *metC*$^+$ *ebg*$^+$ transductants, 1/1 are *tolC*$^+$, but this is too little information for us to make an inference. However, among the *metC*$^+$ *tolC*$^+$transductants, only 1/35 are *ebg*$^+$, which again implies that *ebg* is not in the middle. The data are therefore all consistent with the hypothesis that the gene order is *metC*–*tolC*– *ebg* (or *ebg*–*tolC*–*metC*) and in any case is inconsistent with either of the other genes being in the middle.

Analysis and Applications

10.1 Numbers of phage or bacteria in a suspension are usually so large that single colonies or plaques would be impossible to observe without suitable dilution. The usual dilutions are 100-fold, in which 0.1 ml of the suspension is mixed with 9.9 ml of dilution buffer; or 10-fold, in which 1 ml of suspension is mixed with 9 ml of suspension buffer. Usually *serial dilutions* are necessary, in which the suspension is diluted once into dilution buffer, the resulting suspension mixed thoroughly and diluted a second time, the resulting suspension mixed thoroughly and diluted a third time, and so forth. The serial dilution factors multiply, so, for example, two 100-fold and two 10-fold dilutions yield an overall dilution factor of $100 \times 100 \times 10 \times 10 = 10^6$.

(a) A suspension of bacteriophage λ is serially diluted through four dilutions of 100-fold each and one dilution of 10-fold. From the final dilution, a volume of 0.1 ml is mixed with a great excess of growing bacteria and then spread over nutrient agar in a petri dish and incubated overnight. The next day, 36 plaques are visible. Estimate the number of viable phage per milliliter (the phage *titer*) in the original undiluted suspension.

(b) A suspension of *E. coli* is serially diluted through two dilutions of 100-fold each and two dilutions of 10-fold each. From the final dilution, a volume of 0.1 ml is spread over nutrient agar in a petri dish and incubated overnight. The next day, 42 colonies are visible. Estimate the number of viable bacteria per milliliter in the original undiluted suspension.

10.2 You are given a suspension of bacteria and told that it contains 5×10^7 viable cells per milliliter. What combination of 100-fold and 10-fold serial dilutions would you carry out so that 0.1 ml of the final dilution would contain approximately 50 viable cells?

10.3 Phage T2 (related to phage T4) normally forms small, clear plaques on a lawn of *E. coli strain* B. A mutant of strain B called B/2 ("B bar 2") is unable to adsorb T2 phage particles, so no plaques are formed. T2*h* is a phage with a host-range mutation that is capable of adsorbing to both strains B and B/2, and it forms normal-looking plaques. If *E. coli* strains B and B/2 are mixed in equal proportions and used to generate a lawn, what will be the appearance of plaques made by T2 and T2*h*?

10.4 In the specialized transduction of a *gal*$^-$ *bio*$^-$ strain of *E. coli* using bacteriophage λ from a *gal*$^+$ *bio*$^+$ lysogen, what medium would select for *gal*$^+$ transductants without selecting for *bio*$^+$?

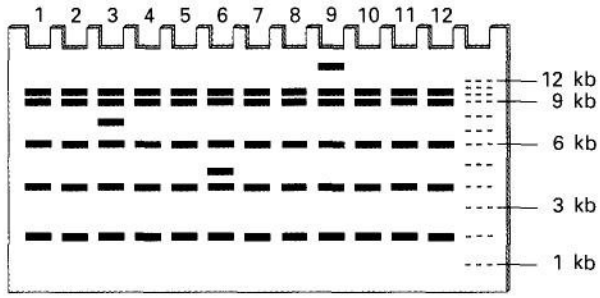
10.5 Given that bacteriophage λ has a genome size of 50 kilobase pairs, what is the approximate genetic length of λ prophage in minutes? (*Hint*: There are 100 minutes in the entire *E. coli* genetic map.)

10.6 How could you obtain a lysate of bacteriophage P1 in which some of the transducing particles contained bacteriophage λ?

10.7 A temperate bacteriophage has gene order *A B C att D E F*. What is the gene order in the prophage?

10.8 Why are λ specialized transducing particles generated only by inducing a lysogen to produce phages rather than by lytic infection?

10.9 A strain of *Escherichia coli*, that was originally isolated from a single colony and had been stored for 50 years in agar in a tightly sealed glass vial, was taken out of the agar and streaked onto nutrient agar to purify single colonies. Each colony represents a separate lineage descended from the original strain. Genomic DNA from 12 colonies was extracted and cleaved with *Eco*RI, the fragments were separated by electrophoresis and transferred to a membrane filter, and the filter was hybridized with a probe for the transposable element known as insertion sequence IS5. (There are no *Eco*RI cleavage sites within the sequence of IS5.) The result is shown in the accompanying diagram.



(a) Explain why some bands are identical across all strains.
(b) Explain why some bands are different.
(c) Estimate the rate of IS5 transposition per lineage per year.
(d) Estimate the rate of IS5 transposition per element per lineage per year.

10.10 If *leu*$^+$ *str-r* recombinants are desired from the cross Hfr *leu*$^+$ *str-s* × F$^-$ *leu*$^-$ *str-r*, on what kind of medium should the matings pairs be plated? Which are the selected and which the counterselected markers?

10.11 A bacteriophage has a genome consisting of a small linear molecule of double-stranded DNA that contains a unique (one per genome) cleavage site for a restriction enzyme in a gene that is essential for phage maturation and lysis. In most strains (type A) of the host bacterium, the phage forms plaques with nearly 100% efficiency (one plaque per infecting phage). However, in a few strains (type B), infection results in almost no plaques. When phage are isolated from the few plaques on type B, they subsequently form plaques with nearly 100% efficiency on type B. DNA analysis of these plaque-forming phage DNA reveals that a cytosine in the unique restriction site has been methylated. These phage also form plaques on type A, but afterwards the progeny phage again have low plaque-forming ability on type B. Explain these results.

10.12 If 10^6 phages are mixed with 10^6 bacteria and all phages adsorb, what fraction of the bacteria remain uninfected?

10.13 In *Escherichia coli*, the genes *lac*$^+$, *gal*$^+$, *trp*$^+$, and *his*$^+$ are necessary for utilization of lactose, utilization of galactose, synthesis of tryptophan, and synthesis of histidine, respectively, and *str-r* and *str-s* are alleles for streptomycin resistance and sensitivity, respectively. You wish to carry out the cross Hfr *lac*$^+$ *gal*$^+$ *trp*$^+$ *his*$^+$ *str-s* × F$^+$ *lac gal trp his str-r*, using each of the markers in turn as selected markers and streptomycin resistance as a counterselected marker. What medium should you use for each of the following purposes?
(a) To select for *lac*$^+$ and counterselect for *str-r*.
(b) To select for *gal*$^+$ and counterselect for *str-r*.
(c) To select for *trp*$^+$ and counterselect for *str-r*.
(d) To select for *his*$^+$ and counterselect for *str-r*.

10.14 In the previous problem, what medium should you use to determine each of the following?
(a) Which of the *lac*$^+$ *str-r* colonies are *gal*$^+$.
(b) Which of the *lac*$^+$ *str-r* colonies are *trp*$^+$.
(c) Which of the *lac*$^+$ *str-r* colonies are *his*$^+$.
(d) Which of the *lac*$^+$ *str-r* colonies are *trp*$^+$ and *his*$^+$.

10.15 An Hfr strain transfers genes in alphabetical order. When tetracycline sensitivity is used for counterselection, the number of *h*$^+$ *tet-r* colonies is 1000-fold lower than the number of *h*$^+$ *str-r* colonies found when streptomycin sensitivity is used for counterselection. Suggest an explanation for the difference.

10.16 An Hfr strain transfers genes in the order *a b c*. In an Hfr *a*$^+$ *b*$^+$ *c*$^+$ *str-s* × F$^-$ *a*$^-$ *b*$^-$ *c*$^-$ *str-r* mating, do all *b*$^+$ *str-r* recombinants receive the *a*$^+$ allele? Are all *b*$^+$ *str-r* recombinants also *a*$^+$? Why or why not?

10.17 If the genes in a bacterial chromosome are in alphabetical order and an Hfr cell transfers genes in the order *g h i . . . d e f*, what types of F' plasmids could be derived from the Hfr?

10.18 In an Hfr *lac*$^+$ *met*$^+$ *str-s* × F$^-$ *lac*$^-$ *met*$^-$ *str-r* mating, *met* is transferred much later than *lac*. If cells are plated on minimal medium containing glucose and streptomycin, what fraction of the cells are expected to be *lac*$^+$? If cells are plated on minimal medium containing lactose, methionine, and streptomycin, what fraction of the cells are expected to be *met*$^+$?

10.19 In the mating

Hfr *met*$^-$ *his*$^+$ *leu*$^+$ *trp*$^+$ × F$^-$ *met*$^+$ *his*$^-$ *leu*$^-$ *trp*$^-$

the *met* marker is known to be transferred very late. After a short time, the mating cells are interrupted and the cell suspensions plated on four different growth media. The amino acids in the growth media and the number of colonies observed on each are as follows:

histidine + tryptophan	250 colonies
histidine + leucine	50 colonies
leucine + tryptophan	500 colonies
histidine	10 colonies

What is the purpose of the *met*$^-$ mutation in the Hfr strain? What is the order of transfer of the genes? Why is the number of colonies so small for the medium containing only histidine?

10.20 *Salmonella enterica* is closely related to *Escherichia coli*. It can be infected with the F plasmid, which can integrate into the chromosome to produce Hfr strains. These can be mated with F$^-$ *E. coli* to study the order and time of entry of genetic markers. The data shown here pertain to times of entry of four genetic markers in crosses of *E. coli* Hfr × *E. coli* F$^-$ and *S. enterica* Hfr × *E. coli* F$^-$.

	ile	***met***	***pro***	***arg***
E. coli Hfr × *E. coli* F$^-$	28	20	6	22
S. enterica Hfr × *E. coli* F$^-$	4	22	47	18

(a) How do the genetic maps of *E. coli* and *S. enterica* compare with respect to these genes?
(b) What are the origin and direction of transfer in each Hfr?
(c) How fast does an *S. enterica* Hfr transfer chromosomal DNA relative to an *E. coli* Hfr?

GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by-chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

•With about 100 bacteriophage, 200 plasmid, and 50 bacterial genomes completely sequenced, merely finding the information that one seeks is not always as straightforward as it may seem. The National Center for Biotechnology Information (NCBI) maintains a site to facilitate the use of databases and software for analyzing data about genetics, molecular biology, and biochemistry. Browse the site map at the keyword **NCBI** and find the annotated maps of the *E. coli* and bacteriophage λ genomes.

• Available **Hfr** strains in *E. coli* can be found by searching at this site. Simply enter *Hfr* in the box that says "sex" and you will obtain the list. Clicking on any entry—say, *3310 Paris Hfr*—will open a link to more detailed information. The entry will give you the origin of transfer of the Hfr (in this case, 96.80 minutes), the direction of transfer (1 = clockwise, 0 = counterclockwise; in this case, counterclockwise), any mutant genes that are present in the strain, and one or more relevant references. Check out the "classic" Hfr strains, Hfr Hayes and Hfr Cavalli.

10.21 An Hfr donor of genotype a^+ b^+ c^+ *str-s* is mated with an F^- recipient of genotype *a b c str-r*; the order of transfer is *a b c*. None of the genes is transferred early, the distance between *a* and *b* is the same as that between *b* and *c*, and none of the markers is near *str*. Recombinants are selected as usual by plating on medium that lacks particular nutrients and contains streptomycin. Which of the following statements are true? Explain. (More than one are true.)

(a) Number of a^+ *str-r* colonies > number of c^+ *str-r* colonies.
(b) Number of b^+ *str-r* colonies < number of c^+ *str-r* colonies.
(c) Number of a^+ b^+ *str-r* colonies < number of b^+ *str-r* colonies.
(d) Number of a^+ b^+ *str-r* colonies ≈ number of b^+ *str-r* colonies.
(e) Most a^+ c^+ *str-r* colonies will also be b^+.
(f) Most b^- c^+ *str-r* colonies will also be a^-.
(g) Number of a^+ b^+ c^- *str-r* colonies < number of a^+ b^- c^- *str-r* colonies.

10.22 Bacterial cells of genotype pur^- pro^+ his^+ were transduced with P1 bacteriophage grown on bacteria of genotype pur^+ pro^- his^-. Transductants containing pur^+ were selected and tested for the unselected markers *pro* and *his*. The numbers of pur^+ colonies with each of four genotypes are as follows:

pro^+ his^+	102
pro^- his^+	25
pro^+ his^-	160
pro^- his^-	1

What is the gene order?

10.23 For generalized transduction, the theoretical relationship between map distance in minutes and frequency of cotransduction is

$$\text{map distance} = 2 - 2 \times (\text{cotransduction frequency})^{1/3}$$

(a) What map distance in minutes corresponds to 75 percent cotransduction? to 50 percent cotransduction? to 25 percent cotransduction?
(b) What is the cotransduction frequency between genetic markers separated by half a minute? one minute? two minutes? greater than two minutes?
(c) If genes *a*, *b*, and *c* are in alphabetical order and the cotransduction frequencies are 30 percent for *a*–*b* and 10 percent for *b*–*c*, what is the expected cotransduction frequency between *a* and *c*?

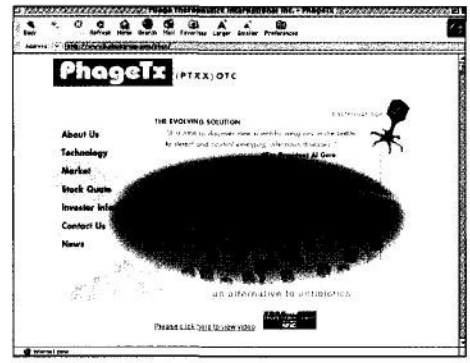
10.24 An ampicillin-resistant (Amp-r) strain of *E. coli* isolated from nature is infected with λ, and a suspension of phage progeny is obtained. This phage suspension is used to infect a culture of laboratory *E. coli* that is lysogenic for λ, sensitive to the antibiotic, and unable to undergo homologous recombination. Because the λ repressor is present in the lysogen, infecting λ molecules cannot replicate and are gradually diluted out of the culture by growth and continued division of the cells. However, rare Amp-r cells are found among the lysogens. Explain how they might have arisen.

10.25 On continued growth of a culture of a λ lysogen, rare cells become nonlysogenic by spontaneous loss of the prophage. Such loss is a result of prophage excision without subsequent development of the phage. The lysogen is said to have been "cured." A λ phage that carries a gene for tetracycline resistance is used to lysogenize a Tet-s cell. At a later time, a cell is isolated that is Tet-r but

• When bacteriophages were first discovered early in the twentieth century, they opened up the possibility of **phage therapy**, or the use of specific kinds of bacteriophage to kill pathogenic bacteria. This avenue remained unexplored after the early 1940s when antibiotics were developed and began to be widely used in the treatment of bacterial infections. Today, with the spread of bacteria resistant to most or all available antibiotics causing increasing concern, phage therapy is the subject of renewed research. At this keyword site you can learn about the historical and ecological context of this research and find a realistic assessment of its prospects for success.

• The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 10 and you will be linked automatically.

• The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 10.



no longer contains a λ prophage. Explain how the cured cell could remain Tet-r.

10.26 A time-of-entry experiment is carried out in the mating

$$\text{Hfr } a^+ b^+ c^+ d^+ \textit{ str-s} \times \text{F}^- a^- b^- c^- d^- \textit{ str-r}$$

where the spacing between the genes is equal. The data in the accompanying table are obtained. What are the times of entry of each gene? Suggest one possible reason for the low frequency of d^+ *str-r* recombinants.

Time of mating in minutes	Number of recombinants of indicated genotype per 100 Hfr			
	a^+ *str-r*	b^+ *str-r*	c^+ *str-r*	d^+ *str-r*
0	0.01	0.006	0.008	0.0001
10	5	0.1	0.01	0.0004
15	50	3	0.1	0.001
20	100	35	2	0.001
25	105	80	20	0.1
30	110	82	43	0.2
40	105	80	40	0.3
50	105	80	40	0.4
60	105	81	42	0.4
70	103	80	41	0.4

10.27 You have reason to believe that a number of *rII* mutants of bacteriophage T4 are deletion mutations. You cross them in all possible combinations in *E. coli* strain B and plate them on *E. coli* strain K12(λ) to determine whether r^+ recombinants are formed. The formation of r^+ recombinants indicates that the mutations can recombine and so, if they are deletions, they must be nonoverlapping. The results are given in the accompanying table, in which *a* through *f* indicate an *rII* mutation and + indicates the formation of r^+ recombinant progeny in the cross.

	a	*b*	*c*	*d*	*e*	*f*
a	–	–	–	–	–	–
b	–	–	+	+	–	
c	–	+	–	+		
d	–	–	+			
e	–	+				
f	–					

Assemble a deletion map for these mutations, using a line to indicate the DNA segment that is deleted in each mutant.

10.28 Seven *rII* mutants, *t* through *z*, thought to carry point mutations, are crossed to the six deletion mutants in Problem 10.27 and scored for the ability to produce r^+ recombinants. The results are given in the accompanying table.

	a	*b*	*c*	*d*	*e*	*f*
t	–	–	–	+	+	+
u	–	+	–	+	+	+
v	+	+	+	–	+	+
w	+	+	+	+	+	–
x	–	+	+	+	+	–
y	–	–	+	+	+	–
z	–	+	+	–	–	+

Using the deletion end points to define genetic intervals along the *rII* gene, position each point mutation within an interval. Refine the deletion end points if required by the data.

10.29 You cross the *rII* point mutations *t* through *z* in Problem 10.28 in all possible combinations in K12(λ) to assess complementation. The results are shown in the accompanying complementation matrix.

	t	*u*	*v*	*w*	*x*	*y*	*z*
t	–	–	+	–	–	–	+
u		–	+	–	–	–	+
v			–	+	+	+	–
w				–	–	–	+
x					–	–	+
y						–	+
z							–

The mutations *v* and *z* complement deletion *c* but not deletion *e*. The mutations *v* and *z* also fail to complement a mutant known to be defective for *rIIB* function but not for *rIIA* function. Which mutants are in the *rIIA* cistron and which in the *rIIB* cistron? Where does the boundary between the two genes lie? Assemble a complementation matrix for the deletion mutants, filling in all the squares if possible.

10.30 P2 and P4 are bacteriophages of *E. coli*. They have the following properties: (1) When P2 phage infects a bacterium, the bacterium eventually bursts, yielding about 100 P2 progeny. (2) When P4 infects a bacterium, the bacterium survives because P4 is a defective phage. (3) When P2 phage and P4 phage coinfect the same bacterium, lysis of the bacterium yields 100 P4 progeny and no P2 progeny, because the presence of P4 inhibits the growth of P2. Suppose 3×10^8 P2 and 2×10^8 P4 are added to 10^8 bacterial cells.

(a) How many bacteria will not be infected with either phage?
(b) How many bacteria will survive?
(c) How many bacteria will produce P2 progeny?
(d) How many bacteria will produce P4 progeny?

(*Hint:* If the average number of phage per cell in a mixed suspension is *x*, then the probability that a cell remains uninfected is e^{-x}.)

Challenge Problems

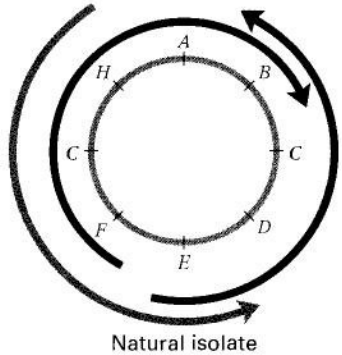
10.31 Bacteriophage 363 can be used for generalized transduction of genetic markers in *E. coli*. François Jacob used phage 363 to demonstrate that the lysogenic state of bacteriophage λ can be genetically transferred from one strain to another so long as the experiment is carried out at 20°C, a temperature at which λ phage do not multiply. Jacob studied four strains of bacteria:

A: $thr^- lac^- gal^-$
B: $thr^- lac^- gal^-$ (λ)
C: $thr^+ lac^+ gal^+$
D: $thr^+ lac^+ gal^+$ (λ)

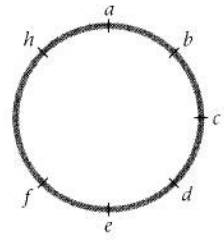
In these designations, the symbol (λ) means that the strain is lysogenic for phage λ. Jacob made two kinds of crosses. Phage grown on strain D were used to transduce strain A, and phage grown on strain C were used to transduce strain B. The results are given in the table below.

Give a genetic explanation of these results. What genetic element is being transduced to give or to remove lysogeny?

10.32 In studying a novel type of antibiotic resistance, a geneticist carries out crosses between three Hfr derivatives of a natural isolate of *Escherichia coli* and an F^- laboratory strain carrying the mutant alleles shown in the illustration. The natural isolate carries a duplication of the wildtype allele of the gene *C*, in which the duplicate copies are on opposite sides of the chromosome. The origin, extent, and direction of transfer in three Hfr crosses are shown by the arrows. Draw the genetic map that will result if the geneticist assumes that only one copy of each wildtype allele was present in the Hfr strains.



Natural isolate



Laboratory strain of *E. coli*

Donor strain	Recipient strain	Thr⁺ Colonies tested	Thr⁺ Colonies lysogenic	Lac⁺ Colonies tested	Lac⁺ Colonies lysogenic	Gal⁺ Colonies tested	Gal⁺ Colonies lysogenic
D	A	400	0	400	0	400	24
C	B	400	400	400	400	400	368

Problem 10.31

10.33 The accompanying illustration shows a procedure known as *inverse PCR*, in which unknown sequences *flanking* a known sequence can be amplified. In the first step, genomic DNA is cleaved with a restriction enzyme (in this example, *Eco*RI) that does not cut within the known sequence. The resulting solution of restriction fragments is diluted and then treated with DNA ligase to join the ends together, forming circular DNA. (The solution is diluted to ensure that the ends of single fragments join together, rather than the ends of two or more different fragments.) Then PCR is carried out using oligonucleotide primers whose 3' ends face *outward*, toward the unknown sequence (rather then inward, as in the usual PCR procedure). The result is amplification of the unknown fragment. Suppose that oligonucleotide primers of length 6 are adequate for the PCR (in reality, they are too short for most purposes), and suppose that known sequence is a transposable element whose ends are the inverted repeats

```
5'-GGTAAA . . . TTTACC-3'
3'-CCATTT . . . AAATGG-5'
```

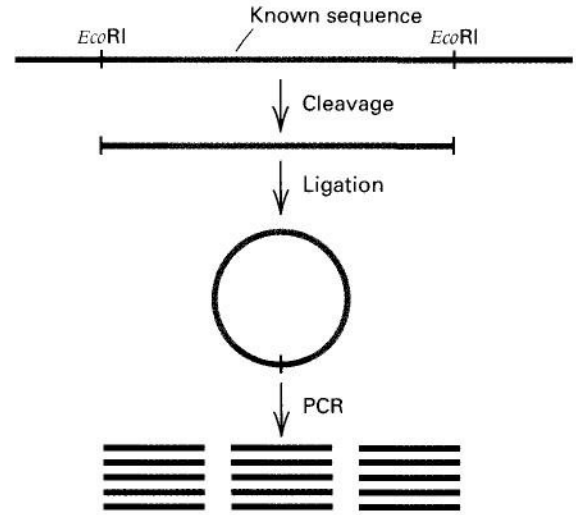
(a) What primers would normally be used to amplify the known sequence?

(b) Are two different primers needed? Why or why not?

(c) What primers would be used to amplify the unknown sequence with inverse PCR?

(d) Would the PCR product be cleaved with *Eco*RI, and if so, into what fragments?

(e) If the PCR product were sequenced, how would one know which part originally flanked the known sequence on the left and which part originally flanked the known sequence on the right?



Further Reading

Adelberg, E. A., ed. 1966. *Papers on Bacterial Genetics.* Boston: Little, Brown.

Campbell, A. 1976. How viruses insert their DNA into the DNA of the host cell. *Scientific American*, December.

Clowes, R. D. 1975. The molecules of infectious drug resistance. *Scientific American*, July.

Davies, J. 1995. Vicious circles: Looking back on resistance plasmids. *Genetics* 139: 1465.

Drlica, K., and M. Riley. 1990. *The Bacterial Chromosome.* Washington, DC: American Society for Microbiology.

Edgar, R. S., and R. H. Epstein. 1965. The genetics of a bacterial virus. *Scientific American*, February.

Hopwood, D. A., and K. E. Chater, eds. 1989. *Genetics of Bacterial Diversity.* New York: Academic Press.

Komano, T. 1999. Shufflons: Multiple Inversion Systems and Integrons. *Annual Review of Genetics* 33: 171.

Losick, R., and D. Kaiser. 1997. Why and how bacteria communicate. *Scientific American*, February.

Low, K. B., and R. Porter. 1978. Modes of genetic transfer and recombination in bacteria. *Annual Review of Genetics* 12: 249.

Mushegian, A. 1999. The minimal genome concept. *Current Opinion in Genetics & Development* 9: 709.

Neidhardt, F. C., R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, eds. 1998. Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (2 volumes). 2d ed. Washington, DC: American Society for Microbiology.

Novick, R. P. 1980. Plasmids. *Scientific American*, December.

Robertson, B. D., and T. F. Meyer. 1992. Genetic variation in pathogenic bacteria. *Trends in Genetics* 8: 422.

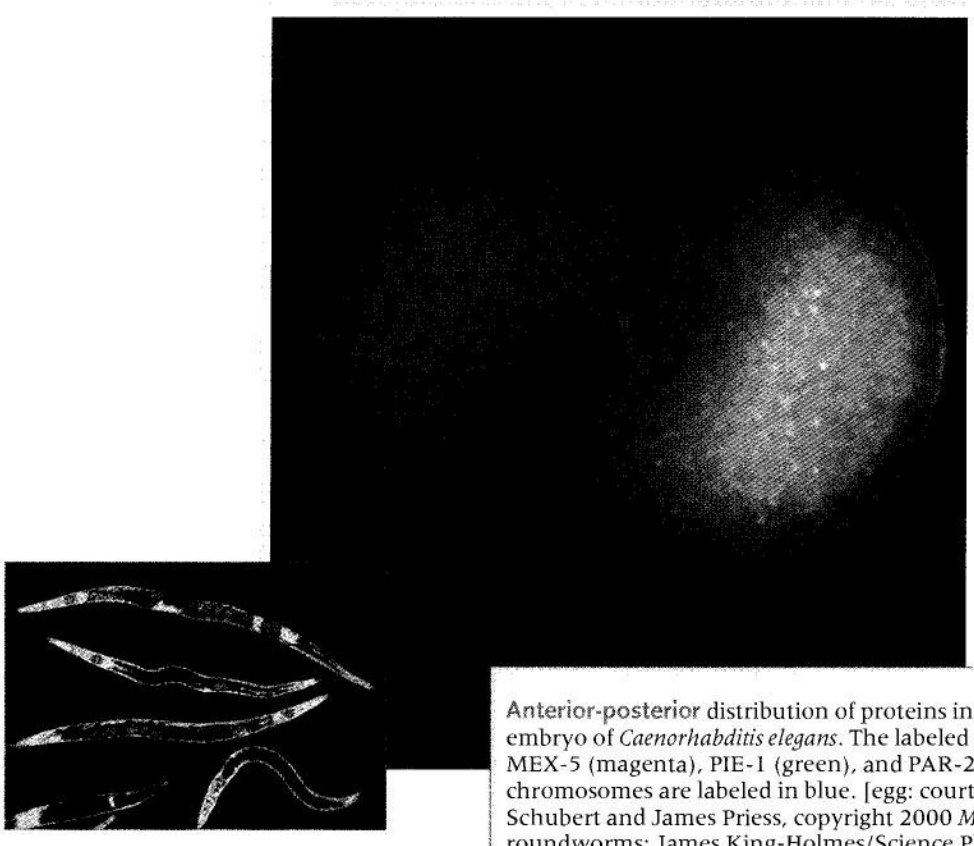
Snyder, L., and W. Champness. 1977. *Molecular Genetics of Bacteria.* Washington, DC: American Society for Microbiology.

Zinder, N. 1958. Transduction in bacteria. *Scientific American*, November.

CHAPTER

11

Molecular Biology of Gene Expression



Anterior-posterior distribution of proteins in the one-cell embryo of *Caenorhabditis elegans*. The labeled proteins are MEX-5 (magenta), PIE-1 (green), and PAR-2 (red); the chromosomes are labeled in blue. [egg: courtesy Charlotte Schubert and James Priess, copyright 2000 *Molecular Cell* 15: 4; roundworms: James King-Holmes/Science Photo Library/ Photo Researchers, Inc.]

CHAPTER OUTLINE

11.1 **Amino Acids, Polypeptides, and Proteins**
Protein Folding
What Are the Minimal Genetic Functions Needed for Life?

11.2 **Colinearity Between Coding Sequences and Polypeptides**

11.3 **Transcription**
Overview of RNA Synthesis
Types of RNA Polymerase
Mechanism of Transcription
Genetic Evidence for Promoters and Terminators

11.4 **Messenger RNA**

11.5 **RNA Processing in Eukaryotes**
5' Capping and 3' Polyadenylation
Splicing of Intervening Sequences
Mutations in Introns
Exon Shuffle in the Origin of New Genes

11.6 **Translation**
Initiation by mRNA Scanning
Elongation
Release

11.7 **Complex Translation Units**
Polysomes
Polycistronic mRNA

11.8 **The Standard Genetic Code**
Genetic Evidence for a Triplet Code
How the Code Was Cracked
Features of the Standard Code
Transfer RNA and Aminoacyl-tRNA Synthetase Enzymes
Redundancy and Wobble
Nonsense Suppression

PRINCIPLES

- In gene expression, information in the nucleotide sequence of DNA is used to dictate the linear order of amino acids in a polypeptide by means of an RNA intermediate.
- Transcription of an RNA from one strand of the DNA is the first step in gene expression.
- In eukaryotes, the RNA transcript is modified and may undergo splicing to make the messenger RNA (mRNA).
- The messenger RNA is translated on ribosomes in groups of three nucleotides (codons), each specifying an amino acid through base pairing with molecules of transfer RNA, each "charged" (covalently bonded) with one amino acid.
- Ribosomes are particles consisting of ribosomal RNA and numerous proteins combined in small and large ribosomal subunits. The subunits come together with other factors to initiate polypeptide synthesis, to translocate the ribosome progressively along the mRNA, and to release the finished polypeptide when a stop codon is encountered.
- Nearly all organisms use the standard genetic code with 61 codons for 20 amino acids and 3 stop codons, but exceptions are found in certain protozoa and in the genetic codes of mitochondrial and other organellar DNA.

CONNECTIONS

One Gene, One Enzyme
George W. Beadle and Edward L. Tatum 1941
Genetic Control of Biochemical Reactions in Neurospora

Messenger "Light"
Sydney Brenner, François Jacob, and Matthew Meselson 1961
An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis

Uncles and Aunts
Francis H. C. Crick, Leslie Barnett, Sydney Brenner, and R. J. Watts-Tobin 1961
General Nature of the Genetic Code for Proteins

Now we come to **gene expression,** the molecular processes by which the information contained in genes is converted into polypeptide chains that determine the metabolic and developmental capabilities of cells and organisms. As we saw in Chapter 1, the information transfer from DNA to protein is accomplished through an RNA intermediate:

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Protein}$$

and this sequence is called the "central dogma" of gene expression. In somewhat more detail, the principal steps in gene expression can be summarized as follows:

1. RNA molecules are synthesized enzymatically by an *RNA polymerase,* which uses the nucleotide sequence of a segment of a single strand of DNA as a template to create a complementary RNA strand in a stepwise manner. The overall process by which the segment corresponding to a particular gene is selected and an RNA molecule is made is *transcription.*
2. In eukaryotes, the RNA usually undergoes chemical modification in the nucleus called *RNA processing.*
3. The nucleotide sequence of the processed RNA molecule (the *messenger RNA*) is used in the process of *translation* to direct the sequential joining of amino acids in a particular order to create a polypeptide chain. The amino acid sequence in a polypeptide chain is therefore specified by the coding sequence in a gene.

11.1 Amino Acids, Polypeptides, and Proteins

Proteins are responsible for catalyzing most intracellular chemical reactions (enzyme proteins), for regulating gene expression (regulatory proteins), and for determining many features of the structures of cells, tissues, and organisms (structural proteins). A protein is composed of one or more **polypeptide chains,** each of which is a series of covalently joined amino acids. There are 20 different amino acids commonly found in polypeptides, and they can be joined in any order and in any number (typically 100–1000).

Each amino acid contains a carbon atom (the α carbon) to which is attached one carboxyl group ($-$COOH), one amino group ($-NH_2$), and a side chain commonly called an **R group** (Figure 11.1). The R groups are generally chains or rings of carbon atoms bearing various distinguishing atoms. The simplest R groups are those of glycine ($-$H) and of alanine ($-CH_3$). For reference, the chemical structures of all 20 amino acids are shown in Figure 11.2. For each amino acid, the R group is indicated by a gold rectangle.

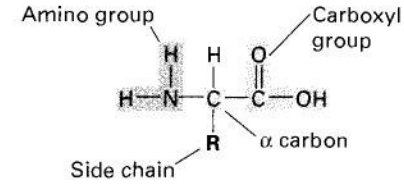


Figure 11.1 The general structure of an amino acid.

Each link in a polypeptide chain is formed when the carboxyl group of one amino acid becomes joined with the amino group of the next amino acid in line; the resulting chemical bond is a covalent bond called a **peptide bond** (Figure 11.3A). Thus a polypeptide chain has an ordered array of side chains jutting off a backbone that consists of α-carbon atoms alternating with peptide groups (Figure 11.3B).

The two ends of every polypeptide molecule are distinct. One end has a free amino ($-NH_2$) group and is called the **amino terminus;** the other end has a free carboxyl ($-$COOH) group and is the **carboxyl terminus.** Polypeptides are synthesized by adding successive amino acids to the carboxyl end of the growing chain. Conventionally, the amino acids of a polypeptide chain are numbered starting at the amino terminus.

Protein Folding

Most polypeptide chains are highly folded, and a wide variety of three-dimensional shapes have been observed. The manner of folding is determined primarily by the sequence of amino acids—in particular, by noncovalent interactions between the side chains—so each polypeptide chain tends to fold into a unique three-dimensional shape as it is being synthesized. In some cases, protein folding is assisted by interactions with other proteins in the cell called *chaperones.*

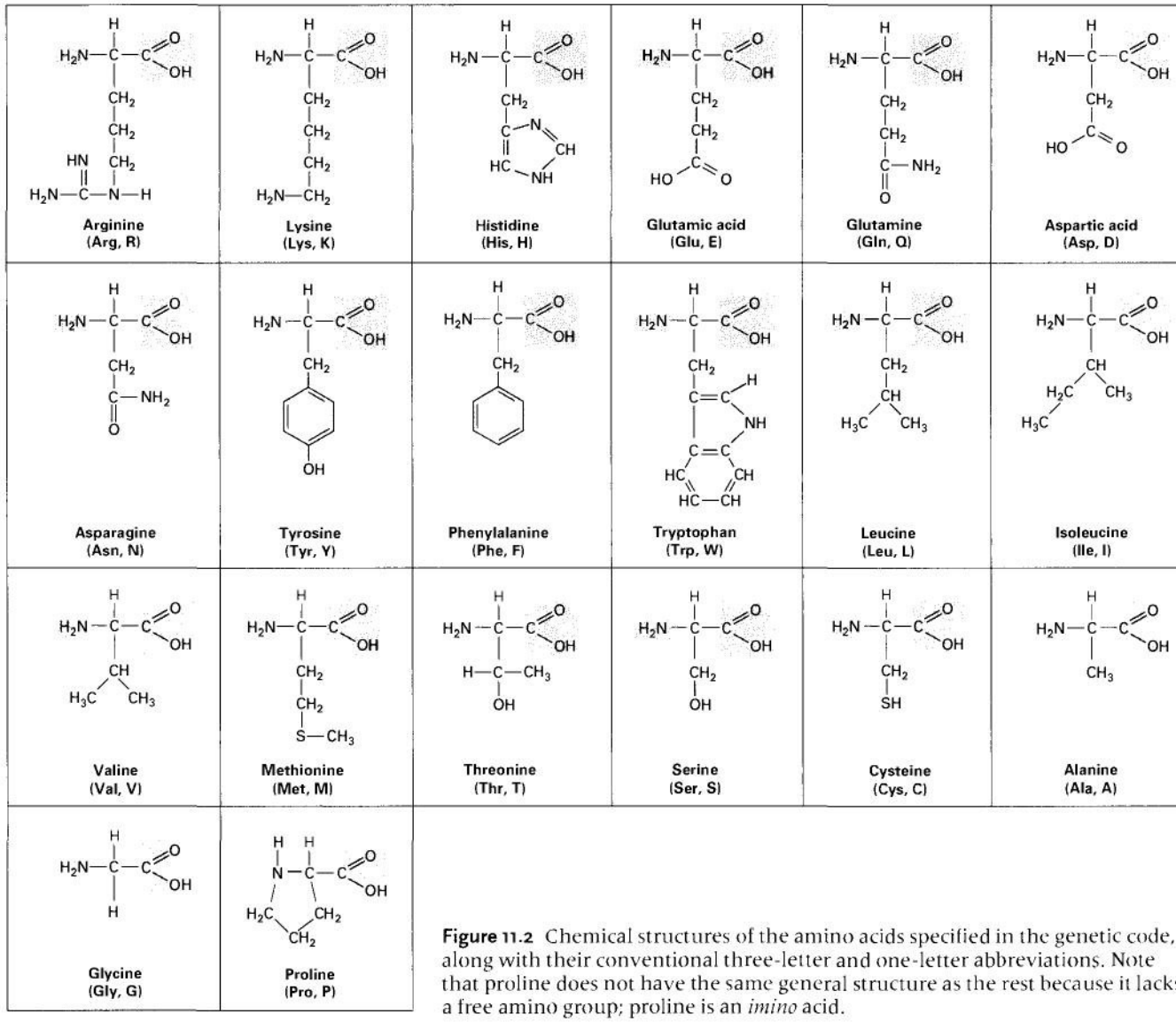


Figure 11.2 Chemical structures of the amino acids specified in the genetic code, along with their conventional three-letter and one-letter abbreviations. Note that proline does not have the same general structure as the rest because it lacks a free amino group; proline is an *imino* acid.

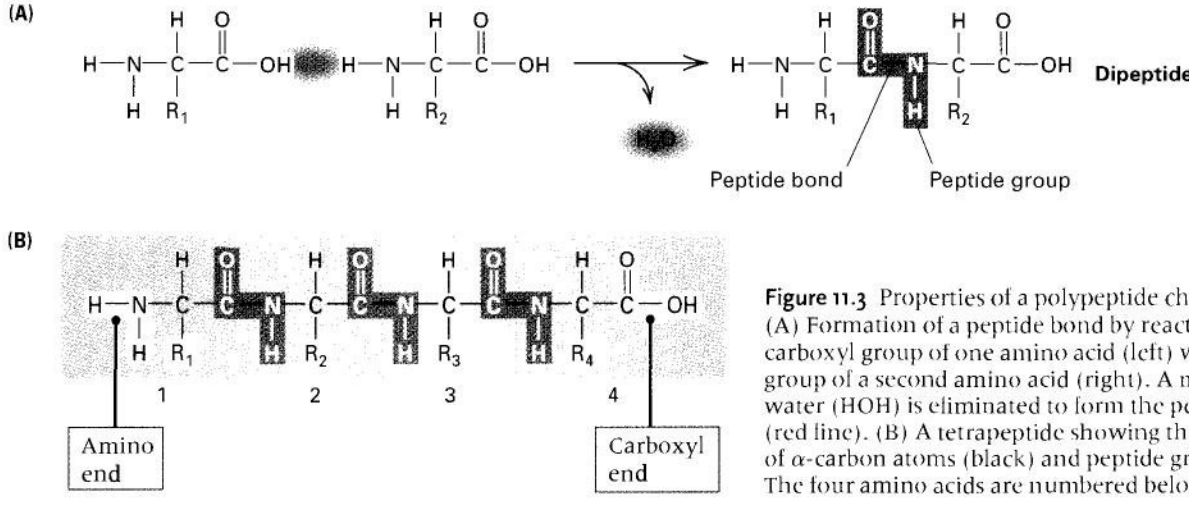


Figure 11.3 Properties of a polypeptide chain. (A) Formation of a peptide bond by reaction of the carboxyl group of one amino acid (left) with the amino group of a second amino acid (right). A molecule of water (HOH) is eliminated to form the peptide bond (red line). (B) A tetrapeptide showing the alternation of α-carbon atoms (black) and peptide groups (blue). The four amino acids are numbered below.

Generally speaking, protein molecules are folded so that amino acids with charged, hydrophilic side chains tend to be on the surface of the protein (in contact with water) and those with uncharged, hydrophobic side chains tend to be internal (hidden from water).

Specific folded configurations also result from hydrogen bonding between peptide groups. Two fundamental polypeptide structures are the *alpha* (α) *helix* and the *beta* (β) *sheet* (**Figure 11.4**). An α helix is formed by hydrogen bonding between peptide groups that are close together in the polypeptide backbone. In an α helix, often represented as a coiled ribbon, the backbone is twisted so that the N−H in each peptide group is hydrogen-bonded with the C = O in the peptide group located four amino acids farther along the helix. The helical twist may be right-handed or left-handed, but right-handed α helices are more common. (Both α helices in Figure 11.4 are right-handed.)

In contrast, a β sheet is formed by hydrogen bonding between peptide groups in distant parts of the polypeptide chain, or even in different polypeptide chains. In a β sheet, often represented as parallel "flat" ribbons, the backbones of the interacting polypeptide chains are held flat and relatively rigid (forming a "sheet"), because alternate N−H groups in one backbone are hydrogen-bonded with alternate C = O groups in the backbone of the adjacent chain. In each polypeptide chain, alternate C = O and N−H groups are free to form hydrogen bonds with their counterparts in a different chain on the opposite side, so a β sheet can consist of multiple aligned segments in the same or different polypeptides. The orientation of the backbones in a β sheet may be antiparallel (adjacent backbones reversed in orientation relative to their amino and carboxyl ends) or parallel, but antiparallel is more common. In Figure 11.4, the three-stranded β sheet in the middle is antiparallel, as is the two-stranded β sheet that brings the amino and carboxyl ends together.

Besides the α helices and β sheets, other types of interactions also are important in

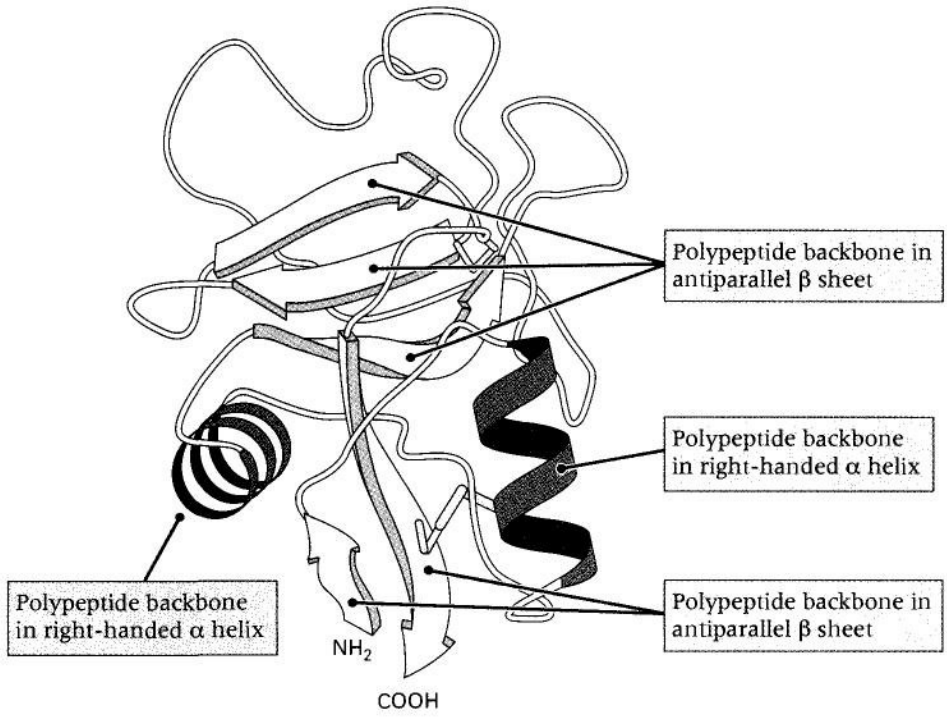


Figure 11.4 A "ribbon" diagram of the path of the backbone of a polypeptide, showing the ways in which the polypeptide is folded. Arrows represent β sheets, each of which is held to its neighboring β sheet by hydrogen bonds. Helical regions are shown as coiled ribbons. The polypeptide chain in this example is a mannose-binding protein. [Adapted from William I. Weis, Kurt Drickamer, and Wayne A. Hendrickson. 1992. *Nature* 360: 127.]

protein folding. For example, covalent bonds may form between the sulfur atoms of pairs of cysteines in different parts of the polypeptide. However, the rules of folding are so complex that, except for the simplest proteins, the final shape of a protein cannot usually be predicted from the amino acid sequence alone.

Many protein molecules consist of more than one polypeptide chain. When this is the case, the protein is said to contain **protein subunits.** The subunits may be identical or different. For example, hemoglobin, the oxygen carrier of blood, consists of four subunits: two copies of a polypeptide designated the α chain, and two copies of a different polypeptide designated the β chain. (The use of α and β as names of the polypeptide chains has nothing to do with the α helices and β sheets that are found within the polypeptides.)

What Are the Minimal Genetic Functions Needed for Life?

The bacterium *Mycoplasma genitalium* is a parasite associated with ciliated epithelial cells of the genital and respiratory tracts of primates, including human beings. It is thought to have the smallest genome of any free-living organism. It belongs to a large group of bacteria (the mycoplasmas) that lack a cell wall and that parasitize a wide range of plant and animal hosts. The *M. genitalium* genome is only about 3.5 times larger than that of the bacteriophage T4. It is a circular DNA molecule 580 kb in length that includes only 471 genes. Analysis of this genome enables us to identify what is probably a minimal set of genes necessary for a free-living cell.

The cellular processes in which the gene products of *M. genitalium* participate are summarized in Table 11.1. A substantial fraction of the genome is devoted to the synthesis of macromolecules such as DNA, RNA, and protein; another substantial fraction supports cellular processes and energy metabolism. There are very few genes for biosynthesis of small molecules. However, genes that encode proteins for salvaging and/or for transporting small molecules make up a significant fraction of the total, which underscores the fact that the bacterium is parasitic. The remaining genes are largely devoted to forming the cellular envelope and to helping the organism evade the immune system of the host.

Table 11.1 Summary of functions of 471 genes of *Mycoplasma genitalium*

Function	Number of genes	Percent*
DNA replication	32	10
Transcription	12	4
Translation	101	32
Cell envelope	17	5
Transport of small molecules	34	11
Energy metabolism	31	10
ATP–proton force generation	8	
Glycolysis	10	
Cell processes (cell division, secretion, stress response)	21	7
Nucleoside and nucleotide synthesis	19	6
Salvage (degradative pathways)	10	
Other metabolism (lipids, cofactors, amino acids, intermediary metabolism)	18	6
Regulatory	7	2
Other	27	8
Hypothetical or unknown	152	

*Percent among all genes with identified functions. Data from C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, and 23 other authors, *Science*, 1995. 270: 397.

11.2 Colinearity Between Coding Sequences and Polypetides

It took half a century to learn that genes specify the amino acid sequence of polypeptides. In the early 1900s, Archibald Garrod suggested that hereditary human diseases, such as phenylketonuria, result from inborn errors of metabolism (Chapter 1). Support for this idea came in the 1940s when George Beadle and Edward Tatum, using *Neurospora crassa,* demonstrated that genes govern the ability of the

CoNnection

One Gene, One Enzyme

George W. Beadle and Edward L. Tatum 1941
Stanford University, Stanford, California
Genetic Control of Biochemical Reactions in Neurospora

How do genes control metabolic processes? The suggestion that genes control enzymes was made very early in the history of genetics, most notably by the British physician Archibald Garrod in his 1908 book Inborn Errors of Metabolism. *But the precise relationship between genes and enzymes was still uncertain. Perhaps each enzyme is controlled by more than one gene, or perhaps each gene contributes to the control of several enzymes. The classic experiments of Beadle and Tatum showed that the relationship is usually remarkably simple: One gene codes for one enzyme. The pioneering experiments united genetics and biochemistry, and for the "one gene, one enzyme" concept, Beadle and Tatum were awarded a Nobel Prize in 1958 (Joshua Lederberg shared the prize for his contributions to microbial genetics). Because we now know that some enzymes contain polypeptide chains encoded by two (or occasionally more) different genes, a more accurate statement of the principle is "one gene, one polypeptide." Beadle and Tatum's experiments also demonstrate the importance of choosing the right organism.* Neurospora *had been introduced as a genetic organism only a few years earlier, and Beadle and Tatum realized that they could take advantage of the ability of this organism to grow on a simple medium composed of known substances.*

These preliminary results appear to us to indicate that the approach may offer considerable promise as a method of learning more about how genes regulate development and function.

From the standpoint of physiological genetics the development and functioning of an organism consist essentially of an integrated system of chemical reactions controlled in some manner by genes. . . . In investigating the roles of genes, the physiological geneticist usually attempts to determine the physiological and biochemical bases of already known hereditary traits. . . . There are, however, a number of limitations inherent in this approach. Perhaps the most serious of these is that the investigator must in general confine himself to the study of non-lethal heritable characters. Such characters are likely to involve more or less non-

fungus to synthesize amino acids, purines, and vitamins. That genes control metabolism by specifying the amino acid sequence in polypeptides was first shown in the early 1950s by the discovery that the allele for sickle-cell anemia brings about a change in the charge of the hemoglobin molecule by causing a substitution of an uncharged valine for a negatively charged glutamic acid at residue number 6 in the β-globin chain.

Most genes contain the information for the synthesis of only one polypeptide chain. Furthermore, the linear order of nucleotides in a gene determines the linear order of amino acids in a polypeptide. This point was first proved by studies of the tryptophan synthase gene *trpA* in *E. coli*, a gene in which many mutations had been obtained and accurately mapped. The effects of numerous mutations on the amino acid sequence of the enzyme were determined by directly analyzing the amino acid sequences of the wildtype and mutant enzymes. Each mutation was found to result in a single amino acid substituting for the wildtype amino acid in the enzyme; more important, *the order of the mutations in the genetic map was the same as the order of the mutant amino acids in the polypeptide chain* (**Figure 11.5**). This attribute of genes and polypeptides is called **colinearity,** which means that the sequence of base pairs in DNA determines the sequence of amino acids in the polypeptide in a colinear, or point-to-point, manner. Colinearity is found almost universally in prokaryotes. However, we will see later that in eukaryotes, noncoding DNA sequences interrupt the coding sequences in most genes; in these genes, the order of mutations along a gene (but not their spacing) correlates with the respective amino acid substitutions.

essential so-called "terminal" reactions. . . . A second difficulty is that the standard approach to the problem implies the use of characters with visible manifestations. Many such characters involve morphological variations, and these are likely to be based on systems of biochemical reactions so complex as to make analysis exceedingly difficult. . . . Considerations such as those just outlined have led us to investigate the general problem of the genetic control of development and metabolic reactions by reversing the ordinary procedure and, instead of attempting to work out the chemical bases of known genetic characters, to set out to determine if and how genes control known biochemical reactions. The ascomycete *Neurospora* offers many advantages for such an approach and is well suited to genetic studies. Accordingly, our program has been built around this organism. The procedure is based on the assumption that x-ray treatment will induce mutations in genes concerned with the control of known specific chemical reactions. If the organism must be able to carry out a certain chemical reaction to survive on a given medium, a mutant unable to do this will obviously be lethal on this medium. Such a mutant can be maintained and studied, however, if it will grow on a medium to which has been added the essential product of the genetically blocked reaction. . . . Among approximately 2000 strains [derived from single cells after x-ray treatment], three mutants have been found that grow essentially normally on the complete medium and scarcely at all on the minimal medium. One of these strains proved to be unable to synthesize vitamin B_6 (pyridoxine). A second strain turned out to be unable to synthesize vitamin B_1 (thiamine). A third strain has been found to be unable to synthesize para-aminobenzoic acid. . . . These preliminary results appear to us to indicate that the approach may offer considerable promise as a method of learning more about how genes regulate development and function. For example, it should be possible, by finding a number of mutants unable to carry out a particular step in a given synthesis, to determine whether only one gene is ordinarily concerned with the immediate regulation of a given specific chemical reaction.

Source: Proceedings of the National Academy of Sciences of the USA 27: 499–506

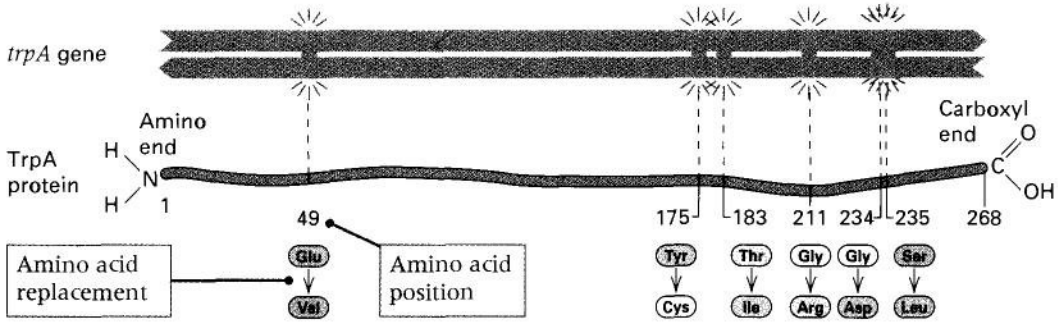


Figure 11.5 Correlation of the positions of mutations in the genetic map of the *E. coli trpA* gene with positions of amino acid replacements in the TrpA protein.

11.3 Transcription

The first step in gene expression is **transcription,** the synthesis of an RNA molecule copied from the segment of DNA that constitutes the gene. The basic features of the production of RNA are described in this section.

Overview of RNA Synthesis

The essential chemical characteristics of the enzymatic synthesis of RNA resemble those of DNA synthesis discussed in Chapter 6.

1. The precursors in the synthesis of RNA are the four ribonucleoside 5'-triphosphates: adenosine triphosphate (ATP), guanosine triphosphate (GTP), cytidine triphosphate (CTP), and uridine triphosphate (UTP). They differ from the DNA precursors only in that the sugar is ribose rather than deoxyribose and in that the base uracil (U) replaces thymine (T) (**Figure 11.6**).
2. The enzyme used in transcription is **RNA polymerase** rather than DNA polymerase. The RNA polymerase binds to a DNA sequence of 20–200 nucleotides in length called a **promoter** and then initiates transcription at a nucleotide in or near the promoter called the *transcription start site.*
3. In the synthesis of RNA, a sugar–phosphate bond is formed between the 3'-hydroxyl group of one nucleotide and the 5'-triphosphate of the next nucleotide in line (**Figure 11.7**, parts A and B). This is the same chemical bond as in the synthesis of DNA, but the enzyme is different and the polymer being synthesized is RNA.
4. The linear order of nucleotides in an RNA molecule is determined by the linear order of nucleotides in the DNA template. Each nucleotide added to the growing end of the RNA chain is chosen for its ability to base-pair with the DNA template strand. Thus the bases C, T, G, and A in a DNA strand cause G, A, C, and U, respectively, to be added to the growing end of an RNA molecule.
5. Nucleotides are added only to the 3'-OH end of the growing chain; as a result, the 5' end of a growing RNA molecule bears a triphosphate group. Note that the 5'-to-3' direction of RNA chain growth is the same as that in DNA synthesis.

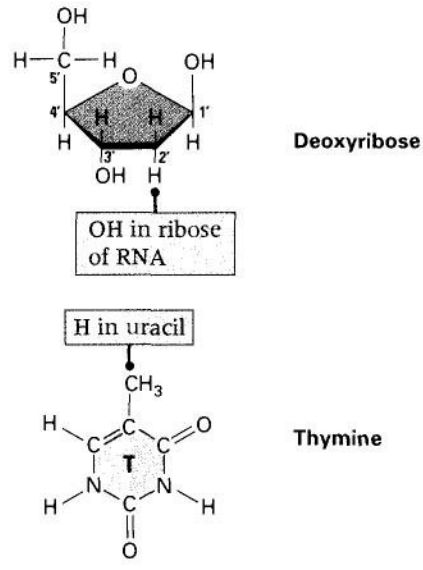


Figure 11.6 Differences between the structures of ribose and deoxyribose and between those of uracil and thymine.

A significant difference between DNA polymerase and RNA polymerase is that *RNA polymerase is able to initiate chain growth without a primer.* Furthermore:

> Each RNA molecule produced in transcription derives from a single strand of DNA, because in any particular region of the DNA, only one strand serves as a template for RNA synthesis.

The implications of this statement are shown in Figure 11.7C.

Types of RNA Polymerase

RNA polymerases are large, multisubunit complexes whose active form is called the **RNA polymerase holoenzyme.** Bacterial cells have only one RNA polymerase, the holoenzyme of which is symbolized by the formula $\beta\beta'\alpha_2\varphi\sigma$, which means that it contains six polypeptide chains (two copies of the α chain and one of each of the others). The molecular mass of the holoenzyme is 450,000 daltons, and its widest dimension is 150Å (**Figure 11.8**). This is about the same length as a stretch of 45 nucleotides in duplex DNA. But in transcriptional initiation, the holoenzyme actually contacts 70–90 bp in the promoter region, which means that the DNA must wrap around the holoenzyme. Once transcription begins, the region of contact is reduced to about 35 nucleotides, centered on the nucleotide being added. The *processivity* of RNA polymerase (the number of nucleotides transcribed without dissociating from the template) is impressive: more that 10^4 nucleotides in

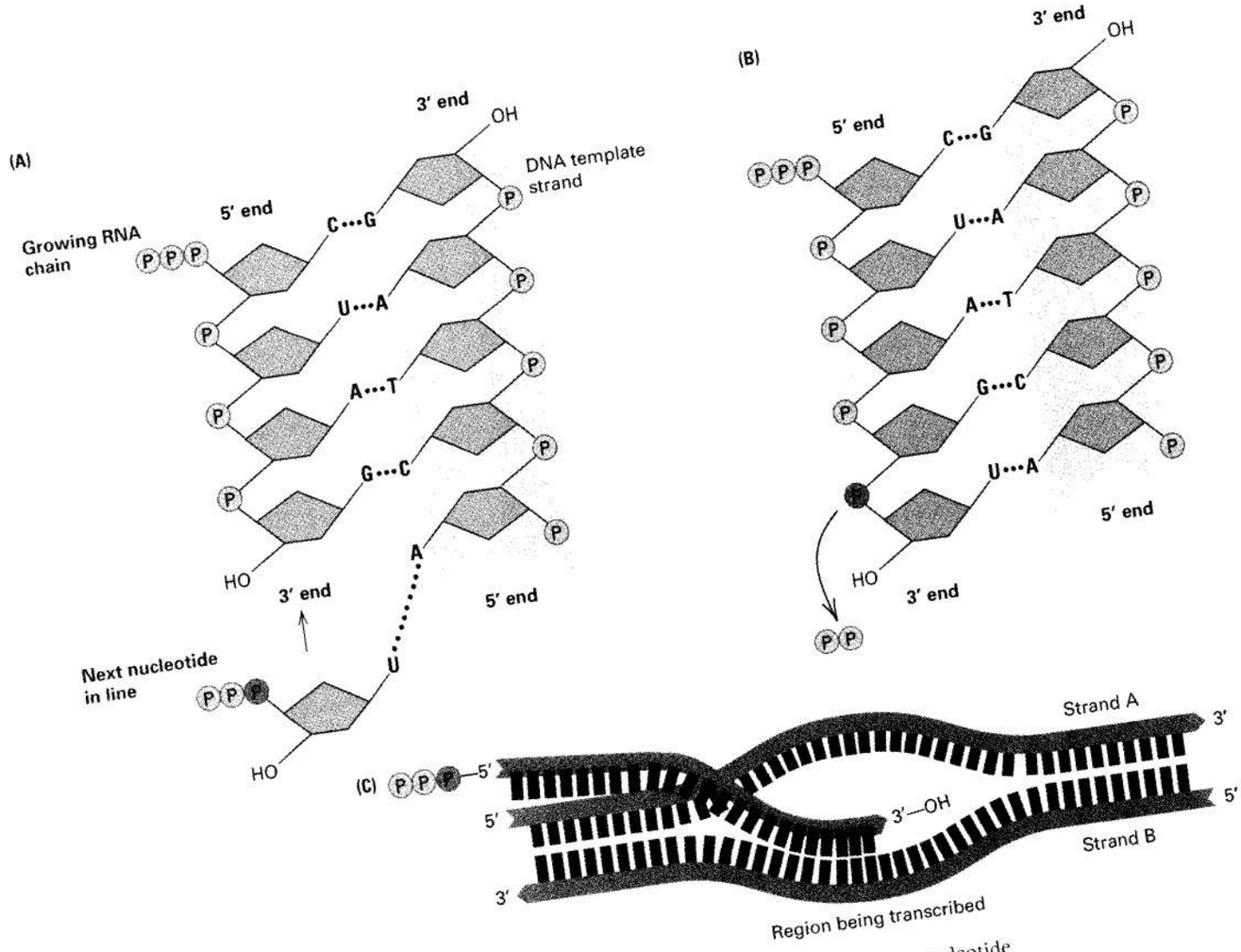


Figure 11.7 RNA synthesis. (A) The polymerization step in RNA synthesis. The incoming nucleotide forms hydrogen bonds (red dots) with a DNA base. (B) The —OH group in the growing RNA chain reacts with the orange P in the next nucleotide in line. (C) Geometry of RNA synthesis. RNA is copied from only one strand of a segment of a DNA molecule—in this example, strand B—without the need for a primer. In this region of the DNA, RNA is not copied from strand A. However, in a different region (for example, in a different gene) strand A might be copied rather than strand B. Because RNA elongates in the 5'-to-3' direction, its synthesis moves along the DNA template in the 3'-to-5' direction; that is, the RNA molecule is antiparallel to the DNA strand being copied.

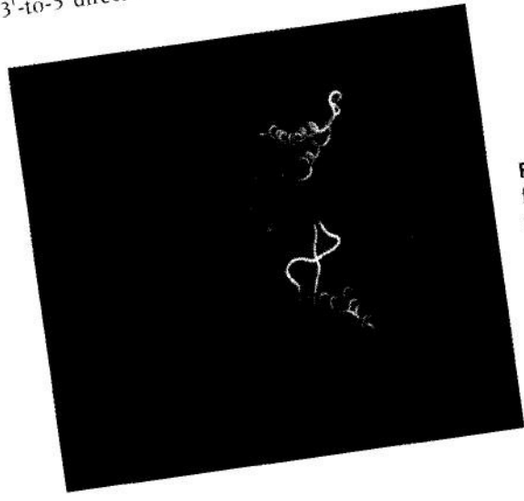


Figure 11.8 Subunit structure of RNA polymerase from the bacterium *Thermus aquaticus*. The complex has a U-shaped channel running through it. DNA to be transcribed passes through this channel, with the promoter region emerging at the bottom right and curving around the upper right. The colored coils highlight some of the structural features of the complex; the pink sphere in the center is a Mg^{2+} ion in the active site. [Courtesy of Rachel Anne Mooney and Robert Landick. 1999. *Cell* 98:687]

prokaryotes and more than 10^6 nucleotides in eukaryotes. Processivity is important, because once the RNA polymerase separates from the template, it cannot resume synthesis.

Eukaryotic RNA polymerases are even larger and include more subunits in the holoenzyme. There are also several different types:

1. ***RNA polymerase I*** is used exclusively in producing the transcript that becomes processed into ribosomal RNA. The promoter region includes the transcription start site.
2. ***RNA polymerase II*** is the workhorse eukaryotic polymerase responsible for transcribing all protein-coding genes as well as the genes for a number of small nuclear RNAs (U1, U2, U3, and so forth) used in RNA processing. The RNA polymerase II promoter is located near the transcription start site but upstream (on the 5' side) of it.
3. ***RNA polymerase III*** is used in transcribing all transfer RNA genes as well as the 5S component of ribosomal RNA. The promoter for RNA polymerase III transcription is located near the transcription start site but downstream (on the 3' side) of it.

Mechanism of Transcription

The mechanism of transcription can be described as consisting of four discrete stages.

1. *Promoter recognition* Many promoter sequences have been isolated and their nucleotide sequences determined. The promoters for polymerases I, II, and III show no commonality, and within each class there is substantial sequence variation. The situation in bacteria is considerably simpler. Although bacterial promoters also differ in sequence, in part because they differ in their polymerase binding affinity, certain sequence patterns or "motifs" are quite frequent. Two such patterns that are often found in promoter regions in *E. coli* are illustrated in **Figure 11.9**. Each pattern is defined by a **consensus sequence** of nucleotides determined from the actual sequences by majority rule: Each nucleotide in the consensus sequence is the nucleotide most often observed at that position in actual sequences. Any particular sequence may resemble the consensus sequence very well or very poorly.

The consensus sequences in some promoter regions in *E. coli* are shown in Figure 11.9, where the transcription start site is numbered as +1. One consensus motif is TTGACA, which is called the −35 motif because it is usually located approximately 35 base pairs upstream from the transcription start site. The other consensus motif is TATAAT, which is usually located near position −10. The −10 sequence is called the **TATA box**; it is similar to motifs found near the −25 position in many eukaryotic RNA polymerase II promoters. The positions of the promoter sequences determine where the RNA polymerase begins synthesis, and an A or G is often the first nucleotide in the transcript.

The strength of the binding of RNA polymerase to different promoters varies greatly, which causes differences in the extent of expression from one gene to an-

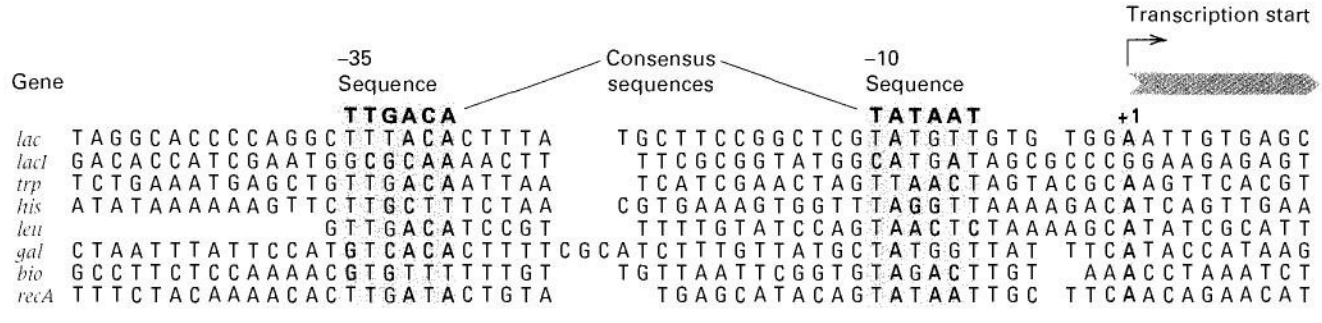


Figure 11.9 Nucleotide sequences in promoter regions of several genes in *E. coli*. The consensus sequences located 10 and 35 nucleotides upstream from the transcription start site (+1) are indicated. Promoters vary tremendously in their ability to promote transcription. Much of the variation in promoter strength results from differences between the promoter elements and the consensus sequences at −10 and −35.

other. Most of the differences in promoter strength result from variations in the −35 and −10 promoter elements and in the spacing between them. Promoter strength among *E. coli* genes differs by a factor of 10^4, and most of the variation can be attributed to the promoter sequences themselves. In general, the more closely the promoter elements resemble the consensus sequence, the stronger the promoter.

Mutations that change the nucleotide sequence in a promoter can alter the strength of the promoter. Changes that result in less resemblance to the consensus sequence lower the promoter's strength, whereas those with greater resemblance to the consensus increase it. Furthermore, some promoters differ greatly from the consensus sequence in the −35 region. These promoters typically require accessory proteins to activate transcription by RNA polymerase. In eukaryotes, in addition to the promoter sequences, there are also other DNA sequences called *enhancers* that interact with the promoter to determine the level of transcription. These are discussed in Chapter 12.

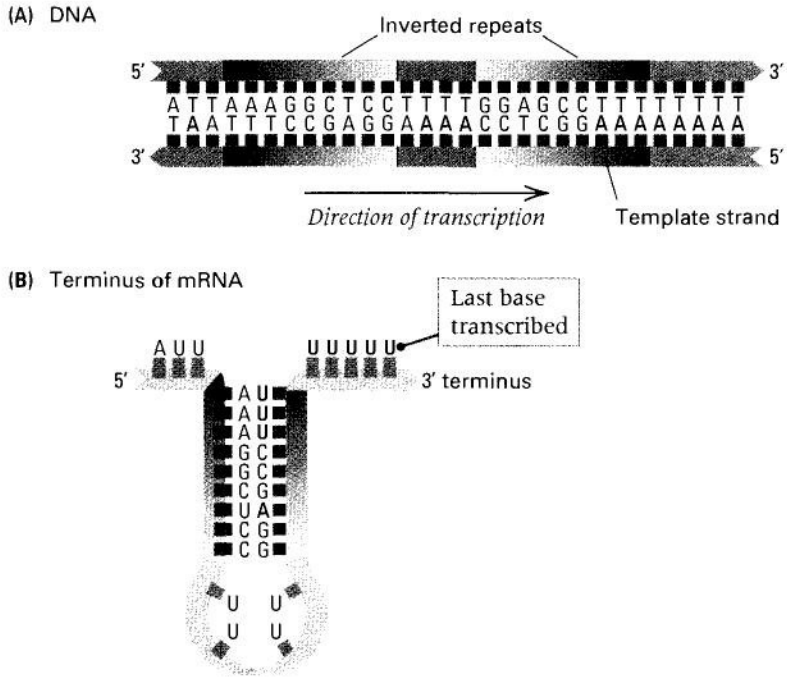


Figure 11.10 (A) Nucleotide sequence of the transcription–termination region for the set of tryptophan-synthesizing genes in *E. coli*. The inverted repeat sequences (blue) are characteristic of termination sites. (B) The 3' terminus of the RNA transcript, folded to form a stem-and-loop structure. The sequence of U's found at the end of the transcript in this and many other prokaryotic genes is shown in red. The RNA polymerase, not shown here, terminates transcription when the loop forms in the transcript.

2. *Chain initiation* After the initial binding step, the RNA polymerase "melts" (locally denatures) the DNA double helix, causing the strand that is to be transcribed to separate from its partner strand and become accessible to the polymerase. The RNA polymerase then initiates RNA synthesis at a nearby transcription start site (conventionally numbered the +1 site, as in Figure 11.9). The first nucleoside triphosphate is placed at this site, and synthesis proceeds in the 5'-to-3' direction.

3. *Chain elongation* RNA polymerase moves progressively along the transcribed DNA strand, adding nucleotides to the growing RNA chain. Only one DNA strand, the **template strand,** is transcribed.

4. *Chain termination* RNA polymerase reaches a chain-termination sequence, and both the newly synthesized RNA molecule and the polymerase are released. Two kinds of termination events are known: those that are self-terminating and depend only on the nucleotide sequence in the DNA template, and those that require the presence of a termination protein. In self-termination, which is the most common case, transcription stops when the polymerase encounters, in the transcribed DNA strand, a particular sequence of nucleotides that is able to fold back upon itself to form a hairpin loop. An example of such a terminator found in *E. coli* is shown in **Figure 11.10**. The hairpin loop alone is not enough for termination of transcription; the run of U's at the end of the hairpin is also necessary. The requirements for transcription termination in eukaryotes are more complex.

Initiation of a second round of transcription need not await completion of the first, because the promoter becomes available once RNA polymerase has polymerized from 50 to 60 nucleotides. For a rapidly transcribed gene, such reinitiation occurs repeatedly, and a gene can be cloaked with numerous RNA molecules in various degrees of completion. The micrograph in **Figure 11.11** shows a region of the



Figure 11.11 Electron micrograph of part of the DNA of the newt *Triturus viridescens*, containing tandem repeats of genes being transcribed into ribosomal RNA. The thin strands forming each feather-like array are RNA molecules. A gradient of lengths can be seen for each rRNA gene. Regions in the DNA between the thin strands are spacer DNA sequences, which are not transcribed. [Courtesy of Oscar Miller.]

DNA of the newt *Triturus* that contains tandem repeats of a ribosomal RNA gene. Each repeat is associated with growing transcripts. The shortest transcripts are at the promoter end of the repeat; the longest are near the terminus.

Genetic Evidence for Promoters and Terminators

The existence of promoters was first demonstrated in genetic experiments with *E. coli* by the isolation of a class of mutations in the lactose operon with the unique feature that the phenotype was Lac$^-$ only when the mutation was in the *cis* configuration (in the same DNA molecule) relative to the *lacZ* and *lacY* structural genes. These were denoted p^- mutations. The relevance of the *cis* configuration can be seen by examining a cell with two copies of the gene *lacZ*—for example, a cell containing an F' *lacZ* plasmid, which contains *lacZ* in the bacterial chromosome as well as *lacZ* in the F' plasmid. Transcription of the *lacZ* gene enables the cell to synthesize the enzyme β-galactosidase. **Table 11.2** shows that a wildtype *lacZ* gene (*lacZ*$^+$) is inactive when it and a p^- mutation are present in the same DNA molecule (either in the chromosome or in an F' plasmid); this can be seen by comparing entries 4 and 5. Analysis of the RNA shows that in a cell with the genotype p^- *lacZ*$^+$, the *lacZ*$^+$ gene is not transcribed, whereas in a cell with the genotype p^+ *lacZ*$^-$, a mutant RNA is produced. The p^- mutations were called *promoter mutations*.

Mutations in *E. coli* have also been instrumental in defining the transcription–termination region. Mutations have been isolated that create a new termination sequence upstream from the normal one. When such a mutation is present, an RNA molecule shorter than the wildtype RNA is made. Other mutations eliminate the terminator, resulting in a longer transcript.

Table 11.2 Effect of promoter mutations on transcription of the *lacZ* gene

Genotype	Transcription of *lacZ*$^+$ gene
1. $p^+ lacZ^+$	Yes
2. $p^- lacZ^+$	No
3. $p^+ lacZ^+ / p^+ lacZ^-$	Yes
4. $p^- lacZ^+ / p^+ lacZ^-$	No
5. $p^+ lacZ^+ / p^- lacZ^-$	Yes

Note: *lacZ*$^+$ is the wildtype gene; *lacZ*$^-$ is a mutant that produces a nonfunctional enzyme.

11.4 Messenger RNA

The RNA molecule produced from a DNA template is the **primary transcript.** Each gene has only one DNA strand that serves as the template strand, but which strand is the template strand can differ from gene to gene along a DNA molecule. Therefore, in

CoNnection

Messenger "Light"

Sydney Brenner,[1] François Jacob,[2] and Matthew Meselson[3] 1961

[1]Cavendish Laboratory, Cambridge, England

[2]Institute Pasteur, Paris, France.

[3]California Institute of Technology, Pasadena, California

An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis

Brenner and Jacob were guest investigators at the California Institute of Technology in 1961. At that time there was great interest in the mechanisms by which genes code for proteins. One possibility, which seemed reasonable at the time, was that each gene produced a different type of ribosome, differing in its RNA, which in turn produced a different type of protein. François Jacob and Jacques Monod had recently proposed an alternative, which was that the informational RNA ("messenger RNA") is actually an unstable molecule that breaks down rapidly. In this model, the ribosomes are nonspecific protein-synthesizing centers that synthesize different proteins according to specific instructions they receive from the genes through the messenger RNA. The key to the experiment is density-gradient centrifugation, which can separate macromolecules made "heavy" or "light" according to their content of ^{15}N or ^{14}N, respectively. (This techinque is described in Chapter 5.) The experiment is a purely biochemical proof of an issue absolutely critical for genetics—that genes code for proteins through the intermediary of a relatively short-lived messenger RNA.

A large amount of evidence suggests that genetic information for protein structure is encoded in deoxyribonucleic acid (DNA) while the actual assembling of amino acids into proteins occurs in cytoplasmic ribonucleoprotein particles called ribosomes. The fact that proteins are not synthesized directly on genes demands the existence of an intermediate information carrier. . . . Jacob and Monod have put forward the hypothesis that ribosomes are non-specialized structures which receive genetic information from the gene in the form of an unstable intermediate or "messenger." We present here the results of experiments on phage-infected bacteria which give direct support to this hypothesis. . . . When growing bacteria are infected with T2 bacteriophage, synthesis of DNA stops immediately, to resume 7 minutes later, while protein synthesis continues at a constant rate; in all likelihood, the protein is genetically determined by the phage. . . . Phage-infected bacteria therefore provide a situation in which the synthesis of a protein is suddenly switched from bacterial to phage control. . . . It is possible to determine experimentally [whether an unstable messenger RNA is produced] in the following way: Bacteria are grown in heavy isotopes so that all cell constituents are uniformly labelled "heavy." They are infected with phage and transferred immediately to a medium containing light isotopes so that all constituents synthesized after infection are "light." The distribution of new RNA and new protein, labelled with radioactive isotopes, is then followed by density gradient centrifugation of purified ribosomes. . . . We may summarize our findings as follows: (1) After phage infection no new ribosomes can be detected. (2) A new RNA with a relatively rapid turnover is synthesized after phage infection. This RNA, which has a base composition corresponding to that of the phage DNA, is added to pre-existing ribosomes, from which it can be detached in a cesium chloride gradient by lowering the magnesium concentration. (3) Most, if not all, protein synthesis in the infected cell occurs in pre-existing ribosomes. . . . The results also suggest that the messenger RNA may be large enough to code for long polypeptide chains. . . . It is a prediction of the messenger RNA hypothesis that the messenger RNA should be a simple copy of the gene, and its nucleotide composition should therefore correspond to that of the DNA. This appears to be the case in phage-infected cells. . . . If this turns out to be universally true, interesting implications for the coding mechanisms will be raised.

The results also suggest that the messenger RNA may be large enough to code for long polypeptide chains

Source: Nature 190: 576–581

an extended segment of a DNA molecule, primary transcripts would be seen growing in either of two directions (**Figure 11.12**), depending on which DNA strand functions as a template in a particular gene. In prokaryotes, the primary transcript serves directly as the **messenger RNA (mRNA)** used in polypeptide synthesis. In eukaryotes, the primary transcript is generally processed before it becomes mRNA.

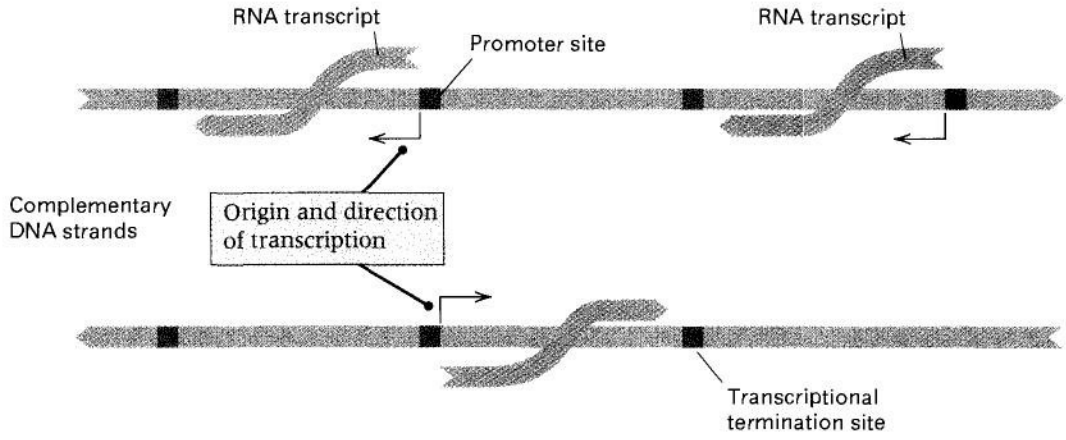


Figure 11.12 A typical arrangement of promoters (green) and termination sites (red) in a segment of a DNA molecule. Promoters are present in both DNA strands. Termination sites are usually located such that transcribed regions do not overlap.

Translation of an mRNA molecule never starts exactly at the 5' end and proceeds straightaway to the other. Initiation of polypeptide synthesis usually begins many nucleotides downstream from the 5' end of the mRNA. The 5' untranslated region is followed by the **coding sequence,** also called the **open reading frame (ORF),** which specifies the order in which the amino acids are present in the polypeptide chain. A typical coding sequence in an mRNA molecule is 500–3000 nucleotides in length, which is translated in nonoverlapping sets of three nucleotides (the *codons*). Beyond the coding region in the mRNA, following the termination codon, is a 3' untranslated region.

In prokaryotes, most mRNA molecules are degraded within a few minutes after synthesis. In eukaryotes, a typical lifetime is several hours, although some last only minutes, and others persist for days. In both kinds of organisms, the degradation enables cells to dispose of molecules that are no longer needed. The short lifetime of prokaryotic mRNA is an important factor in regulating gene activity (Chapter 12).

11.5 RNA Processing in Eukaryotes

The process of transcription is very similar in prokaryotes and eukaryotes, but there are major differences in the relationship between the transcript and the mRNA used for polypeptide synthesis. In prokaryotes, the primary transcript is the mRNA. In contrast, the primary transcript in eukaryotes must be converted into mRNA. This conversion is called **RNA processing** and usually consists of two types of events: modification of the ends and excision of untranslated sequences embedded within the coding sequences. These events are illustrated diagrammatically in **Figure 11.13.**

5' Capping and 3' Polyadenylation

Each end of a eukaryotic transcript is processed (Figure 11.13). The 5' end is altered by the addition of a modified guanosine (7-methyl guanosine) in an uncommon 5'-to-5' (instead of 3'-to-5') linkage; this terminal group is called a **cap.** The 5' cap is necessary for the mRNA to bind with the ribosome to begin protein synthesis. The 3' terminus of a eukaryotic mRNA molecule is usually modified by "trimming" (exonuclease digestion) back to a short sequence that serves as a substrate for the addition of a polyadenylate sequence (the **poly-A tail**) of as many as 200 nucleotides. The poly-A tail helps to determine mRNA stability.

Splicing of Intervening Sequences

The important feature of *splicing* of the primary transcript in eukaryotes is also shown in Figure 11.13. The segments that are excised from the primary transcript are called **introns** or **intervening sequences.** Accompanying the excision of introns is a rejoining of the segments that remain, which are called **exons,** to form the mRNA molecule. The excision of the introns and the joining of the exons constitute **RNA splicing.** The mechanism of RNA splicing is illustrated schematically in **Figure 11.14.** Part A shows the consensus sequence found at the 5' (*donor*) end and at the 3' (*acceptor*) end of

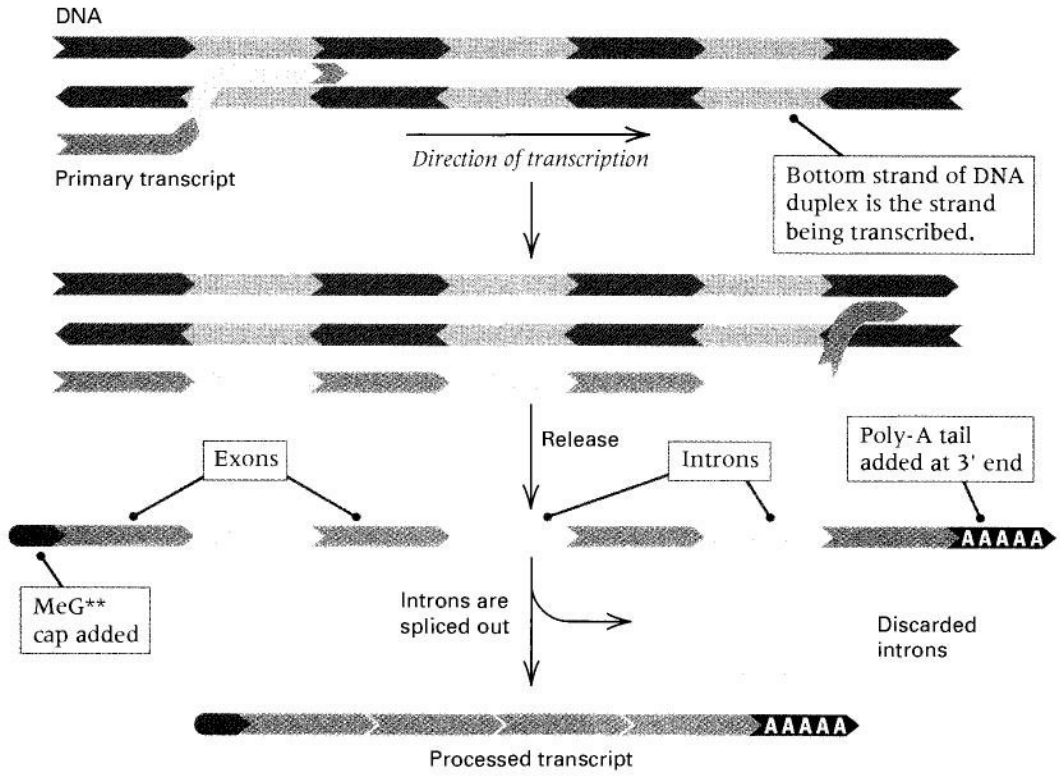


Figure 11.13 In eukaryotes, mRNA is produced by processing of the primary transcript. The primary transcript is capped before it is released from the DNA. MeG denotes 7-methylguanosine (a modified form of guanosine), and the two asterisks indicate two nucleotides whose riboses are methylated. The 3' end is usually modified by the addition of consecutive adenines. Along the way, the introns are excised. These reactions take place within the nucleus.

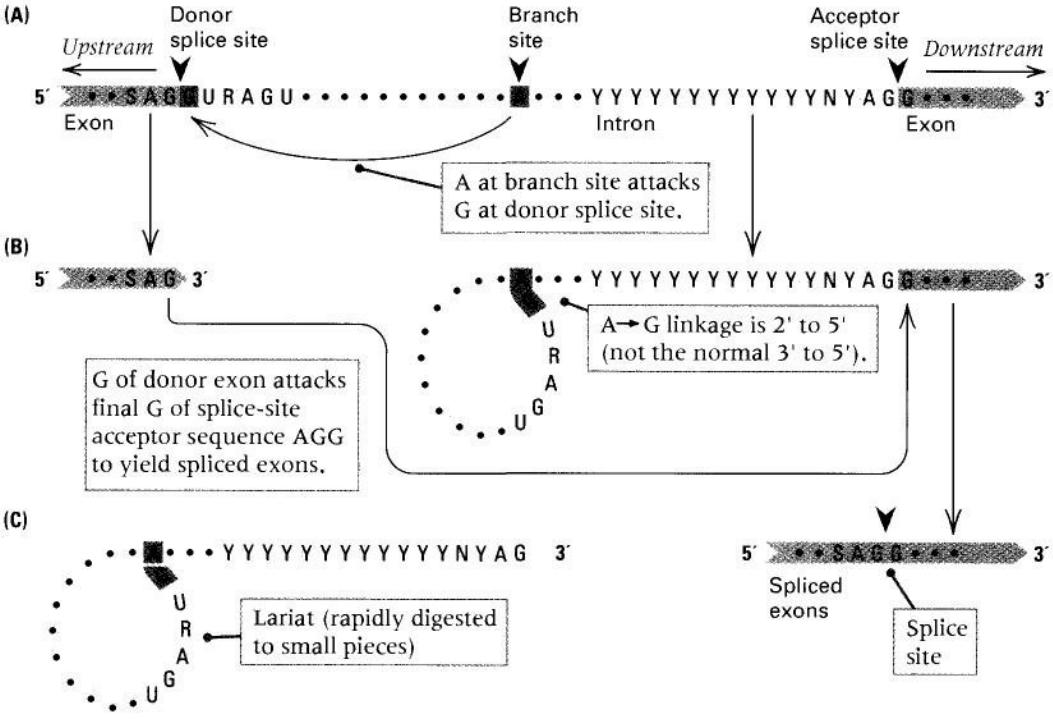


Figure 11.14 Diagram showing removal of one intron from a primary transcript. (A) The A nucleotide at the branch site attacks the terminus of the 5' exon, cleaving the exon–intron junction and forming a loop connected back to the branch site. (B) The 5' exon is later brought to the site of cleavage of the 3' exon, a second cut is made, and the exon termini are joined and sealed. (C) The loop is released as a lariat-shaped structure that is degraded. Because the loop includes most of the intron, the loop of the lariat is usually very much longer than the tail.

most introns. The symbols are N, any nucleotide; R, any purine (A or G); Y, any pyrimidine (C or U); and S, either A or C. In the first step of splicing, the 2'−OH of the adenylate (A) at the branch site, which is located a short distance upstream from a run of prymidines (Y) near the acceptor site, attacks the phosphodiester bond at the donor splice site junction. The attack results in cleavage at the donor splice site and formation of a branched molecule (Figure 11.14B) known as a **lariat** because it has a loop and a tail. The A−G linkage at the "knot" of the lariat is unusual in being 2'-to-5' (instead of the usual 3'-to-5'). In the final step of splicing (Figure 11.14C), the 3'−OH of the guanylate of the donor exon attacks the phosphodiester bond at the acceptor splice site, freeing the lariat intron and joining the donor and acceptor exons together. The lariat intron is rapidly degraded into individual nucleotides by nucleases.

RNA splicing takes place in nuclear particles known as **spliceosomes.** These abundant particles are composed of protein and several specialized small RNA molecules, which are present in the cell as small nuclear ribonucleoprotein particles; the un-

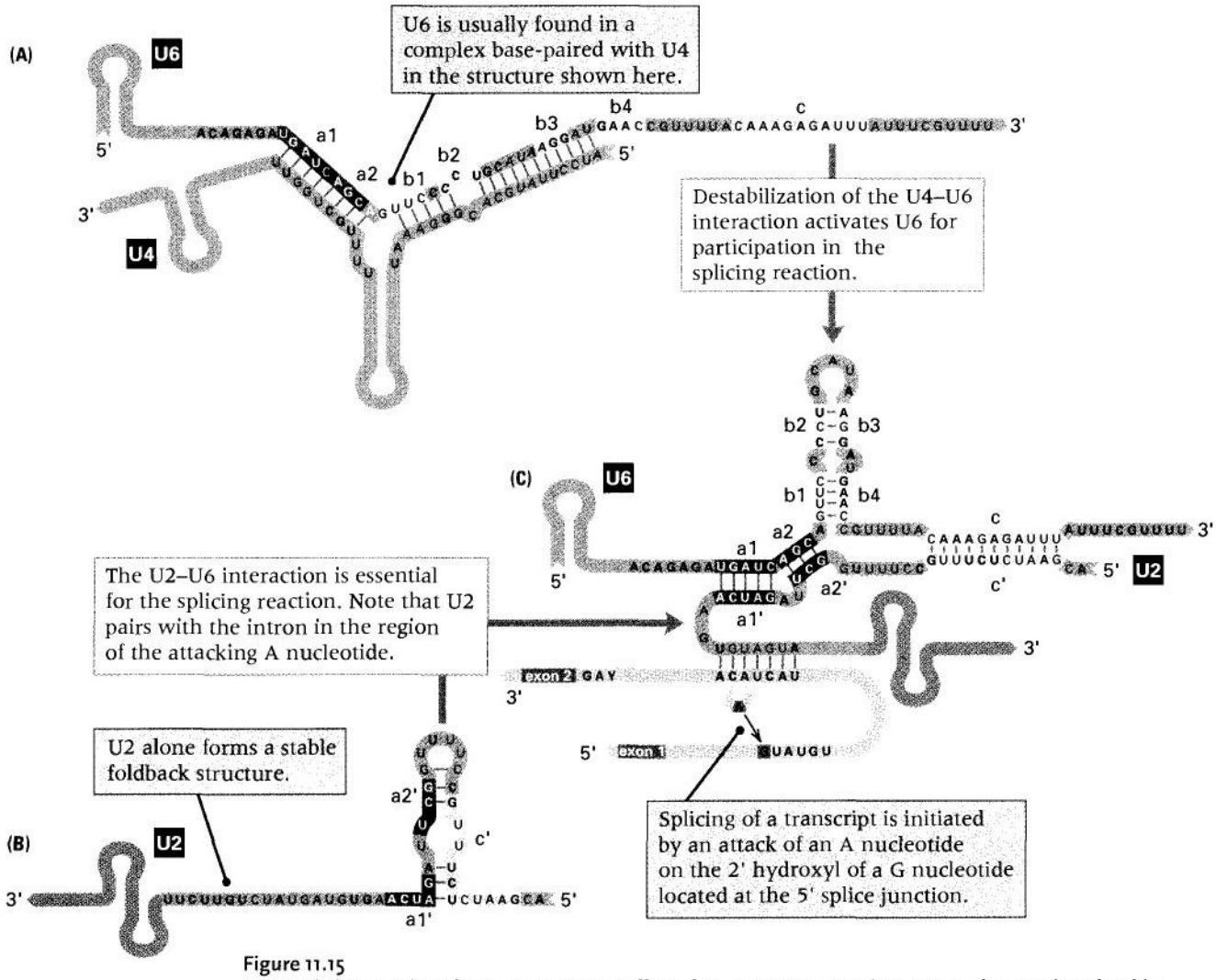


Figure 11.15
Dynamic interactions between some small nuclear RNAs present in snRNPs that are involved in splicing. (A) U6 snRNA is usually found complexed with U4 snRNA. (B) U2 snRNA forms a stable foldback structure on its own. (C) Essential to the splicing reaction is destabilization of the U4−U6 structure and formation of a U2−U6 structure in which U2 is base-paired with part of the intron. An A in the paired region attacks the G at the 5' splice junction, initiating the splicing reaction. The nucleotides in boldface are critical to the structures, judging by their having been conserved in very diverse species. Note that G−U base pairs are allowed in double-stranded RNA. [After H. D. Madhani and C. Guthrie. 1994. *Annual Reviews of Genetics* 28: 1.]

derlined letters give the abbreviation for these particles: **snRNPs** (pronounced *snerps*). The specificity of splicing comes from the five types of short RNA molecules (U1, U2, U4, U5, and U6) that are present in snRNPs. These RNAs contain sequences complementary to the splice junctions, to the branchpoint region of the intron, and/or to one another (**Figure 11.15**); as many as 100 spliceosome proteins may also be required for splicing. The ends of the intron are brought together by U1 RNA, which forms base pairs with nucleotides in the intron at both the 5' and the 3' ends, and U2 binds to the branchpoint region. As illustrated in Figure 11.15C, base pairing in the spliceosome includes U2 and U6, both of which are normally stabilized either by foldback pairing, in the case of U2 (part C) or pairing with U4, in the case of U6 (part A). These interactions bring the branchpoint region near to the donor splice site and allow the A in the branchpoint to attack the G of the donor splice site, freeing the upstream exon and forming the lariat intermediate (Figure 11.14). U5 RNA helps align the two exons and somehow facilitates the final step in splicing, which results in scission of the intron from the downstream exon and ligation of the upstream and downstream exons.

Introns are also present in some genes in organelles, such as mitochondria, but the mechanisms of their excision differ from those of introns in nuclear genes because organelles do not contain spliceosomes. In one class of organelle introns, the intron contains a sequence coding for a protein that participates in removing the intron that codes for it. The situation is even more remarkable in the splicing of a ribosomal RNA precursor in the ciliate *Tetrahymena*. In this case, the splicing reaction is intrinsic to the folding of the RNA precursor; that is, the RNA precursor is *self-splicing* because the folded precursor RNA creates its own RNA-splicing activity. The self-splicing *Tetrahymena* RNA was the first example found of an RNA molecule that could function as an enzyme in catalyzing a chemical reaction; such enzymatic RNA molecules are called *ribozymes*.

The existence and the positions of introns in a particular primary transcript are readily demonstrated by renaturing the transcribed DNA with the fully processed mRNA molecule. The DNA–RNA hybrid can then be examined by electron microscopy. An example of adenovirus mRNA (fully processed) and the corresponding DNA are shown in **Figure 11.16**. The DNA copies of the introns appear as single-stranded loops in the hybrid molecule, because no corresponding RNA sequence is available for hybridization.

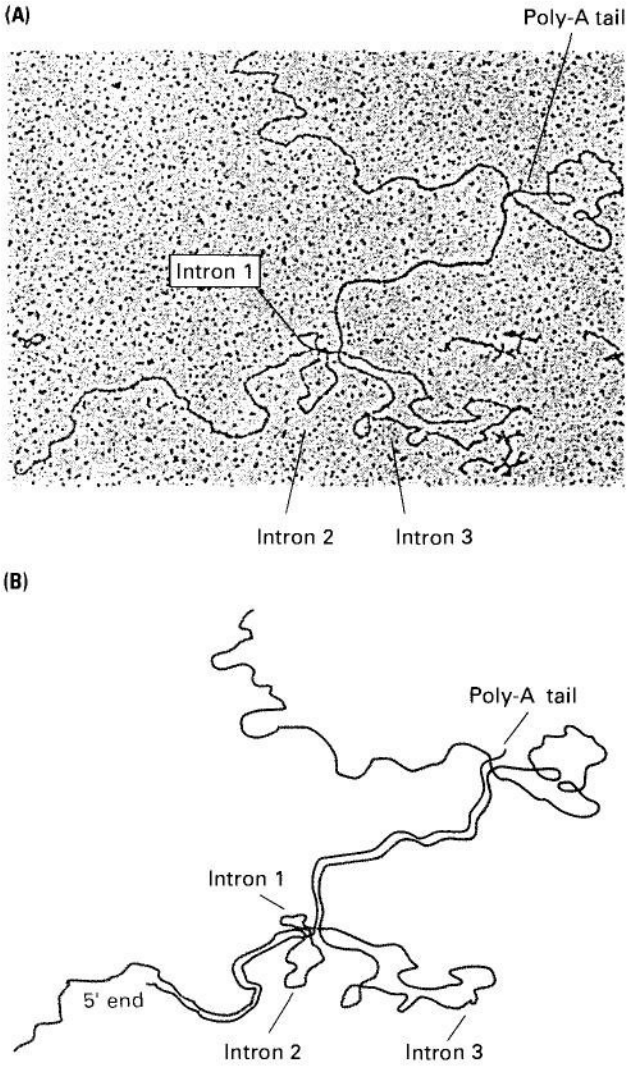


Figure 11.16 (A) An electron micrograph of a DNA–RNA hybrid obtained by annealing a single-stranded segment of adenovirus DNA with one of its mRNA molecules. The loops are single-stranded DNA. (B) An interpretive drawing. RNA and DNA strands are shown in red and blue, respectively. Four regions do not anneal, creating three single-stranded DNA segments that correspond to the introns and the poly-A tail of the mRNA molecule. [Electron micrograph courtesy of Tom Broker and Louise Chow.]

The number of introns per RNA primary transcript varies considerably from one gene to the next. One of the major genes for inherited breast cancer in women (*BRCA1*) contains 21 introns spread across more than 100,000 nucleotides. More than 90 percent of this primary transcript is excised in processing, yielding a processed mRNA of about 7800 nucleotides, which codes for a polypeptide chain of 1863 amino acids. Among the human genes with a simpler intron–exon structure is that for α-globin, which contains two introns. Introns vary greatly in size as well as in number. In human beings and other mammals, most introns range in size from 100 to 10,000 nucleotides, and in the processing of a typical primary transcript, the amount of discarded RNA ranges from about 50 to 90 percent of the primary transcript. Genes in lower eukaryotes, such as yeast, nematodes, and fruit flies, generally have fewer introns than do genes in mammals, and the introns tend to be much smaller.

Mutations in Introns

Many mutations in introns have no detectable effects on gene function. These include small deletions and insertions as well as nucleotide substitutions. In most genes, complete removal of an intron does not impair gene function, and in examples where it does, it is usually not because the interruption of the gene is necessary, but because the intron happens to include nucleotide sequences that regulate the timing or tissue specificity of transcription. Moreover, the nucleotide sequence of a particular intron usually changes extremely rapidly in the course of evolution, which suggests that there are few constraints on the sequence. These observations imply that introns have no essential function in themselves. However, examination of the complete DNA sequence of the euchromatin of *D. melanogaster* reveals that some large introns contain genes within them and that, in a few cases, the contained genes themselves contain other genes. How frequently these nested configurations occur in higher metazoans and plants is unknown.

On the other hand, introns do contain critical splicing signals (Figure 11.14), and mutations that affect any of these sequences have important consequences because they interfere with the splicing reaction. Two possible outcomes are illustrated in **Figure 11.17**. In Figure 11.17A, the intron with the mutated splice site fails to be removed, and it remains in the processed mRNA. The result is the production of a mutant protein with a normal sequence of amino acids up to the splice site but an abnormal sequence afterward. Most introns are long enough that, by chance, they contain a stop sequence that terminates protein synthesis, and once a stop is encountered, the protein grows no further. A second kind of outcome is shown in Figure 11.17B. In this case, splicing does occur, but at an alternative splice site. (The example shows the alternative site downstream from the mutation, but alternative sites can also be upstream.) The alternative site is called a **cryptic splice site** because it is not normally used. The cryptic splice site is usually a poorer match with the consensus sequence and is ignored when the normal splice site is available. The result of using the alternative splice site is again an incorrectly processed mRNA and a mutant protein. In some splice-site mutations, both outcomes (Figure 11.17A and B) can occur: Some transcripts leave the intron unspliced, whereas others are spliced at cryptic splice sites.

Exon Shuffle in the Origin of New Genes

Although introns are not usually essential in regulating gene expression, they may play a role in gene evolution. In some cases, the exons in a gene code for segments of the completed protein that are relatively independent in their folding characteristics. For example, the central exon of the β-globin gene codes for the segment of the protein that folds around an iron-containing molecule of heme. Relatively autonomous folding units in proteins are known as **folding domains,** and the correlation between exons and domains found in some genes suggests that the genes were originally assembled from

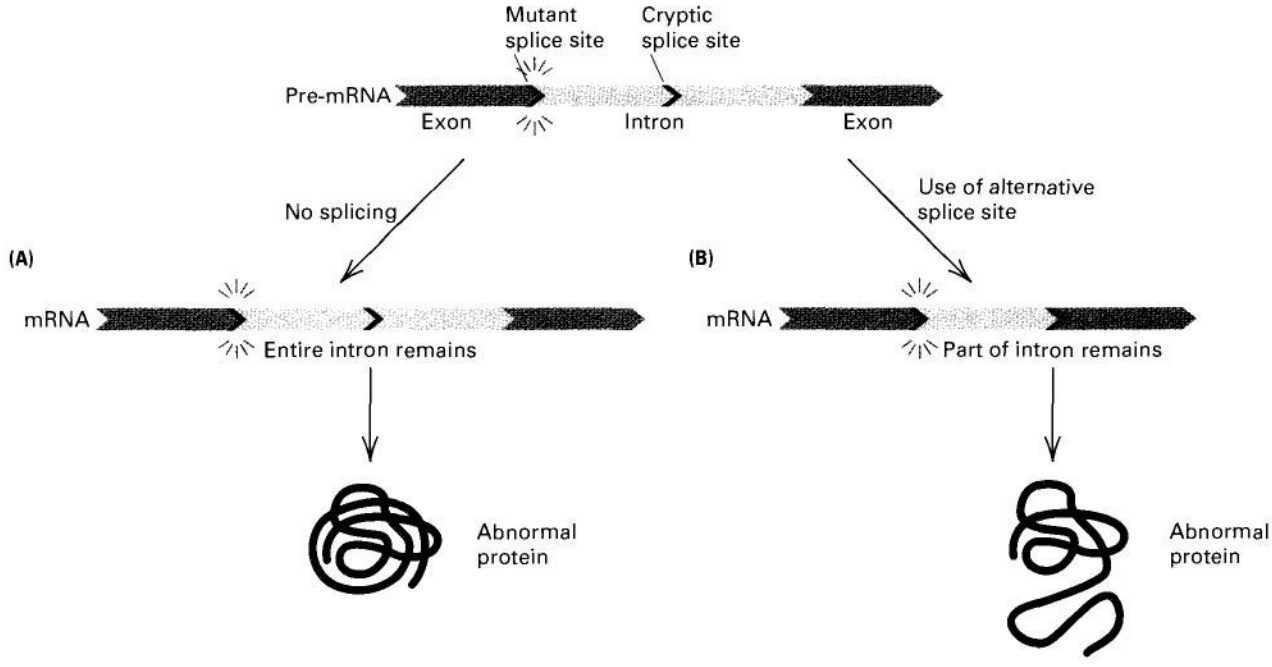


Figure 11.17 Possible consequences of mutation in the donor splice site of an intron. (A) No splicing occurs, and the entire intron remains in the processed transcript. (B) Splicing occurs at a downstream cryptic splice site, and only the upstream part of the original intron remains in the processed transcript. Neither outcome results in a normal protein product.

smaller pieces. In some cases, the ancestry of the exons can be traced. For example, the human gene for the low-density lipoprotein receptor that participates in cholesterol regulation shares exons with certain blood-clotting factors and epidermal growth factor. The model of protein evolution through the combination of different exons is called the **exon-shuffle model.** The mechanism for combining exons from different genes is not known. Although some genes support the model, in other genes the boundaries of the folding domains do not coincide with exons.

The evolutionary origin of introns is unknown. On the one hand, introns may be an ancient feature of gene structure. The existence of self-splicing RNAs means that introns could have existed long before the evolution of the spliceosome mechanism, so some introns may be as old as the genes themselves. Furthermore, the finding that some genes have introns in the same places in both plants and animals suggests that the introns may have been in place before plants and animals became separate lineages. If introns are ancient, then exon shuffling might have been important early in evolution by creating new genes with novel combinations of exons. It has even been suggested that all forms of early life had introns in their genes and that today's prokaryotes, which lack introns, lost their introns in their evolution. On the other hand, there are many examples in which new introns have became inserted into already existing genes, particularly in vertebrate genomes.

11.6 Translation

The synthesis of every protein molecule in a cell is directed by an mRNA originally copied from DNA. Polypeptide synthesis includes two kinds of processes: (1) the information transfer by which the RNA nucleotide sequence determines the amino acid sequence, and (2) the chemical processes by which the amino acids are linked together. The complete series of events constitutes **translation.**

The principal components of the translation apparatus are as follows:

- ***Messenger RNA,*** or ***mRNA*** Messenger RNA is needed to bring the ribosomal subunits together (described below) and to provide the coding sequence of nucleotides that determines the amino acid sequence in the resulting polypeptide chain.
- ***Ribosomes*** These components are particles on which protein synthesis takes place. They move along an mRNA molecule and align successive transfer RNA molecules; the amino acids are attached one by one to the growing polypeptide chain by means of peptide bonds. Ribosomes consist of two subunit particles. In *E. coli*, their sizes are 30S (the small subunit) and 50S (the large subunit). The counterparts in eukaryotes are 40S and 60S. (The S stands for *Svedberg unit*, which measures the rate of sedimentation of a particle in a centrifuge and so is an indicator of size.) Together, the small and large particles form a functional ribosome.
- ***Transfer RNA,*** or ***tRNA*** The sequence of amino acids in a polypeptide is determined by the nucleotide sequence in the mRNA by means of a set of adaptor molecules, the tRNA molecules, each of which is attached to a particular amino acid. Each group of three adjacent nucleotides in the mRNA forms a **codon** that binds to a particular group of three adjacent nucleotides in the tRNA (an **anticodon**), bringing the attached amino acid into line for addition to the growing polypeptide chain.
- ***Aminoacyl tRNA synthetases*** This set of enzymes catalyzes the attachment of each amino acid to its corresponding tRNA molecule. A tRNA attached to its amino acid is called a **charged tRNA** or an **aminoacylated tRNA.**
- ***Initiation, elongation, and release factors*** Polypeptide synthesis can be divided into three stages: (1) initiation, (2) elongation, and (3) release. Each stage requires specialized proteins that are examined below.

In prokaryotes, all of the components for translation are present throughout the cell; in eukaryotes, they are located in the cytoplasm, as well as in mitochondria and chloroplasts.

In overview, the process of translation begins with an mRNA molecule binding to a ribosome. The aminoacylated tRNAs are brought along sequentially, one by one, to the ribosome that is translating the mRNA molecule. Peptide bonds are made between successively aligned amino acids, each time joining the amino group of the incoming amino acid to the carboxyl group of the amino acid at the growing end. Finally, the chemical bond between the last tRNA and its attached amino acid is broken, and the completed polypeptide is removed.

Initiation by mRNA Scanning

We will examine the processes of translation as they occur in eukaryotes, pointing out differences in the prokaryotic mechanism that are significant. In the predominant mode of translation **initiation** in eukaryotes, the 5' cap on the mRNA is instrumental (**Figure 11.18**). First the elongation factor eIF4F binds to the cap and then recruits eIF4A and eIF4B (Figure 11.18A). This creates a binding site for the other components of the initiation complex, which consist of a charged $tRNA^{Met}$ (that serves as an initiator tRNA), bound with elongation factor eIF2, and a small 40S ribosomal subunit together with elongation factors eIF3 and eIF5. These components all come together at the 5' cap and form the 48S initiation complex (Figure 11.18B).

Once the initiation complex has formed, it moves along the mRNA in the 3' direction, scanning for the first occurrence of the nucleotide sequence AUG, which signals the start of polypetide synthesis. When this motif is encountered, the AUG is recognized as the initial methionine codon, and polypeptide synthesis begins. At this point eIF5 causes the release of all the initiation factors and the recruitment of a large 60S ribosomal subunit (Figure 11.18C). This subunit includes three binding sites for tRNA molecules. These sites are called the **E (exit)** site, the **P (peptidyl)** site, and the **A (aminoacyl)** site. Note that at the beginning of polypeptide synthesis, the initiator methionine tRNA is located in the P site and that the A site is the next site in line to be occupied. The tRNA binding is accomplished by hydrogen bonding between bases in the AUG codon in the mRNA and the three-base *anticodon* in the tRNA.

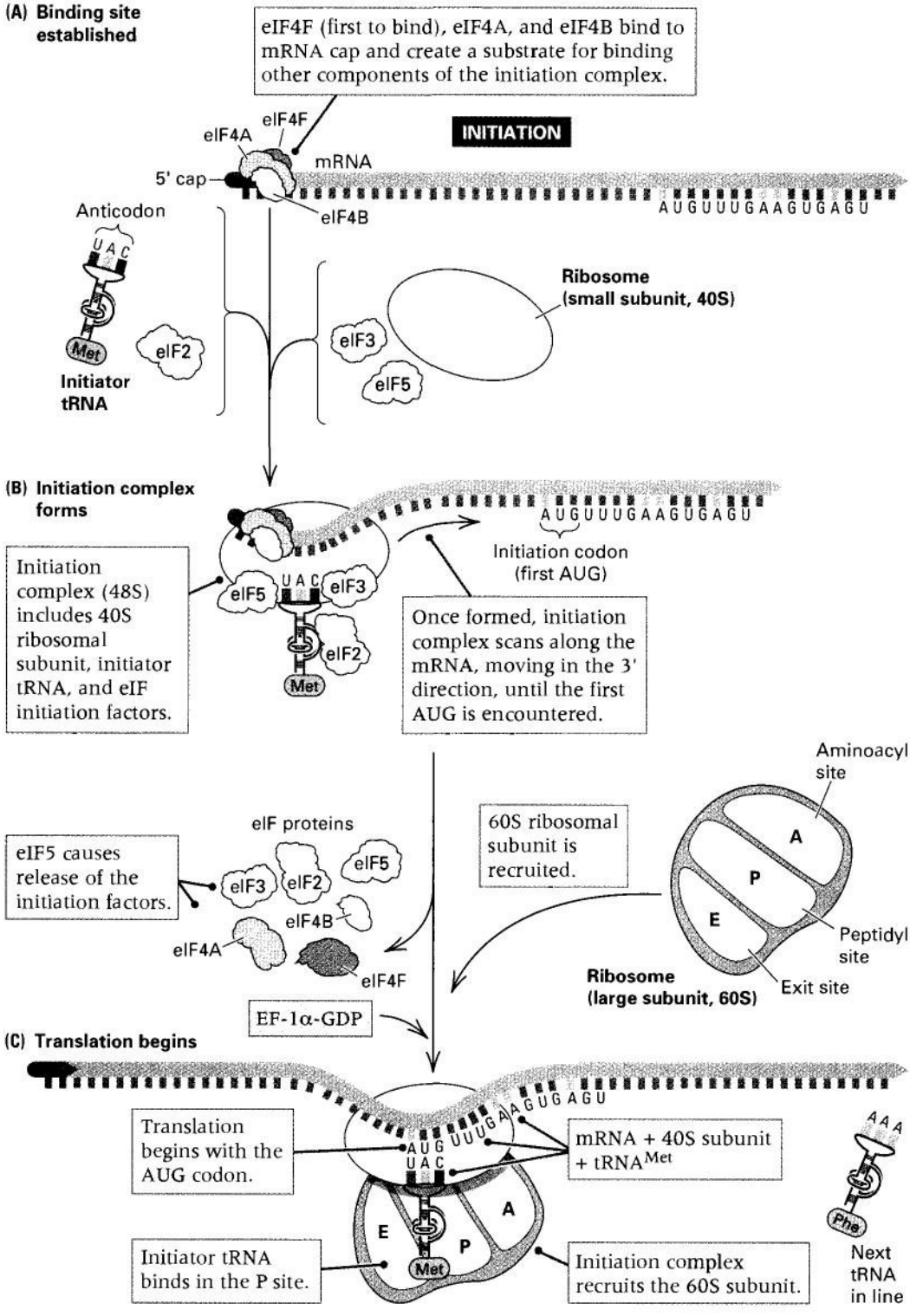


Figure 11.18 Initiation of protein synthesis. (A) The initiation complex forms at the 5' end of the mRNA. (B) This consists of one 40S ribosomal subunit, the initiator $tRNA^{Met}$, and the eIF initiation factors. (C) The initiation complex recruits a 60S ribosomal subunit in which the $tRNA^{Met}$ occupies the P (peptidyl) site of the ribosome. This complex travels along the mRNA until the first AUG is encountered, at which codon translation begins.

Prokaryotic mRNA molecules have no cap, and there is no scanning mechanism to locate the first AUG. In the initiation of translation in *E. coli*, two initiation factors (IF-1 and IF-3) interact with the 30S subunit at the same time that another initiation factor (IF-2) binds with a special initiator tRNA charged with formylmethionine, symbolized $tRNA^{fMet}$. These components come together and combine with an mRNA, but not at the end. The attachment occurs by hydrogen bonding between the 3' end of the 16S RNA present in the 30S subunit and a special sequence, the **ribosome-binding site,** in the mRNA (also called the *Shine-Dalgarno sequence*). Together, the 30S + $tRNA^{fMet}$ + mRNA complex recruits a 50S subunit, in which the $tRNA^{fMet}$ is positioned in the P site and aligned with the AUG initiation codon, just as in Figure 11.18C. In the assembly of the completed ribosome, the initiation factors dissociate from the complex.

One major functional difference between translational initiation in eukaryotes and prokaryotes is that in eukaryotes, because of the scanning mechanism of initiation, a single mRNA can usually encode only one polypeptide chain. In prokaryotic mRNA, in contrast, the ribosome-binding site can be present anywhere near an AUG, so polypeptide synthesis can begin at any AUG that is closely preceded by a ribosome-binding site. Prokaryotes put this feature to good use because, as we shall see later, many of their mRNAs contain more than one ribosome-binding site and initiation codon and hence code for more than one polypeptide, because translation can be initiated at any of the sites.

Elongation

Recruitment of the elongation factor EF-1α-GTP into the initiation complex begins the **elongation** phase of polypeptide synthesis (Figure 11.18C). Elongation consists of three processes executed iteratively:

1. Bringing each new aminoacylated tRNA into line
2. Forming the new peptide bond to elongate the polypeptide
3. Moving the ribosome to the next codon along the mRNA

The first step in elongation is illustrated in **Figure 11.19**, part A. Here the EF-1α, bound with guanosine triphosphate (EF-1α-GTP), plays a key role, bringing the next aminoacylated tRNA into the A site on the 60S subunit, which in this example is $tRNA^{Phe}$. This process requires the hydrolysis of GTP to GDP, and once the GDP is formed, the EF-1α-GDP has low affinity for the ribosome and diffuses away, becoming available for reconversion into EF-1α-GTP.

Once the A site is filled, a **peptidyl transferase** activity catalyzes a concerted reaction in which the bond connecting the methionine to the $tRNA^{Met}$ is transferred to the amino group of the phenylalanine, forming the first peptide bond. Peptidyl transferase activity is not due to a single molecule but requires multiple components of the 60S subunit, including several proteins and the ribosomal RNA (called 28S) in the 60S subunit. Some evidence indicates that the actual catalysis is carried out by the 28S RNA, which would suggest that 28S is an example of a ribozyme at work. The configuration of the subunits in Figure 11.19A is called the **pretranslocation state.**

The next step in chain elongation is **translocation,** which means the movement of the 40S subunit one codon further along the mRNA. The ribosome in the pretranslocation state releases EF-1α-GDP and binds instead with elongation factor EF-2-GTP in the same ribosomal binding site. Hydrolysis of the GTP to GDP yields the energy to shift the tRNAs in the P and A sites to the E and P sites, respectively, as well as to translocate the 40S subunit one codon along the mRNA (Figure 11.19B). The ribosome is thereby converted to the **post-translocation state.**

One cycle of elongation is now completed, and the entire procedure is repeated for the next codon. The 60S subunit shifts one step to the right, the EF-2-GDP is released and replaced with EF-1α-GTP, and the ribosome is reconfigured into the pretranslocation state (Figure 11.19C). At this stage, EF-1α-GTP does its job again, and four events happen, as indicated in Figure 11.19C:

- The next aminoacylated tRNA is brought into line (in this case, $tRNA^{Glu}$).
- The uncharged tRNA is ejected from the E site.

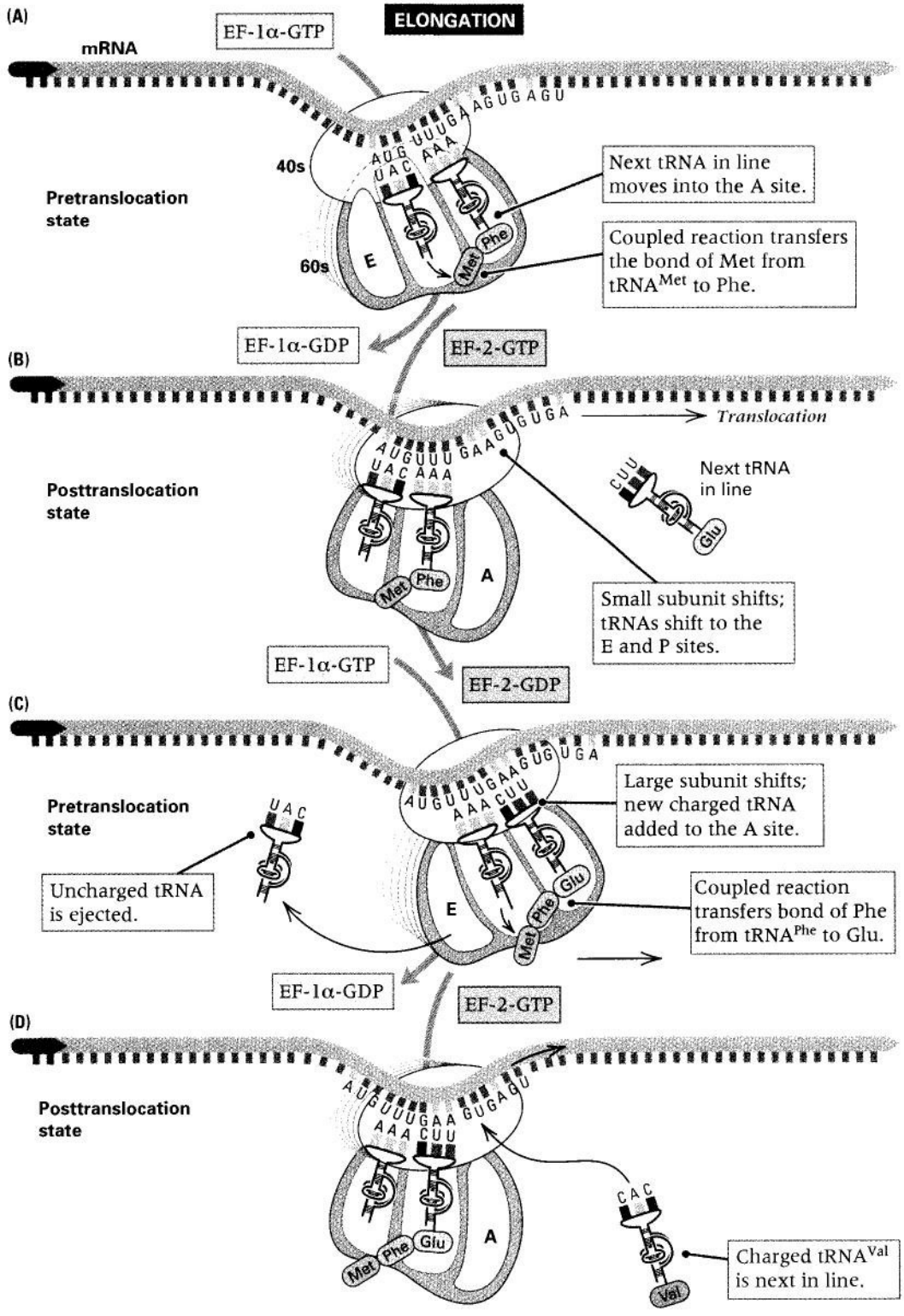


Figure 11.19 Elongation cycle in protein synthesis. (A) Pretranslocation state. (B) Posttranslocation state, in which an uncharged tRNA occupies the E site and the polypeptide is attached to the tRNA in the P site. (C) The function of EF-1α is to release the uncharged tRNA and bring the next charged tRNA into the A site, at which time a peptide bond is formed between the polypeptide and the amino acid held in the A site, in this case Glu. Simultaneously, the 60S subunit is shifted relative to the 40S subunit, re-creating the pretranslocation state. (D) The function of EF-2 is to translocate the 40S ribosome to the next codon, once again generating the posttranslocation state.

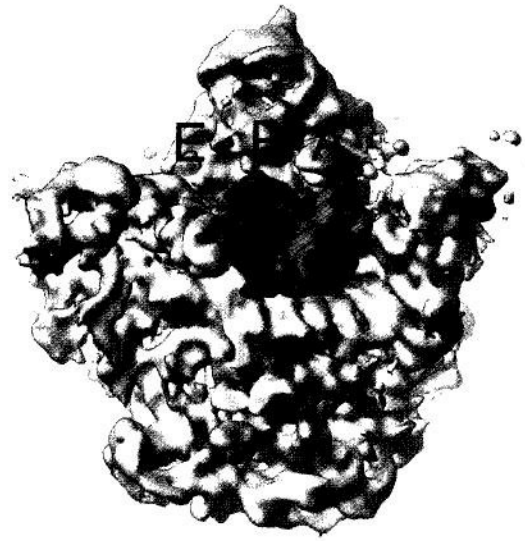


Figure 11.20 Cutaway view of a bacterial ribosome, showing the groove in which the tRNAs are bound and the positions of tRNAs when present in the E (exit) site in gold, the P (peptidyl) site in blue, and the A (aminoacyl) site in green. [Courtesy of Jamie H. Cate. From J. H. Cate, M. M. Yusupov, G. Zh. Yusupova, T N. Earnest, and H. F. Noller. 1999. *Science* 285:2095.]

- In a concerted reaction, the bond connecting the growing polypeptide chain to the tRNA in the P site is transferred to the amino group of the amino acid in the A site, forming the new peptide bond.
- Translocation of the 40S subunit one codon to the right follows release of EF-1α-GDP and binding with EF-2-GTP, returning the ribosome to the posttranslocation state (Figure 11.19D).

Polypeptide elongation consists of the steps C → D → C → D carried out repeatedly until a termination codon is encountered. The elongation cycle happens relatively rapidly. Under optimal conditions, eukaryotes synthesize a polypeptide chain at the rate of about 15 amino acids per second. Elongation in prokaryotes is a little faster (about 20 amino acids per second), but the essential processes are very similar. In *E. coli*, the sizes of the ribosomal subunits are 30S (small) and 50S (large), and the complete ribosome is 70S. **Figure 11.20** shows a large ribosomal subunit from *E. coli*, reconstructed from the x-ray diffraction structure, depicting the locations of the tRNA molecules in their binding pockets: E in gold, P in blue, and A in green. Above the tRNAs is a channel through which the mRNA is moved along as translation progresses. The small subunit fits on top of the large subunit, leaving enough space for the tRNA molecules to bind. In prokaryotes the source of energy for elongation is also GTP hydrolysis. The *E. coli* analogs of EF-1α and EF-2 are EF-Tu and EF-G, respectively.

Release

Compared to initiation and elongation, the termination of polypeptide synthesis—the **release** phase—is simple (**Figure 11.21**). When a stop codon is encountered, the tRNA holding the polypeptide remains in the P site, and a *release factor (RF)* binds with the ribosome. GTP hydrolysis provides the energy to cleave the polypeptide from the tRNA to which it is attached, as well as to eject the release factor and dissociate the 80S ribosome from the mRNA. At this point the 40S and 60S subunits are recycled to initiate translation of another mRNA. Eukaryotes have only one release factor that recognizes all three stop codons: UAA, UAG, and UGA. The situation differs in prokaryotes. In *E. coli*, the release factor RF-1 recognizes the stop codons UAA and UAG, whereas release factor RF-2 recognizes UAA and UGA. A third release factor, RF-3, is also required, but its function is uncertain.

11.7 Complex Translation Units

The unit of translation is rarely one ribosome traversing one mRNA molecule. A single mRNA can be translated simultaneously by several ribosomes moving along it in tandem, and in prokaryotes a single mRNA can code for multiple polypeptide chains. These situations are examined in the following sections.

Polysomes

After about 25 amino acids have been joined together in a polypeptide chain, an AUG initiation codon is completely free of the ribosome, and a second initiation complex can form. The overall configuration is that of two ribosomes moving along the mRNA at

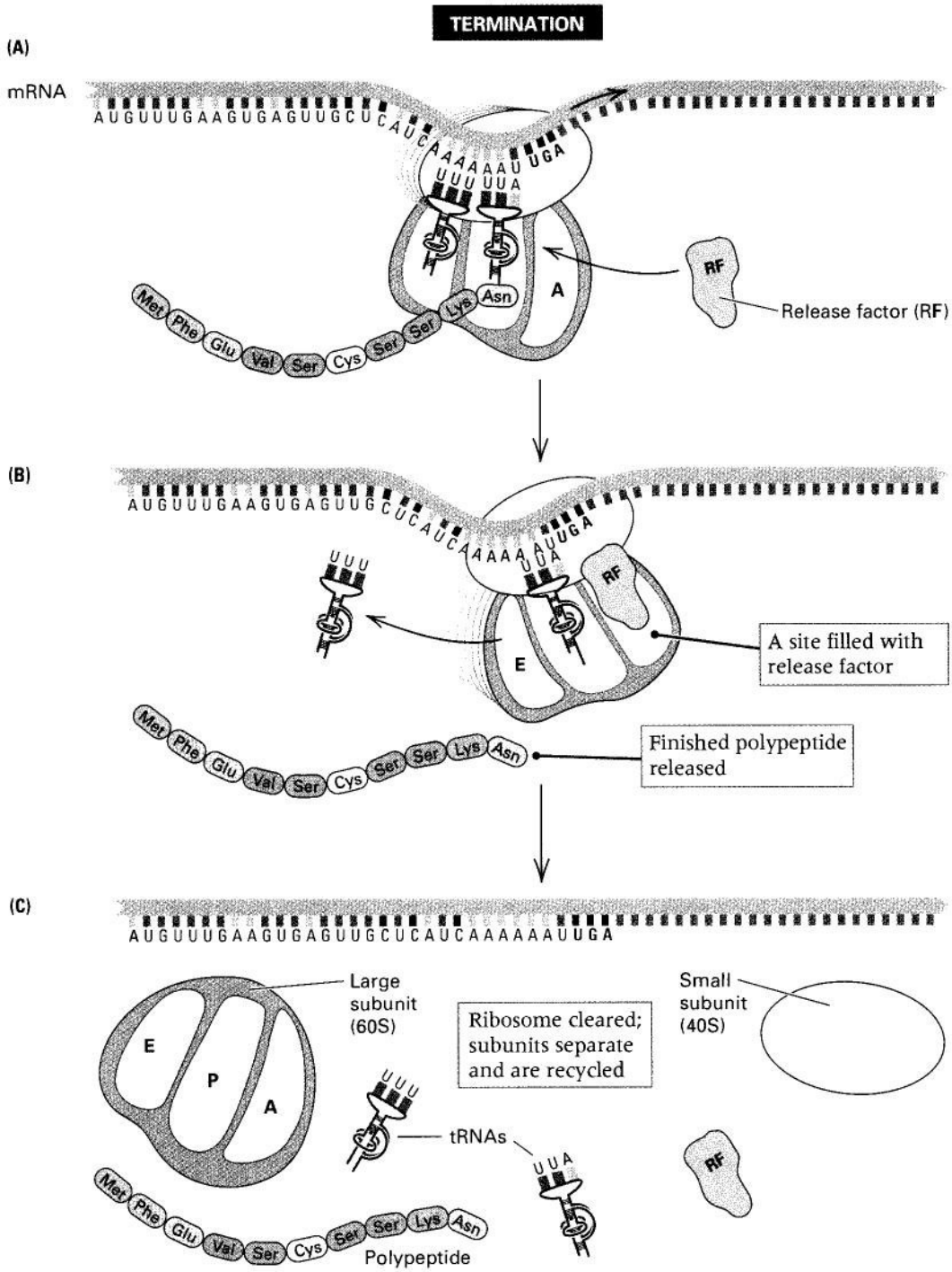


Figure 11.21 Termination of protein synthesis. When a stop codon is reached (A), no tRNA can bind to that site (B), which causes the release of the newly formed polypeptide and the remaining bound tRNA (C).

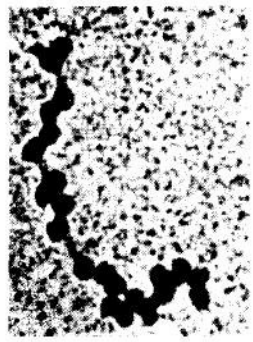


Figure 11.22 Electron micrograph of *E. coli* polysomes. [Courtesy of Barbara Hamkalo.]

the same speed. When the second ribosome has moved along a distance similar to that traversed by the first, a third ribosome can attach to the initiation site. The process of movement and reinitiation continues until the mRNA is covered with ribosomes at a density of about one ribosome per 80 nucleotides (**Figure 11.22**) This large translation unit is called a **polysome,** which is the usual form of the translation unit.

In eukaryotes, transcription and RNA processing take place in the nucleus, whereas translation takes place in the cytoplasm. The processes are spatially and temporally uncoupled. In prokaryotes, because no nuclear envelope separates the DNA from the ribosome, the translational initiation complex can form even before the mRNA is released from the DNA. This allows the simultaneous occurrence, or *coupling,* of transcription and translation. **Figure 11.23** shows an electron micrograph of a DNA molecule with a number of attached mRNA molecules, each associated with ribosomes (Figure 11.23A), and an interpretation (Figure 11.23B). Transcription of DNA is beginning at the upper left of the micrograph. The lengths of the polysomes increase with distance from the transcription initiation site; the mRNA is farther from that site and hence of greater length because the process of transcription has been going on longer.

Polycistronic mRNA

As we have noted, the eukaryotic scanning process for selecting the initiation codon implies that one mRNA can code for one and only one polypeptide chain. In prokaryotes, as shown in **Figure 11.24**, mRNA molecules commonly contain coding sequences for several different polypeptide

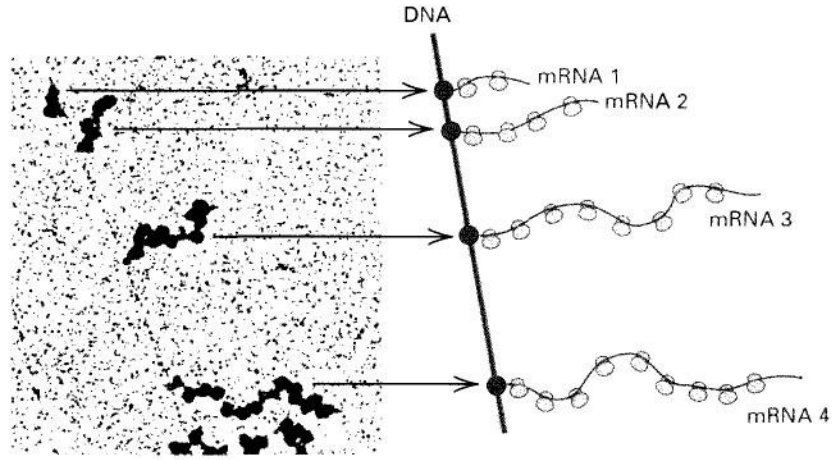


Figure 11.23 Visualization of transcription and translation. The electron micrograph shows transcription of a section of the DNA of *E. coli* and translation of the nascent mRNA. The dark spots are ribosomes, which coat the mRNA. An interpretation of the electron micrograph is at the right. Each mRNA has ribosomes attached along its length. The large red dots are the RNA polymerase molecules; they are too small to be seen in the electron micrograph. The length of each mRNA is equal to the distance that each RNA polymerase has progressed from the transcription-initiation site. [Reprinted with permission from O. L. Miller, B. A. Hamkalo, and C. A. Thomas. 1977. *Science* 169: 392. Copyright 1977 American Association for the Advancement of Science.]

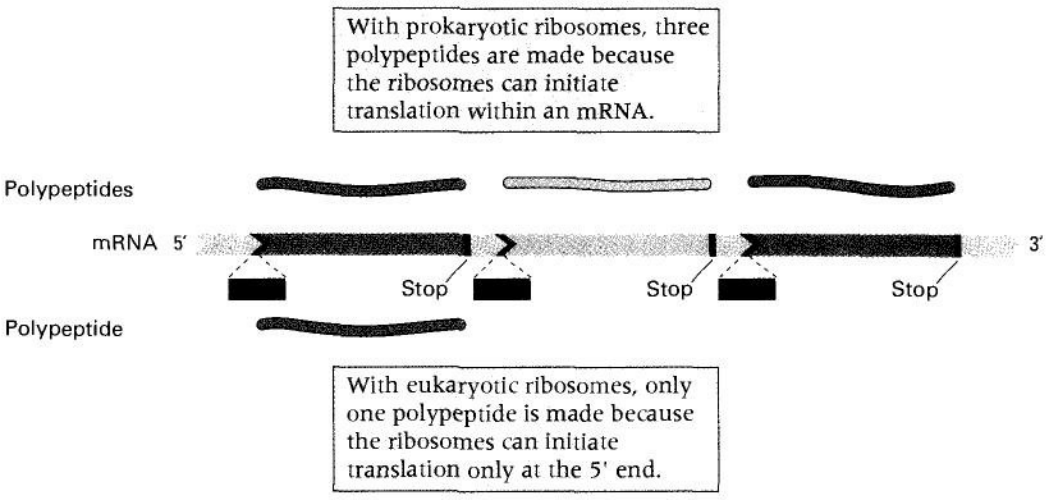


Figure 11.24 Different products are translated from a three-cistron mRNA molecule by the ribosomes of prokaryotes and eukaryotes. The prokaryotic ribosome translates all of the genes, but the eukaryotic ribosome translates only the gene nearest the 5' terminus of the mRNA. Translated sequences are shown in purple, yellow, and orange; stop codons in red; the ribosome binding sites in green; and the spacer sequences in light green.

chains; such a molecule is called a **polycistronic mRNA.** (The word *cistron* means a nucleotide sequence that encodes a single polypeptide chain). Eukaryotic mRNAs are always monocistronic. In a polycistronic mRNA, each polypeptide coding region is preceded by its own ribosome-binding site and AUG initiation codon. After the synthesis of one polypeptide is finished, the next along the way is translated. The genes contained in a polycistronic mRNA molecule often encode the different proteins of a metabolic pathway. For example, in *E. coli*, the ten enzymes needed to synthesize histidine are encoded by one polycistronic mRNA molecule. The use of polycistronic mRNA is an economical way for a cell to regulate the synthesis of related proteins in a coordinated manner. For example, in prokaryotes, the usual way to regulate the synthesis of a particular protein is to control the synthesis of the mRNA molecule that codes for it (Chapter 12). With a polycistronic mRNA molecule, the synthesis of several related proteins can be regulated by a single signal, so that appropriate quantities of each protein are made at the same time; this is termed *coordinate regulation.*

The definitive feature of translation is that it proceeds in a particular direction along the mRNA and the polypeptide:

> The mRNA is translated from an initiation codon to a stop codon in the 5'-to-3' direction. The polypeptide is synthesized from the amino end toward the carboxyl end by the successive addition of amino acids to the carboxyl end.

These directions of synthesis are illustrated schematically in **Figure 11.25**. The convention in writing nucleotide sequences is to place the 5' end at the left, and, in writing amino acid sequences, to place the amino end at the left. This means that an RNA sequence is written from left to right in the same order in which it is transcribed, and an amino acid sequence is written in the same order in which it is translated.

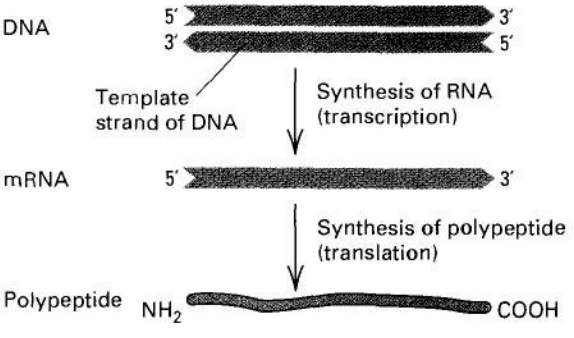


Figure 11.25 Direction of synthesis of RNA with respect to the coding strand of DNA, and of synthesis of protein with respect to mRNA.

CoNnection

Uncles and Aunts

Francis H. C. Crick, Leslie Barnett, Sydney Brenner, and R. J. Watts-Tobin
1961
Cavendish Laboratory, Cambridge, England

General Nature of the Genetic Code for Proteins

No other paper affords a better demonstration of the power of mutational analysis in the hands of clever researchers. The issue was this: How many bases are needed to code for one amino acid in a protein? Crick et al. answered the question by using single-base insertions and deletions in the rIIB *cistron (Seymour Benzer's name for a region of DNA that codes for a single polypeptide chain). In the laboratory they referred to the (+) and (−) mutations as "uncles" and "aunts" because it was not known, in any particular case, whether a mutation was truly an insertion or truly a deletion. The paper is written as if* FC0 *were an insertion mutation, solely for the sake of simplicity. (This guess later proved to be right.) The presentation is unusual in another respect also. The paper starts by assuming its conclusion in order to describe the results and then demonstrates how the conclusion was arrived at. Try to imagine writing it in the usual way. Any such effort would be bound to be less clear. (Incidentally, the suggestion that the use of synthetic polynucleotides would solve the coding problem within a year was not far wrong. By 1968, Marshall W. Nirenberg, Robert W. Holley, and Har Gobind Khorana would be awarded a Nobel Prize for their deciphering of the genetic code table.)*

Fortunately, we have convincing evidence that the coding ratio is in fact 3 or a multiple of 3.

In this article we report genetic experiments which suggest that the genetic code [is one in which] a group of three bases (or, less likely, a multiple of three bases) codes for one amino acid. . . . Our genetic experiments have been carried out on the *B* cistron of the *rII* region of the bacteriophage T4. . . . We report here our work on the mutant *FC0*. This mutant was originally produced by the action of proflavin, . . . which we have previously argued acts as a mutagen because it adds or deletes a base or bases. . . . If an acridine mutant is produced by, say, adding a base, it should revert to "wildtype" by deleting a base. Our work on *FC0* shows that it usually reverts not by reversing the original mutation but by producing a second mutation at a nearby point on the genetic map. . . . A genetic map of 18 suppressors of *FC0* shows that they scatter over a region about, say, one-tenth of the size of the *B* cistron. . . . In

11.8 The Standard Genetic Code

The **genetic code** is the list of all codons and the amino acid that each one specifies. Before the nature of the genetic code was known, scientists reasoned that if all codons were assumed to have the same number of nucleotides, then each codon would have to contain at least three nucleotides. Codons consisting of pairs of nucleotides would be insufficient, because four nucleotides can form only $4^2 = 16$ pairs; triplets of nucleotides would suffice, because four nucleotides can form $4^3 = 64$ triplets. In fact, the genetic code is a triplet code, and all 64 possible codons carry information of some sort. Because only 20 amino acids need be specified, most amino acids are encoded by more than one codon. Furthermore, in the translation of mRNA molecules, the codons do not overlap but are used sequentially.

Genetic Evidence for a Triplet Code

Although theoretical considerations suggested that each codon must contain at least three nucleotides, codons of greater length could not be ruled out. The first widely accepted proof for a triplet code came from genetic experiments using *rII* mutants of bacteriophage T4 that had been induced by replication in the presence of the chemical *proflavin*. These experiments were carried out in 1961 by Francis Crick and collaborators. Proflavin-induced mutations typically resulted in total loss of function. Because proflavin is a large planar molecule, the investigators suspected that it caused mutations that inserted or deleted a base pair by interleaving between base pairs in the double helix. Analysis of the properties of these mutations led directly to the deduction that the code is read in nonoverlapping groups of three nucleotides

all we have isolated about eighty independent *rIIB* mutants, all suppressors of *FC0*, or suppressors of suppressors, or suppressors of suppressors of suppressors. . . . Although we have no direct evidence that the *B* cistron produces a polypeptide chain (probably through an RNA intermediate), in what follows we shall assume this to be so. To fix ideas, we imagine that the string of nucleotides is read, triplet by triplet, from a starting point on the left of the *B* cistron. We now suppose that, for example, the mutant *FC0* was produced by the insertion of an additional base in the wildtype sequence. This addition of a base at the *FC0* site will mean that the reading of all the triplets to the right of *FC0* will be shifted along one base, and will therefore be incorrect. Thus the amino-acid sequence of the protein which the *B* cistron is presumed to produce will be completely altered from that point onwards. This explains why the function of the gene is lacking. . . . We now postulate that a suppressor of *FC0* (for example, *FC1*) is formed by deleting a base. Thus when the *FC1* mutation is present by itself, all triplets to the right of *FC1* will be read incorrectly and thus the function will be absent. However, when both mutations are present in the same piece of DNA, then although the reading of triplets between *FC0* and *FC1* will be altered, the original reading will be restored to the rest of the gene. . . . So far we have spoken as if the evidence supported a triplet code, but this was simply for illustration. . . . Fortunately, we have convincing evidence that the coding ratio is in fact 3 or a multiple of 3. This we have obtained by constructing triple mutants of the form (+ with + with +) or (− with − with −). One must be careful not to make shifts across the "unacceptable" region of *rIIB*, but this we can avoid by a proper choice of mutants. . . . It is possible by various devices, either chemical or enzymatic, to synthesize polyribonucleotides with defined or partially defined sequences. If these will produce specific polypeptides, the coding problem is wide open for experimental attack . . . and the genetic code may well be solved within a year.

Source: Nature 192: 1227–1232

from a fixed point that establishes the **reading frame** of the mRNA. Mutations that delete or add a nucleotide pair shift the reading frame and are called **frameshift mutations.** Figure 11.26 illustrates the profound effect of a frameshift mutation on the amino acid sequence of the polypeptide produced from a mutant mRNA.

The genetic analysis of the structure of the code began with an *rII* (rapid lysis) mutation called FC0, which was arbitrarily designated (+), as though it had an inserted nucleotide pair. [It could just as easily have been designated (−), as though it had a deleted nucleotide pair. Calling it (+) was a lucky guess, however, because when FC0 was sequenced, it did turn out to have a single-base insertion.] If FC0 has a (+) insertion, then it should be possible to revert the FC0 allele to "wildtype" by deletion of a nearby nucleotide. Selection for r^+ revertants was carried out by isolating plaques formed on a lawn of an *E. coli* strain K12 that was lysogenic for phage λ. The basis of the selection is that *rII* mutants are unable to propagate in K12(λ). Analysis of the revertants revealed that each still

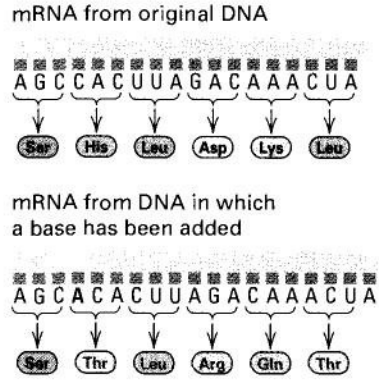


Figure 11.26 The change in the amino acid sequence of a protein caused by the addition of an extra base, which shifts the reading frame. A deleted base also shifts the reading frame.

carried the original FC0 mutation and, in addition, another (intragenic suppressor) mutation that reversed the effects of the FC0 mutation. The suppressor mutations could be separated by recombination from the original mutation by crossing each revertant to wildtype; each suppressor mutation proved to be an *rII* mutation that, by itself, would cause the rapid lysis phenotype. If FC0 had an inserted nucleotide, then the suppressors should all result in deletion of a nucleotide pair; therefore, each suppressor of FC0 was inferred to have a single-nucleotide deletion (−). Three such revertants and their consequences for the translational reading frame, depicted using ordinary three-letter words for simplicity, are illustrated in **Figure 11.27**. The (−) mutations are designated $(-)_1$, $(-)_2$, and $(-)_3$, and those parts of the mRNA translated in the correct reading frame are indicated in green.

Each of the individual (−) suppressor mutations could, in turn, be used to select other "wildtype" revertants, with the expectation that these revertants would carry new suppressor mutations of the (+) variety, because the (−)(+) combination should yield a phage able to form plaques on K12(λ).

Once a large set of (+) and (−) mutations had been isolated, various double-mutant combinations could be brought together by phage recombination. Usually any (+) (−) combination and any (−)(+) combination resulted in a wildtype phenotype, whereas (+)(+) and (−)(−) combinations always resulted in the mutant phenotype. The key result came from the triple mutants (+)(+)(+) and (−)(−)(−), which often had a wildtype phenotype!

The phenotypes of the various (+) and (−) combinations were interpreted in terms of offsetting shifts in the reading frame, as

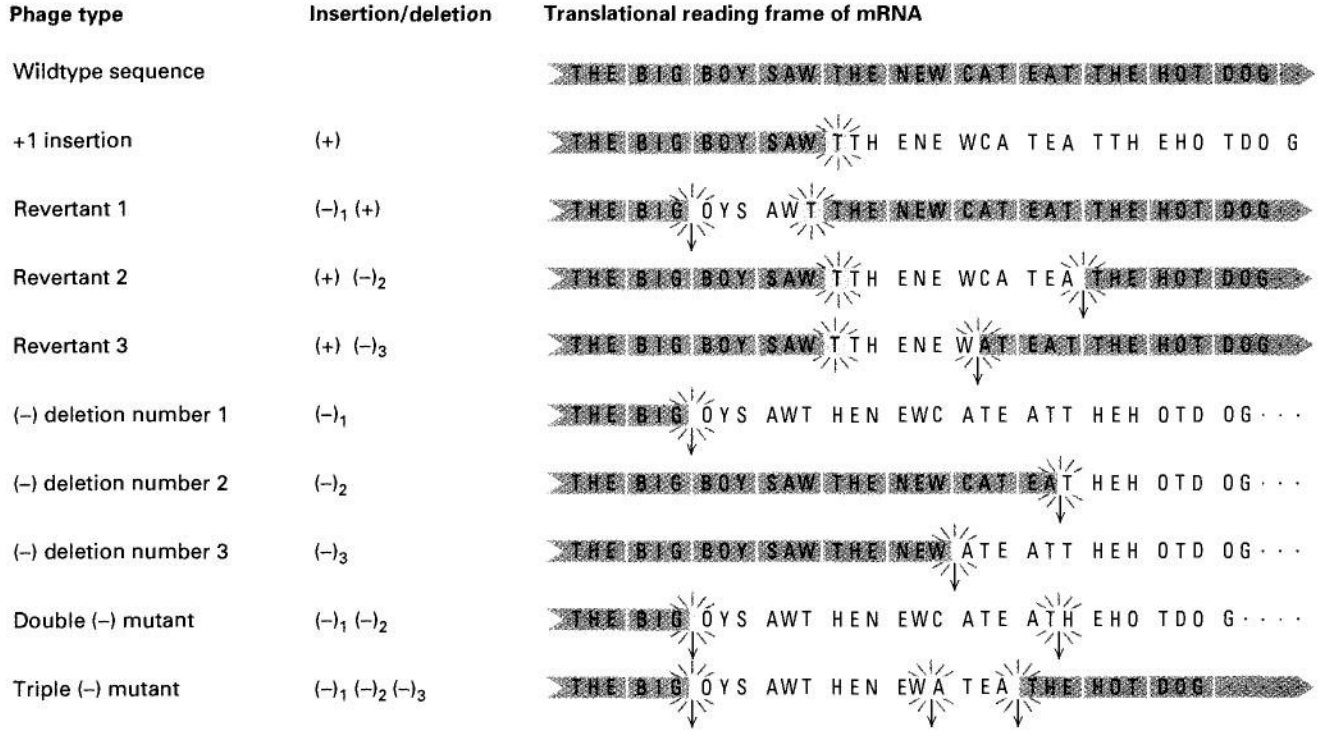


Figure 11.27 Interpretation of the *rII* frameshift mutations, showing that combinations of appropriately positioned single-base insertions (+) and single-base deletions (−) can restore the correct reading frame (green). The key finding was that a combination of three single-base deletions, as shown in the bottom line, also restores the correct reading frame (green). Two single-base deletions do not restore the reading frame. These classic experiments gave strong genetic evidence that the genetic code is a triplet code.

shown in Figure 11.27. The initial FC0 mutation shifts the reading frame from the point of the +1 insertion onward, resulting in incorrect amino acid sequence and a nonfunctional protein. Deletion of a nearby nucleotide pair will restore the reading frame, although the amino acid sequence encoded between the two mutations will be different. In (+)(+) or (−)(−) double mutants, the reading frame is shifted by two nucleotides, and the resulting protein is still nonfunctional. However, in the (+)(+)(+) and (−)(−)(−) triple mutants, the reading frame is restored, though all amino acids encoded between the outside mutations are incorrect. The resulting protein is one amino acid longer for (+)(+)(+) and one amino acid shorter for (−)(−)(−) (Figure 11.27).

The genetic analysis of the insertion and deletion mutations strongly supported the following conclusions:

- Translation of an mRNA starts from a fixed point.
- There is a single reading frame maintained throughout the process of translation.
- Each codon consists of three nucleotides.

Crick and his colleagues also drew other inferences from these experiments. First, in the genetic code, most codons must function in the specification of an amino acid. Second, each amino acid must be specified by more than one codon. They reasoned that if each amino acid had only one codon, then only 20 of the 64 possible codons could be used for coding amino acids. In this case, most frameshift mutations should have affected one of the remaining 44 "noncoding" codons in the reading frame, and hence a nearby frameshift of the opposite-polarity mutation should not have suppressed the original mutation. Consequently, the code was inferred to contain synonymous codons (a property originally called *degeneracy*).

How the Code Was Cracked

Polypeptide synthesis can be carried out in *E. coli* cell extracts obtained by breaking cells open. Various components can be isolated, and a functioning protein-synthesizing system can be reconstituted by mixing ribosomes, tRNA molecules, mRNA molecules, and the protein factors needed for translation. If radioactive amino acids are added to the extract, then radioactive polypeptides are made. Synthesis continues for only a few minutes because the mRNA is rapidly degraded by nucleases in the mixture. The experimental elucidation of the genetic code began with the observation that when the degradation of native mRNA was allowed to go to completion, and the synthetic polynucleotide polyuridylic acid (poly-U) was added to the mixture to serve as an mRNA molecule, a polypeptide consisting only of phenylalanine (Phe−Phe−° Phe. . .) was synthesized. From this simple result, and knowledge that the code is a triplet code, it follows that UUU must be a codon for the amino acid phenylalanine. Variations on this basic experiment identified other codons. For example, when a long sequence of guanines was added at the terminus of the poly-U, the polyphenylalanine was terminated by a sequence of glycines, indicating that GGG is a glycine codon (**Figure 11.28**). A trace of leucine or tryptophan was also present in the glycine-terminated polyphenylalanine. Incorporation of these amino acids was directed by the codons UUG and UGG at the transition point between U and G. When a single

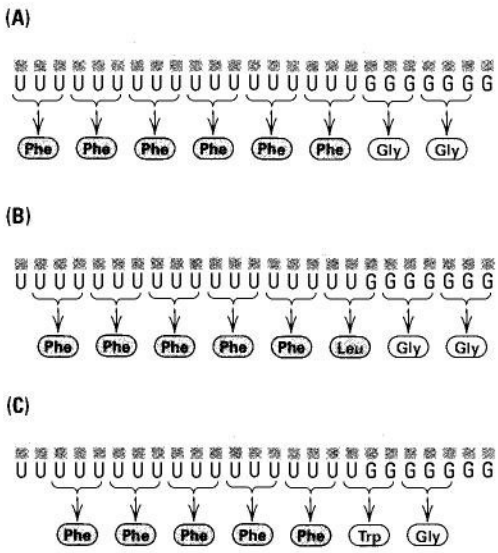


Figure 11.28 Polypeptide synthesis using

5'-UUUU · · · UUGGGGGGG-3'

as an mRNA in three different reading frames, showing the reasons for the incorporation of glycine, leucine, and tryptophan.

guanine was added to the terminus of a poly-U chain, the polyphenylalanine was terminated by leucine. Thus UUG is a leucine codon, and UGG must be a codon for tryptophan. Similar experiments were carried out with poly-A, which yielded polylysine, and with poly-C, which produced polyproline.

Other experiments led to a complete elucidation of the code. The three codons

UAA UAG UGA

were found to be stop signals for translation, and the codon AUG, which encodes methionine, was shown to be the initiation codon. AUG also codes for internal methionines.

Features of the Standard Code

The *in vitro* translation experiments, which originally used components isolated from *E. coli*, have been repeated with components from many species of bacteria, yeast, plants, and animals. The standard genetic code deduced from these experiments is considered to be nearly universal because the same codon assignments can be made for nuclear genes in nearly all organisms that have been examined. However, some minor differences in codon assignments are found in certain protozoa and in the genetic codes of organelles.

The standard code is shown in **Table 11.3**. Altogether, 61 codons specify amino acids. In many cases, several codons specify the same amino acid. This feature confirms the inference from the *rII* frameshift mutations that the genetic code contains synonymous codons. All amino acids except tryptophan and methionine are specified by more than one codon. The pattern of synonymy is not random. For example, with the exception of serine, leucine, and arginine, all codons that correspond to the same amino acid are in the same box in Table 11.3. This means that *synonymous codons usually differ only in the third nucleotide position;* for example, GGU, GGC, GGA, and GGG all code for glycine. Furthermore, in all cases in which two codons code for the same amino acid, the third nucleotide is either A or G (both purines) or T or C (both pyrimidines).

The codon assignments shown in Table 11.3 are completely consistent with all chemical observations and with the amino acid sequences of wildtype and mutant proteins. In virtually every case in which a mutant protein differs by a single amino acid from the wildtype form, the amino acid substitution can be accounted for by a single nucleotide change between the codons corresponding to the two different amino acids. For example, substitution of glutamic acid by valine, which occurs in sickle-cell hemoglobin, results from a change from GAG to GUG in codon 6 of the β-globin mRNA. The presence of synonymous codons implies that some nucleotide substitutions can be "silent" in that they do not alter the amino acid; taking codon 6 in β-globin as an example, GAG → GAA results in a mutant codon that still encodes glutamic acid. Some nucleotide substitutions cause chain termination; in the β-globin example, GAG → UAG would result in a truncated polypeptide only five amino acids in length.

Table 11.3 The standard genetic code

First position (5' end)	Second position U	Second position C	Second position A	Second position G	Third position (3' end)
U	UUU Phe } F UUC Phe } UUA Leu } L UUG Leu }	UCU Ser UCC Ser UCA Ser UCG Ser } S	UAU Tyr } Y UAC Tyr } **UAA Stop** **UAG Stop**	UGU Cys } C UGC Cys } **UGA Stop** UGG Trp W	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu } L	CCU Pro CCC Pro CCA Pro CCG Pro } P	CAU His } H CAC His } CAA Gln } Q CAG Gln }	CGU Arg CGC Arg CGA Arg CGG Arg } R	U C A G
A	AUU Ile AUC Ile AUA Ile } I **AUG Met** M	ACU Thr ACC Thr ACA Thr ACG Thr } T	AAU Asn } N AAC Asn } AAA Lys } K AAG Lys }	AGU Ser } S AGC Ser } AGA Arg } R AGG Arg }	U C A G
G	GUU Val GUC Val GUA Val GUG Val } V	GCU Ala GCC Ala GCA Ala GCG Ala } A	GAU Asp } D GAC Asp } GAA Glu } E GAG Glu }	GGU Gly GGC Gly GGA Gly GGG Gly } G	U C A G

Note: Each amino acid is given its conventional abbreviation in both the single-letter and the three-letter format. The codon AUG, which codes for methionine (boxed) is generally used for initiation. The codons are conventionally written with the 5' base on the left and the 3' base on the right.

Transfer RNA and Aminoacyl-tRNA Synthetase Enzymes

The decoding operation by which the nucleotide sequence within an mRNA molecule becomes translated into the amino acid sequence of a protein is accomplished by aminoacylated, or charged, tRNA molecules, each of which is linked to the correct amino acid by an **aminoacyl-tRNA synthetase.**

The tRNA molecules are small, single-stranded nucleic acids ranging in size from about 70 to 90 nucleotides. Like all RNA molecules, they have a 3'-OH terminus, but the opposite end terminates with a 5'-monophosphate rather than a 5'-triphosphate, because tRNA molecules are cut from a larger primary transcript. Internal complementary base sequences form short double-stranded regions, causing the molecule to fold into a structure in which open loops are connected to one another by double-stranded stems (**Figure 11.29**, part A). In two dimensions, a tRNA molecule is often drawn as a planar cloverleaf. Its three-dimensional structure is more complex, as is shown in Figure 11.29B. Note that the TψC loop and the DHU loop are in close proximity. When viewed from either side, the folded structure roughly resembles the diagrammatic representations of the tRNAs that appear in Figures 11.18, 11.19, and 11.21.

Particular regions of each tRNA molecule are used in the decoding operation. One region is the anticodon sequence, which consists of three nucleotides that can form base pairs with a codon sequence in the mRNA. No normal tRNA molecule has an anticodon complementary to any of the stop codons (UAG, UAA, and UGA), which is why these codons are stop signals. A second critical site is at the 3' terminus of the tRNA molecule, where the amino acid attaches. A specific aminoacyl-tRNA synthetase matches the amino acid with the anticodon. At least one, and usually only one, aminoacyl-synthetase exists for each amino acid. To make the correct attachment, the synthetase must be able to distinguish one tRNA molecule from another. The necessary distinction is provided by recognition regions that encompass many parts of the tRNA molecule.

Figure 11.30 shows the three-dimensional structure of the seryl-tRNA synthetase

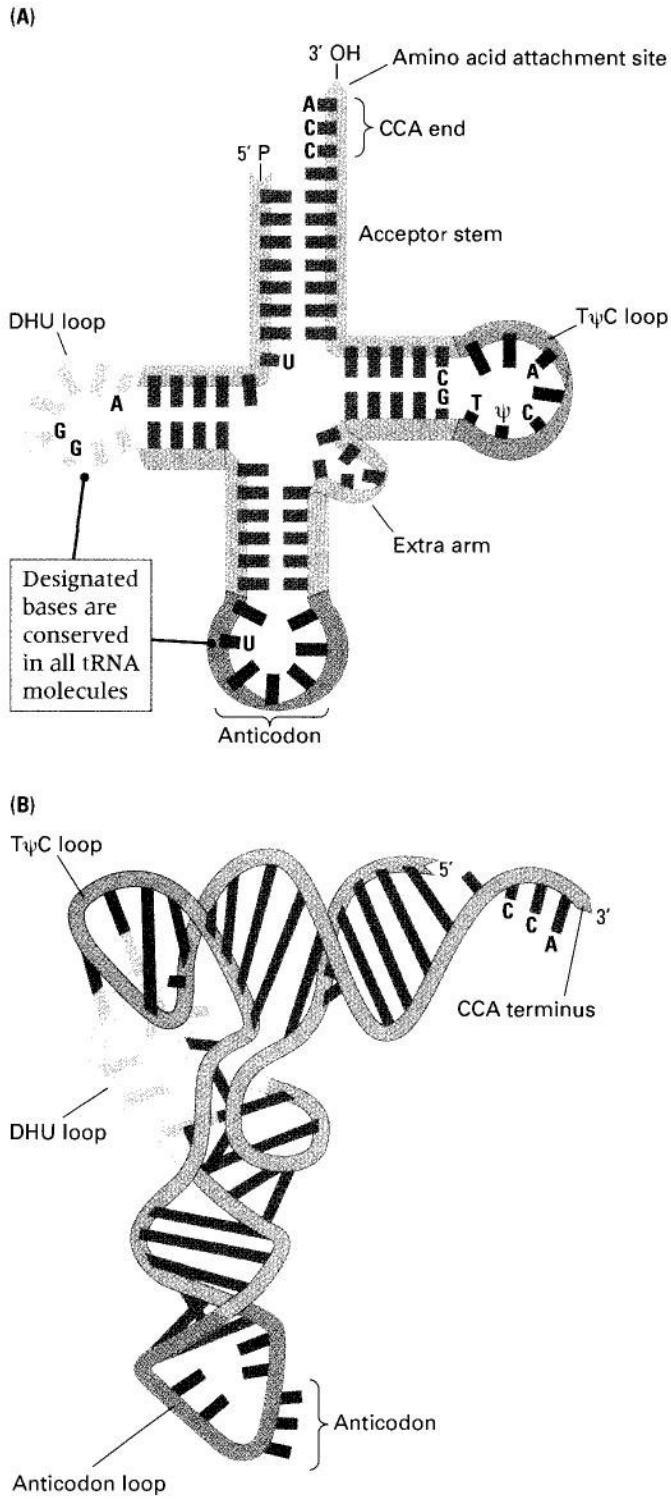


Figure 11.29 (A) A tRNA "cloverleaf" configuration. The heavy black letters indicate a few bases that are conserved in the sequence of all tRNA molecules. The labeled loop regions are those found in all tRNA molecules. DHU refers to a base, dihydrouracil, found in one loop; the Greek letter Ψ is a symbol for the unusual base pseudouridine. (B) A schematic diagram of the three-dimensional structure of yeast $tRNA^{Phe}$. [(B) courtesy of Sung-Hou Kim.]

complexed with its tRNA. On binding with the tRNA, a part of the protein makes contact with the variable part of the TψC loop of the tRNA and guides the acceptor stem into the active pocket of the enzyme. These interactions depend primarily on recognition of the shape of the $tRNA^{Ser}$ through contacts with the backbone; they depend only secondarily on interactions that are specific to the anticodon.

Redundancy and Wobble

Several features of the genetic code suggest that normal Watson–Crick base pairing is not the only feature of codon–anticodon binding. First, the code is highly redundant. Second, the identity of the third nucleotide of a codon is often unimportant. For some amino acids—for example, proline (Pro), threonine (Thr), and glycine (Gly)—any nucleotide in the third position will do; and for other amino acids—for example, histidine (His), glutamine (Gln), and tyrosine (Tyr)—either purine (A or G) or either pyrimidine (U or C) in the third position will do. Third, the number of distinct tRNA molecules that have been isolated from a single organism is less than the number of codons; because all codons are used, the anticodons of some tRNA molecules must be able to pair with more than one codon. Experiments with several purified tRNA molecules showed this to be the case.

To account for these observations, the **wobble** concept was advanced in 1966 by Francis Crick. He proposed that the first two nucleotides in a codon form base pairs with the tRNA anticodon according to the usual rules (A–U and G–C) but that the base at the 5' end of the anticodon is less spatially constrained than the first two and can form hydrogen bonds with more than one base at the 3' end of the codon. He suggested the pairing rules given in **Table 11.4**. Evidence has confirmed the wobble concept and indicates that the pairings given in Table 11.4 are largely true for *E. coli*.

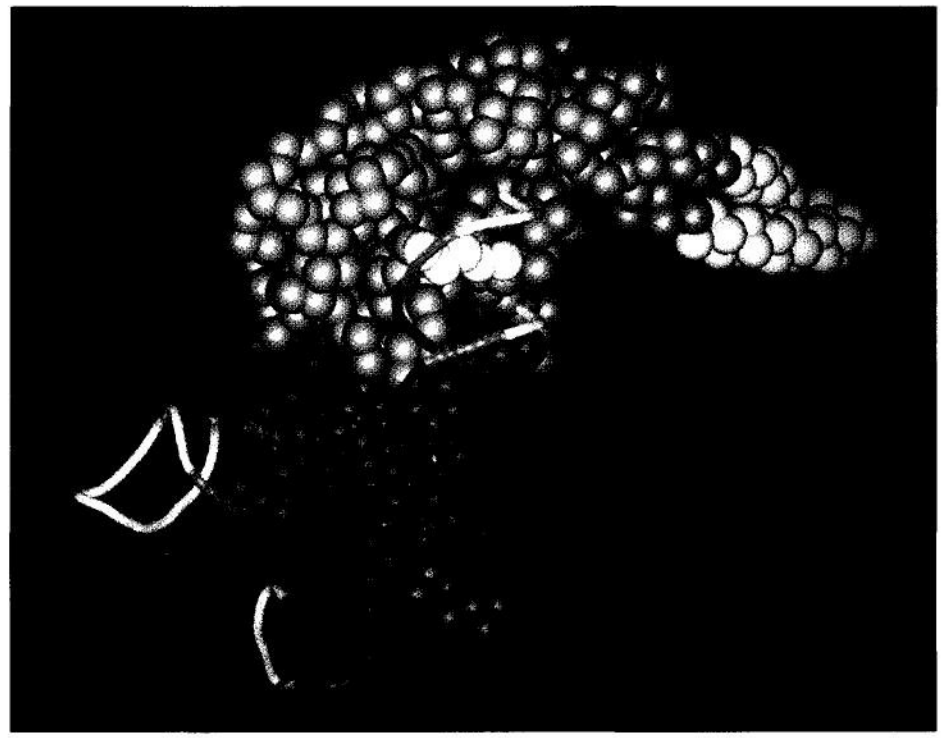


Figure 11.30 Three-dimensional structure of seryl-tRNA synthetase (solid spheres) complexed with its tRNA. Note that there are many points of contact between the enzyme and the tRNA. The molecules are from *Thermus thermophilus*. [Courtesy of Stephen Cusack. Reprinted with permission from S. Cusack, V. Biou, A. Yaremchuk, and M. Tukalo, *Science* 263: 1404. Copyright 1996 American Association for the Advancement of Science.]

Table 11.4 Allowed pairings due to wobble

First base in anticodon (5' position)	Allowed bases in third codon position (3' position)
A	U
C	G
U	A or G
G	C or U
I	A or C or U

On the other hand, analysis of tRNAs in the yeast *Saccharomyces cerevisiae* has indicated that wobble is more restricted in yeast than in *E. coli*. **Table 11.5** summarizes the wobble rules for *E. coli* and yeast. The yeast rules may hold for other eukaryotes as well. In yeast, single tRNAs can recognize the pairs of related codons ending in U or C. However, separate tRNAs are needed for codons that end in A or G. Thus at least three tRNAs are required for amino acids such as proline and glycine, which are specified by a set of four codons. A total of 46 tRNAs are needed to decode mRNA molecules in yeast. As noted, there are no tRNAs corresponding to the stop codons.

Nonsense Suppression

A nucleotide substitution that creates a stop codon results in premature chain termination during translation. Such a mutation is called a **nonsense mutation.** For whimsical historical reasons, the stop codons are referred to as the *amber* (UAG), *ochre* (UAA), and *umber* (UGA). A remarkable observation was that some T4 phage mutants bearing nonsense mutations in the gene encoding the major head protein were able to propagate in some bacterial strains but not in others. How could this happen? On further analysis, it turned out that the strains able to support growth of the phage nonsense mutants carry suppressor mutations that act by changing the way the mRNA is read, not by changing the nucleotide sequence of the phage gene. The suppressor mutations proved to be mutant tRNA genes. These suppressor tRNA genes act by reading a stop codon as though it were a signal for a specific amino acid. The amino acid is inserted at that position, and translation continues. So long as the inserted amino acid is compatible with the function of the protein, the effects of the original mutations are suppressed, and plaques can be produced. In *E. coli*, "amber suppressors" recognize and suppress only amber mutations because the anticodon of the altered tRNA pairs only with the UAG codon; on the other hand, "ochre suppressors" suppress both ochre and amber mutations because the mutant tRNA can recognize and suppress both UAG and UAA. In yeast there is more specificity because of the more stringent wobble rules: Ochre suppressors suppress only ochre mutations, and amber suppressors suppress only amber mutations.

We will illustrate nonsense suppression by examining a chain-termination codon formed by mutation of the tyrosine codon UAC to the stop codon UAG (**Figure 11.31**, part A versus B). Part C shows how this mutation can be suppressed by a mutant leucine tRNA molecule. In *E. coli*, $tRNA^{Leu}$ has the anticodon 3'-AAC-5', which pairs with the codon 5'-UUG-3'. A suppressor mutation in the $tRNA^{Leu}$ gene produces an altered tRNA with the anticodon AUC; this tRNA molecule is still charged with leucine but responds to the stop codon UAG rather than to the normal leucine codon UUG. Thus, in a cell that contains this suppressor

Table 11.5 Wobble rules for tRNAs of *E. coli* and *Saccharomyces cerevisiae*

Third position of codon	First position of anticodon	
	E. coli	**Yeast (*S. cerevisiae*)**
U	A, G, or I	G or I
C	G or I	G or I
A	U or I	U*
G	C or U	C

Note: I indicates inosine, which is structurally similar to adenosine except that the $-NH_2$ is replaced with $-OH$. U* indicates a modified uridine.

Source: Data from C. Guthrie and J. Abelson, 1982. *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, edited by J. N. Strathern, E. W. Jones, and J. R. Broach, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, p. 487.

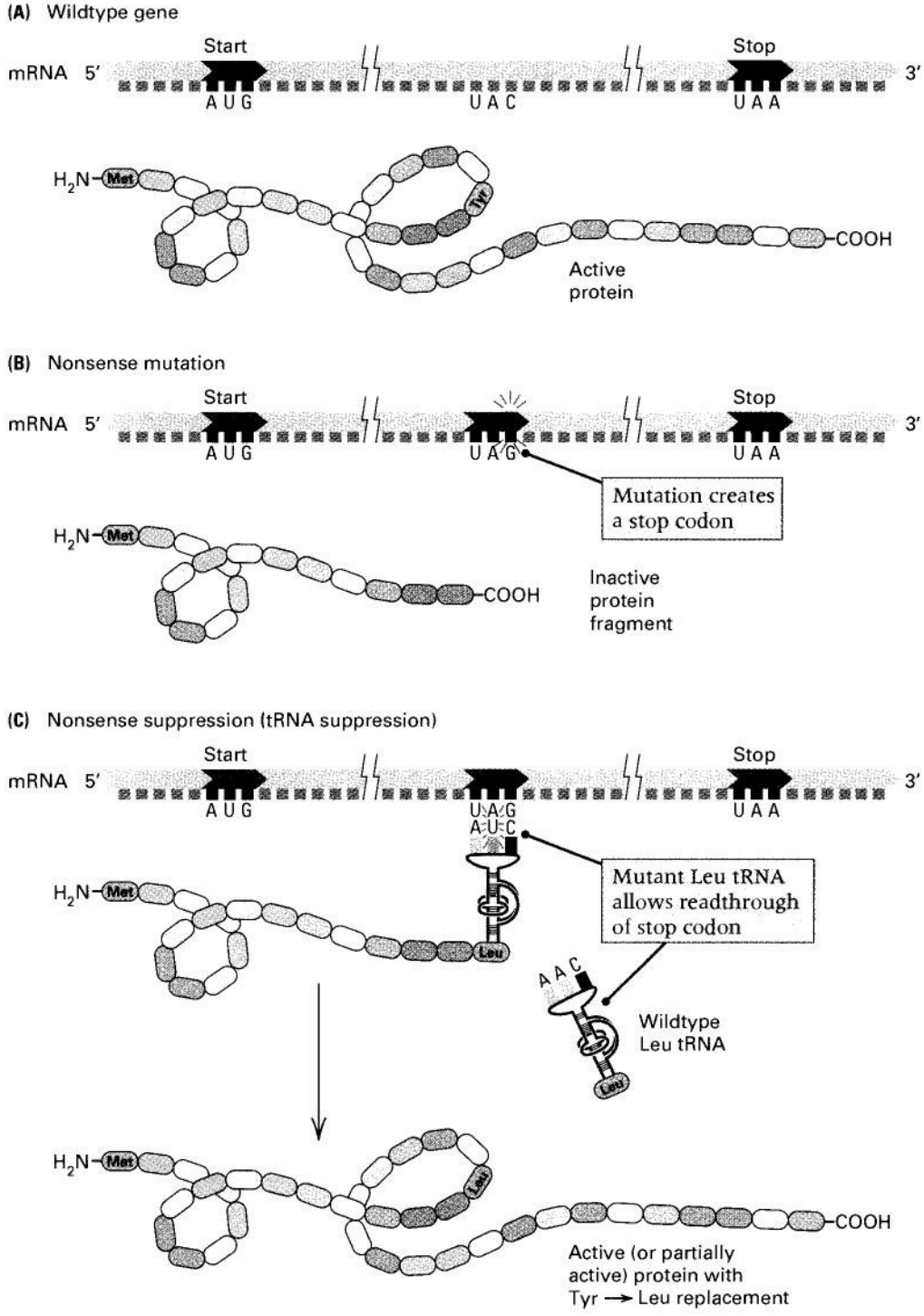


Figure 11.31 The mechanism of suppression by a nonsense suppressor tRNA molecule. (A) The wildtype gene. (B) A UAC → UAG chain-termination ("nonsense") mutation leads to an inactive, prematurely terminated protein. (C) A mutation in the $tRNA^{Leu}$ gene produces an altered tRNA molecule, which has an anticodon complementary to a UAG stop codon but can still be charged with leucine. This tRNA molecule allows the protein to be completed, but with a leucine at the site of the original tyrosine. Suppression will be achieved if the substitution restores activity to the protein.

tRNA, the mutant protein is completed, and the resulting protein will be functional provided that it can tolerate leucine instead of tyrosine. Many suppressor tRNA molecules of this type have been observed. Each suppressor is effective against only some nonsense mutations, because the resulting amino acid replacement may not yield a functional protein.

In *E. coli*, there are three classes of tRNA suppressors: those that suppress only UAG (amber suppressors), those that suppress both UAA and UAG (ochre suppressors), and those that suppress only UGA. They share the following properties:

1. The original mutant gene still contains the mutant nucleotide sequence (UAG in Figure 11.31).

2. The suppressor tRNA suppresses all chain-termination mutations with the same stop codon, provided the amino acid inserted is an acceptable amino acid at the site.
3. A cell can survive the presence of a tRNA suppressor *only* if the cell contains two or more copies of the same tRNA gene. Taking the example in Figure 11.31, if only one $tRNA^{Leu}$ gene were present in the genome and if it were mutated, then the normal leucine codon UUG would no longer be read as a sense codon, and all polypeptide chains would terminate wherever a UUG codon occurred. However, *multiple copies of most tRNA genes exist,* so if one copy is mutated to yield a suppressor tRNA, a normal copy nearly always remains.
4. Any chain-termination codon can be translated by a suppressor tRNA mutation that recognizes that codon. For example, translation of UAG by insertion of an amino acid would prevent termination of all wildtype mRNA reading frames terminating in UAG. However, the anticodon of the suppressor tRNA usually binds rather weakly to the stop codon, so the stop codon often results in termination anyway.

Suppressor tRNA mutations are very useful in genetic analysis, because they allow nonsense mutations to be identified through their ability to be suppressed. This is important because nonsense alleles usually result in a truncated and completely inactive protein and so are considered true loss-of-function alleles. Suppressor tRNAs have been widely used for genetic analysis in prokaryotes and yeast, because their other harmful effects are tolerated by the organism. In most metazoans, suppressor tRNAs have such severe harmful effects that they are of limited usefulness.

Chapter Summary

The normal flow of information from a gene to its product is from DNA to RNA to protein. The properties of the different protein products of genes are determined by the sequence of amino acids of the polypeptide chain and by the way in which the chain is folded. Each protein-coding gene is usually responsible for the synthesis of a single polypeptide chain.

Gene expression begins with the enzymatic synthesis of an RNA molecule complementary to one strand of the DNA. This process is called transcription, and it is carried out by the enzyme RNA polymerase. This enzyme joins ribonucleoside triphosphates by the same chemical reaction used in DNA synthesis. RNA polymerase differs from DNA polymerase in that a primer is not needed to initiate synthesis. Transcription is initiated when RNA polymerase binds to a promoter sequence. The product of transcription is the primary transcript. In prokaryotes, the primary transcript is used directly as a messenger RNA (mRNA) in polypeptide synthesis. In eukaryotes, the primary transcript is processed into mRNA: Noncoding sequences (introns) are removed, the exons are spliced together, and the termini are modified by formation of a 5' cap and usually by addition of a poly-A tail at the 3' end.

After mRNA is formed, polypeptide chains are synthesized by translation of the mRNA molecule. Translation includes initiation, elongation, and release of the polypeptide chain. In eukaryotes, the initiation complex scans from the 5' cap to locate the first AUG; in prokaryotes, the initiation complex forms at any ribosome-binding site in the mRNA. In elongation, the process of translocation ratchets the 40S subunit along the mRNA as each new peptide bond is formed at the tRNA-binding site P. In the posttranslocation state, the 40S subunit is one notch (codon) farther along the mRNA than the 60S subunit. Then the pretranslocation state is restored as the 60S subunit catches up with the 40S subunit, the used tRNA is ejected from the E site, and the next charged tRNA moves into line at the A site. Elongation makes possible the successive reading of the nucleotide sequence of an mRNA molecule in groups of three nucleotides (the codons). When a stop codon is encountered, one or more release factors bind with the ribosome and cause release of the finished polypeptide and dissociation of the ribosome from the mRNA.

Several ribosomes can translate an mRNA molecule simultaneously, forming a polysome. In prokaryotes, translation often begins before synthesis of mRNA is completed (coupled transcription–translation); in eukaryotes, this does not occur because mRNA is made in the nucleus, whereas the ribosomes are located in the cytoplasm. Prokaryotic mRNA molecules are often polycistronic, encoding several different polypeptides.

The standard genetic code features 64 codons, of which 61 correspond to the 20 amino acids and 3 (UAA, UAG, and UGA) are stop codons. The code is highly redundant: Many amino acids have several codons. The codons in mRNA are recognized by tRNA molecules through base pairing between the 3-base anticodon in the tRNA and the complementary codon. When used in polypeptide synthesis, each tRNA molecule possesses a terminally bound amino acid (aminoacylated, or charged, tRNA). The correct amino acid is attached to each type of tRNA by specific enzymes called aminoacyl-tRNA synthetases. Nonsense suppressors are mutant tRNA molecules (usually mutated in the anticodon) that interact with a chain-termination ("nonsense") codon, allowing the insertion of an amino acid and continuation of polypeptide synthesis.

Key Terms

amino terminus
aminoacyl-tRNA synthetase
aminoacylated tRNA
anticodon
A (aminoacyl) site
cap
carboxyl terminus
charged tRNA
coding sequence
codon
colinearity
consensus sequence
cryptic splice site
elongation
E (exit) site
exon
exon-shuffle model
folding domain
frameshift mutation
gene expression
genetic code
initiation
intervening sequence
intron
lariat structure
messenger RNA (mRNA)
nonsense mutation
open reading frame (ORF)
peptide bond
peptidyl transferase
poly-A tail
polycistronic mRNA
polypeptide chain
polysome
posttranslocation state
pretranslocation state
primary transcript
promoter
protein subunit
P (peptidyl) site
R group
reading frame
release
ribosome
ribosome-binding site
RNA polymerase
RNA polymerase holoenzyme
RNA processing
RNA splicing
snRNP
spliceosome
TATA box
template strand
transcription
transfer RNA (tRNA)
translation
translocation
wobble

Review the Basics

- What is meant by the term *gene expression?* Would you make a distinction between gene expression and gene regulation? Why or why not?
- Would you regard an original text and its translation into another language as "colinear"? Explain your answer.
- Suppose that a duplex DNA molecule undergoes two double-stranded breaks that tightly flank the promoter of a gene and that the promoter region is inverted prior to enzymatic repair of the breaks. Would you expect the inverted promoter to be able to recruit the transcription complex? What, if anything, would be wrong with the transcript of the gene?
- Is the DNA strand that serves as the template for RNA polymerase transcribed in the 5'-to-3' or the 3'-to-5' direction? Which end of the mRNA molecule is translated first? Which end of the polypeptide encoded in the mRNA is synthesized first?
- In a eukaryotic cell, four general types of RNA molecules are used in gene expression. What are these types of RNA called? Which type is not involved in gene expression in prokaryotic cells, and why not?
- What is a primary transcript? How does a primary transcript in eukaryotes differ from that in prokaryotes?
- How do prokaryotes and eukaryotes differ in the mechanism for selecting an AUG codon as a start for polypeptide synthesis?
- What role do release factors play in translation?
- What does it mean to say that the standard genetic code is redundant? Which (if any) amino acids are encoded by one codon? by two? by three? by four? by five? by six?
- Give an example of a genetic system that does not use the standard genetic code.
- What is a frameshift mutation? Explain how *rII* recombinants containing multiple, single-nucleotide frameshift mutations were used to show that the messenger RNA is translated in consecutive groups of three nucleotides (codons).

Guide to Problem Solving

Problem 1 The following is the nucleotide sequence of a strand of DNA.

```
TACGTCTCCAGCGGAGATCTTTTCCGGTCGCAACTGAGGTTGATC
```

The strand is transcribed from left to right and codes for a small peptide.

(a) Which end is the 3' end and which the 5' end?
(b) What is the sequence of the complementary DNA strand?
(c) What is the sequence of the transcript?
(d) What is the amino acid sequence of the peptide?

(e) In which direction along the transcript does translation occur?

(f) Which is the amino ($-NH_2$) and which the carboxyl ($-COOH$) end of the peptide?

(a) 3'-TACGTCTCCAGCGGAGACCTTTTCCGGTCGCAACTGAGGTTGATC-5'
(b) 5'-ATGCAGAGGTCGCCTCTGGAAAAGGCCAGCGTTGACTCCAACUAG-3'
(c) 5'-AUGCAGAGGUCGCCUCUGGAAAAGGCCAGCGUUGACUCCAACUAG-3'
(d) MetGlnArgSerProLeuGluLysAlaSerValAspSerAsn

Answer to Problem 1(a)–1(d)

Answer Because the DNA strand is transcribed from left to right, and the first nucleotides incorporated into an RNA transcript form its 5' end, the 3' end of the DNA strand must be at the left. The original strand (answer to Problem 1**(a)**), the complementary strand (answer to Problem 1**(b)**), the transcript (answer to Problem 1**(c)**), and the peptide (answer to Problem 1**(d)**) are shown at top right.

(e) Translation goes from 5'-to-3' along the mRNA, so in this example, translation also goes from left to right.

(f) The amino end of the peptide is synthesized first (Met) and the carboxyl end last (Asn).

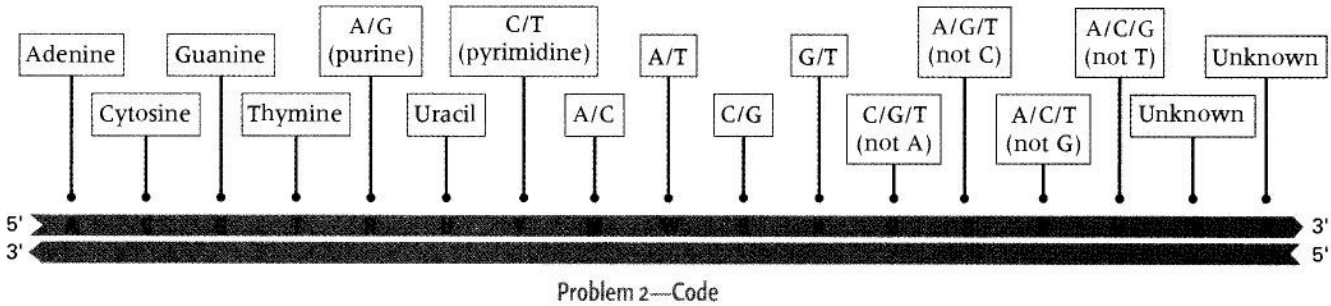
Problem 2 The International Union of Biochemistry and Molecular Biology (IUBMB) has designated a single-letter code for abbreviating the nucleotide bases that allows for ambiguous assignments. The code is shown in the diagram at the bottom of the page. The same code is used for DNA as for RNA. For ambiguous nucleotides, T and U are regarded as equivalent. Assuming standard Watson–Crick pairing between the two nucleotide strands shown, complete the sequence of the bottom strand, using the appropriate symbol from the standard ambiguity code.

Answer The ambiguity codes are very useful not only for designating uncertain nucleotides in DNA sequences but also for summarizing the redundancies in the genetic code (see Problem 3). The pairing relationships are straightforward for A, T, G, C, and U, but for ambiguous nucleotides one has to enumerate the possibilities and then select the symbol that expresses these ambiguities. One peculiar feature is that some symbols pair with themselves. For example, W (A or T) in one strand must also have a W (T or A) in the other strand, where the convention is that the paired nucleotides, though ambiguous, must obey the Watson–Crick pairing rules. All of the pairings can be worked out in this way, and the results are shown in the answer diagram at the bottom of the page. There are two symbols—namely N and X—in use for "any nucleotide," so these can be paired however it is convenient.

Problem 3 Rewrite the genetic code table using as many as possible of the single-letter codes for ambiguous bases established by the International Union of Biochemistry and Molecular Biology (IUBMB), as shown in Problem 2.

Answer This problem requires that you examine the standard genetic code and select the proper symbol for ambiguous nucleotides. The version of the genetic code that results is shown here. It has considerably fewer entries than the standard format, and it shows the general structure of the code at a glance.

First nucleotide in codon	Second nucleotide in codon: T	C	A	G
T	TTY Phe **F** TTR Leu **L**	TCN Ser **S**	TAY Tyr **Y** **TAR Stop**	TGY Cys **C** **TGA Stop** TGG Trp **W**
C	CTN Leu **L**	CCN Pro **P**	CAY His **H** CAR Gln **Q**	CGN Arg **R**
A	ATH Ile **I** **ATG Met** **M**	ACN Thr **T**	AAY Asn **N** AAR Lys **K**	AGY Ser **S** AGR Arg **R**
G	GTN Val **V**	GCN Ala **A**	GAY Asp **D** GAR Glu **E**	GGN Gly **G**



Problem 2—Code



Problem 2—Answer

Analysis and Applications

11.1 What possible amino acids are specified by a codon that consists of

(a) All pyrimidines?

(b) All purines?

11.2 Using the ambiguous-nucleotide designations of the International Union of Biochemistry and Molecular Biology (IUBMB), indicate how many codons are of the form:

(a) BBB.

(b) DDD.

(c) BDM.

11.3 Using the IUBMB symbols, indicate which amino acids may correspond to the codons:

(a) MMM.

(b) SSS.

11.4 What are the formulas of the stop codons in terms of the IUBMB symbols?

11.5 Each of the 61 codons that specify amino acids can have three possible nucleotide substitutions in the third position of the codon, yielding $61 \times 3 = 183$ possible mutations. How many of these are synonymous mutations, which specify the same amino acid as the nonmutant codon?

11.6 A certain protein has an amino acid composition such that no nucleotides in any of the codons can be abbreviated using the IUBMB symbols for ambiguous nucleotides. What amino acids can the protein contain?

11.7 What are the translation initiation and stop codons in the genetic code? In a random sequence of four ribonucleotides, what is the probability that any three adjacent nucleotides will be a start codon? A stop codon? In an mRNA molecule of random sequence, what is the average distance between stop codons?

11.8 A part of the *template* strand of a DNA molecule that codes for the 5' end of an mRNA has the sequence 3'-TTTTACGGGAATTAGAGTCGCAGGATG-5'. What is the amino acid sequence of the polypeptide encoded by this region, assuming that the normal start codon is needed for initiation of polypeptide synthesis?

11.9 A single codon in a double-stranded DNA molecule undergoes an inversion.

(a) If the original codon in the mRNA is 5'-AUG-3', what is the codon in the transcript from the inversion?

(b) Would the inversion of any codon conserve the amino acid that is specified?

(c) What codons would, upon inversion, yield nonsense codons?

(d) Which amino acids would inversion of each of the nonsense codons specify?

11.10 Some RNA molecules contain sequences that are complementary to one another and can fold back upon themselves to form regions of double-stranded RNA in which the complementary bases are paired. These "stem-and-loop" or "hairpin" secondary structures are quite stable. If the RNA molecule shown here forms a stem-and-loop structure, with the "loop" consisting of the five bases immediately preceding the dashes, what nucleotides are needed at the 3' end to form the stem?

5'-UACGGCUUCGAUGACAUGCG_____-3'

11.11 Poly-U codes for polyphenylalanine. If a G is added to the 5' end of the molecule, the polyphenylalanine has a different amino acid at the amino terminus, and if a G is added to the 3' end, there is a different amino acid at the carboxyl terminus. What are the amino acids?

11.12 The synthetic polymer, poly-A, is used as an mRNA molecule in an *in vitro* protein-synthesizing system that does not need a special start codon. Polylysine is synthesized. A single guanine nucleotide is added to one end of the poly-A. The resulting polylysine has a glutamic acid at the amino terminus. Was the G added to the 3' or the 5' end of the poly-A?

11.13 What polypeptide products are made when the alternating polymer GUGU . . . is used in an *in vitro* protein-synthesizing system that does not need a start codon?

11.14 What polypeptide products are made when the alternating polymer GUCGUC . . . is used in an *in vitro* protein-synthesizing system that does not need a start codon?

11.15 Some codons in the genetic code were determined experimentally by the translation of random polymers. If a ribonucleotide polymer is synthesized that contains 3/4 A and 1/4 C in random order, which amino acids would the resulting polypeptide contain, and in what frequencies?

11.16 How many different sequences of nine ribonucleotides would code for the amino acids Met–His–Thr? For Met–Arg–Thr? Using the symbol Y for any pyrimidine, R for any purine, and N for any nucleotide, what are the sequences?

11.17 Prior to the demonstration that messenger RNA is translated in consecutive, nonoverlapping groups of three nucleotides (codons), the possibility of an *overlapping code* had to be considered. Such overlapping codes could be rejected because they impose constraints on which consecutive pairs of amino acids could be found in proteins. To understand why, suppose that the standard triplet codons were translated with an overlap of two. (In other words, the last two nucleotides of any codon are also the first two nucleotides of the next codon in line.) Which amino acids could follow:

(a) Lys? **(c)** Met?

(b) Tyr? **(d)** Trp?

11.18 If the standard genetic code were translated with an overlap of two nucleotides (that is, the last two nucleotides of any codon are also the first two nucleotides of the next codon in line), which amino acids could be at the carboxyl end of a polypeptide chain?

11.19 If the standard genetic code were translated with an overlap of one nucleotide (that is, the last nucleotide of any codon is also the first nucleotide of the next codon in line), which amino acids could follow:

(a) Trp?
(b) Met?

11.20 If the standard genetic code were translated with an overlap of one nucleotide (that is, the last nucleotide of any codon is also the first nucleotide of the next codon in line), which amino acids could not be found at the carboxyl end of a polypeptide chain?

11.21 If the standard genetic code were translated with the codons overlapping, what amino acids could follow Met at the amino end of a polypeptide chain if:

(a) The overlap consisted of two nucleotides?
(b) The overlap consisted of one nucleotide?

11.22 What codons could pair with the anticodon 5'-IAU-3'? (I stands for inosine.) What amino acid would be incorporated?

11.23 Two possible anticodons could pair with the codon UGG, but only one is actually used. Identify the possible anticodons, and explain why one of them is not used.

11.24 Two *E. coli* genes, *A* and *B*, are known from mapping experiments to be very close to each other. A deletion mutation is isolated that eliminates the activity of both *A* and *B*. Neither the A nor the B protein can be found in the mutant, but a novel protein is isolated in which the amino-terminal 30 amino acids are identical to those of the *B* gene product and the carboxyl-terminal 30 amino acids are identical to those of the *A* gene product.

(a) With regard to the 5'-to-3' orientation of the nontranscribed DNA strand, is the order of the genes *AB* or *BA*?
(b) Can you make any inference about the number of bases deleted in the coding region?

11.25 The nontranscribed sequence at the beginning of a gene reads

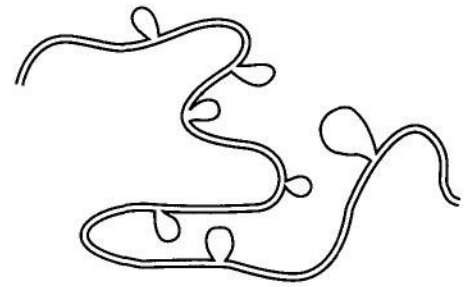
5'-ATGCATCCGGGCTCATTAGTCT . . . -3'

Two mutations are studied. Mutation X has an insertion of a G immediately after the underlined G, and mutation Y has a deletion of the red A. What is the amino acid sequence of each of the following?

(a) the wildtype polypeptide
(b) the polypeptide in mutant X
(c) the polypeptide in mutant Y
(d) the polypeptide in a recombinant organism containing both mutations

11.26 Protein synthesis occurs with high fidelity. In prokaryotes, incorrect amino acids are inserted at the rate of approximately 10^{-3} (that is, one incorrect amino acid per 1000 translated). What is the probability that a polypeptide of 300 amino acids has exactly the amino acid sequence specified in the mRNA?

11.27 A DNA fragment containing a particular gene is isolated from a eukaryotic organism. This DNA fragment is mixed with the corresponding mRNA isolated from the organism, denatured, renatured, and observed by electron microscopy. Heteroduplexes of the type shown in the accompanying figure are observed. How many introns does this gene contain?



11.28 If the DNA molecule shown here is transcribed from left to right, what are the sequence of the mRNA and the amino acid sequence? What are the sequence of the mRNA and the amino acid sequence if the segment in red is inverted?

5'-AGACTTCAGGCTCAACGTGGT-3'
3'-TCTGAAGTCCGAGTTGCACCA-5'

11.29 You are given the nontemplate-strand nucleotide sequence of a part of an exon of an active gene.

5'-TAACGTATGCTTGACCTCCAAGCAATCGATGCCAGCTCAAGG-3'

Assuming the standard genetic code, what is the amino acid sequence in the polypeptide chain? What tells you that you have identified the correct reading frame?

11.30 An organism is discovered in which the promoter sequence in the template strand is 3'-TTTTT-5', the transcript begins with the first nucleotide following the promoter sequence, and transcription terminates immediately prior to the sequence 3'-GGGGGG-5' in the template strand. The primary transcript is capped and used directly for the mRNA, and translation is initiated by scanning from the 5' end.

(a) What mRNA sequence would be transcribed from the DNA sequence below?
(b) What polypeptide chain would result?

5'-CACCGAAAAATACCATGTCAAACAGCGTATGGTAGCAGTGCCCCCCCAGT-3'
3'-GTGGCTTTTTATGGTACAGTTTGTCGCATACCATCGTCACGGGGGGGTCA-5'

Problem 11.30

GeNETics on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

•A **sequence logo** is a graphical representation of a set of nucleotide sequences that are binding sites for proteins or other macromolecules. The logo uses letters of differing size to display the relative frequencies of alternative bases at each position in the sequence so that information about variation in the sequence motif is not lost, as it is in a conventional consensus sequence. This keyword site contains logos for many sequence motifs that are important in gene expression. Check out the logos for yeast TATA boxes as well as the splice donor and splice acceptor sites in human RNA processing to judge how strongly the "consensus sequences" are conserved.

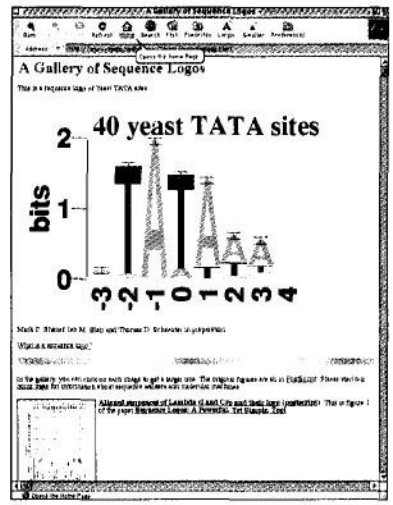
• The keyword **structures** will link you to a site offering a three-dimensional display of over 200,000 small biological molecules, including the common amino acids. Using the mouse, you can tilt and rotate the structures to analyze the spatial relationships among the atoms. At this keyword site you can browse through the amino acids to gain an appreciation of their relative sizes and shapes, and you can examine a peptide bond by selecting alanine–alanine in the category biomolecules. (The comparisons are facilitated if your computer system can support two browser windows running side by side.)

• The gene *CCA1* in yeast is critical for the formation of all mature **transfer RNA** molecules that are able to be charged with the correct amino acid. Search this keyword site for the gene *CCA1* to find out what reaction is catalyzed by the gene product and how this reaction changes the sequence of the tRNA. You can also locate the gene on the genetic map and retrieve its DNA and amino acid sequence.

• **Cystic fibrosis** is among the most common of the serious inherited human disorders. It results from any of a number of mutations causing a defect in the cystic fibrosis transmembrane conductance regulator (CFTR), one of a family of membrane chloride channels. The gene is designated *CFTR*. The mutant allele of *CFTR* that is most frequently encountered in patients is ΔF508, a deletion of three nucleotides that deletes a phenylalanine at position 508 in the *CFTR* polypeptide chain. Although the ΔF508 deletion accounts for about two-thirds of the mutant *CFTR* alleles in the human population, more than 700 mutant alleles have been described. Most of these are listed at this keyword site. Note the great variety of different types of mutations that have been described: promoter mutation, amino acid replacement, frameshift mutation, splicing defect, deletion, and so forth.

• The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 11 and you will be linked automatically.

• The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 11.



Challenge Problems

11.31 For two different frameshift mutations in the second codon of a gene, the amino terminal sequences of the mutant proteins are

Mutant 1: Met–Lys
Mutant 2: Met–Ile–Val

Mutant 1 has a single-nucleotide addition and terminates in a UAG codon, and mutant 2 has a single-nucleotide deletion and terminatest in a UAA codon. Furthermore, the first five amino acids of the wildtype protein are known to be Met-(Asn, Val, Ser, Lys), where the parentheses mean that the order of the amino acids is un-

known. Using the information provided by the frameshift mutations, determine the first five codons in the wildtype gene as well as the nature of each frameshift mutation.

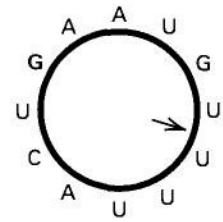
11.32 An organism is discovered whose RNA polymerase II has the following properties:

(1) The DNA sequence in the template strand 3'-TATAATA-5' serves as the promoter.
(2) Transcription begins at the nucleotide immediately following the promoter.
(3) Transcription continues until the transcript includes the sequence 5'-GGGGG-3', at which point the polymerase *switches to the other DNA strand* and continues transcribing (still adding nucleotides only to the 3' end of the growing chain).
(4) Transcription terminates immediately after transcribing 5'-TATATA-3' from the template strand.

The primary transcript is immediately capped and used as the mRNA, and translation is initiated by scanning from the 5' end.

(a) What mRNA would be produced from the DNA sequence shown below?
(b) What polypeptide chain results from the sequence?

11.33 A prokaryote is discovered with the unusual property that some of its mRNA molecules are circular, but they are translated according to the standard genetic code using any 5'-AUG-3' for initiation. For the circular mRNA shown here, deduce:



(a) The amino acid sequence of the wildtype polypeptide chain.
(b) The amino acid sequence of a mutant in which the red G was substituted with a U.
(c) The amino acid sequence of a mutant in which a C was inserted at the position indicated by the arrow.
(d) Why the polypeptide in part **(b)** is longer than that in part **(c)**.
(e) The amino acid sequence of the double mutant.

```
5'-CCGTATATATTATGATCAATATGCATGCTCTCGGGGGTCACACT-3'
3'-GGCATATATAATACTAGTTATACGTACGAGAGCCCCCAGTGTGA-5'
```

Problem 11.32

Further Reading

Beadle, G. W. 1948. Genes of men and molds. *Scientific American*, September.

Blumenthal, T. 1995. *Trans*-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends in Genetics* 11: 132.

Chambon, P. 1981. Split genes. *Scientific American*, May.

Crick, F. H. C. 1966. The genetic code. *Scientific American*, October.

Garrett, R. A., S. R. Douthwaite, A. Liljas, A. T. Matheson, P. B. Moore, and H. F. Noller. 2000. *The Ribosome: Structure, Function, Antibiotics, and Cellular Interaction*. Washington, DC: American Society for Microbiology.

Hanke, J., D. Brett, I. Zastrow, A. Aydin, S. Delbruck. G. Lehmann, F. Luft, J. Reich, and P. Bork. 1999. Alternative splicing of human genes: More the rule than the exception? *Trends in Genetics* 15: 389.

Haseltine, W. A. 1997. Discovering genes for new medicines. *Scientific American*, March.

Henkin, T. M. 1996. Control of transcription termination in prokaryotes. *Annual Review of Genetics* 30: 35.

Hill, W. E., and A. Dahlberg, eds. 1990. *The Ribosome: Structure, Function, and Evolution*. Washington, DC: American Society for Microbiology.

Jackson, R . J., and M. Wickens. 1997. Translational controls impinging on the 5'-untranslated region and initiation factor proteins. *Current Opinion in Genetics & Development* 7: 233.

Kim, J. L., D. B. Nikolov, and S. K. Burley. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* 3656: 520.

Kim, Y., J. H. Geiger, S. Hahn, and P. B. Sigler. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature* 365: 512.

Lee, M. S., and P. A. Silver. 1997. RNA movement between the nucleus and the cytoplasm. *Current Opinion in Genetics & Development* 7: 212.

Neidhardt, F. C., R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, eds. 1996. Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (2 volumes). 2d ed. Washington, DC: American Society for Microbiology.

Nirenberg, M. 1963. The genetic code. *Scientific American*, March.

Ross, J. 1996. Control of messenger RNA stability in higher eukaryotes. *Trends in Genetics* 12: 171.

Taylor, J. H., ed. 1965. *Selected Papers on Molecular Genetics*. New York: Academic Press.

Wickens, M., P. Anderson, and R. J. Jackson. 1997. Life and death in the cytoplasm: Messages from the 3′ end. *Current Opinion in Genetics & Development* 7: 220.

CHAPTER 12

Molecular Mechanisms of Gene Regulation



In the early development of the *Drosophila* wing (the wing imaginal disk), the expression of the *vestigial* gene occurs in two distinct patterns. These patterns, shown at left by the contrasting colors, are owing to its activation in different cell types by different enhancers. If either or both mechanisms of activation are defective, the wings fail to develop normally. [Disks: courtesy of Stephen Paddock and Sean Carroll; *D. melanogaster*: Oliver Meckes/Photo Researchers. Inc.]

CHAPTER OUTLINE

12.1 Transcriptional Regulation in Prokaryotes
Inducible and Repressible Systems of Negative Regulation
Positive Regulation

12.2 Lactose Metabolism and the Operon
System of Gene Regulation
Lac⁻ Mutants
Inducible and Constitutive Synthesis and Repression
The Repressor
The Operator Region
The Promoter Region
The Operon System of Transcriptional Regulation
Positive Regulation of the Lactose Operon

12.3 Regulation of the Tryptophan Operon
Attenuation

12.4 Regulation in Bacteriophage Lambda

12.5 Transcriptional Regulation in Eukaryotes
Galactose Metabolism in Yeast
Transcriptional Activator Proteins
Transcriptional Enhancers and Transcriptional Silencers
The Eukaryotic Transcription Complex
Chromatin-Remodeling Complexes
Alternative Promoters

12.6 Epigenetic Mechanims of Transcriptional Regulation
Cytosine Methylation
Methylation and Transcriptional Inactivation
Cosuppression Through Transcriptional Silencing
Genomic Imprinting in the Female and Male Germ Lines

12.7 Regulation Through RNA Processing and Decay
Alternative Splicing
Messenger RNA Stability
Cosuppression Through RNA Turnover
RNA Interference (RNAi)

12.8 Translational Control

12.9 Programmed DNA Rearrangements
Gene Amplification
Antibody and T-Cell Receptor Variability
Mating Type Interconversion
Transcriptional Control of Mating Type

PRINCIPLES

- Control of transcription is an important level of gene regulation, with mechanisms ranging from complete negative regulation ("on unless turned off") to complete positive regulation ("off unless turned on").
- Most genes are subject to multiple, overlapping regulatory mechanisms.
- In prokaryotes, the genes coding for the enzymes in a metabolic pathway are often clustered in the genome and controlled coordinately by a regulatory protein that binds with an operator region at the 5' end of the cluster. This type of gene organization is known as an operon.
- In eukaryotes, most genes are not organized into operons. Genes at dispersed locations in the genome are coordinately controlled by one or more enhancer DNA sequences located near each gene that interact with transcriptional activator proteins.
- The transcription complex in eukaryotes consists of numerous protein components that are recruited to the promoter of a gene whose chromatin has been suitably reconfigured.
- Epigenetic mechanisms of transcriptional control are hereditary changes in gene expression mediated by modification of the DNA bases (usually cytosine methylation) or by binding with regulatory proteins.
- Gene expression can also be regulated at the level of RNA processing, alternative patterns of splicing, or transcript or mRNA stability.
- Programmed DNA rearrangements underlie some genetic processes, such as the origin of antibody variability.

CONNECTIONS

Operator? Operator?
François Jacob, David Perrin, Carmen Sanchez, and Jacques Monod 1960
The Operon: A Group of Genes Whose Expression Is Coordinated by an Operator [original in French]

Sex-Change Operations
James B. Hicks, Jeffrey N. Strathern, and Ira Herskowitz 1977
The Cassette Model of Mating-type Interconversion

A vertebrate animal—for example, a mouse—contains approximately 200 different cell types that are specialized in their functions. These differences are correlated with patterns of gene expression, because most genes differ in their level of expression according to cell type or stage in the cell cycle. The activity of genes is also keyed to the functions of the cell; for instance, the genes for hemoglobin are expressed at high levels only in precursors of the red blood cells. The control of synthesis of particular gene products is called *gene regulation*.

In many cases, gene activity is regulated at the level of transcription, either through signals originating within the cell itself or in response to external conditions. For example, many gene products are needed only on occasion, and transcription can be regulated in an on–off manner that enables such products to be present only when external conditions demand them. The flow of genetic information is regulated in other ways also. Control points for gene expression include

1. ***Transcriptional regulation*** of the synthesis of RNA transcripts by controlling initiation or termination.
2. ***RNA processing***, or regulation through RNA splicing or alternative patterns of splicing.
3. ***Translational control*** of polypeptide synthesis.
4. ***Stability of mRNA***, because mRNAs that persist in the cell have longer-lasting effects than those that are degraded rapidly.
5. ***Posttranslational control***, which includes a great variety of mechanisms that affect enzyme activity, activation, stability, and so on.
6. ***DNA rearrangements***, in which gene expression changes depending on the position of DNA sequences in the genome.

The regulatory systems of prokaryotes and eukaryotes differ from each other in many details. Prokaryotes are generally free-living unicellular organisms that grow and divide indefinitely as long as environmental conditions are suitable and the supply of nutrients is adequate. Their regulatory systems are often geared to provide the maximum growth rate in a particular environment. In contrast, the cells in a developing multicellular organism modulate their growth rate as they undergo dramatic, coordinated differentiation in morphology and metabolism. In an adult animal, growth and division of most cell types has ceased, and each type of cell needs to maintain its identity through time.

12.1 Transcriptional Regulation in Prokaryotes

In bacteria and phages, on–off gene activity is often controlled through transcription. Under conditions when a gene product is needed, transcription of the gene is turned "on"; under other conditions, transcription is turned "off." The term *off* should not be taken literally. In bacteria, few examples are known of a system being switched off completely. When transcription is in the "off" state, a basal level of gene expression nearly always remains, often averaging one transcriptional event or fewer per cell generation; hence "off" really means that there is very little synthesis of the gene product. Extremely low levels of expression are also found in certain classes of genes in eukaryotes, including many genes that participate in embryonic development. Regulatory mechanisms other than the on–off type also are known in both prokaryotes and eukaryotes; in these examples, the level of expression of a gene may be modulated in gradations from high to low according to conditions in the cell.

In bacterial systems, when several enzymes act in sequence in a single metabolic pathway, usually either all or none of these enzymes are produced. This **coordinate regulation** results from control of the synthesis of one or more polycistronic mRNA molecules encoding all of the gene products that function in the same metabolic pathway. This type of regulation is not found in eukaryotes because, as we saw in Chapter 11, eukaryotic mRNA is monocistronic.

Inducible and Repressible Systems of Negative Regulation

The molecular mechanisms of regulation usually fall into either of two broad categories: *negative regulation* and *positive regulation*. In a system subject to **negative regulation** (Figure 12.1A), the default state is "on," and transcription takes place until it is

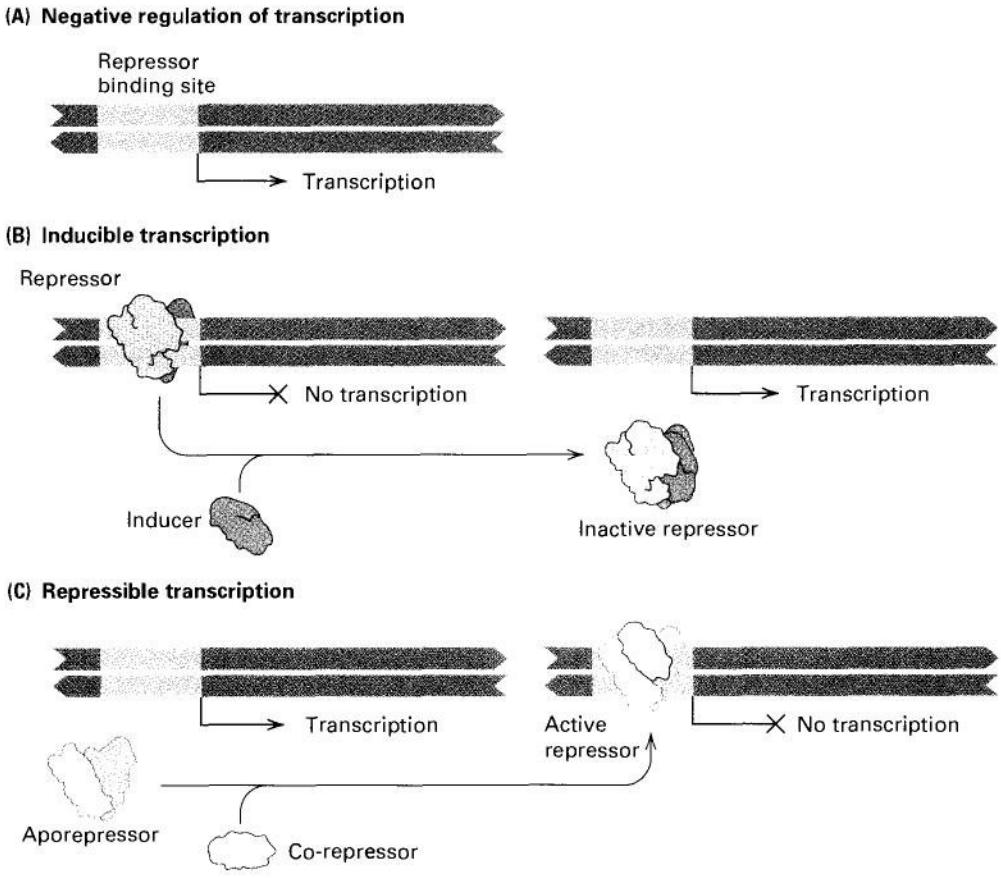


Figure 12.1 (A) Negative regulation of transcription means that the default state of transcription is "on" unless a repressor turns it "off." (B) In an inducible system, the repressor is a protein whose DNA binding is inactivated by the inducer. (C) In a repressible system, the repressor is formed by the interaction between an aprorepressor protein and a co-repressor.

turned "off" by a **repressor** protein that binds to the DNA upstream from the transcriptional start site. A negatively regulated system may be either *inducible* (Figure 12.1B) or *repressible* (Figure 12.1C), depending on how the active repressor is formed. In **inducible transcription,** a repressor DNA-binding protein normally keeps transcription in the "off" state. In the presence of a small molecule called the **inducer,** the repressor binds preferentially with the inducer and loses its DNA-binding capability, allowing transcription to occur. Many degradative (catabolic) pathways are inducible and use the initial substrate of the degradative pathway as the inducer. In this way, the enzymes used for degradation are not synthesized unless the substrate is present in the cell.

In **repressible transcription** (Figure 12.1C), the default state is "on" until an active repressor is formed to turn it "off." In this case the regulatory protein is called an **aporepressor,** and it has no DNA-binding activity on its own. The active repressor is formed by the combination of the aporepressor and a small molecule known as the **co-repressor.** Presence of the co-repressor thereby results in the cessation of transcription. Repressible regulation is often found in the control of the synthesis of enzymes that participate in biosynthetic (anabolic) pathways; in these cases the final product of the pathway is frequently the co-repressor. In this way, the enzymes of the biosynthetic pathway are not synthesized until the concentration of the final product becomes too low to cause repression.

Positive Regulation

Note that in negative regulation (Figure 12.1A), the default state of transcription is "on," and a repressor is required to turn it

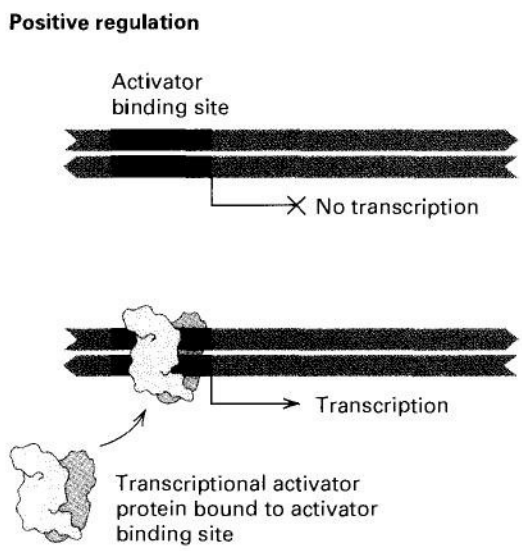


Figure 12.2 In positive regulation, the default state of transcription is "off." Transcription is stimulated by the binding of a transcriptional activator protein.

"off." Contrast this with a positively regulated system (**Figure 12.2**), in which the default state is "off," and binding with a regulatory protein is required to turn it "on." Such a regulatory protein is called a **transcriptional activator protein.** Negative and positive regulation are not mutually exclusive, and some systems are both positively and negatively regulated, utilizing two regulators to respond to different conditions in the cell. Negative regulation is more common in prokaryotes, positive regulation in eukaryotes.

Some genes exhibit **autoregulation,** which means that the protein product of a gene regulates its own transcription. In negative autoregulation, the protein inhibits transcription, and high concentrations of the protein result in less transcription of the mRNA. This mechanism automatically adjusts the steady-state level of the protein in the cell. In positive autoregulation, the protein stimulates transcription: As more protein is made, transcription increases to the maximum rate. Positive autoregulation is a common way for weak induction to be amplified. Only a weak signal is necessary to get production of the protein started, but then the positive autoregulation takes over and stimulates further production to the maximum level.

Next we examine two classical systems of regulation found in the bacterium *Escherichia coli.* These serve as specific examples of the general concepts introduced in Figures 12.1 and 12.2. We shall see that in the real world, most genes have overlapping mechanisms of control that include both positive and negative regulatory elements.

12.2 Lactose Metabolism and the Operon System of Gene Regulation

Gene regulation was first studied in detail in the system in *E. coli* that is responsible for degradation of the sugar lactose, and some of the key features of gene regulation were originally discovered in this system. It is a classic example of negative regulation in which transcription is inducible (Figure 12.1B). The first insights into the regulatory mechanisms came from a genetic analysis of mutations affecting lactose metabolism. The following sections describe the properties of these mutations and the interpretations that emerged from their analysis. The mechanisms of regulation inferred from the mutants has been abundantly confirmed by direct molecular studies.

Lac$^-$ Mutants

In *E. coli,* two proteins are necessary for the metabolism of lactose. They are the enzyme ***β*-galactosidase,** which cleaves lactose (a *β*-galactoside) to yield galactose and glucose; and a transporter molecule, **lactose permease,** which is required for the entry of lactose into the cell. The existence of two different proteins in the lactose-utilization system was first shown by a combination of genetic experiments and biochemical analysis.

First, hundreds of mutants unable to use lactose as a carbon source, designated Lac$^-$ mutants, were isolated. Some of the mutations were in the *E. coli* chromosome, and others were in an F′ *lac,* a plasmid carrying the genes for lactose utilization. By performing F′ × F$^-$ matings, investigators constructed partial diploids with the genotypes F′ *lac*$^-$/*lac*$^+$ and F′ *lac*$^+$/*lac*$^-$. (The

genotype of the plasmid is given to the left of the slash and that of the chromosome to the right.) It was observed that all of these diploids always had a Lac$^+$ phenotype (that is, they made β-galactosidase and permease); thus none produced an inhibitor that prevented functioning of the *lac* genes. Other partial diploids were then constructed in which both the F' *lac* plasmid and the chromosome carried a *lac*$^-$ allele. These were tested for the Lac$^+$ phenotype, with the result that all of the mutants initially isolated could be placed into two complementation groups, called *lacZ* and *lacY*, a result that implies that the *lac* system consists of at least two genes. In particular, the partial diploids

F' *lacZ*$^-$*lacY*$^+$/*lacZ*$^+$*lacY*$^-$ and
F' *lacZ*$^+$*lacY*$^-$/*lacZ*$^-$*lacY*$^+$

had a Lac$^+$ phenotype (complementation), producing both β-galactosidase and permease. However, the genotypes

F' *lacZ*$^-$*lacY*$^+$/*lacZ*$^-$*lacY*$^+$ and
F' *lacZ*$^+$*lacY*$^-$/*lacZ*$^+$*lacY*$^-$

had a Lac$^-$ phenotype (noncomplementation). The F' *lacZ*$^-$*lacY*$^+$/*lacZ*$^-$*lacY*$^+$cells were unable to synthesize the β-galactosidase, hence *lacZ* encodes β-galactosidase; and the F' *lacZ*$^+$*lacY*$^-$/*lacZ*$^+$*lacY*$^-$cells were unable to synthesize the permease, hence *lacY* encodes the permease. (A third gene encoding a β-galactoside transacetylase was later discovered. It was not included among the early mutants because it is not essential for growth on lactose.) A final important result—that the *lacZ* and *lacY* genes are adjacent—was deduced from the high frequency of cotransduction observed in genetic mapping experiments.

Inducible and Constitutive Synthesis and Repression

The on–off nature of the lactose-utilization system is shown by the following observations:

1. If a culture of Lac$^+$ *E. coli* is growing in a medium that does not include lactose or any other β-galactoside, then the intracellular concentrations of β-galactosidase and permease are exceedingly low, roughly 1–2 molecules of each per bacterial cell. However, if lactose is present in the growth medium, then the number of each of these molecules is about 10^3-fold higher.
2. If lactose is added to a Lac$^+$ culture growing in a lactose-free medium (also lacking glucose, a point that will be discussed shortly), then both β-galactosidase and permease are synthesized nearly simultaneously, as shown in **Figure 12.3**. Analysis of the total mRNA present in the cells before and after the addition of lactose shows that almost no *lac* mRNA (the polycistronic mRNA that codes for β-galactosidase and permease) is present before lactose is added and that the addition of lactose triggers synthesis of *lac* mRNA.

These two observations led to the conclusions that transcription of the lactose genes is inducible and that lactose (or a derivative of lactose) is the inducer. Some analogs of lactose are also inducers, such as a sulfur-containing analog denoted IPTG (isopropylthiogalactoside); this synthetic inducer is

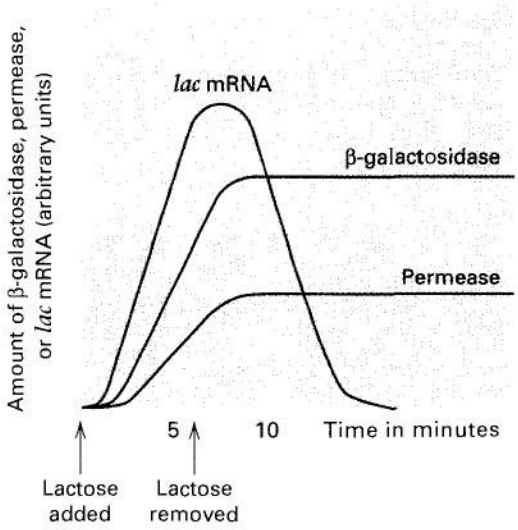


Figure 12.3 The "on–off" nature of the *lac* system. The *lac* mRNA appears soon after lactose or another inducer is added; β-galactosidase and permease appear at nearly the same time but are delayed with respect to mRNA synthesis because of the time required for translation. When lactose is removed, no more *lac* mRNA is made, and the amount of *lac* mRNA decreases because of the degradation of mRNA already present. Both β-galactosidase and permease are stable proteins: Their amounts remain constant even when synthesis ceases. However, their concentration per cell gradually decreases as a result of repeated cell divisions.

convenient for experiments because it induces but is not cleaved by β-galactosidase, so IPTG is stable in the cell whether or not β-galactosidase is present.

The key to understanding induction came from mutants with defective regulation. In one class of regulatory mutants, the *lac* mRNA was synthesized even in the absence of inducer. Because of their constant synthesis, these mutants were termed **constitutive.** The constitutive mutants fell into two classes, *lacI*$^-$ and *lacO*^{c}. Mutants were also obtained that failed to produce *lac* mRNA even when the inducer was present. These uninducible mutants fell into two classes, *lacI*^{s} and *lacP*$^-$. The characteristics of the mutants are shown in **Table 12.1** and discussed in the following sections.

The Repressor

In Table 12.1, genotypes 3 and 4 show that *lacI*$^-$ mutations are recessive. In the absence of inducer, a *lacI*$^+$ cell does not synthesize *lac* mRNA, whereas the mRNA is produced in a *lacI*$^-$ mutant. These results suggest that

> The *lacI* gene is a regulatory gene whose product is the repressor protein that keeps the system turned off. Because the repressor is necessary to shut off mRNA synthesis, regulation by the repressor is negative regulation.

How the *lacI* repressor protein binds to the DNA and prevents synthesis of *lac* mRNA will be explained shortly. A *lacI*$^-$ mutant lacks the repressor protein and hence is constitutive. Wildtype copies of the repressor protein are present in a *lacI*$^+$/*lacI*$^-$ partial diploid, so transcription is repressed. It is important to note that the single *lacI*$^+$ gene prevents synthesis of *lac* mRNA from both the F' plasmid and the chromosome. Therefore, the repressor protein must be diffusible within the cell, because it can shut off mRNA synthesis from both DNA molecules present in the partial diploid. Genetic mapping experiments placed the *lacI* gene very close to the *lacZ* gene and established the gene order *lacI lacZ lacY*.

Referring to Table 12.1 again, genotypes 7 and 8 indicate that the *lacI*^{s} mutations are dominant and act to shut off mRNA synthesis from both the F' plasmid and the chromosome whether or not the inducer is present (the superscript in *lacI*^{s} signifies *super-repressor*). The *lacI*^{s} mutations result in repressor molecules that fail to recognize and bind the inducer, and thus repress *lac* mRNA synthesis even in the presence of inducer.

Table 12.1 Characteristics of partial diploids containing several combinations of *lacI*, *lacO* and *lacP* alleles

	Genotype	Synthesis of *lac* mRNA	Lac phenotype
1.	F' *lacO*^{c} *lacZ*$^+$/*lacO*$^+$ *lacZ*$^+$	Constitutive	+
2.	F' *lacO*$^+$ *lacZ*$^+$/*lacO*^{c} *lacZ*$^+$	Constitutive	+
3.	F' *lacI*$^-$ *lacZ*$^+$/*lacI*$^+$ *lacZ*$^+$	Inducible	+
4.	F' *lacI*$^+$ *lacZ*$^+$/*lacI*$^-$ *lacZ*$^+$	Inducible	+
5.	F' *lacO*^{c} *lacZ*$^-$/*lacO*$^+$ *lacZ*$^+$	Inducible	+
6.	F' *lacO*^{c} *lacZ*$^+$/*lacO*$^+$ *lacZ*$^-$	Constitutive	+
7.	F' *lacI*^{s} *lacZ*$^+$/*lacI*$^+$ *lacZ*$^+$	Uninducible	−
8.	F' *lacI*$^+$ *lacZ*$^+$/*lacI*^{s} *lacZ*$^+$	Uninducible	−
9.	F' *lacP*$^-$ *lacZ*$^+$/*lacP*$^+$ *lacZ*$^+$	Inducible	+
10.	F' *lacP*$^+$ *lacZ*$^+$/*lacP*$^-$ *lacZ*$^+$	Inducible	+
11.	F' *lacP*$^+$ *lacZ*$^-$/*lacP*$^-$ *lacZ*$^+$	Uninducible	−
12.	F' *lacP*$^+$ *lacZ*$^+$/*lacP*$^-$ *lacZ*$^-$	Inducible	+

The Operator Region

Entries 1 and 2 in Table 12.1 show that *lacO*^{c} mutants are dominant. However, the dominance is evident only in certain combinations of *lac* mutations, as can be seen by examining the partial diploids shown in entries 5 and 6. Both combinations are Lac$^+$ because a functional *lacZ* gene is present. However, in the combination shown in entry 5, synthesis of β-galactosidase is inducible even though a *lacO*^{c} mutation is present. The difference is that in entry 5, the *lacO*^{c} mutation is present in the same DNA molecule as the *lacZ*$^-$ mutation, whereas in entry 6, *lacO*^{c} is contained in the same DNA molecule as *lacZ*$^+$. The key feature of these results is that

> A *lacO*^{c} mutation causes constitutive synthesis of β-galactosidase only when the *lacO*^{c} and *lacZ*$^+$ alleles are contained in the same DNA molecule.

The *lacO*^{c} mutation is said to be ***cis*-dominant,** because only genes in the *cis* configuration (in the same DNA molecule as that containing the mutation) are expressed in dominant fashion. Confirmation of this conclusion comes from an important biochemical observation: The mutant enzyme coded by the *lacZ*$^-$ sequence is synthesized constitutively in a *lacO*^{c} *lacZ*$^-$/*lacO*$^+$ *lacZ*$^+$ partial diploid (entry 5), whereas the wildtype enzyme (coded by the *lacZ*$^+$ sequence) is synthesized only if an inducer is added. All *lacO*^{c} mutations are located between the *lacI* and *lacZ* genes; hence the gene order of the four genetic elements of the *lac* system is

lacI lacO lacZ lacY

An important feature of all *lacO*^{c} mutations is that they cannot be complemented (a characteristic feature of all *cis*-dominant mutations); that is, a *lacO*$^+$ allele cannot alter the constitutive activity of a *lacO*^{c} mutation. This observation implies that the *lacO* region does not encode a diffusible product and must instead define a site in the DNA that determines whether synthesis of the product of the adjacent *lacZ* gene is inducible or constitutive. The *lacO* region is called the **operator.** In the next section, we will see that the operator is in fact a binding site in the DNA for the repressor protein.

The Promoter Region

Entries 11 and 12 in Table 12.1 show that *lacP*$^-$ mutations, like *lacO*^{c} mutations, are *cis*-dominant. The *cis*-dominance can be seen in the partial diploid in entry 12. The genotype in entry 11 is uninducible, in contrast to the partial diploid of entry 12, which is inducible. The difference between the two genotypes is that in entry 11, the *lacP*$^-$ mutation is in the same DNA molecule with *lacZ*$^+$, whereas in entry 12, the *lacP*$^-$ mutation is combined with *lacZ*$^-$. This observation means that the presence of *lacP*$^-$ mutation makes the adjacent genes uninducible: A *lacP*$^-$ mutation prevents a *cis*-located wildtype *lacZ*$^+$ gene from being transcribed. The *lacP*$^-$ mutations map between *lacI* and *lacO*, and the order of the five genetic elements of the *lac* system is

lacI lacP lacO lacZ lacY

As expected because of the *cis*-dominance of *lacP*$^-$ mutations, they cannot be complemented; that is, a *lacP*$^+$ allele on another DNA molecule cannot supply the missing function to a DNA molecule carrying a *lacP*$^-$ mutation. Thus *lacP*, like *lacO*, must define a site that determines whether synthesis of *lac* mRNA will take place. Because synthesis does not occur if the site is defective or missing, *lacP* defines an essential site for mRNA synthesis. The *lacP* region is called the **promoter.** It is, in fact, the site in the DNA at which RNA polymerase binds to initiate transcription.

The Operon System of Transcriptional Regulation

The genetic regulatory mechanism of the *lac* system was first explained by the **operon model** of François Jacob and Jacques Monod, which is illustrated in **Figure 12.4.** (The figure uses the abbreviations *i*, *o*, *p*, *z*, *y*, and *a* for *lacI*, *lacO*, *lacP*, *lacZ*, *lacY*, and *lacA*.) The operon model has the following features:

1. The lactose-utilization system consists of two kinds of components—*structural genes* (*lacZ* and *lacY*), which encode proteins needed for the transport and metabolism of lactose, and *regulatory elements* (the repressor gene *lacI*, the promoter *lacP*, and the operator *lacO*).

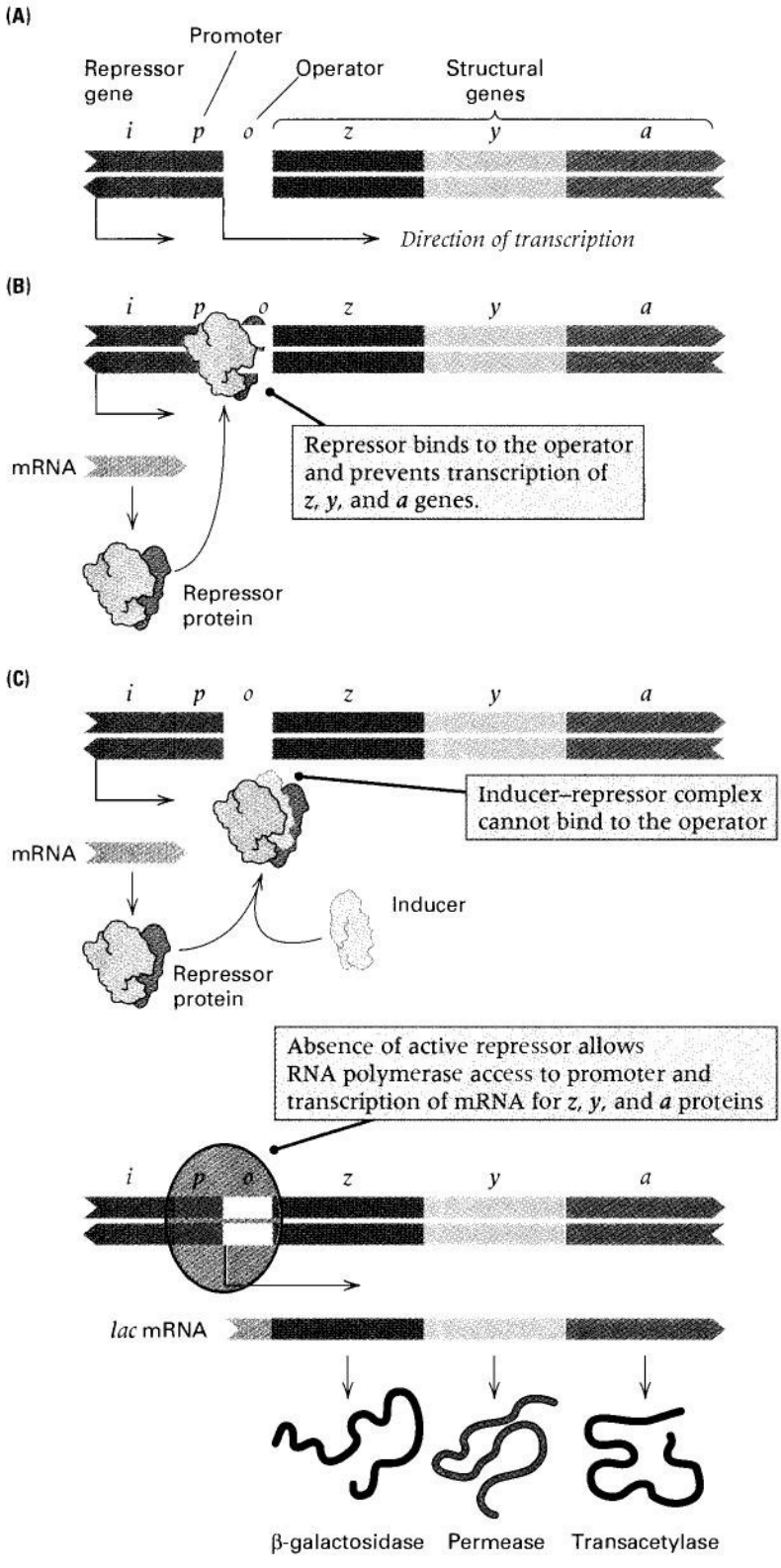


Figure 12.4 (A) A map of the *lac* operon, not drawn to scale. The *p* and *o* sites are actually much smaller than the other regions and together comprise only 83 base pairs. (B) The *lac* operon in the repressed state. (C) The *lac* operon in the induced state. The inducer alters the shape of the repressor so that the repressor can no longer bind to the operator. The common abbreviations *i*, *p*, *o*, *z*, *y*, and *a* are used instead of *lacI*, *lacO*, and so on. The *lacA* gene is not essential for lactose utilization.

2. The products of the *lacZ* and *lacY* genes are coded by a single polycistronic mRNA molecule. (A third protein, encoded by *lacA*, is also translated from the mRNA. This protein is the enzyme transacetylase; it is used in the metabolism of certain β-galactosides other than lactose and will not be of further concern here.) The linked structural genes, together with *lacP* and *lacO*, constitute the *lac* **operon.**
3. The promoter mutations ($lacP^-$) eliminate the ability to synthesize *lac* mRNA.
4. The product of the *lacI* gene is a repressor, which binds to a sequence of nucleotides in the DNA constituting the repressor-binding site, which is called the *operator*.
5. When the repressor is bound to the operator, initiation of transcription of *lac* mRNA by RNA polymerase is prevented.
6. Inducers stimulate mRNA synthesis by binding to, and inactivating, the repressor. In the presence of an inducer, the operator is not bound by the repressor, and the promoter is available for the initiation of mRNA synthesis.

Note that regulation of the operon requires that the *lacO* operator either overlap or be adjacent to the promoter of the structural genes, because binding with the repressor prevents transcription. Proximity of *lacI* to *lacO* is not strictly necessary, because the *lacI* repressor is a soluble protein and is therefore diffusible throughout the cell. The presence of inducer has a profound effect on the DNA-binding properties of the repressor; the inducer–repressor complex has an affinity for the operator that is approximately 10^3 smaller than that of the repressor alone.

The ratio of the numbers of copies of β-galactosidase, permease, and transacetylase is 1.0 : 0.5 : 0.2 when the operon is induced. These differences are partly due to the order of the genes in the mRNA: Downstream cistrons are less likely to be translated owing to failure of reinitiation when an upstream cistron has finished translation.

The operon model is supported by a wealth of experimental data and explains many of the features of the *lac* system, as well as numerous other negatively regulated genetic systems in prokaryotes. One aspect of the regulation of the *lac* operon—the effect of glucose—has not yet been discussed. Examination of this feature indicates that the *lac* operon is also subject to positive regulation, as we will see in the next section.

Positive Regulation of the Lactose Operon

The function of β-galactosidase in lactose metabolism is to form glucose by cleaving lactose. (The other cleavage product, galactose, also is ultimately converted into glucose by the enzymes of the galactose operon.) Glucose is the preferred source of carbon and energy, and when both glucose and lactose are present in the growth medium, transcription of the *lac* operon does not take place, even in *lacI*$^-$ mutants, until virtually all of the glucose in the medium has been consumed. This observation reveals a second level of regulation of the lac operon, which is responsive to the concentration of glucose.

The inhibitory effect of glucose on expression of the *lac* operon is indirect, mediated by the small molecule *cyclic adenosine monophosphate* (cAMP) (**Figure 12.5**). Cyclic AMP is synthesized by the enzyme *adenyl cyclase*, and the concentration of cAMP is regulated indirectly by glucose metabolism. When bacteria are growing in a medium containing glucose, the cAMP concentration in the cells is quite low. In a medium containing glycerol (or any other carbon source that does not use the glycolytic pathway for its metabolism), or when the bacteria are starved of an energy source, the cAMP concentration is high (**Table 12.2**). Glucose levels help regulate the cAMP concentration in the cell, and cAMP regulates the activity of the *lac* operon as well as several other operons that control degradative metabolic pathways.

E. coli and many other bacterial species contain a protein, the *cyclic AMP receptor protein* (CRP), encoded by the gene *crp*. Mutations in either *crp* or the adenyl cyclase gene prevent synthesis of *lac* mRNA, which indicates that both CRP protein and cAMP are required for *lac* mRNA synthesis. CRP and cAMP bind to one another, forming a complex denoted **cAMP–CRP** that serves as a transcriptional activator that must bind DNA in the region of the *lac* promoter in order for transcription to occur (**Figure 12.6**). The requirement for cAMP–CRP is independent of the *lacI* repression system, because *crp*$^-$ and adenyl cyclase mutants are unable to make *lac* mRNA, even in *lacI*$^-$ or *lacO*^{c} mutants. Unlike the repressor, which negatively regulates the *lac* operon, the

Table 12.2 Concentration of cyclic AMP in cells growing in media with the indicated carbon sources

Carbon source	cAMP concentration
Glucose	Low
Glycerol	High
Lactose	High
Lactose + glucose	Low
Lactose + glycerol	High

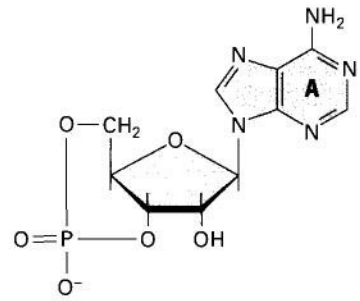


Figure 12.5 Structure of cyclic AMP.

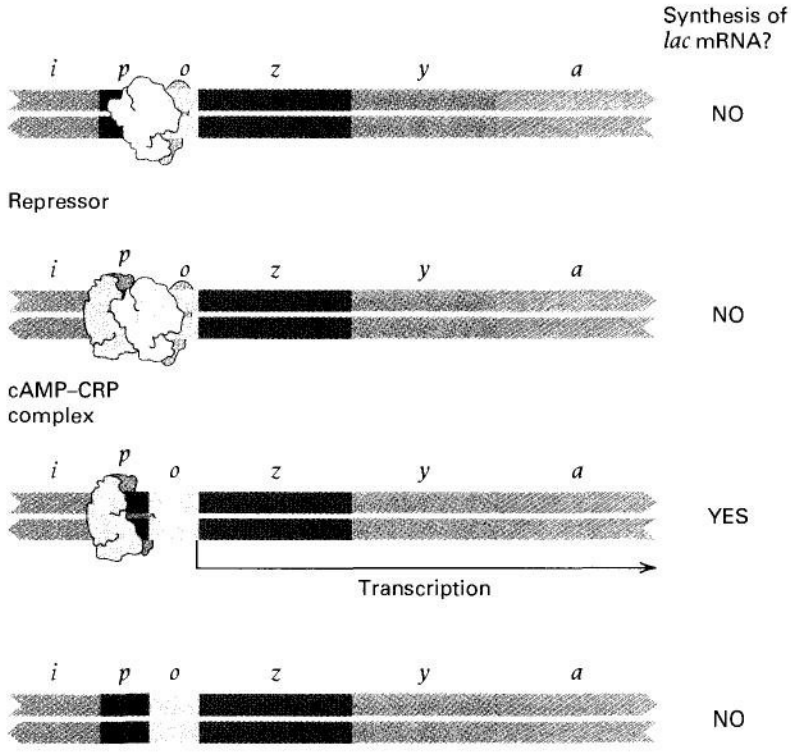


Figure 12.6 Four regulatory states of the *lac* operon. The *lac* mRNA is synthesized only if cAMP–CRP is present and the repressor is absent.

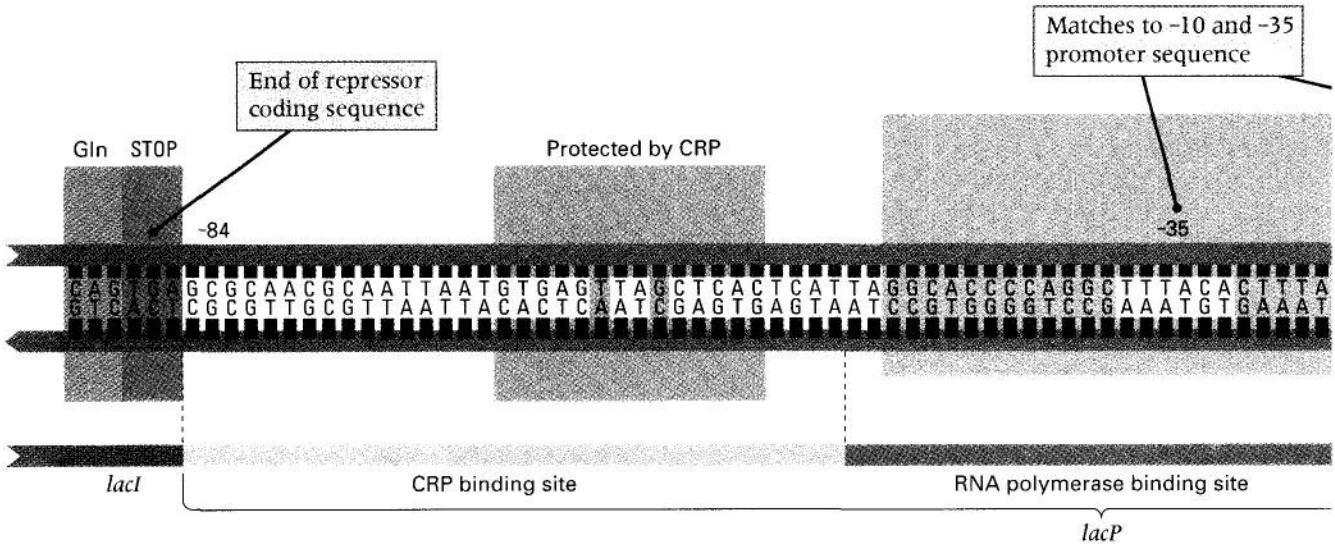


Figure 12.7 (above and facing page) The nucleotide sequence of the regulatory region of the *lac* operon, showing regions protected from DNase digestion by the binding of various proteins. The end of the *lacI* gene is shown at the extreme left; the ribosome binding site is the site at which the ribosome binds to the *lac* mRNA. The consensus sites for CRP binding and for RNA polymerase binding are indicated along the bottom.

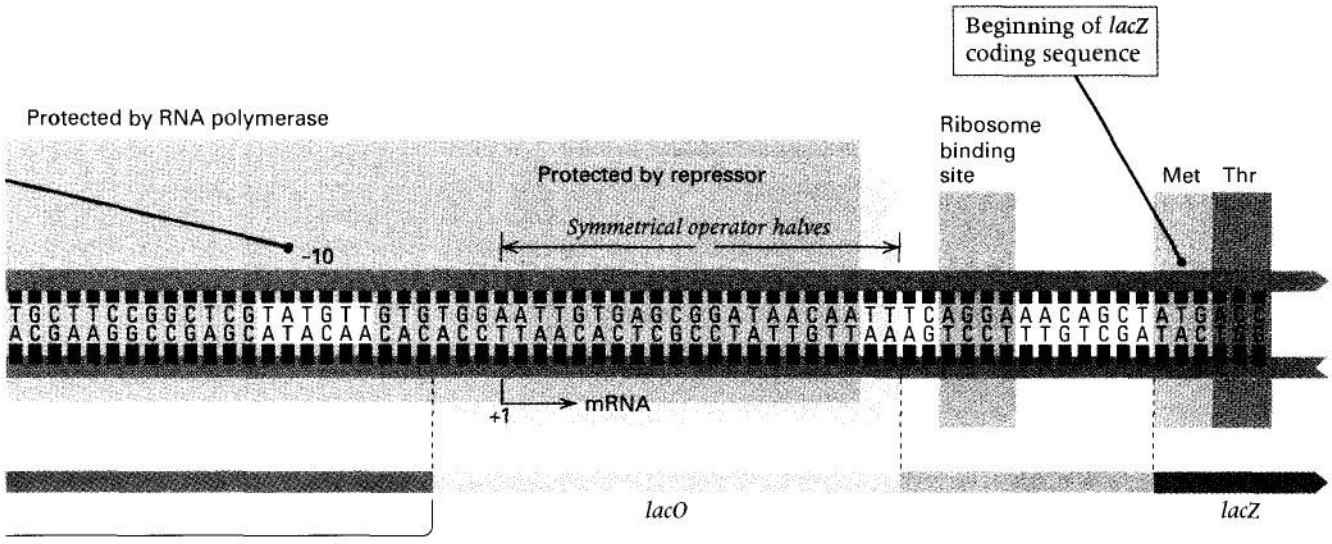
cAMP–CRP complex positively regulates the operon.

Experiments carried out *in vitro* with purified *lac* DNA, *lac* repressor, cAMP–CRP, and RNA polymerase have established two further points:

1. In the absence of the cAMP–CRP complex, RNA polymerase binds only weakly to the promoter, but its binding is stimulated when cAMP–CRP is also bound to the DNA. The weak binding rarely leads to initiation of transcription, because the correct interaction between RNA polymerase and the promoter does not occur.
2. If the repressor is bound to the operator, then RNA polymerase cannot stably bind to the promoter.

These results explain how lactose and glucose function together to regulate transcription of the *lac* operon. The relationship of these elements to one another, to the start of transcription, and to the base sequence in the region is depicted in **Figure 12.7.**

A great deal is also known about the three-dimensional structure of the regulatory states of the *lac* operon. **Figure 12.8** shows a 93-bp loop of DNA that forms in the operator region when it is in contact with the repressor. This loop corresponds to the *lac* operon region −82 to +11, numbering as in Figure 12.7. The DNA region in red corresponds, on the right-hand side, to the operator region centered at +11 and, on the left-hand side, to a second repressor-binding site immediately upstream of the CRP binding site. The *lac* repressor tetramer (violet) is shown bound to these sites. The DNA loop is formed by the region between the repressor-binding sites and includes, in medium blue, the CRP binding site, to which the CRP protein (dark blue) is shown bound. The DNA regions in green are the −10 and −35 sites in the *lacP* promoter indicated in Figure 12.7. In this configuration, the *lac* operon is not transcribed. Removal of the repressor opens up the loop and allows transcription to occur. Full repression of the *lac* operon requires the multiple repressor-binding sites (there is a third repressor-binding site around position +400 in the *lacZ* coding sequence); the operator at +11 identified in the classic genetic studies (Figure 12.7), by itself, accounts for only about 3 percent of the total repression.



12.3 Regulation of the Tryptophan Operon

The tryptophan (*trp*) operon of *E. coli* serves as an example of negative regulation in which transcription is repressible (Figure 12.1C). The *trp* operon contains structural genes for enzymes that synthesize the amino acid tryptophan. When adequate tryptophan is present in the growth medium, transcription of the operon is repressed, and when the supply of tryptophan is insufficient, transcription takes place.

Tryptophan is synthesized in five steps, each requiring a particular enzyme. The genes coding for these enzymes are adjacent and are in the same linear order in the *E. coli* chromosome as the order in which the enzymes function in the biosynthetic pathway. The genes are called *trpE*, *trpD*, *trpC*, *trpB*, and *trpA*, and the enzymes are translated from a single polycistronic mRNA molecule. The *trpE* coding region is the first one translated. Upstream (on the 5' side) of *trpE* are the promoter, the operator,

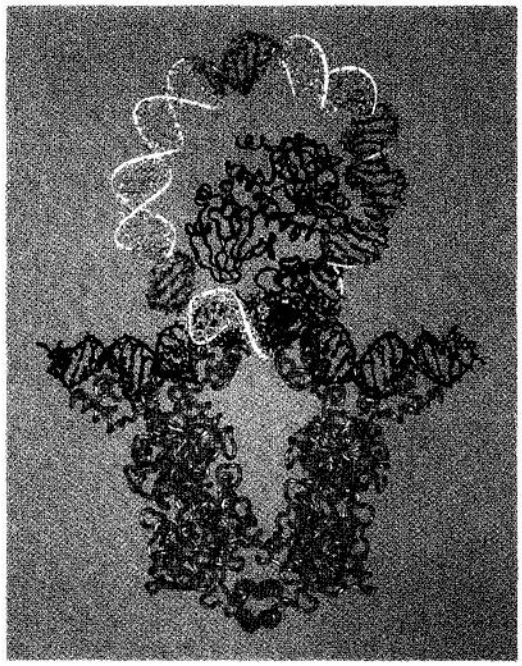


Figure 12.8 Structure of the *lac* operon repression loop. The *lac* repressor, shown in violet, binds to two DNA regions (red) consisting of the symmetrical operator region indicated in Figure 12.7 and a second region immediately upstream from the CRP binding site. Within the loop is the CRP binding site (medium blue), shown bound with CRP protein (dark blue). The −10 and −35 promoter regions are in green. [Courtesy of Mitchell Lewis. Reprinted with permission from M. Lewis, G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu, *Science* 271: 1247. Copyright 1996 American Association for the Advancement of Science.]

CoNnection

Operator? Operator?

François Jacob, David Perrin, Carmen Sanchez, and Jacques Monod 1960
Institute Pasteur, Paris, France
The Operon: A Group of Genes Whose Expression Is Coordinated by an Operator [original in French]

How is gene expression controlled? Before Jacob and Monod and their collaborators addressed this question experimentally, it was all a matter of speculation. Prior to this report, the researchers had previously discovered the i (lacI) *gene that controls expression of the β-galactosidase* (z) *and permease* (y) *genes needed for lactose utilization. They also had strong reason to believe that* lacI *produces a regulatory protein. How does the regulatory protein work? Here they give evidence that it works by directly binding to a DNA "operator" adjacent to the genes it regulates. Furthermore, the* z *and* y *genes are adjacent and are controlled coordinately by the same "operator" upstream from* z. *The discovery was immediately recognized as fundamental. Jacob and Monod, along with André Lwoff, were awarded the Nobel Prize in 1965. We now know that coordinate regulation via operons is restricted to bacteria. However, the underlying principle—that regulatory genes often control their target genes by direct binding to DNA—is valid for all organisms.*

The analysis of different bacterial systems leads to the conclusion that, in the synthesis of certain proteins, there is a dual genetic determination involving two types of genes with distinct functions: one (the gene for structure) is responsible for the structure of the protein molecule, and the other (the regulatory gene) governs the expression of the former through the intermediary action of a repressor. The regulatory genes that have so far been identified show the remarkable property of exercising a *coordinated effect*, each governing the expression of several genes for structure, closely linked together, and corresponding to enzyme proteins belonging to the *same biochemical pathway*. To explain this effect, it seems necessary to invoke a new type of genetic entity, called an "operator," which would be (a) adjacent to the group of genes and would control their activity; and (b) would be sensitive to the repressor produced by a particular regulatory gene. In the presence of the repressor, the expression of the group of genes would be inhibited through the mediation of the repressor. This hypothesis leads to some distinctive predictions concerning mutations that could affect the structure of the operator. (1) Certain mutations affecting an operator would be manifested by the loss of the capacity to synthesize the proteins determined by the group of linked genes "coordinated" by that operator. . . . (2) Other mutations, for example involving a loss of sensitivity (affinity) of the operator for the corresponding repressor, would be manifested by the constitutive synthesis of the proteins determined by the coordinated genes. . . . We have studied certain mutations affecting the metabolism of lactose in *Escherichia coli* that act simultaneously on the synthesis of β-galactosidase [the product of the *z* gene] and galactoside permease [the product of the *y* gene]. . . . The *i* gene is the regulatory gene synthesizing a repressor specific for the system. The genes *i*, *z* and *y* are closely linked. . . . Constitutive mutants (o^c) have now been isolated. [In partially diploid genotypes] only the allele of *z* or *y* that is *cis* with respect to o^c is constitutively expressed. . . . Other mutants have been isolated that have lost the ability to synthesize both the permease and the β-galactosidase. . . . These mutants are recessive. . . . Genetic analysis shows that these mutations (o^o) are extremely closely linked to the o^c mutations and that the order of the *lac* region is *i*–*o*–*z*–*y*. . . . The remarkable properties of the o^c and o^o mutations are inexplicable according to the "classical" concept of the genes for structure [*z* and *y*] and distinguish them equally from mutations affecting the regulatory gene *i*. On the other hand, they conform to the predictions arising from the hypothesis of the operator.

It seems necessary to invoke a new type of genetic entity, called an "operator," which would be (a) adjacent to the group of genes and would control their activity; and (b) would be sensitive to the repressor produced by a particular regulatory gene.

Source: Comptes Rendus des Séances de l'Academie des Sciences 250: 1727–1729. Translated in E. A. Adelberg, 1966. *Papers on Bacterial Genetics*. Boston: Little, Brown.

and two regions called the *leader* and the *attenuator*, which are designated *trpL* and *trpa* (not *trpA*), respectively (**Figure 12.9**). The repressor gene, *trpR*, is located quite far from this operon.

The product of the *trpR* gene is the aporepressor protein. Mutations in either this gene or the operator cause constitutive initiation of transcription of *trp* mRNA. The role of tryptophan is that of a co-repressor,

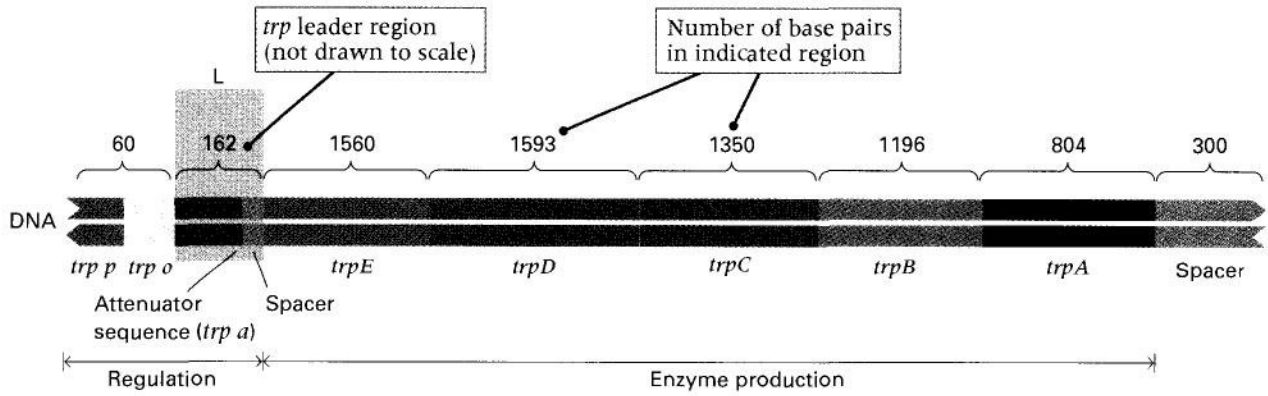


Figure 12.9 The *E. coli trp* operon. For clarity, the regulatory region is enlarged with respect to the coding region. The actual sizes of the regions are indicated by the numbers of base pairs. Region *L* is the leader.

which binds with the *trp* aporepessor to form the active repressor. Binding of the repressor to the *trp* operator shuts off synthesis. This basic on–off regulatory mechanism is outlined in **Figure 12.10**. When there is not enough tryptophan, the aporepressor adopts a three-dimensional conformation unable to bind with the *trp* operator, and the operon is transcribed (Figure 12.10A). On the other hand, when tryptophan is present at high enough concentration, some molecules bind with the aporepressor and cause it to change conformation into the active repressor. The active repressor

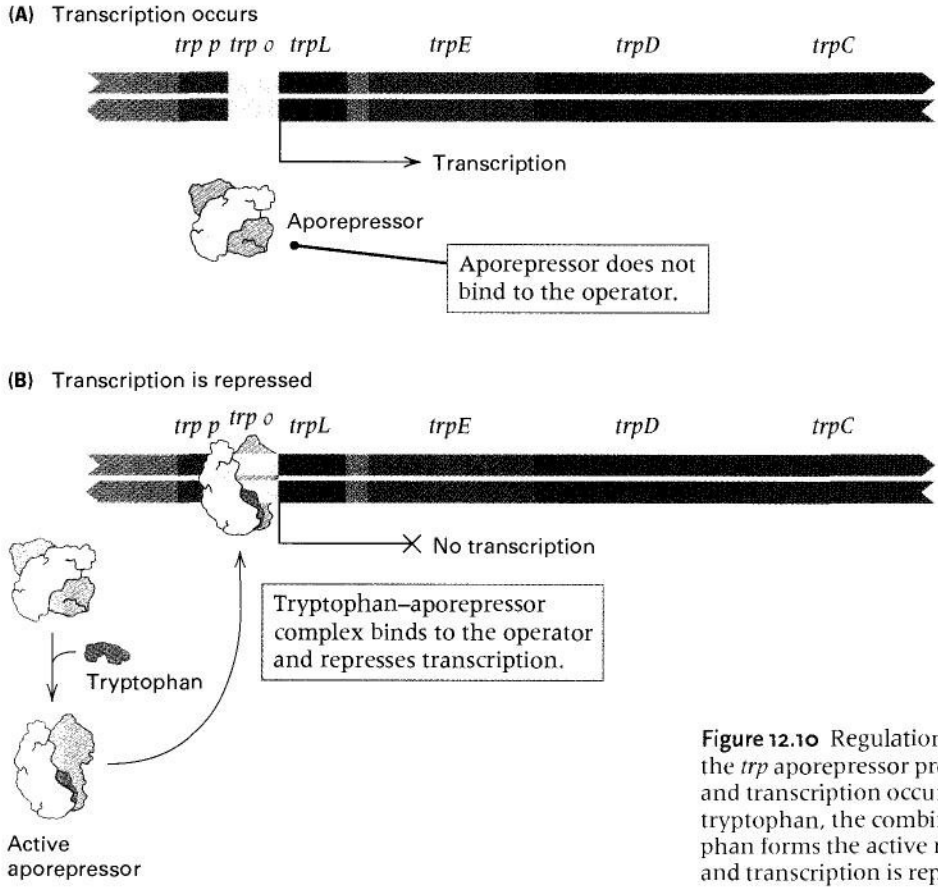


Figure 12.10 Regulation of the *E. coli trp* operon. (A) By itself, the *trp* aporepressor protein does not bind to the operator, and transcription occurs. (B) In the presence of sufficient tryptophan, the combination of aporepressor and tryptophan forms the active repressor that binds to the operator, and transcription is repressed.

binds with the *trp* operator and prevents transcription (Figure 12.10B). Thus only when tryptophan is present in sufficient amounts is the active repressor molecule formed.

The mechanism of repressible transcription in Figure 12.10 provides a coarse level of regulation of the *trp* operon in that transcription can be either "off" or "on." There is another level of regulation that provides for fine tuning, which is employed when some tryptophan is present in the growth medium, but not enough to sustain optimal growth. Under these conditions, it is advantageous for cells to synthesize tryptophan, but at less than the maximum possible rate. The *trp* operon responds to this situation by means of a regulatory mechanism in which the amount of transcription in the derepressed state is controlled quantitatively by the concentration of tryptophan in the cell. This regulatory mechanism is called *attenuation*, and it is found in many operons responsible for amino acid biosynthesis.

Attenuation

The **attenuation** mechanism controls whether transcription, once started, will continue through the operon or be terminated prematurely. In wildtype cells, transcription of the *trp* operon is often initiated. However, in the presence of even small amounts of tryptophan, most of the mRNA molecules terminate in a specific 28-base region within the leader sequence (Figure 12.9). The result of termination is an RNA molecule containing only 140 nucleotides that stops short of the genes coding for the *trp* enzymes. The 28-base region in which termination occurs is called the

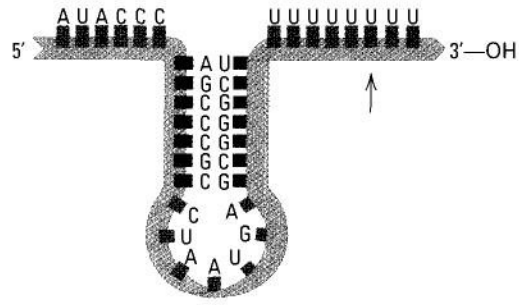


Figure 12.11 The terminal region of the *trp* attenuator sequence. The arrow indicates the final uridine in attenuated RNA. Nonattenuated RNA continues past this point. The nucleotides in red letters form a stem by base-pairing within the RNA.

attenuator. The nucleotide sequence of the attenuator region contains the usual features of a termination site, including a potential stem-and-loop configuration in the mRNA followed by a sequence of eight uridines (**Figure 12.11**).

In *E. coli*, termination of transcription is determined by whether a small peptide encoded in the leader sequence can be translated. This coding sequence, shown in **Figure 12.12**, specifies a **leader polypeptide** 14 amino acids in length, and it includes two adjacent tryptophan codons at positions 10 and 11. When there is sufficient charged tryptophan tRNA to allow translation of these codons, the nascent transcript adopts a conformation in which the attenuator is exposed, and transcription is terminated. On the other hand, when there is insufficient charged tryptophan tRNA to allow translation of the leader polypeptide, the ribosome stalls at the tryptophan

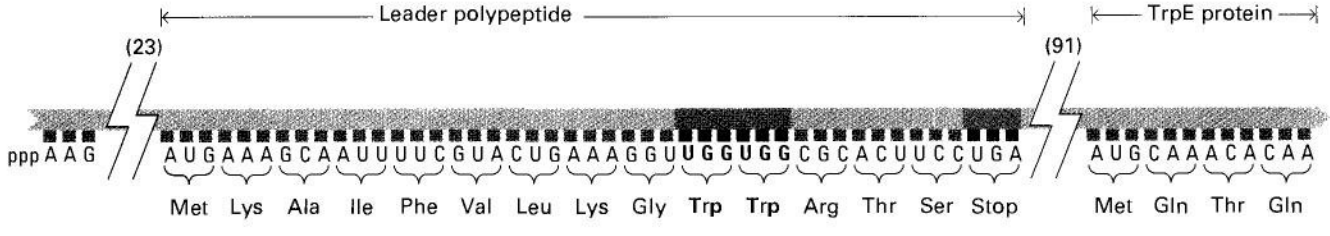
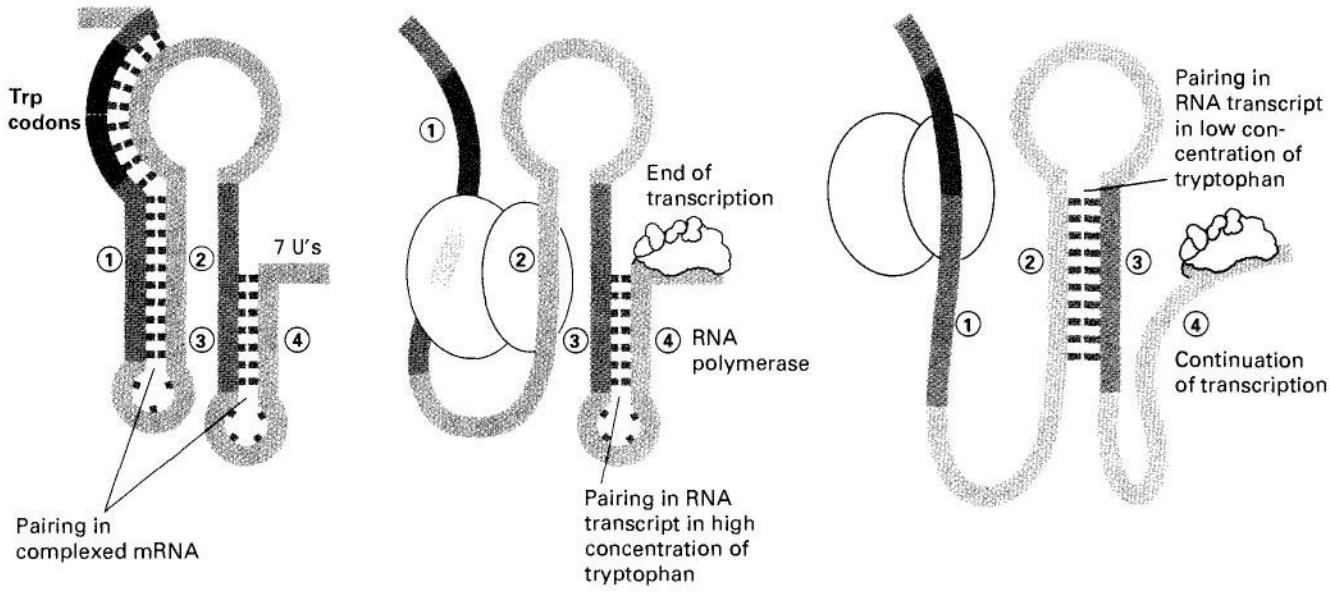


Figure 12.12 The sequence of nucleotides in the *trp* leader mRNA, showing the leader polypeptide, the two tryptophan codons (red), and the beginning of the TrpE protein. The numbers 23 and 91 are the numbers of nucleotides in the sequence that, for clarity, are not shown.

codons; in this case the attenuator is hidden, and transcription continues through the entire operon.

The mechanism of attenuation is diagrammed in **Figure 12.13**. Part A shows the leader RNA molecule, including the two tryptophan codons in the leader polypeptide. Region 2 has a nucleotide sequence that enables it to pair either with region 1 or with region 3. In the purified RNA, region 1 pairs with region 2, and region 3 pairs with region 4. Part B shows the configuration in a cell in which there is sufficient $tRNA^{Trp}$ to allow translation of the leader polypeptide. The ribosome translocates beyond the Trp codons and blocks region 2, so the pairing that forms is between region 3 and region 4; this creates the transcriptional terminator, with termination occurring at the run of uridines that follows region 4. Part C shows what happens when the ribosome stalls at the Trp codons as a result of insufficient $tRNA^{Trp}$. In this case, region 2 preferentially pairs with region 3, which disrupts the conformation of the terminator, allowing transcription to continue through the rest of the operon. The fine tuning of this system takes place at intermediate concentrations of tryptophan, when the fraction of nascent transcripts that are completed depends on how frequently translation is stalled, which in turn depends on the intracellular concentration of charged tryptophan tRNA.

Some bacteria have evolved a mechanism of regulation of the *trp* operon in which the intracellular concentration of tryptophan is detected in a manner that does not depend on translation. A particularly elegant example is found in *Bacillus subtilis*. In this oganism, the *trp* operon is not repressible. There is no aporepressor, nor is there a short polypeptide encoded in the leader region. Regulation of transcription is controlled solely through attenuation, which in this case is mediated by a protein denoted TRAP (*trp* RNA-binding attenuation protein). Each monomer of the TRAP protein binds a molecule of tryptophan and, in so doing, undergoes a confor-



(A) Free mRNA. Base pairs between 1 and 2 and between 3 and 4.

(B) High concentration of tryptophan. Ribosome reaches region 2 and pairing of 3–4 causes termination of transcription.

(C) Low concentration of tryptophan. Ribosome stalled in region 1 at Trp codons permits pairing of 2–3 and transcription is not terminated

Figure 12.13 The explanation for attenuation in the *E. coli trp* operon. The tryptophan codons are highlighted in red.

mational change that enables it to bind with other monomers. The active form of TRAP is a radially symmetric 11-mer, which binds specifically with nucleotide sequences in the nascent mRNA in the regions corresponding to 1 and 2 in Figure 12.13B. The resulting structure, with regions 1 and 2 wrapped around the outside of the TRAP 11-mer, is shown in **Figure 12.14**. When regions 1 and 2 are trapped in this manner, region 3 pairs with region 4, much as is shown in Figure 12.13B, forming the transcriptional terminator. In the absence of the TRAP multimer, the terminator is not formed and transcription continues. The system is sensitive to the intracellular concentration of tryptophan: If the level is high enough to allow TRAP multimers to form, transcription is terminated, but if it is too low for TRAP multimers to form, the complete *trp* operon is transcribed.

Many operons responsible for amino acid biosynthesis (for example, the leucine, isoleucine, phenylalanine, and histidine operons) are regulated by attenuators that function by forming alternative paired regions in the transcript. In the histidine operon, the coding region for the leader polypeptide contains seven adjacent histidine codons (**Figure 12.15**, part A). In the phenylalanine operon, the coding region for the leader polypeptide contains seven

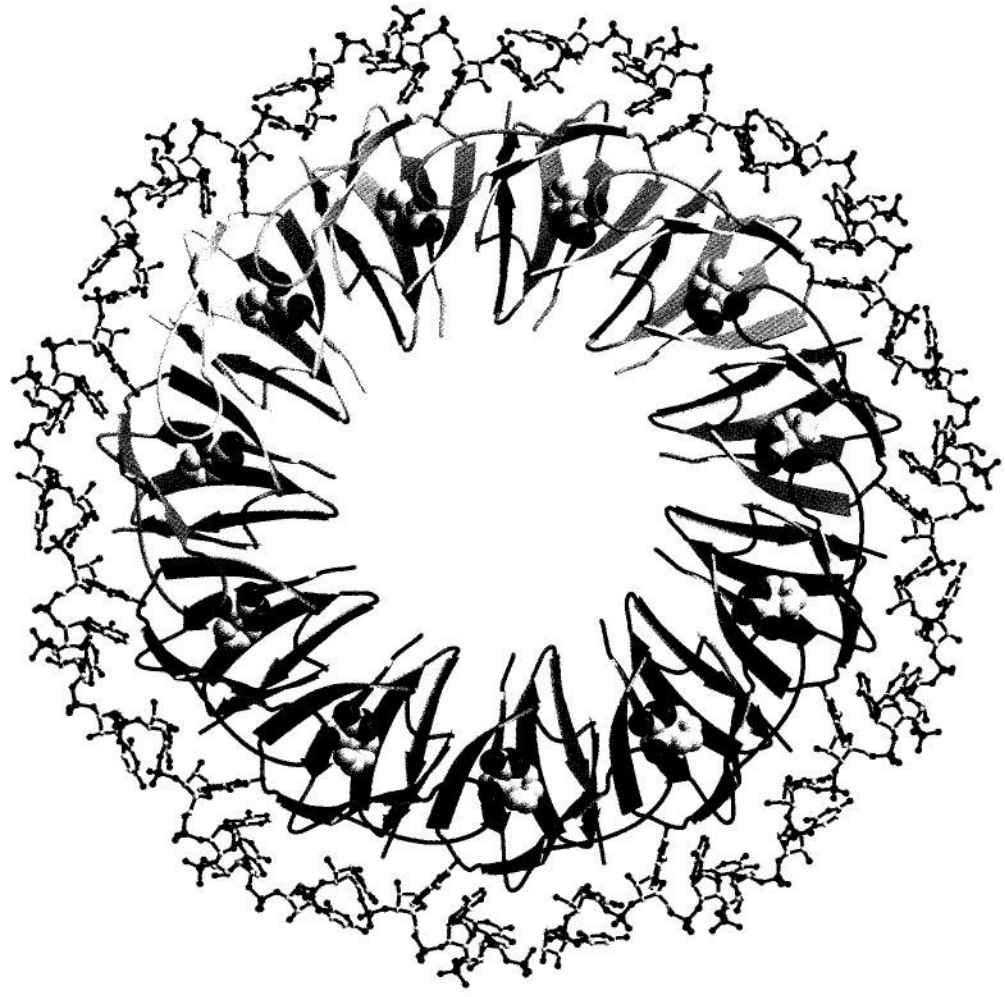


Figure 12.14 Structure of TRAP protein (the tryptophan RNA-binding attenuation protein from *B. subtilis*) bound to *trp* RNA. The protein is an 11-mer, each subunit is shown in a different color, and the RNA is shown in a ball-and-stick model. The solid spheres in the inside of the TRAP subunits are tryptophan molecules. [Courtesy of Alfred Antson and Paul Gollnick. Reprinted by permission from *Nature* 401:235, A. Antson, P. Gollnick, E. J. Dodson, G. Dodson, R. B. Greaves, X. P. Chen. Copyright 1999 Macmillan Magazines Ltd.]

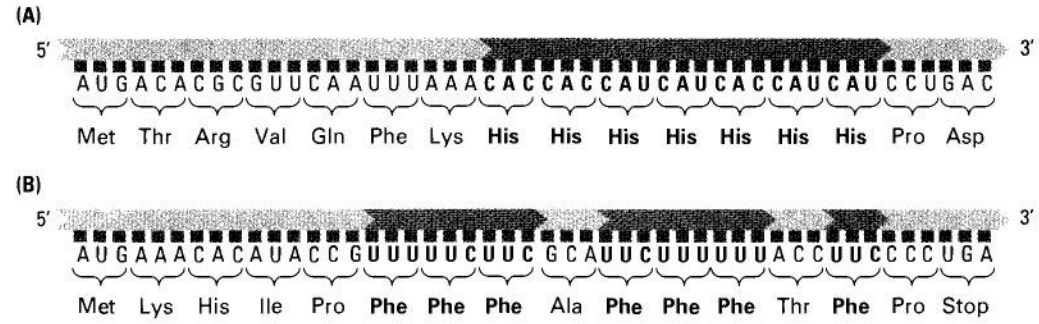


Figure 12.15 Amino acid sequence of the leader peptide and nucleotide sequence of the corresponding segment of mRNA from the histidine operon (A) and the phenylalanine operon (B). The repetition of these amino acids is emphasized in red.

phenylalanine codons divided into three groups (Figure 12.15B). This pattern, in which codons for the amino acid produced by enzymes of the operon are present at high density in the leader peptide mRNA, is characteristic of operons in which attenuation is coupled with translation. Through these codons, the cell monitors the level of aminoacylated tRNA charged with the amino acid that is the end product of each amino acid biosynthetic pathway.

Regulation of the *lac* and *trp* operons exemplifies some of the important mechanisms that control transcription of genes in prokaryotes. In the following section, we will see that similar mechanisms are used in the control of genes in bacteriophages.

12.4 Regulation in Bacteriophage Lambda

When Jacob and Monod first proposed the operon model featuring negative regulation by a repressor protein, they suggested that the model could account not only for regulation in inducible and repressible operons for metabolic enzymes but also for the lysogenic cycle of temperate bacteriophages. They proposed that λ bacteriophage was kept quiescent and prevented from replicating within bacterial lysogens by a repressor. This explanation ultimately proved to be correct, although the biochemical mechanism of attaining the repressed, lysogenic state is more complicated than they imagined.

When λ bacteriophage infects *E. coli*, each infected cell can undergo one of two possible outcomes:

1. A lytic infection, resulting in lysis and production of phage particles, or
2. A lysogenic infection, resulting in integration of the λ molecule into the *E. coli* chromosome and formation of a lysogen.

Because of this dichotomy, λ normally produces turbid (not completely clear) plaques on a lawn of *E. coli*. The initial infection and lysis produce a cleared region in the bacterial lawn, but a few lysogens grow within the cleared region, partially repopulating the cleared zone and producing a turbid plaque.

Mutations in regulatory genes in λ were first identified in phage mutants that give clear rather than turbid plaques. The mutants proved to fall into four classes: *λvir*, cI^-, cII^-, and $cIII^-$. The characteristics of these mutants are shown in **Table 12.3**. The genetic positions of the *cI*, *cII*, and *cIII*

Table 12.3 Characteristics of mixed infections containing several combinations of *λvir*, cI^-, cII^-, and $cIII^-$ mutants

Infecting phages	Clear or turbid plaques
1. $\lambda vir + \lambda^+$	Clear
2. $cI^- + cI^+$	Turbid
3. $cII^- + cII^+$	Turbid
4. $cIII^- + cIII^+$	Turbid

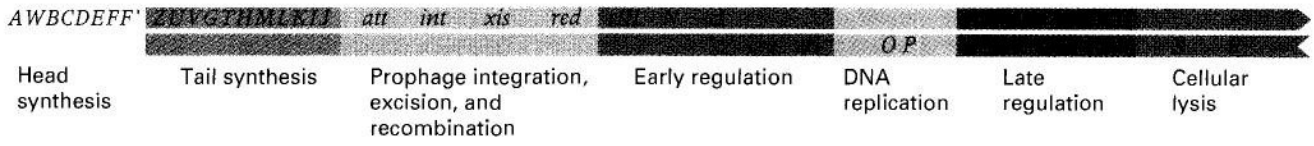


Figure 12.16 Genetic map of λ bacteriophage. The map is drawn to emphasize the functional organization of genes within the phage genome and to call attention to the regulatory features.

regions are shown in the simplified genetic map of λ bacteriophage in **Figure 12.16**, in which the genes are grouped by functional categories. Recall from Chapter 10 that upon infection, the λ DNA molecule circularizes, bringing the *R* and *A* genes adjacent to one another.

Among the mutants in Table 12.3, the "clear" mutants proved to be analogous to *lacI* and *lacO* mutants in *E. coli*. The *λvir* mutant is dominant to the wildtype λ^+ in mixed infection, as indicated by the combination of infecting phage designated 1 in Table 12.3; this combination of phage carries out a productive infection and prevents lysogeny by the wildtype λ^+. The *λvir* mutant is therefore analogous to the $lacO^c$ mutation. However, *λvir* proves to be a *double* mutant, bearing mutations in two different operators, O_L and O_R, as depicted in **Figure 12.17**. The cI^- mutations are recessive, as can be seen in entry 2 in Table 12.3. These cI^- mutations are analogous to $lacI^-$ mutations in that the cI^+ gene encodes the λ repressor, which is diffusible. The cII^+ and $cIII^+$ genes encode not for repressor but, rather, for proteins needed in establishing lysogeny. The cII^- and $cIII^-$ mutations are also recessive in mixed infections (entries 3 and 4 in Table 12.3).

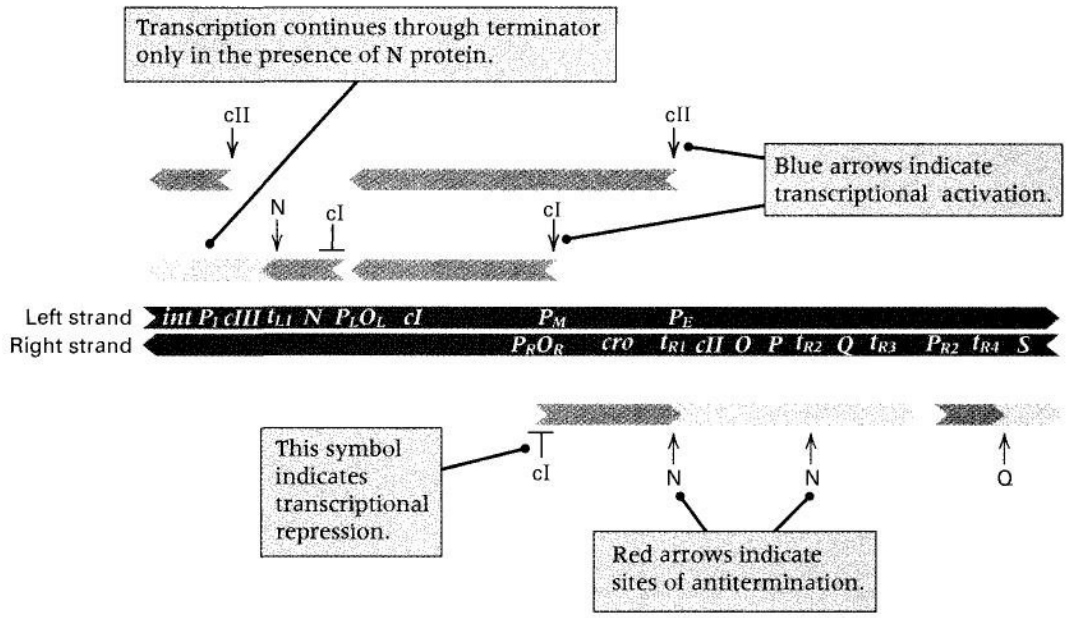


Figure 12.17 Genetic and transcriptional map of the control region of bacteriophage λ as expressed in the early stages of a lysogenic infection. The green arrows show the origin, direction, and extent of transcription. Light green arrows indicate portions of transcripts that are synthesized as a result of antitermination activity of the N or Q proteins. The sites of antitermination activity are indicated with red arrows pointing to the interior of a transcript. Blue arrows pointing to the origin of a transcript indicate transcriptional activation by cI or cII proteins, and the sites of transcriptional repression by cI protein are indicated.

The molecular basis on which the decision between the lysogenic and the lytic cycle is determined can be explained with reference to Figure 12.17. Upon infection of *E. coli* by λ, the λ molecule circularizes, and RNA polymerase binds at P_L and P_R and initiates transcription of the *N* and *cro* genes. The N protein acts to prevent termination of the transcripts from P_L and P_R, thus allowing production of cII protein; cII protein activates transcription at P_E and P_I, thus allowing production of the cI and int (integration) proteins. The cI protein shuts down further transcription from P_L and P_R and stimulates transcription at P_M, increasing its own synthesis in a positive autoregulatory loop. Lysogeny is achieved if the concentration of cI protein reaches a level high enough to prevent transcription from P_L and P_R and to allow the int protein to catalyze site-specific recombination between the circular λ molecule and the *E. coli* chromosome at their respective attachment (*att*) sites. The result is the formation of a λ lysogen.

The alternative pathway is lytic development, which takes place when the cro protein dominates. The cro protein also can bind to O_R and, in doing so, blocks transcription from P_M. If this occurs, then the concentration of repressor cannot rise to the level required to block transcription from P_L and P_R. Transcription continues from P_R and P_{R2}, and N and Q proteins prevent termination in the rightward (and also leftward) transcripts. Because the λ DNA molecule is in a circular configuration, rightward transcription moves through genes *S* and *R* and thence through the head and tail genes (*A* through *J* in Figure 12.16). The production of proteins needed for cellular lysis and formation of phage particles ensues, followed by phage assembly and cellular lysis to release phage.

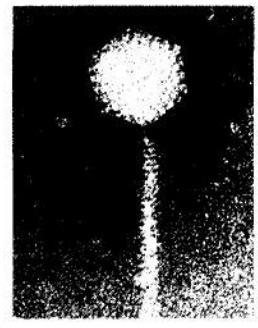
In effect, the regulation of lysogeny in λ involves a sort of "race" between the production of the cro and cI proteins: The "winner" is whichever protein first achieves a concentration high enough to dominate the outcome of infection. The underlying molecular mechanism is that the cro and cI proteins compete for binding to O_L and O_R (each operator has three protein-binding subsites that participate in the competition). If cro wins the contest, the lytic cycle results in that particular cell; if cI wins, a lysogen is formed. Hence

> The cI and cro proteins function as a *genetic switch*: cI turns on lysogeny, and cro turns on the lytic cycle.

Additional factors that determine whether cI or cro controls the outcome of a particular infection include the nutritional state of the infected cell. This in turn influences the stabilities and levels of the regulatory proteins. Thus the regulation of λ, an apparently "simple" system, is actually quite complex.



Bacteriophage T2



Bacteriophage λ

Plaques formed by bacteriophage λ, shown here at the right, are much smaller than those formed by bacteriophage T2, shown above. [Cabisco/Visuals Unlimited, Inc.]

12.5 Transcriptional Regulation in Eukaryotes

Many eukaryotic genes are **housekeeping genes** that encode essential metabolic enzymes or cellular components and are expressed constitutively at relatively low levels in all cells. Other genes, which differ in their expression according to cell type or stage of the cell cycle, are often regulated at the level of transcription. Typically, levels of expression of eukaryotic genes may differ 2 to 10 fold between the uninduced and induced levels, in contrast to the more dramatic differences seen in prokaryotes, in which the ratio between the uninduced and

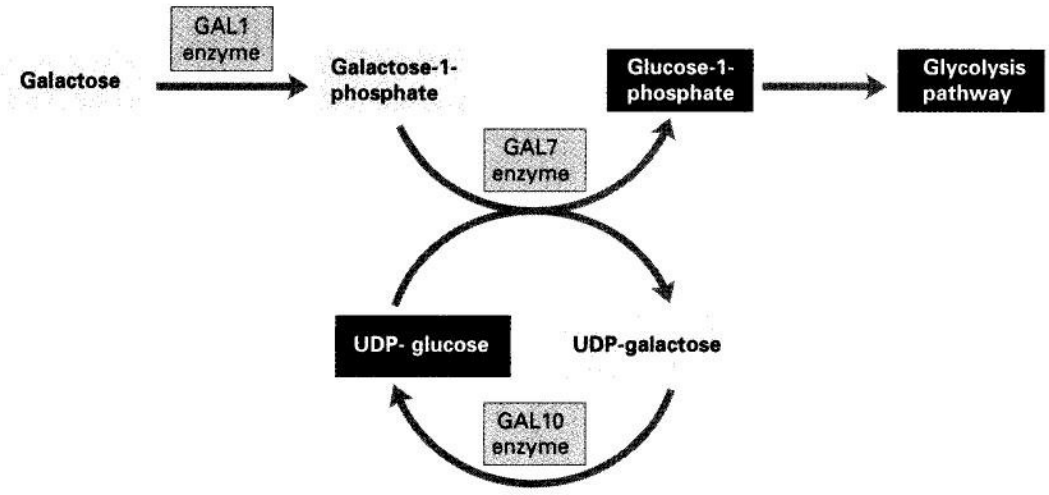


Figure 12.18 Metabolic pathway by which galactose is converted to glucose-1-phosphate in the yeast *Saccharomyces cerevisiae*.

induced levels may be as great as 1000 fold.

Galactose Metabolism in Yeast

We introduce transcriptional regulation in eukaryotes by examining the control of galactose metabolism in yeast and comparing it with the *lac* operon in *E. coli*. The first steps in the biochemical pathway for galactose degradation are illustrated in **Figure 12.18**. Three enzymes, encoded by the genes *GAL1*, *GAL7*, and *GAL10*, are required for conversion of galactose to glucose-1-phosphate. These three genes are tightly linked, as shown in **Figure 12.19**. Despite the tight linkage of the three genes, the genes are not part of an operon: The mRNAs are monocistronic. The *GAL1* and *GAL10* mRNAs are synthesized from divergent promoters lying between the genes, and *GAL7* mRNA is synthesized from its own promoter. The mRNAs are synthesized only when galactose is present as inducer; the genes are thus inducible.

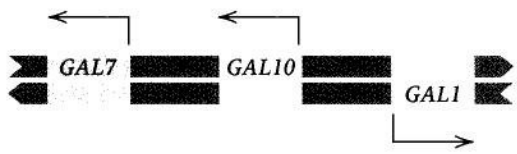


Figure 12.19 The linked *GAL* genes of *Saccharomyces cerevisiae*. Arrows indicate the transcripts produced. The *GAL1* and *GAL10* transcripts come from divergent promoters, *GAL7* from its own promoter.

Constitutive and uninducible mutants have been observed. In two types of mutants, *gal80* and *GAL81*^{c}, the mutants synthesize *GAL1*, *GAL7*, and *GAL10* mRNAs constitutively. Another type of mutant, *gal4*, is uninducible: It does not synthesize the mRNAs whether or not galactose is present. The characteristics of the mutants are shown in **Table 12.4**. The terms *cis* and *trans* are of no help in interpreting these results, because the regulatory genes are unlinked to the genes they regulate: *GAL1*, *GAL7*, and *GAL10* are on chromosome II, *GAL80* is on chromosome XIII, and *GAL4* and *GAL81* are on chromosome XVI.

The *gal80* mutation is recessive (entry 1 of Table 12.4). Thus, superficially, it behaves like a *lacI*$^{-}$ mutation. The wildtype *GAL80* allele does indeed encode a protein that negatively regulates transcription, and although the GAL80 protein is called a "repressor," it acts not by binding to an operator but by binding to a transcriptional activator protein and rendering it inactive. The activator that GAL80 binds to is the product of the *GAL4* gene, denoted GAL4.

The GAL4 protein is required for transcription of all three *GAL* genes. The *gal4* mutation is recessive, and in homozygous mutants the *GAL* genes are all uninducible. The GAL4 protein is a positive regulatory

Table 12.4 Characteristics of diploids containing various combinations of *gal80*, *gal4*, and *GAL81*^{c} mutations

Genotype	Synthesis of *GAL1*, *GAL7*, and *GAL10* mRNAs	Gal phenotype
1. *gal80 GAL1/ GAL80 GAL1*	Inducible	+
2. *gal4 GAL1/ GAL4 GAL1*	Inducible	+
3. *GAL81*^{c} *GAL1/ GAL81 GAL1*	Constitutive	+

protein that activates transcription of the three *GAL* genes individually, starting at a different site upstream from each gene. The GAL4 protein bound with its target site in the DNA is shown in **Figure 12.20**, in which the GAL4 protein (a dimer) is shown in blue and the DNA molecule in red. The small yellow spheres represent ions of zinc, which are essential components in the DNA binding.

The GAL80 "repressor" protein acts by binding to GAL4 protein and sequestering it so that GAL4 is not free to activate transcription. The inducer (galactose) binds with GAL80 protein and causes it to release GAL4, which then activates transcription.

The constitutive mutation, *GAL81*^{c}, is dominant; it results in constitutive synthesis of all three mRNAs. However, because it does not map near *GAL1*, *GAL7*, and *GAL10* and is a single mutation, it cannot be an operator mutation comparable to *lacO*^{c}. The mystery of *GAL81*^{c} was solved when it was shown that the mutation does not define a separate regulatory gene but, rather, maps to a position within the *GAL4* gene. The mutation yields a GAL4 protein unable to bind to GAL80 protein. Hence the mutant GAL4 protein cannot be sequestered by GAL80 and is free to activate transcription in the absence of galactose, whether or not wildtype GAL4 protein is also present.

The main point of these comparisons is that the superficial similarity between the constitutive and uninducible mutations in the *lac* operon of *E. coli* and the *GAL* genes of yeast is not indicative of similar molecular regulatory mechanisms. Some physiological similarities remain in that, in both prokaryotes and eukaryotes, the genes for a particular metabolic (or developmental) pathway are expressed in a coordinated manner in response to a signal. The principle at work is that alternative molecular mechanisms are often employed to achieve similar ends.

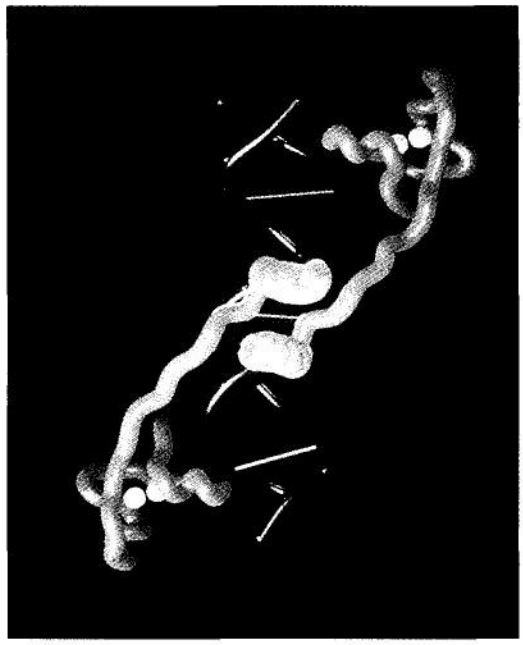


Figure 12.20 Three-dimensional structure of the GAL4 protein (blue) bound to DNA (red). The protein is composed of two polypeptide subunits held together by the coiled regions in the middle. The DNA-binding domains are at the extreme ends, and each physically contacts three base pairs in the major groove of the DNA. The zinc ions in the DNA-binding domains are shown in yellow. [Courtesy of Dr. Stephen C. Harrison. Reprinted by permission from *Nature* 356: 408–414, S. C. Harrison, R. Marmorstein, M. Carey, and M. Ptashne. Copyright 1992 Macmillan Magazines Ltd.]

Transcriptional Activator Proteins

The GAL4 protein is an example of a *transcriptional activator protein*, which must bind with an upstream DNA sequence in order to prepare a gene for transcription. Some transcriptional activator proteins work by direct interaction with one or more components in the transcription complex, and in this way they recruit the transcription complex to the promoter of the gene to be activated. Other transcriptional activator proteins may initiate transcription by an already assembled transcription complex. In either case, the activator proteins are essential for the transcription of genes that are positively regulated.

Many transcriptional activator proteins can be grouped into categories on the basis of characteristics shared by their amino acid sequences. For example, one category has a *helix–turn–helix* motif, which consists of a sequence of amino acids forming a pair of α-helices separated by a bend; the helices are so situated that they can fit neatly into the grooves of a double-stranded DNA molecule. The helix–turn–helix motif is the basis of the DNA-binding ability, although the sequence specificity of the binding results from other parts of the protein.

A second large category of transcriptional activator proteins includes a DNA-binding motif that is called a *zinc finger* because the folded structure incorporates a zinc ion. An already familiar example is the GAL4 transcriptional activator protein in yeast. The protein functions as a dimer composed of two identical GAL4 polypeptides oriented with their zinc-binding domains at the extreme ends (Figure 12.20 shows the zinc ions in yellow). The DNA sequence recognized by the protein is a symmetrical sequence, 17 base pairs in length, which includes a CCG triplet at each end that makes direct contact with the zinc-containing domains.

A more detailed illustration of the DNA-binding domain of the GAL4 protein is shown in **Figure 12.21**. Each of two zinc ions (Zn^{2+}) is chelated by bonds with four cysteine residues in characteristic positions at the base of a loop that extends for an additional 841 amino acids beyond those shown. The amino acids marked by a red asterisk are the sites of mutations that result in mutant proteins unable to activate transcription. Replacements at amino acid positions 15 (Arg → Gln), 26 (Pro → Ser), and 57 (Val → Met) are particularly interesting because they provide genetic evidence that zinc is necessary for DNA binding. In particular, the mutant phenotypes can be rescued by extra zinc in the growth medium, because the molecular defect reduces the ability of the zinc finger part of the molecule to chelate zinc. Extra zinc in the medium overcomes the defect and restores the ability of the mutant activator protein to attach to its particular binding sites in the DNA.

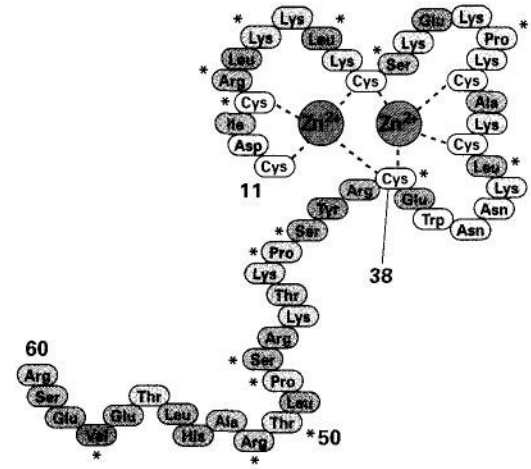


Figure 12.21 DNA-binding domain of the GAL4 transcriptional activator protein in yeast. The four cysteine residues bind a zinc ion and form a peptide loop called a zinc finger. The zinc finger is a motif commonly found in DNA-binding proteins. The amino acids marked by an asterisk have been identified via mutations as sites at which amino acid replacements can abolish the DNA-binding activity of the protein. The result is that the target genes cannot be activated.

Transcriptional Enhancers and Transcriptional Silencers

Some transcriptional activator proteins bind with particular DNA sequences known as **enhancers.** Enhancer sequences are typically rather short (usually fewer than 20 base pairs) and are found at a variety of locations around the gene they regulate. Most enhancers are upstream of the transcrip-

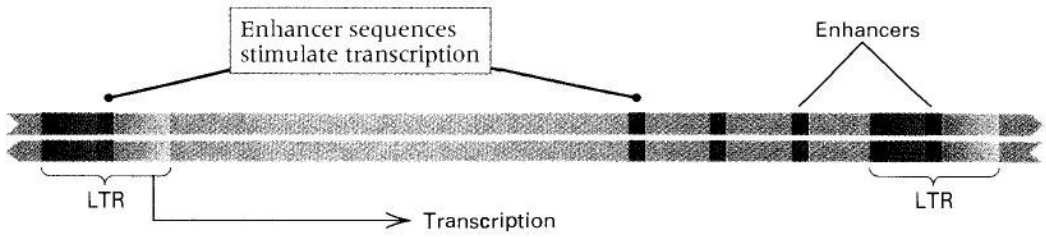


Figure 12.22 Positions, in the mouse mammary tumor virus, of enhancers (orange) that allow transcription of the viral sequence to be induced by glucocorticoid steroid hormone. LTR stands for the long terminal repeated sequences found at the extreme ends of the virus.

tional start site (sometimes many kilobases away), others are in introns within the coding region, and a few are even located at the 3′ end of the gene. One of the most thoroughly studied enhancers is in the mouse mammary tumor virus and determines transcriptional activation by the glucocorticoid steroid hormone. The consensus sequence of the enhancer is AGAQCAGQ, in which Q stands for either A or T. The virus contains five copies of the enhancer positioned throughout its genome (**Figure 12.22**), providing five binding sites for the hormone–receptor complex that activates transcription.

Enhancers are essential components of gene organization in eukaryotes because they enable genes to be transcribed only when proper transcriptional activators are present. Some enhancers respond to molecules outside the cell—lor example, steroid hormones that form receptor–hormone complexes. Other enhancers respond to molecules that are produced inside the cell (for example, during development), and these enhancers enable the genes under their control to participate in cellular differentiation or to be expressed in a tissue-specific manner. Many genes are under the control of several different enhancers, so they can respond to a variety of different molecular signals, both external and internal.

Some genes are also subjected to regulation by transcriptional **silencers,** which are short nucleotide sequences that are targets for DNA-binding proteins that, once recruited to the site, promote the assembly of large protein complexes that prevent transcription of the silenced genes. Examples of such silencing complexes are the *Drosophila* PcG (Polycomb group) proteins, which silence certain genes during development, and the complex in yeast that includes the Rap1p, SIR, ORC, and histone 4 proteins, which is involved in silencing the unexpressed mating-type cassettes (described in Section 12.9).

The Eukaryotic Transcription Complex

Many enhancers activate transcription by means of **DNA looping,** which refers to physical interactions between relatively distant regions along the DNA. The mechanism is illustrated in **Figure 12.23**. The factors necessary for transcription include a transcriptional activator protein that interacts with at least one protein subunit of the transcription complex to recruit the transcription complex to the gene.

The **basal transcription factors** are proteins in the transcription complex that are used widely in the transcription of many different genes. The basal transcription factors in eukaryotes have been highly conserved in evolution. A minimal set necessary for accurate transcription *in vitro* includes TFIIB, TFIID, TFIIE, TFIIF, TFIIH, Pol II, and (less critically) TFIIA. (TF in these designations stands for *transcription factor.*) These components can assemble *in vitro* in stepwise fashion on a promoter. The first step is recruitment of TFIID, itself a complex of proteins that includes a **TATA-box-binding protein (TBP),** which binds with the promoter in the region of the TATA box, and about 10 other proteins, called **TBP-associated factors (TAFs),** which are the components that respond specifically to activator proteins. The Pol II

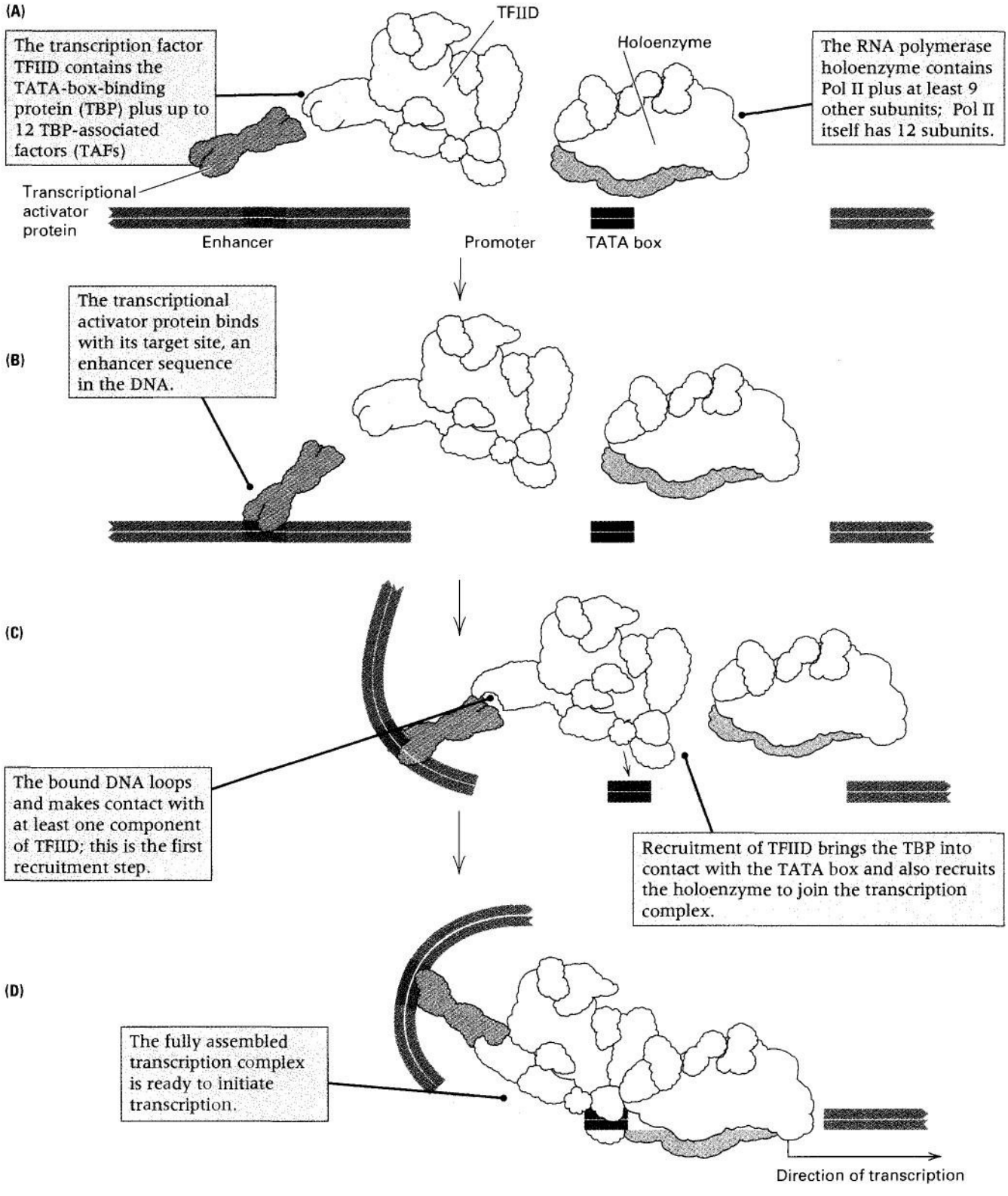


Figure 12.23 Transcriptional activation by recruitment. (A) Relationship between enhancer and promoter and the protein factors that bind to them. (B) Binding of the transcriptional activator protein to the enhancer. (C) Bound transcriptional activator protein makes physical contact with a subunit in the TFIID complex, which contains the TATA-box-binding protein, and attracts ("recruits") the complex to the promoter region. (D) The Pol II holoenzyme and any remaining general transcription factors are recruited by TFIID, and the transcription complex is fully assembled and ready for transcription. In the cell, not all of the Pol II is found in the holoenzyme, and not all of the TBP is found in TFIID. In this illustration, transcription factors other than those associated with TFIID and the holoenzyme are not shown.

RNA polymerase includes 12 protein subunits and is found in a complex called the **Pol II holoenzyme,** which also includes most of the basal transcription factors (TFIIB, TFIIF, and TFIIH) as well as other proteins.

It is not yet clear whether the transcription complex is recruited to the promoter and assembled stepwise, as *in vitro* studies suggest, or recruited in the form of one or more large preassembled complexes, which the composition of the Pol II holoenzyme suggests is the case. For simplicity, Figure 12.23 shows recruitment of one preassembled complex that includes TFIID, which in turn recruits the preassembled Pol II holoenzyme. To activate transcription (Figure 12.23B), the transcriptional activator protein binds to an enhancer in the DNA and to one of the TAF subunits in the TFIID complex. This interaction attracts ("recruits") the TFIID complex to the region of the promoter (Figure 12.23C). Attraction of the TFIID to the promoter also recruits the Pol II holoenzyme (Figure 12.23D), as well as any remaining general transcription factors, and once brought together, the transcriptional complex is ready for transcription to begin.

Experimental evidence for transcriptional activation by recruiting preassembled complexes has come from studies of a number of artificial proteins created by fusing a protein with a DNA-binding domain to one of the protein subunits in TFIID. Such fusion proteins act as transcriptional activators wherever they bind to DNA (provided that a promoter is nearby), because the TFIID is "tethered" to the DNA-binding domain and so the "recruitment" of TFIID is automatic. Similarly, fusion proteins that are tethered to a subunit of the holoenzyme can recruit the holoenzyme to the promoter. In this case, TFIID and the remaining general transcription factors are also attracted to the promoter, and the transcriptional complex is assembled. These experiments suggest that a transcriptional activator protein can activate transcription by interacting with subunits of either the TFIIA complex or the holoenzyme.

As Figure 12.23 suggests, the fully assembled transcription complex in eukaryotes is a very large structure. A real example, taken from early development in *Drosophila,* is shown in **Figure 12.24**. In this case, the enhancers, located a considerable distance upstream from the gene to be activated, are

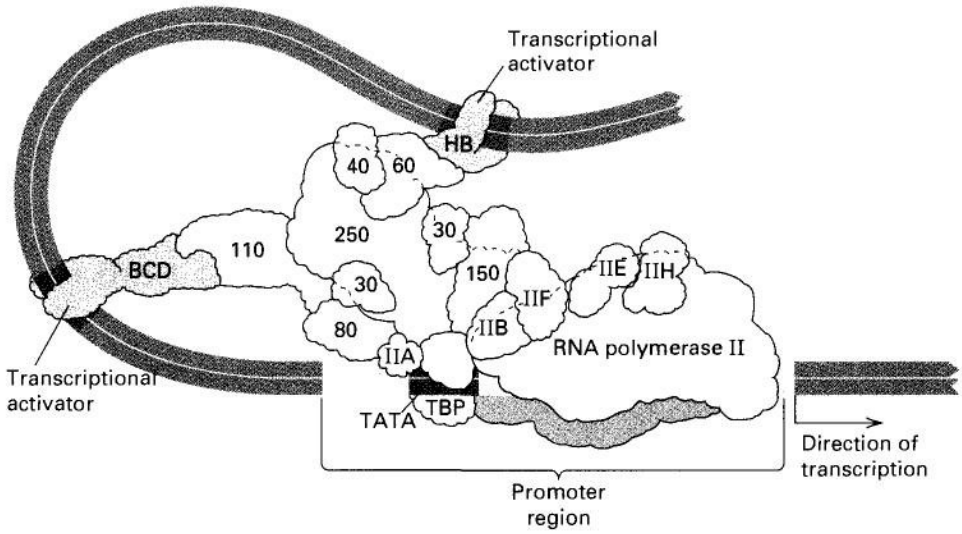


Figure 12.24 An example of transcriptional activation during *Drosophila* development. The transcriptional activators in this example are bicoid protein (BCD) and hunchback protein (HB). The numbered subunits are TAFs (TBP-associated factors) that, together with TBP (TATA-box-binding protein) correspond to TFIID. BCD acts through a 110-kilodalton TAF, and HB through a 60-kilodalton TAF. The transcriptional activators act via enhancers to cause recruitment of the transcriptional apparatus. The fully assembled transcription complex includes TBP and TAFs, RNA polymerase II, and general transcription factors TFIIA, TFIIB, TFIIE, TFIIF, and TFIIH.

bound by the transcriptional activator proteins BCD and HB, which are products of the genes *bicoid (bcd)* and *hunchback (hb)*, respectively; these transcriptional activators function in establishing the anterior–posterior axis in the embryo. (Early *Drosophila* development is discussed in Chapter 14.) Note the position of the TATA box in the promoter of the gene. The TATA box binding is the function of the TBP. The functions of a number of other components of the transcription complex have also been identified. For example, the TFIIH contains both helicase and kinase activity to melt the DNA and to phosphorylate RNA polymerase II.

Phosphorylation allows the polymerase to leave the promoter and elongate mRNA. The looping of the DNA effected by the transcriptional activators is an essential feature of the activation process. Transcriptional activation in eukaryotes is a complex process, especially when compared to the prokaryotic RNA polymerase, which consists of only six polypeptide chains ($\beta\beta'\alpha_2\omega\sigma$).

Chromatin-Remodeling Complexes

Eukaryotic DNA is typically found in the form of chromatin packaged with nucleosomes (Chapter 8). Special mechanisms are required for transcriptional activator proteins and the transcription complex to acquire access to the DNA. The existence of such mechanisms is implied by the observation that the components of transcription sufficient to transcribe purified DNA *in vitro* are unable to initiate transcription of purified chromatin. The nucleosomes in chromatin must prevent the transcription complex from either binding to DNA or using it as a template.

Several different multiprotein complexes have been identified that can restructure chromatin and that enable it to be transcribed. These are known as **chromatin-remodeling complexes (CRCs).** Among them are the SWI–SNF and RSC complexes in yeast and the NURF, CHRAC, and ACF complexes in *Drosophila*. SWI–SNF is named after the mutations, *swi* and *snf*, that led to its discovery, and the other abbreviations stand for *remodels the structure of chromatin* (RSC), *nucleosome remodeling factor* (NURF), *chromatin accessibility complex* (CHRAC), and *ATP-dependent chromatin assembly and remodeling factor* (ACF). All of these complexes use energy derived from ATP to restructure chromatin, and although the detailed mechanisms of chromatin remodeling are unknown, the complexes do differ in their activities. For SWI–SNF, for example, secondary mutations that suppress *swi* and *snf* map to the histone genes, the presence of the SWI–SNF complex renders the DNA in mononucleosomes susceptible to cleavage by DNAse, and purified SWI–SNF can restructure chromatin *in vitro*. RSC has similar, but not identical, properties. Furthermore, SWI–SNF is found as an integral component of the Pol II holoenzyme, and it enables the holoenzyme to disrupt nucleosomes. In contrast, NURF facilitates the binding of transcription factors to chromatin, CHRAC acts on nucleosomal spacing and increases the sensitivity of the DNA in chromatin to endonuclease activity, and ACF helps assemble properly spaced nucleosomal arrays and facilitates the binding of transcriptional activator proteins. Similar kinds of chromatin-remodeling complexes have been identified in higher eukaryotes, and they are probably present in all eukaryotes.

The molecular mechanism of chromatin remodeling is unknown, and because there are several distinct types of CRCs, there may be several mechanisms. In one general class of models, the CRC disrupts nucleosome structure without displacing the nucleosomes, rendering the DNA accessible to transcriptional activator proteins, the TATA-box-binding protein, and other components of the transcription complex. In another general class of models, the CRC repositions the nucleosomes along the DNA, making key DNA-binding sites accessible. An example is illustrated in **Figure 12.25**. Part A shows a transcriptionally inactive chromatin conformation, with the DNA-binding sites for a transcriptional activator protein (TAP) and TATA-box-binding protein (TBP) sequestered in nucleosomes and unavailable. Recruitment of a CRC to the site results in repositioning of the nucleosomes (part B), which renders the binding sites accessible (part C). In this chromatin configuration, TAP and TBP can bind with the DNA and recruit the rest of the transcription complex.

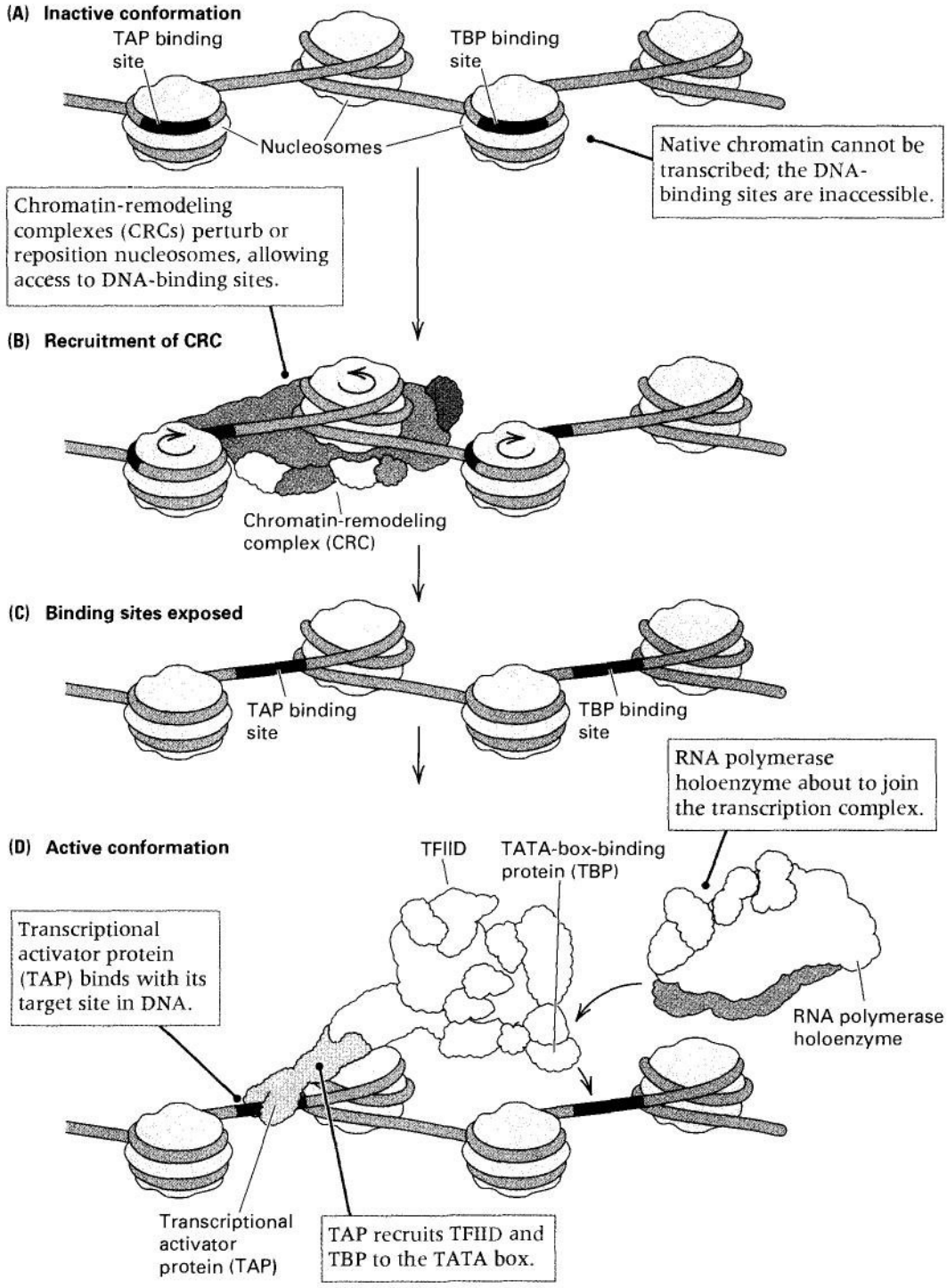


Figure 12.25 Function of chromatin-remodeling complexes. (A) Native chromatin may conceal key DNA-binding sites. (B) A chromatin-remodeling complex either repositions the nucleosomes along the DNA or chemically modifies the histones. (C) DNA-binding sites become accessible. (D) The transcription complex is recruited to the site.

Alternative Promoters

Some eukaryotic genes have two or more promoters that are active in different cell types. The different promoters result in different primary transcripts that contain the same protein-coding regions. An example from *Drosophila* is shown in **Figure 12.26**. The gene codes for alcohol dehydrogenase, and its organization in the genome, shown in part A, includes three protein-coding regions interrupted by two introns. Transcription in larvae (Figure 12.26B) uses a different promoter from that used in transcription in adults (Figure 12.26C). The adult transcript has a longer 5' leader sequence, but most of this sequence is eliminated in RNA splicing. Alternative promoters make possible the independent regulation of transcription in larvae and adults.

The versatility of some enhancers results from their ability to interact with two different promoters in a competitive fashion; that is, at any one time, the enhancer can stimulate one promoter or the other, but not both. An example of this mechanism is illustrated in **Figure 12.27**, in which *P1*

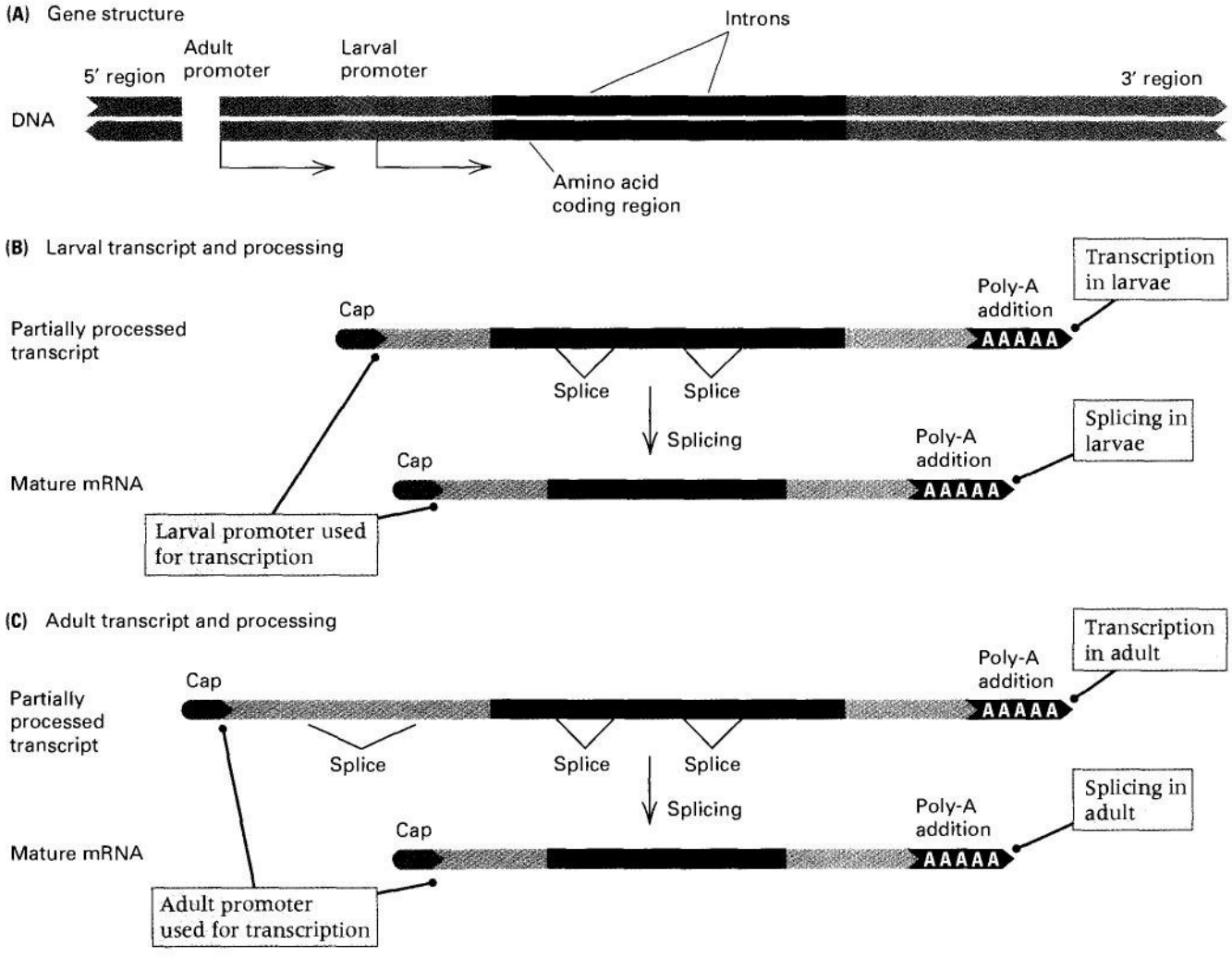


Figure 12.26 Use of alternative promoters in the gene for alcohol dehydrogenase in *Drosophila*. (A) The overall gene organization includes two introns within the amino acid coding region. (B) Transcription in larvae uses the promoter nearest the 5' end of the coding region. (C) Transcription in adults uses a promoter farther upstream, and much of the larval leader sequence is removed by splicing.

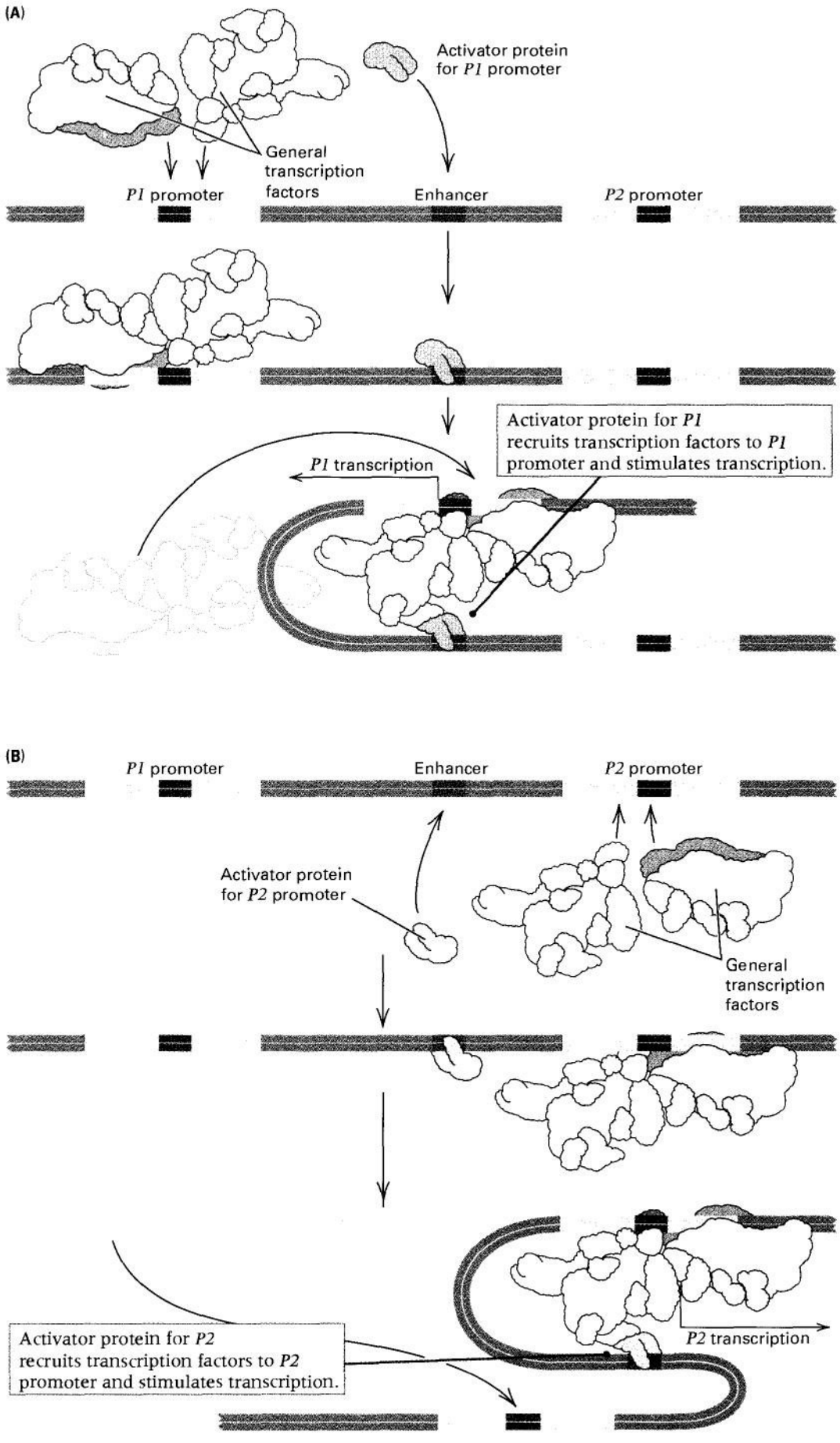


Figure 12.27 Genetic switching regulated by competition for an enhancer. Promoters *P1* and *P2* compete for a single enhancer located between them. When complexed with an appropriate transcriptional activator protein, the enhancer binds preferentially with promoter *P1* (A) or promoter *P2* (B). Binding to the promoter recruits the transcription complex. If either promoter is mutated or deleted, then the enhancer binds with the alternative promoter. The location of the enhancer relative to the promoters is not critical.

and *P2* are alternative promoters that compete for an enhancer located between them. When complexed with a transcriptional activator specific for promoter *P1*, the enhancer binds preferentially with promoter *P1* and stimulates transcription (Figure 12.27A). When complexed with a different transcriptional activator specific for promoter *P2*, the enhancer binds preferentially with promoter *P2* and stimulates transcription from it (Figure 12.27B). In this way, competition for the enhancer serves as a switch mechanism for expression of the *P1* or *P2* promoter. This regulatory mechanism is present in chickens and results in a change from the production of embryonic β-globin to that of adult β-globin in development. In this case, the embryonic globin gene and the adult gene compete for a single enhancer, which in the course of development changes its preferred binding from the embryonic promoter to the adult promoter. In human beings, enhancer competition appears to control the developmental switch from the fetal γ-globin to that of the adult β-globin polypeptide chains. In persons in whom the β-globin promoter is deleted or altered in sequence and unable to bind with the enhancer, there is no competition for the enhancer molecules, and the γ-globin genes continue to be expressed in adult life when normally they would not be transcribed. Adults with these types of mutations have fetal hemoglobin instead of the adult forms. The condition is called *high-F disease* because of the persistence of fetal hemoglobin, but the clinical manifestations are very mild.

12.6 Epigenetic Mechanisms of Transcriptional Regulation

The prefix *epi* means "besides" or "in addition to"; **epigenetic** therefore refers to heritable changes in gene expression that are due not to changes the DNA sequence itself, but to something "in addition to" the DNA sequence, usually either chemical modification of the bases, or protein factors bound with the DNA. In this section we discuss some examples of epigenetic regulation. We shall see that there is a great deal yet to be learned about the molecular mechanisms by which epigenetic modifications are established and maintained.

Cytosine Methylation

In most higher eukaryotes, a proportion of the cytosine bases are modified by the addition of a methyl (CH_3) group to the number-5 carbon atom (**Figure 12.28**). The cytosines are incorporated in their normal, unmodified form in the course of DNA replication, and then the methyl group is added by an enzyme called **DNA methylase.** In mammals, cytosines are modified preferentially in 5'-CG-3' dinucleotides. Many mammalian genes have CG-rich regions upstream of the coding region that provide multiple sites for methylation; these are called **CpG islands,** where the "p" represents the phosphate group in the polynucleotide backbone.

When a CG dinucleotide that is methylated in both strands undergoes DNA replication, each of the resulting daughter molecules contains one parental strand with a methylated CG and one newly synthesized strand with an unmethylated CG. The DNA methylase recognizes the hall-methylation in these molecules and methylates the cytosines in the newly synthesized strands. Methylation of CG dinucleotides in the sequence CCGG can be detected by the use of the restriction enzymes *Msp*I and *Hpa*II. Both enzymes cleave the sequence CCGG. However, *Msp*I cleaves regardless of whether the interior C is methylated, whereas *Hpa*II cleaves only

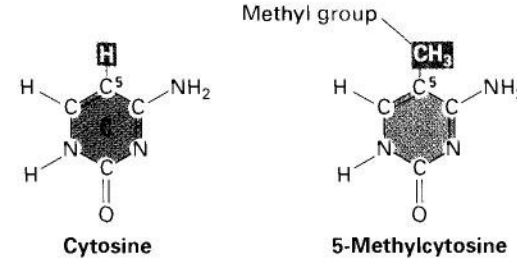


Figure 12.28 Structures of cytosine and 5-methylcytosine.

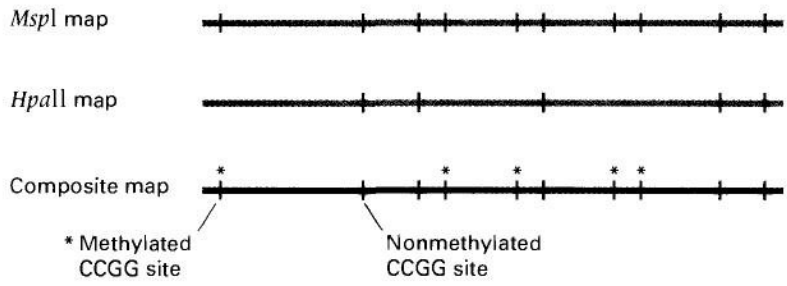


Figure 12.29 Detection of methylated cytosines in CCGG sequences by means of restriction enzymes. The enzyme *Msp*I cleaves all CCGG sites regardless of methylation, whereas *Hpa*II cleaves only nonmethylated sites. The positions of the methylated sites are determined by comparing the restriction maps.

unmethylated DNA. Therefore, *Msp*I restriction sites that are not cleaved by *Hpa*II are sites at which the interior C is methylated (**Figure 12.29**).

Methylation and Transcriptional Inactivation

A number of observations suggest that heavy methylation is associated with genes for which the rate of transcription is low. One example is the inactive X chromosome in mammalian cells, which is extensively methylated. In fact, in adult mammals, the majority of CpG dinucleotides in all chromosomes are methylated in somatic cells. The unmethylated CpGs are usually associated with the promoters of active housekeeping genes. The widespread methylation of inactive genes in adult somatic cells is thought to minimize accidental, low-level transcription from them.

Methylation is also widespread in the genomes of higher plants, where it is also correlated with transcriptional silencing. One example is found in the *Ac* transposable element in maize. Certain *Ac* elements lose activity of the transposase gene without any change in DNA sequence. These elements prove to have heavy methylation in a region particularly rich in the CG dinucleotides. Return to normal activity of the methylated *Ac* elements coincides with loss of methylation through the action of demethylating enzymes in the nucleus.

Methylation is also associated with transcriptional silencing in fungi. For example, in the phenomenon of *RIP (repeat-induced point mutation)* in *Neurospora*, discussed in Chapter 7, repeated sequences are subjected to hypermutation, and many of the C-bearing nucleotides in the repeats become methylated, thereby inhibiting transcription. A related process affecting repeated sequences occurs in the sexual cycle in the fungi *Ascobolus immersus* and *Coprinus cinereus*. This process is called **methylation induced premeiotically (MIP),** and it results in heavy methylation of cytosines in repeated sequences, leading to transcriptional silencing. In all of these examples, the methylation can be maintained through cell division; however, the mechanism of maintenance is not well understood.

Although there is a very strong correlation between heavy methylation and transcriptional silencing, heavy methylation may result from an earlier epigenetic signal that marks a gene for silencing and that recruits the methylase. If there is such an earlier signal, then it implies that methylation is the result of gene inactivity as well as its mechanism. In any case, treatment of cells with the cytosine analog *azacytidine* reverses methylation and can restore transcriptional activity. For example, in cell culture, some lineages of rat pituitary tumor cells express the gene for prolactin, whereas other related lineages do not. The gene is methylated in the nonproducing cells but is not methylated in the producers. Reversal of methylation in the nonproducing cells via azacytidine results in prolactin expression.

Cosuppression Through Transcriptional Silencing

Methylation is also implicated in a unique type of gene silencing in higher plants known as **transcriptional cosuppression.** This is a phenomenon in which extra copies of a gene result in the transcriptional silencing—not only of the extra copies but of all copies present in the genome. Transcriptional cosuppression was originally discovered in attempts to genetically engineer plants by introducing extra copies of genes into the genome using techniques that are discussed in Chapter 13. Flower color in petunia (*Petunia hybrida*) provides an excellent example. The normal red or purple flower color results from a flavonoid pigment known as *anthocyanin*, which is synthesized via a metabolic pathway in which the rate-limiting step is catalyzed by the enzyme *chalcone synthase* (CHS) encoded in the gene *Chs*. Given these facts, it is reasonable to suppose that an extra copy of the *Chs* gene (a *transgene*) would increase the level of CHS enzyme and thus the amount of pigment, thereby yielding a darker flower color. In the actual experiment, the flower color of the genetically engineered plants was much lighter! The reason was soon shown to be reduced transcription: In the transgenic plant, the total level of *Chs* mRNA was about 50-fold lower than that in control plants. Furthermore, in crosses, the reduced pigmentation pattern segregated along with the *Chs* transgene. In progeny lineages in which the transgene was missing, the level of pigmentation gradually returned to normal over several generations. Clearly, the presence of the *Chs* transgene somehow results in epigenetic transcriptional silencing of all copies of *Chs* present in the genome. Furthermore, the silenced *Chs* genes are heavily methylated. The mechanism of transcriptional cosuppression is not known, but it is known that homology between the promoters of the endogenous genes and the transgene is necessary. (Plants also exhibit a type of post-transcriptional silencing, discussed below, that does not require promoter homology.)

A type of transcriptional cosuppression has also been identified in *Drosophila*, but in this case the silencing cannot be due to methylation because *Drosophila* DNA is not methylated (the organism has no DNA methylase). In this case, the transcriptional silencing is associated with the PcG (Polycomb group) of transcriptional silencer proteins, which are recruited to the endogenous genes as well as to the homologous transgene.



Geneticists were quite surprised when they added additional copies of the *chalcone synthase* gene to petunias because instead of making the flowers darker, it made them lighter, and in extreme cases white. This was the discovery of cosuppression. This plot has wildtype and mutant flowers growing together. [Bill Kamin/Visuals Unlimited.]

Genomic Imprinting in the Female and Male Germ Lines

Mammals feature an unusual type of epigenetic silencing known as **genomic imprinting**, a process with the following characteristics:

- Imprinting occurs in the germ line.
- It affects at most a few hundred genes (many of them located in clusters).
- It is accompanied by heavy methylation (though the primary signal for imprinting is unknown).
- Imprinted genes are differentially methylated in the female and male germ lines.
- Once imprinted and methylated, a silenced gene remains transcriptionally inactive during embryogenesis.
- Imprints are erased early in germ-line development, then later re-established according to sex-specific patterns.

Although mammalian gametes are extensively methylated, most of the DNA is demethylated in preimplantation development, except for imprinted genes that retain their sex-specific patterns of methylation. The embryonic DNA is remethylated beginning after implantation, gradually attaining the heavy methylation levels found in adult somatic cells. In the germ line, the original imprints are erased when the DNA is globally demethylated, and remethylation takes place later in germ-line development. All remethylated genes acquire identical patterns of methylation in the germ line of both sexes, except for those few that have sex-specific patterns of imprinting and differential methylation. The imprinted genes undergo methylation during oocyte growth prior to ovulation in females, and in males probably around the time of birth. Because the methylation associated with imprinting is retained throughout embryonic development, any gene that is imprinted in either the female or the male germ line has, effectively, only one active copy in the embryo.

The epigenetic, sex-specific gene silencing associated with imprinting is dramatically evident in a pair of syndromes characterized by neuromuscular defects, mental retardation, and other abnormalities. These are *Prader–Willi syndrome* and *Angelman syndrome*. Both conditions are associated with rare, spontaneous deletions that include chromosomal region *15q11*. If the deletion takes place in the father, the result is Prader–Willi syndrome, whereas if it takes place in the mother, the result is Angelman syndrome. The reason is that *15q11* includes at least three genes (*SNRPN*, *necdin*, and *UBE3A*) that are imprinted and differentially methylated in the gametes. **Figure 12.30**A shows the pattern of imprinting of these three genes in a normal embryo. *SNRPN* and *necdin* are imprinted in the egg, *UBE3A* in the sperm. In the

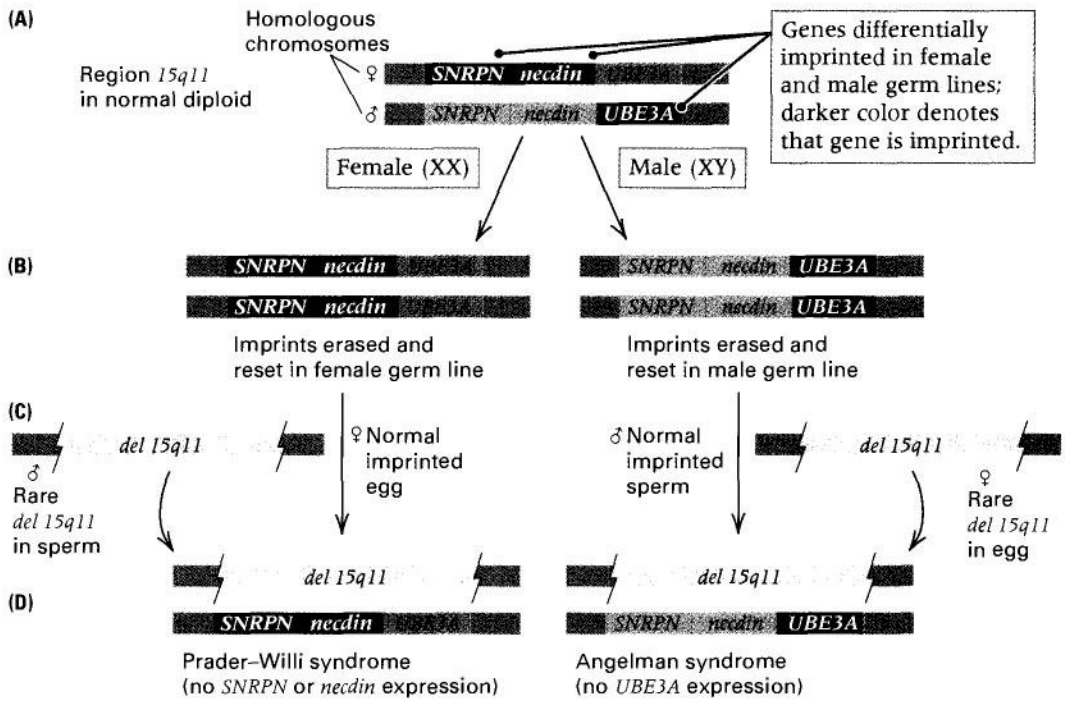


Figure 12.30 Imprinting of genes in chromosomal region *15q11* results in different neuromuscular syndromes, depending on which parent contributes a *15q11* deletion and which an imprinted chromosome. (A) Pattern of imprinting in a normal diploid. The maternal chromosome is at the top, the parental chromosome at the bottom. Imprinted and transcriptionally inactive genes are indicated. (B) In the germ line, the imprints are erased and reset in either female-specific or male-specific patterns. (C) An individual who inherits a maternally imprinted chromosome along with a *15q11* deletion has Prader–Willi syndrome, whereas one who inherits a paternally imprinted chromosome along with a *15q11* deletion has Angelman syndrome. Other genes in the region, not shown, may also be imprinted.

embryo, therefore, *UBE3A* is transcriptionally active in the maternal chromosome, and *SNRPN* and *necdin* in the paternal chromosome. In the germ line of female and male embryos, shown in Figure 12.30B, the imprints are erased and reset according to sex: In the female both homologs have *SNRPN* and *necdin* imprinted, whereas in the male both homologs have *UBE3A* imprinted. If a normal, imprinted female gamete is fertilized by a sperm with a *15q11* deletion, the embryo has no transcriptionally active copy of either *SNRPN* or *necdin* and develops Prader–Willi syndrome. On the other hand, if a normal, imprinted male gamete fertilizes an egg with a *15q11* deletion, the embryo has no transcriptionally active copy of *UBE3A* and develops Angelman syndrome. These syndromes demonstrate not only the epigenetic control of gene expression by imprinting but also differential imprinting in the sexes and the clustering of imprinted genes in the genome.

Why is there imprinting? One suggestion is that it evolved in early mammals with polyandry (each female mating with a series of males). In such a situation, it is to a male's benefit to silence genes that conserve maternal resources at the expense of the fetus, because this strategy maximizes the father's immediate reproduction. But it is to a female's benefit to silence genes that allocate resources to the fetus at the expense of the mother, because this strategy maximizes the female's long-term reproduction. This hypothesis is supported by the fact that some imprinted genes do affect the allocation of resources between mother and fetus in the direction that would be predicted. On the other hand, many genes that are imprinted have no obvious connection to maternal–fetal conflict.

12.7 Regulation Through RNA Processing and Decay

Although transcriptional control of gene expression is of cardinal importance, transcription is by no means the only level at which gene activity can be regulated. In this section we consider some mechanisms that act at the level of primary-transcript splicing or at the level of mRNA stability.

Alternative Splicing

Even when the same promoter is used to transcribe a gene, different cell types can produce different quantities of the protein (or even different proteins) because of differences in the mRNA produced in processing. The reason is that the same transcript can be spliced differently from one cell type to the next. The different splicing patterns may include exactly the same protein-coding exons, in which case the protein is identical, but the rate of synthesis differs because the mRNA molecules are not translated with the same efficiency. In other cases, the protein-coding part of the transcript has a different splicing pattern in each cell type, and the resulting mRNA molecules code for proteins that are not identical even though they share certain exons.

The α-amylase gene in the mouse provides an example of alternative splicing in the salivary gland and the liver, resulting in mRNAs that are translated with different efficiencies even though they contain the same open reading frame. The initial part of the primary transcript is shown in **Figure 12.31**. The coding sequence begins 50 base pairs inside exon 2 and is formed by joining exon 3 and subsequent exons. In the salivary gland, the primary transcript is processed such that exon S is joined with exon 2 (that is, exon L is removed as part of introns 1 and 2). In the liver, exon L is joined with exon 2, and exon S is removed along with intron 1 and the leader L. The exons S and L become alternative 5′ leader sequences of the amylase mRNA, and the alternative mRNAs are translated at different rates so that the mouse salivary gland produces more of the enzyme than the liver.

Messenger RNA Stability

A short-lived mRNA produces fewer protein molecules than a long-lived mRNA, so features that affect the rate of mRNA stability affect the level of gene expression. One route of degradation is the *deadenylation-dependent pathway*, which begins with enzymatic trimming of length of the poly-A tail on the mRNA. When the poly-A tail is trimmed to a length of 25–60 nucleotides, the mRNA becomes susceptible to a decapping enzyme that removes the 5′ cap and

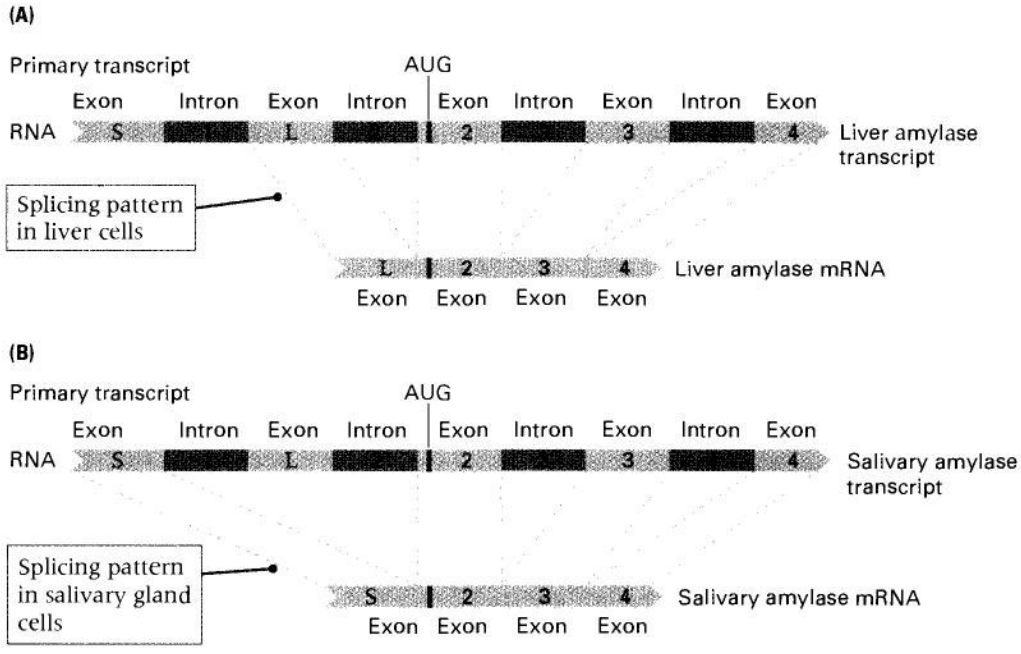


Figure 12.31 Production of distinct amylase mRNA molecules by different splicing events in cells of the salivary gland and liver of the mouse. The leader and the introns are distinguished from the exons by color. Exons S and L form part of the untranslated 5' end of the mRNA in salivary glands and liver, respectively. The coding sequence begins at the AUG codon in exon 2. (A) Splicing in the liver. Exon L is joined with exons 2, 3, and 4. (B) Splicing in the salivary gland. Exon S is joined with exons 2, 3, and 4.

renders the molecule unable to initiate translation; from this state the mRNA is rapidly degraded by exonucleases. An alternative pathway is the *deadenylation-independent pathway,* which is initiated either with decapping or with endonuclease cleavage of the mRNA, after which digestion goes to completion by exonuclease activity. The deadenylation-independent pathway is particularly active for mRNAs that contain early chain-termination codons or unspliced introns, and it prevents the accumulation of truncated polypeptides in the cell.

Cosuppression Through RNA Turnover

We have already examined transcriptional cosuppression, in which the presence of a transgene introduced by genetic engineering results in methylation-associated transcriptional silencing of both the transgene and endogenous copies of the gene. This type of cosuppression is mediated by homology between the promoter sequences of the silenced genes. Plants also exhibit another type of cosuppression in which the genes are actively transcribed, but no RNA accumulates because the transcripts are rapidly degraded. This type of gene silencing is called **posttranscriptional cosuppression;** it is mediated by homology between the transcripts of the silenced genes, and it is highly sensitive to promoter strength and gene dosage, which suggests that the silencing is triggered by the accumulation of transcripts above some threshold level. Remarkably, in grafts between plants in which silencing occurs and host plants in which it does not, the silencing is transmissible from graft cells to host cells, and it moves progressively through host tissues via the phloem, possibly through the movement of the RNA molecules that trigger the silencing.

The molecular mechanism of posttranscriptional gene silencing is unknown, but there are some ideas about why it evolved. A key observation is that a plant RNA virus that carries a sequence homologous to a transgene can be the target of posttranscriptional silencing. For example, a potato virus X genome that includes a beta-glucuronidase gene is able to infect wildtype plants but not plants that also carry a beta-glucuronidase transgene. This

suggests that posttranscriptional silencing (and probably also transcriptional cosuppression) originally evolved as a molecular mechanism to defend against invading viruses and transposable elements that are transcribed above a threshold level.

RNA Interference (RNAi)

Probably related to posttranscriptional gene silencing is a phenomenon known as **RNA interference (RNAi),** in which the introduction of a few hundred nucleotide pairs of *double-stranded RNA* triggers degradation of RNA transcripts containing homologous sequences. Originally discovered in *C. elegans*, RNAi has been shown to occur in a wide variety of organisms, including trypanosomes, planaria, *Drosophila*, and amphibians. Although the silenced genes are apparently transcribed at the normal level, the transcripts are degraded before they leave the nucleus. The effect is highly specific and requires only a few double-stranded RNA molecules per cell to trigger the silencing. As in the case of posttranscriptional cosuppression discussed in the previous section, the RNAi effect can be transmitted from cell to cell.

The discovery of RNAi has created a great deal of excitement in genetics because of its potential in research as well as practical applications. In effect, RNAi affords a method of producing the equivalent of knockout mutations in organisms without well-developed systems of mutagenesis and genetic manipulation.

12.8 Translational Control

Because transcription and translation are uncoupled in eukaryotes, gene expression can be regulated at the level of translation separately from transcription. The principal types of translational control are

- Inability of an mRNA molecule to be translated except under certain conditions
- Regulation of the overall rate of protein synthesis
- Aborted translation of the principal open reading frame because of the presence of a smaller open reading frame upstream in the mRNA

An important example of translational regulation is that of activating previously untranslated cytoplasmic mRNAs. This mechanism is prominent in early development, when newly fertilized eggs synthesize at a rapid rate many new proteins, virtually all of which derive from preexisting cytoplasmic mRNAs. In a few cases the molecular mechanism of mRNA activation is known. For example, in *Drosophila*, the mRNAs for the genes *bicoid*, *Toll*, and *torso* become activated because of the cytoplasmic elongation of their poly-A tail.

A dramatic example of translational control is the extension of the lifetime of silk fibroin mRNA in the silkworm. During cocoon formation, the silk gland predominantly synthesizes a single type of protein, silk fibroin, in very large amounts. The amount of fibroin is increased by three different mechanisms. First, the silk-gland cells become highly polyploid, accumulating thousands of copies of each chromosome. Second, transcription of the fibroin gene is initiated at a strong promoter, which results in the creation of about 10^4 fibroin mRNA molecules per gene copy in a period of a few days. Third, the fibroin mRNA molecule has a very long lifetime. In contrast to a typical eukaryotic mRNA molecule, which has a lifetime of about 3 hours, fibroin mRNA survives for several days, during which each mRNA molecule is translated repeatedly to yield 10^5 fibroin molecules. Thus each fibroin gene copy yields about 10^9 protein molecules in the few days during which the cocoon is formed.

12.9 Programmed DNA Rearrangements

The introduction of changes in DNA sequence is an uncommon mechanism of gene regulation, but there are a number of important examples. In some cases the changes are reversible, but in others they are permanent. A permanent change in DNA sequence implies that in the cell lineage in which it takes place, the genotype becomes permanently altered. Such irreversible changes take place only in somatic cells, not in the germ line, so they are not genetically transmitted.

Gene Amplification

In some cases, the number of gene copies increases temporarily by a process of **gene amplification.** An example is found in the development of the oocytes of the toad *Xenopus laevis*. The formation of an egg from its oocyte precursor is a complex process that requires a high rate of protein synthesis and consequently a great many (about 10^{12}) ribosomes. Although they are already tandemly duplicated to about 600 copies, the number of rRNA genes is insufficient to produce so many ribosomes. In the development of the oocyte, the number of rRNA gene copies increases to about 2.4 million (a 4000-fold amplification). This amplification takes place over a 3-week period in oocyte development and comes about from the formation of a large number of extrachromosomal, replicating rolling circles of the rDNA genes. The molecular mechanism of excision of the circles from the chromosome and formation of the rolling circles is not known. When the oocyte is mature, no more rRNA needs to be synthesized, and the extrachromosomal copies of the rRNA cease to replicate and begin to be degraded. Amplification of rRNA genes during oogenesis occurs in many organisms, including insects, amphibians, and fish.

Antibody and T-Cell Receptor Variability

Perhaps the most impressive examples of programmed DNA rearrangement take place in the bone marrow–derived (B) cells and thymus–derived (T) cells that play key roles in the vertebrate immune system. B cells produce and secrete circulating antibodies that combine with foreign antigens and mark them for destruction, whereas T cells have specific receptors that recognize foreign antigens at the cell surface. Both antibodies and T-cell receptors have highly variable amino acid sequences from one B cell or T cell to the next, and much of this variability is generated by DNA rearrangements that assemble a relatively small number of coding sequences into a very large number of possible combinations, one in each B cell or T cell. We will examine the process of DNA splicing that generates diversity in antibodies. A similar process accounts for the variability of T-cell receptors.



Normal T-cell function requires the enzyme adenosine deaminase (ADA). These two girls suffer from severe combined immunodeficiency disease (SCID) resulting from a mutant form of ADA. They were among the first persons to receive gene therapy. Doctors removed some of the white blood cells from each girl, cultured them, added genetically engineered viruses containing working copies of the ADA gene, then treated each girl with her own engineered cells. Years after the initial therapy, each girl still has a functional subset of T cells that produce ADA. [Ted Thai/Time Magazine.]

A normal mammal is capable of producing more than 10^8 different antibodies, only a fraction of which are produced at any one time. Each B cell can produce a single type of antibody, but the antibody is not secreted until the cell has been stimulated by the appropriate antigen. Once stimulated, the B cell undergoes successive mitoses and eventually produces a lineage of genetically identical cells that secrete the antibody. Antibody secretion may continue even if the antigen is no longer present.

There are five distinct classes of antibodies known as IgG, IgM, IgA, IgD, and IgE (Ig stands for **immunoglobulin**). These classes perform specialized functions in the immune response and exhibit certain structural differences. However, each contains two types of polypeptide chains differing in size: A large subunit called the *heavy (H) chain* and a small subunit called the *light (L) chain*. Immunoglobulin G (IgG) is the most abundant class of antibodies and has the simplest molecular structure (**Figure 12.32**). An IgG molecule consists of two heavy and two light chains held together by disulfide bridges (two covalently joined sulfur atoms) and has the overall shape of the letter Y. The sites for antibody

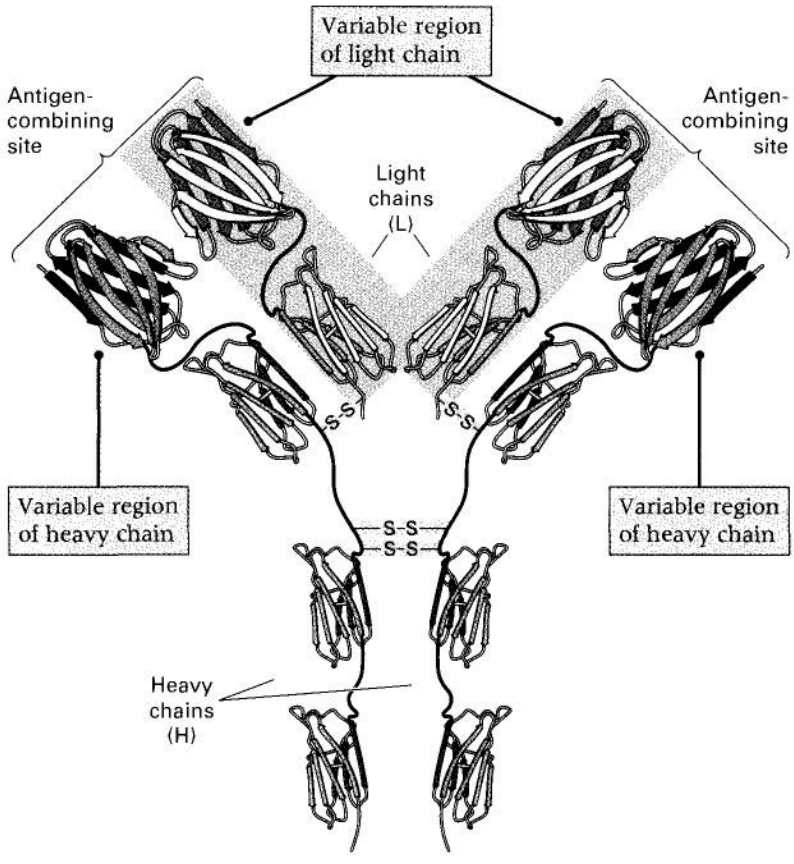


Figure 12.32 Structure of the immunoglobulin G (IgG) molecule, showing the light chains (L, shaded blue) and heavy chains (H, shaded yellow). Variable and constant regions are indicated.

specificity that combine with the antigen are located in the upper half of the arms above the fork of the Y. Each IgG molecule with a different specificity has a different amino acid sequence for the heavy and light chains in this part of the molecule. These specificity regions are called the **variable antibody regions** of the heavy and light chains. The remaining regions of the polypeptide are the **constant antibody regions,** which are called constant because they have virtually the same amino acid sequence in all IgG molecules.

Initial understanding of the genetic mechanisms responsible for variability in the amino acid sequences of antibody polypeptide chains came from cloning a gene for the light chain of IgG. The critical observation was made by comparing the nucleotide sequence of the gene in embryonic cells with that in mature antibody-producing B cells. In the genome of a B cell that was actively producing the antibody, the DNA segments corresponding to the constant and variable regions of the light chain were found to be very close together. However, in embryonic cells, these same DNA sequences were located far apart. Similar results were obtained for the variable and constant regions of the heavy chains: Segments encoding these regions were close together in B cells but widely separated in embryonic cells.

Complete DNA sequencing of the genomic region that codes for antibody proteins revealed not only the reason for the different gene locations in B cells and embryonic cells but also the mechanism for the origin of antibody variability. Cells in the germ line contain a small number of genes corresponding to the constant region of the light chain, and these are close together along the DNA. Separated from them, but on the same chromosome, is another cluster consisting of a much larger number of genes that correspond to the variable region of the light chain. In the differentiation of a B cell, one gene for the constant region is spliced (cut and joined) to one gene for the variable region, and this splicing produces a complete light-chain antibody gene. A similar splicing mechanism yields the constant and variable regions of the heavy chains.

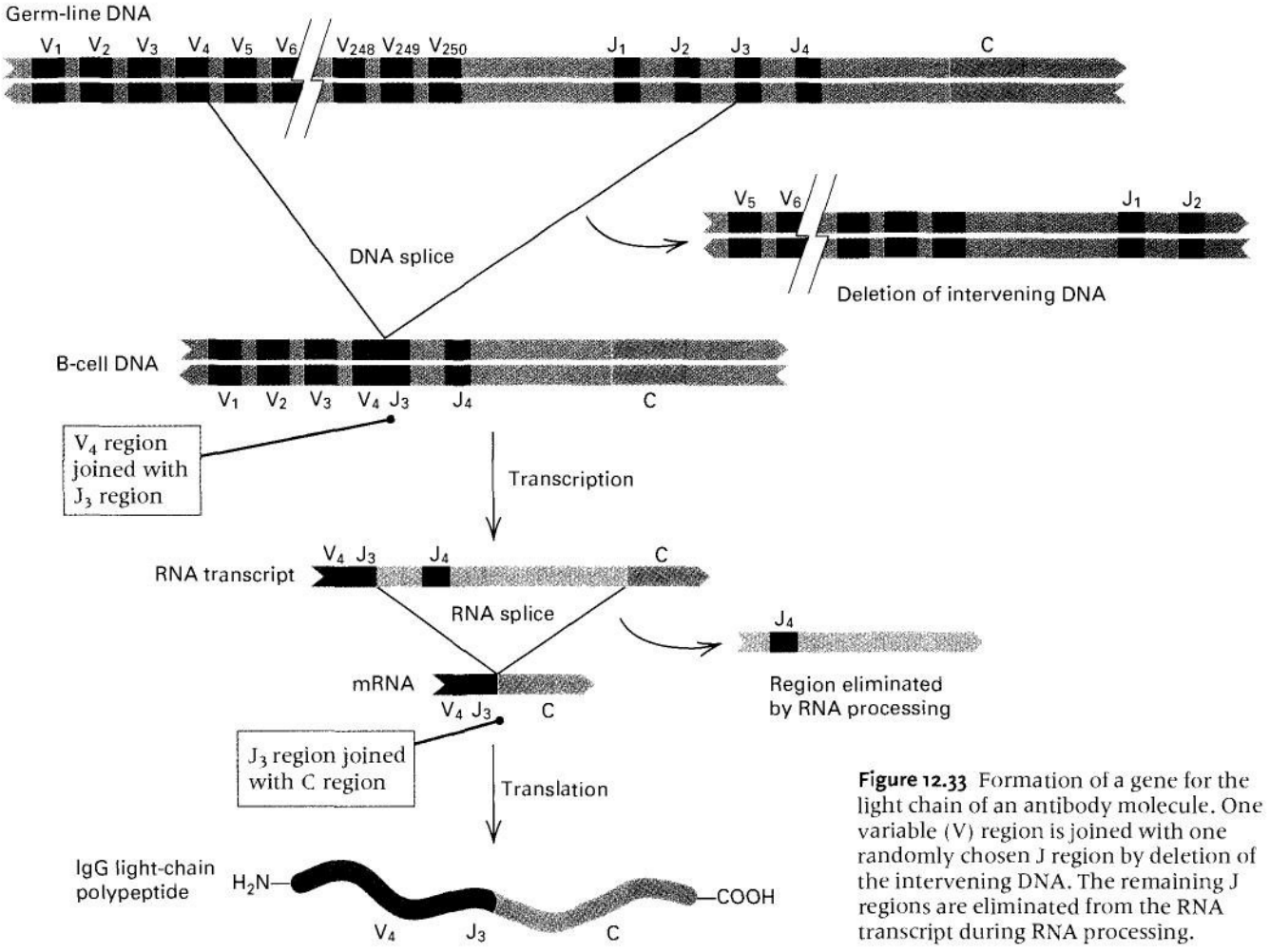


Figure 12.33 Formation of a gene for the light chain of an antibody molecule. One variable (V) region is joined with one randomly chosen J region by deletion of the intervening DNA. The remaining J regions are eliminated from the RNA transcript during RNA processing.

The formation of a finished antibody gene is somewhat more complicated than this overview implies, because light-chain genes consist of three parts and heavy-chain genes consist of four parts. DNA splicing in the origin of a light chain is illustrated in **Figure 12.33**. For each of two parts of the variable region, the germ line contains multiple coding sequences called the **variable (V)** and **joining (J) antibody regions.** In the differentiation of a B cell, a deletion makes possible the joining of one of the V regions with one of the J regions. The DNA joining process is called **V–J joining,** and it can create many combinations of the V and J regions. When transcribed, this joined V–J sequence forms the 5′ end of the light-chain RNA transcript. Transcription continues on through the DNA region coding for the **constant (C) antibody region** of the gene. RNA splicing subsequently joins the V–J region with the C region, creating the light-chain mRNA. Combinatorial joining also takes place in the genes for the antibody heavy chains, but in this case the DNA splicing joins the heavy-chain counterparts of V and J with a third set of sequences, called D (for diversity), located between the V and J clusters.

The level of antibody variability created by combinatorial joining can be calculated. In mice, the light chains are formed from combinations of about 250 V regions and 4 J regions, giving $250 \times 4 = 1000$ different possible chains. For the heavy chains, there are approximately 250 V, 10 D, and 4 J regions, producing $250 \times 10 \times 4 = 10{,}000$ possible combinations. Because any light chain can combine with any heavy chain, there are at least $1000 \times 10{,}000 = 10^7$ possible types of antibodies. The number of DNA sequences used for antibody production is quite small (about 500), but the number of possible antibodies is very large.

The value of 10^7 different antibody types is an underestimate, because there are two additional sources of antibody variability:

1. The junction for V–J (or V–D–J) splicing in combinatorial joining can be between different nucleotides of a particular V–J combination in light chains (or a particular V–D or D–J combination in heavy chains). The different splice junctions can generate different codons in the spliced gene. For example, a particular combination of V and J sequences can be spliced in five different ways. At the splice junction, the V sequence contains the nucleotides CATTTC, and the J sequence contains CTGGGTG. The splicing event determines the codons for amino acids 97, 98, and 99 in the completed antibody light chain, and it can occur in any of the following ways, the last two of which result in altered amino acid sequences.

Spliced sequence	97	98	99
CACTGGGTG	His	Trp	Val
CATTGGGTG	His	Trp	Val
CATTGGGTG	His	Trp	Val
CATTTGGTG	His	Leu	Val
CATTTCGTG	His	Phe	Val

   In this manner, variability in the junction of V–J joining can result in polypeptides that differ in amino acid sequence.
2. The V regions are susceptible to a high rate of *somatic mutation*, which occurs in B-cell development. These mutations allow different B-cell clones to produce different polypeptide sequences, even if they have undergone exactly the same V–J joining. The mechanism for the hypermutation is unknown.

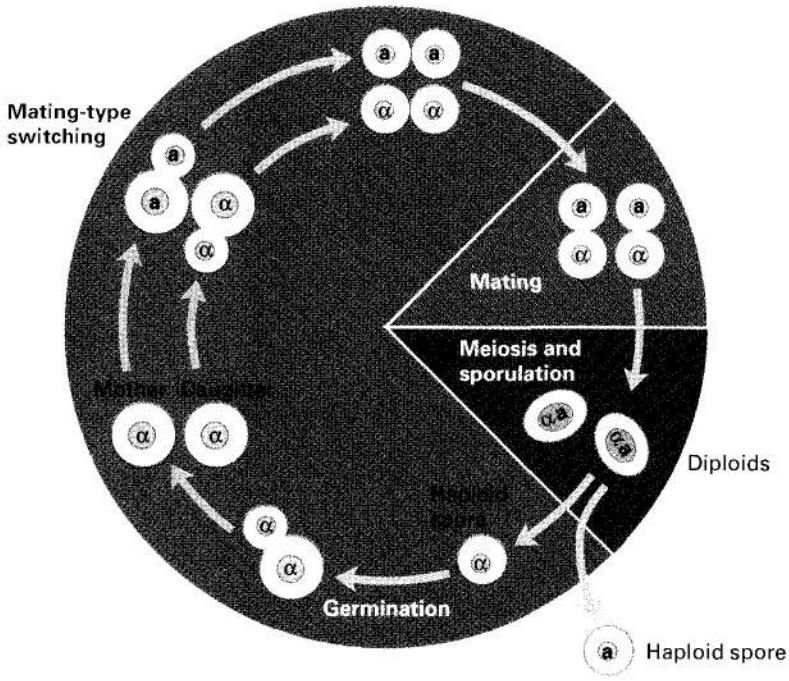


Figure 12.34 Mating-type switching in the yeast *Saccharomyces cerevisiae*. Germination of a spore (in this example, one of mating type α) forms a mother cell and a bud that grows into a daughter cell. In the next division, the mother cell and its new daughter cell switch to the opposite mating type (in this case, **a**). The result is two α and two **a** cells. Cells of opposite mating type can fuse to form **a**α diploid zygotes. In a similar fashion, germination of an **a** spore will result in switching to the α mating type.

Mating Type Interconversion

Rearrangement of DNA sequences underlies the phenomenon of **mating-type interconversion** in yeast. As we saw in Chapter 5, *S cerevisiae* has two mating types, denoted **a** and α. Mating between haploid **a** and haploid α cells produces the **a**α diploid, which can undergo meiosis to produce four-spored asci containing haploid **a** and α spores in the ratio 2 : 2. If a single yeast spore of either the **a** or the α genotype is cultured in isolation from other spores, then mating between progeny cells would not be expected because the daughter cells would all have the same mating type. However, *S. cerevisiae* has a mating system called **homothallism,** in which some cells undergo a conversion into the opposite mating type that allows matings between cells in what would otherwise be a pure culture.

The outlines of mating-type interconversion are shown in **Figure 12.34**. An original haploid spore (in this example, α) undergoes germination to produce two progeny cells. Both the mother cell (the original parent) and the daughter cell have mating type α, as expected from a normal mitotic division. However, in the next cell division, a switching (interconversion) of mating type takes place in both the mother cell and its *new* progeny cell, in which the original α mating type is replaced with the **a** mating type. After this second cell division is complete, the α and **a** cells are able to undergo mating because they now are of opposite mating types. Fusion of the nuclei

produces the **a**α diploid, which undergoes mitotic divisions and later sporulation to produce **a** and α haploid spores again.

The genetic basis of mating-type interconversion is DNA rearrangement, as outlined in **Figure 12.35**. The gene that controls mating type is the *MAT* gene in chromosome III, which can have either of two allelic forms, **a** or α. If the allele in a haploid cell is *MAT***a**, then the cell has mating type **a**; if the allele is *MAT*α, then the cell has mating type α. However, both genotypes normally contain both **a** and α genetic information in the form of unexpressed **cassettes** present in the same chromosome. The *HML*α cassette, about 200 kb distant from the *MAT* gene, contains the α DNA sequence; and the *HMR***a** cassette, about 150 kb distant from *MAT* on the other side, contains the **a** DNA sequence. When mating-type interconversion occurs, a specific endonuclease, encoded by the *HO* gene elsewhere in the genome, is produced and cuts both strands of the DNA in the *MAT* region. The double-stranded break initiates a process in which genetic information in the unexpressed cassette containing the opposite mating type becomes inserted into *MAT*. In this process, the DNA sequence in the donor cassette is duplicated, so the mating type becomes converted, but the same genetic information is retained in unexpressed form in the cassette. The terminal regions of *HML*, *MAT*, and *HMR* are identical (they are shown in light blue and dark blue in Figure 12.35), and these regions are critical in pairing of the regions and interconversion. The unique part of the α region is 747 base pairs in length; that of the **a** region is 642 base pairs long. The molecular details of the conversion process are similar to those of the double-strand gap mechanism of recombination discussed in Chapter 6.

Figure 12.35 illustrates two sequential mating-type interconversions. In the first, an α cell (containing the *MAT*α allele) undergoes conversion into **a**, using the DNA sequence contained in the *HMR***a** cassette. The converted cell has the genotype *MAT***a**. In a later generation, a descendant **a** cell may become converted into mating type α, using the unexpressed DNA sequence contained in *HML*α. This cell has the genotype *MAT*α. Mating-type switches can occur repeatedly in the lineage of any particular cell.

Transcriptional Control of Mating Type

Although mating-type *interconversion* is controlled by DNA rearrangements, mating type itself is controlled at the level of transcription. Both *MAT***a** (mating type **a**) and *MAT*α (mating type α) express a set of haploid-specific genes. They differ in that *MAT***a** expresses a set of **a**-specific genes and *MAT*α expresses a set of α-specific genes. The haploid-specific genes expressed in cells of both mating types include *HO*, which encodes the HO endonuclease used in mating type interconversion, and *RME1*,

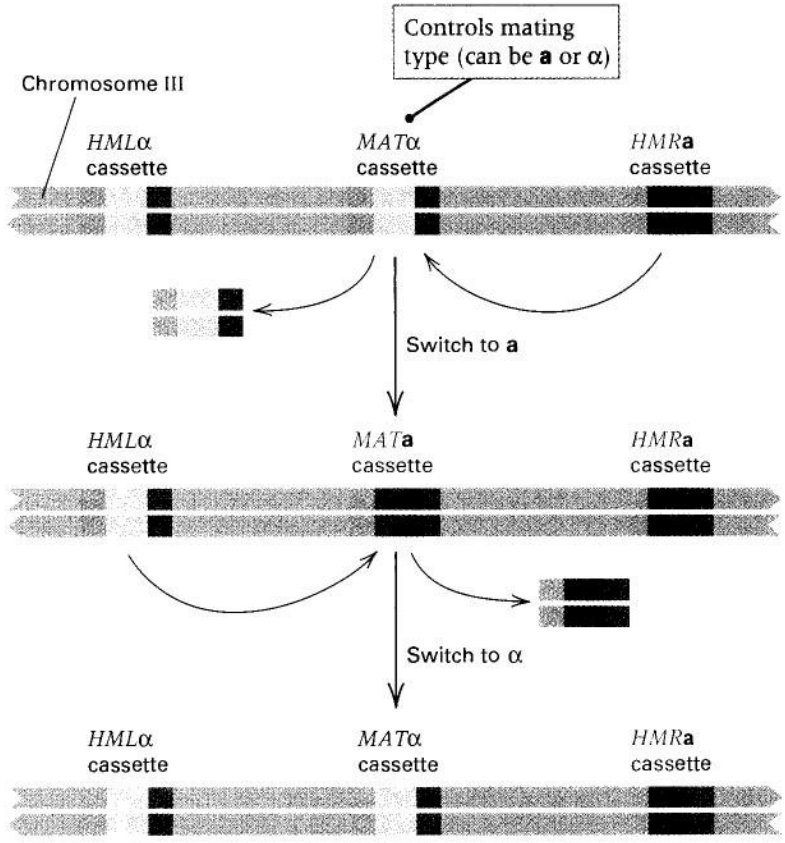


Figure 12.35 Genetic basis of mating-type interconversion. The mating type is determined by the DNA sequence present at the *MAT* locus. The *HML* and *HMR* loci are cassettes that contain unexpressed mating-type genes, either α or **a**. In the interconversion from α to **a**, the α genetic information present at *MAT* is replaced with the **a** genetic information from *HMR***a**. In the switch from **a** to α, the **a** genetic information at *MAT* is replaced with the α genetic information from *HML*α.

CoNnection

Sex-Change Operations

James B. Hicks, Jeffrey N. Strathern, and Ira Herskowitz 1977
University of Oregon, Eugene, Oregon
The Cassette Model of Mating-type Interconversion

*Mating in yeast requires cells of opposite mating types, **a** and α, to come together and fuse. Both **a** and α cells release signaling substances into the medium that prepare the opposite cell type for mating. In the **a**α diploid cell, genes specific for the diploid phase of the life cycle are expressed, and those specific for the haploid phase of the life cycle are turned off. Remarkably, yeast cells can change their mating type. In homothallic cells, the switch can take place in every generation. In heterothallic cells, it takes place at a frequency of 10^{-6}. This paper proposed a very bold hypothesis, later confirmed experimentally, that all yeast cells contain, in addition to the information for the expressed mating type at the* MAT *locus, both α and **a** genetic information in unexpressed cassettes at* HML *and* HMR *syntenic to, but distinct from,* MAT. *The* HO *gene that distinguishes homothallic from heterothallic yeast codes for a site-specific endonuclease that cleaves within the mating-type locus and initiates the information-transfer process from either* HMLα *or* HMR**a**. *This is the physical basis of mating-type interconversion.*

Studies of mating-type interconversion in the yeast *Saccharomyces cerevisiae* have led us to propose a new mechanism of gene control involving mobile genes. . . . The mating-type locus of *S. cerevisiae* exists in two states, **a** or α, which control the ability of yeast cells to mate and sporulate. . . . It is clear that the **a** and α alleles are distinct entities, as they are codominant—an **a**/α diploid differs from a [rarely formed] **a**/**a** or α/α diploid. . . . The **a** allele thus is not simply the absence of α, and the α allele is not simply the absence of **a**. . . .

The mating-type locus is viewed as analogous to a playback head of a tape recorder which can give expression to whatever cassette of information is plugged into it.

In homothallic strains, changes to opposite mating types occur frequently, as often as every generation. These strains carry a dominant nuclear gene (*HO*), unlinked to the mating-type locus. . . . *HO* cells that have sustained a change in mating type are fully capable of continuing to change mating type. However, when the *HO* gene is removed by genetic crosses, the new mating type is stable. . . . Cells with a defect at the α mating-type locus can be converted to functional **a** cells. However, these **a** cells are then observed to switch to become functional α cells. In other words, a functional α mating-type locus can be restored through the mating-type interconversion process. We explain this recovery by proposing that yeast cells contain an additional copy (or copies) of the mating-type locus information. Specifically, we propose that yeast cells contain a silent (unexpressed) copy of **a** information and a silent copy of α information and that the *HO* gene activates this information by inserting it (or a copy) into the mating-type locus. Genetic studies have revealed the existence of two loci [*HML*α and *HMR***a**] in addition to *HO* which are necessary for mating-type interconversion and which we propose are the silent α and **a** information. . . . To summarize, we propose that cell type in *S. cerevisiae* is regulated by a locus [*MAT*] into which various blocs of information can be inserted. The mating-type locus is viewed as analogous to a playback head of a tape recorder which can give expression to whatever cassette of information is plugged into it.

Source: DNA Insertion Elements, Plasmids, and Episomes, eds. Ahmad I. Bukhari, James A. Shapiro, and Sankar L. Adhya. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory, pp. 457–462

which encodes a repressor of meiosis-specific genes. The functions of the expressed genes that differ in the mating types include (1) secretion of a mating peptide that arrests cells of the opposite mating type before DNA synthesis and prepares them for cell fusion, and (2) production of a receptor for the mating peptide secreted by the opposite mating type. Therefore, when **a** and α cells are in proximity, they prepare each other for mating and undergo fusion.

Regulation of mating type at the level of transcription takes place according to the regulatory interactions diagrammed in **Figure 12.36**. These regulatory interactions were originally proposed on the basis of the phenotypes of various types of mutants, and most of the details have been confirmed by direct molecular studies. The symbols **a***sg*, α*sg*, and *hsg* represent the **a**-specific genes, the α-specific genes, and the haploid-specific genes, respectively; each set of

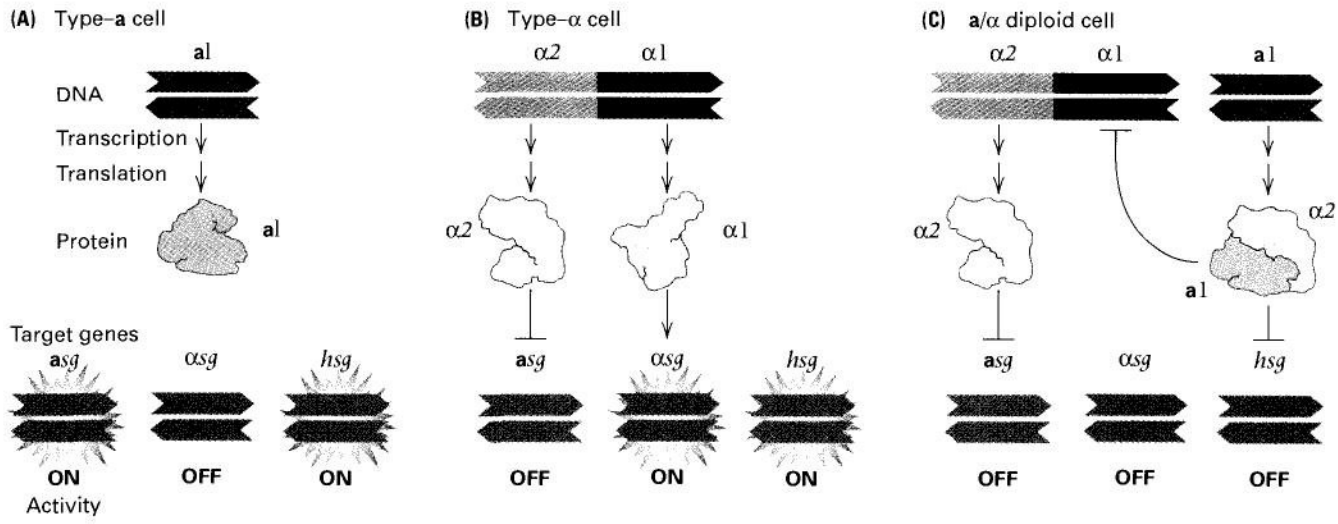
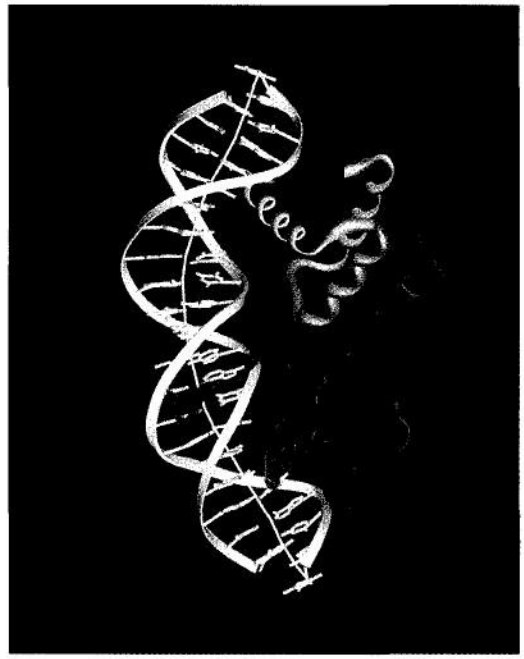


Figure 12.36 Transcriptional regulation of mating type in yeast. The symbols **a**sg, αsg, and *hsg* denote sets of **a**-specific genes, α-specific genes, and haploid-specific genes, respectively. Sets of genes represented with a "sunburst" are "on," and those unmarked are "off." (A) In an **a** cell, the **a**1 peptide is inactive, and the sets of genes manifest their default states of activity (**a***sg* and *hsg* on and α*sg* off), so the cell is an **a** haploid. (B) In an α cell, the α2 peptide turns the **a***sg* off and the α1 peptide turns the α*sg* on, so the cell is an α haploid. (C) In an **a**/α diploid, the α2 and **a**1 peptides form a complex that turns the *hsg* off, the α2 peptide turns the **a***sg* off, and the α*sg* manifest their default activity of off, so physiologically the cell is non-**a**, non-α, and nonhaploid (that is, it is a normal diploid).

genes is represented as a single segment, and lack of a "sunburst" indicates that transcription does not take place. In a cell of mating type **a** (Figure 12.36A), the *MAT***a** region is transcribed and produces a polypeptide called **a**1. By itself, **a**1 has no regulatory activity, and in the absence of any regulatory signal, **a***sg* and *hsg*, but not α*sg*, are transcribed. In a cell of mating type α (Figure 12.36B), the *MAT*α region is transcribed, and two regulatory proteins denoted α1 and α2 are produced: α1 is a *positive regulator* of the α-specific genes, and α2 is a *negative regulator* of the **a**-specific genes. The result is that α*sg* and *hsg* are transcribed, but transcription of **a***sg* is turned off. Both α1 and α2 bind with particular DNA sequences upstream from the genes that they control.

In the diploid (Figure 12.36C), both *MAT***a** and *MAT*α are transcribed, but the only polypeptides produced are **a**1 and α2. The reason is that the **a**1 and α2 polypeptides combine to form a negative regulatory protein that represses transcription of the α1 gene in *MAT*α and of the haploid-specific genes. The α2 polypeptide acting alone is a negative regulatory protein that turns off **a***sg*. Because α1 is not produced,



Structure of the **a**1/α2 protein bound with DNA. The **a**1 subunit is shown in blue, the α2 subunit in red. Contact with the DNA target results in a sharp bend in the DNA. [Courtesy of Cynthia Wolberger. Reprinted with permission from C. Wolberger, T. Li, M. R. Stark, and A. D. Johnson, *Science* 270: 262. Copyright 1995 American Association for the Advancement of Science.]

transcription of *αsg* is not turned on. The overall result is that the *αsg* are not turned on because α1 is absent, the **a***sg* are turned off because α2 is present, and the *hsg* are turned off by the **a**1/α2 complex. This ensures that meiosis can occur (because expression of *RME1* is turned off) and that mating-type switching ceases (because the HO endonuclease is absent). Thus the homothallic **a**α diploid is stable and can undergo meiosis. The result is that the **a**α diploid does not transcribe either the mating-type-specific set of genes or the haploid-specific genes.

The repression of transcription of the haploid-specific genes mediated by the **a**1/α2 protein is an example of negative control of the type already familiar from the *lac* and *trp* systems in *E. coli*. The interesting twist in the yeast example is that the α2 protein has a regulatory role of its own in repressing transcription of the **a**-specific genes. Why does the α2 protein, on its own, not repress the haploid-specific genes as well? The answer lies in the specificity of its DNA binding. By itself, the α2 protein has low affinity for the target sequences in the haploid-specific genes. However, the **a**1/α2 heterodimer has both high affinity and high specificity for the target DNA sequences in the haploid-specific genes.

Chapter Summary

The processes of gene regulation generally differ between prokaryotes and eukaryotes, but transcriptional regulation plays a key role in both types of organisms. Transcriptional regulation may be either negative ("on" unless specifically turned "off") or positive ("off" unless specifically turned "on"). A negative regulatory system may be inducible (when the repressor protein is inactivated by combining with the inducer) or repressible (when the repressor is formed by the combination of aporepressor protein and the co-repressor). Positively regulated systems are typically controlled by transcriptional activator proteins.

Lactose metabolism in *E. coli* is a classic example of an inducible system. When lactose is present, the genes for the enzymes required to metabolize lactose are transcribed; when lactose is absent, transcription is repressed. Two proteins needed to utilize lactose (permease, required for the entry of lactose into the cell, and β-galactosidase, the enzyme that cleaves lactose) are tightly linked and transcribed into a single polycistronic mRNA molecule, *lac* mRNA. Immediately adjacent to the promoter for *lac* mRNA is a regulatory sequence of bases called the operator. A repressor protein, the product of the *lacI* gene, binds with the operator and prevents transcription. The inducer binds to the repressor and inactivates it. Therefore, in the absence of lactose the permease and β-galactosidase genes are not transcribed, but in the presence of lactose there is no active repressor, and the *lac* promoter becomes accessible to RNA polymerase. The operator, the promoter, and the structural genes are adjacent to one another and together constitute the *lac* operon. Repressor mutations have been isolated that inactivate the repressor protein, and operator mutations are known that prevent recognition of the operator by an active repressor; such mutations cause continuous production of *lac* mRNA and are said to be constitutive.

The *lac* operon as well as other sugar-utilization operons are also subject to positive regulation through binding with the cAMP–CRP complex. (The cAMP stands for cyclic AMP, and CRP for *cyclic AMP receptor protein*.) When glucose is present in the growth medium, the level of cAMP is low, and not enough cAMP–CRP is formed to allow transcription of the operons. Only when glucose is absent is the concentration of cAMP sufficient to produce enough cAMP–CRP to make transcription possible.

Many biosynthetic enzyme systems exemplify the repressible type of negative transcriptional control. In the *trp* operon in *E. coli*, transcription of the operon is controlled by the concentration of tryptophan in the growth medium. When excess tryptophan is present, it binds with the *trp* aporepressor to form the active repressor that prevents transcription. Regulation of the *trp* operon is also subject to fine tuning by attenuation, in which transcription is initiated continually but the transcript forms a hairpin structure that results in premature termination. The frequency of termination of transcription is determined by the availability of charged tryptophan tRNA: With decreasing concentrations of tryptophan, termination occurs less often, thereby allowing transcription of the operon and production of the enzymes for tryptophan synthesis. Attenuators also regulate operons for the synthesis of other amino acids.

Bacteriophage λ adopts either the quiescent lysogenic state or the lytic cycle as a result of competition between two repressors, cI and cro. If cI repressor dominates, the λ genome becomes integrated into the bacterial chromosome and a lysogen is formed, in which case cI continues to be synthesized and prevents transcription of all other λ genes. If cro repressor dominates, cI repressor is no longer synthesized, and the lytic cycle ensues. Although there are elements of positive transcriptional regulation in the regulatory circuitry, the predominant mode of regulation is negative.

Many genes in eukaryotes are also regulated at the level of transcription. Transcriptional activators bind to DNA sequences known as enhancers, which are usually

short sequences that may be present at a variety of positions around the genes they regulate. In the recruitment model of transcriptional activation, a transcriptional activator protein interacts directly with one or more protein components of the transcription complex, such as TFIID, a complex that includes the TATA-box-binding protein (TBP) and multiple TBP-associated factors (TAFs), or the Pol II holoenzyme. The fully assembled transcriptional apparatus in eukaryotes consists of a complex assemblage of TFIID, RNA polymerase II, and numerous other basal transcription factors. A prerequisite to formation of an active transcription complex in native chromatin is that the nucleosomes be modified or repositioned, a function of several chromatin-remodeling complexes.

In some organisms, epigenetic mechanisms of regulation also occur, in which the DNA is modified either chemically or by protein binding in such a way as to cause a genetically transmissible change in pattern of gene expression. Epigenetic transcriptional silencing of a gene is often associated with heavy cytosine methylation. In plants, transgenes often trigger epigenetic transcriptional cosuppression, in which both the transgene and homologous endogenous genes are transcriptionally repressed. Imprinting in mammals is an epigenetic, sex-specific modification of a limited number of genes that takes place in the female and male germ lines and that influences gene activity in embryogenesis. The imprints are erased early in germ-line development but then re-established according to their sex-specific patterns.

Some genes contain alternative promoters used in different tissues; other genes use a single promoter, but the transcripts are spliced in different ways. Alternative splicing can result in mRNA molecules that are translated with different efficiencies, or even in different proteins if there is alternative splicing of the protein-coding exons. In one type of posttranscriptional cosuppression in plants, which depends on transcript homology, the transcripts of the cosuppressed genes are rapidly degraded. Transcript degradation is also characteristic of RNA interference (RNAi), a widespread phenomenon in which just a few molecules of short, double-stranded RNA trigger destruction of all transcripts containing homologous sequences. Regulation can also be at the level of translation—for example, through the activation of untranslated mRNAs.

Genetic regulatory mechanisms also occasionally include changes in DNA copy number or DNA sequence. In many animals, rDNA genes are temporarily amplified during oocyte development. Programmed DNA rearrangements also occur, for example, in generating antibody and T-cell-receptor variability and in mating-type interconversion in yeast.

Key Terms

aporepressor
attenuation
attenuator
autoregulation
basal transcription factors
β-galactosidase
cAMP–CRP complex
cassette
chromatin-remodeling complexes (CRC)
cis-dominant
constant (C) antibody region
constitutive mutant
coordinate regulation
co-repressor
CpG island
DNA looping
DNA methylase
enhancer
epigenetic
gene amplification
genomic imprinting
homothallism
housekeeping genes
immunoglobulin
inducer
inducible transcription
joining (J) antibody region
lactose permease
leader polypeptide
mating-type interconversion
methylation induced premeiotically (MIP)
negative regulation
operator
operon
operon model
Pol II holoenzyme
positive regulation
posttranscriptional cosuppression
promoter
repressible transcription
repressor
RNA interference (RNAi)
silencer
TATA-box-binding protein (TBP)
TBP-associated factors (TAFs)
transcriptional activator protein
transcriptional cosuppression
variable (V) antibody region
V–J joining

Review the Basics

- What is positive regulation of transcription? What is negative regulation of transcription? What is the role of the repressor in each case? Give an example of each type of regulation.
- What is autoregulation? Distinguish between positive and negative autoregulation. Which would be used to amplify a weak induction signal? Which would be used to prevent overproduction?
- What class of *lac* mutants demonstrated that the presence of lactose in the growth medium was not necessary for expression of the genes for lactose utilization?

- How does an operon result in coordinate control of the genes included? Are operons usually found in eukaryotic organisms?
- In what sense does attentuation provide a "fine-tuning" mechanism for operons that control amino acid biosynthesis?
- An operon containing genes that encode the enzymes in a metabolic pathway for the synthesis of an amino acid includes a short, open reading frame in the leader sequence. This short, open reading frame contains multiple codons for the amino acid synthesized by the pathway. What does this observation suggest about regulation of the operon?
- What is a transcriptional activator protein? a transcriptional enhancer? a chromatin remodeling complex? What role do these elements play in eukaryotic gene regulation?
- How does the possibility of alternative splicing affect the generality of the statement that one gene encodes one polypeptide chain?
- What is meant by the term *epigenetic regulation?* Explain how epigenetic regulation can be mediated through cytosine methylation.
- What phenomenon is explained by cosuppression? Is there only one mechanism of cosuppression? Explain.
- Would you expect DNA splicing of the type observed in antibody formation to be a reversible process or an irreversible process? Why?

Guide to Problem Solving

Problem 1 The LacY permease is able to transport lactobionic acid as well as lactose, but LacZ cleaves lactobionic acid very inefficiently. Certain mutant forms of LacZ have an altered substrate specificity that allows the mutant enzyme to cleave lactobionic acid. These mutants are able to grow in medium containing lactobionic acid as the sole source of carbon and energy. However, they cannot grow unless they also have a $lacO^c$ mutation or unless IPTG (isopropylthiogalactoside) is present in the medium when the operator is wildtype. How can you explain these results?

Answer In order to grow on lactobionic acid, even the mutant LacZ strain must have both *lacZ* and *lacY* expressed. Evidently lactobionic acid is not an inducer of the *lac* operon, and therefore a $lacO^+$ mutant cannot grow in lactobionic acid unless IPTG, an inducer of the *lac* operon, is also present to induce transcription of the operon. The finding with $lacO^c$ mutants is consistent with this explanation, because when LacY and the mutant LacZ are produced constitutively, the strains can grow in lactobionic acid.

Problem 2 The *arabinose* operon encodes the enzymes necessary for degradation of the 5-carbon sugar arabinose. It is one of the few degradative operons in *E. coli* that are controlled primarily by positive regulation of transcription. The protein AraC, encoded in the gene *araC*, works as follows. In the absence of arabinose, AraC binds to the operator and an initiator region and, by causing formation of a DNA loop, prevents transcription. In the presence of arabinose, AraC has a high affinity only for the initiator, where it binds and induces transcription of the operon in a polycistronic mRNA that includes coding regions for the degradative enzymes AraB, AraA, and AraD (collectively called AraBAD).The operon is also positively regulated by the cyclic AMP receptor protein (CRP), in much the way the *lac* operon is regulated by CRP. Fill each box in the table shown here with a + or a − sign. A + (or −) in the column for AraC or CRP means that under the specified growth condition, the AraC or CRP protein is (or is not) bound with the arabinose initiator; a + (or −) in the column for AraBAD means that under the specified growth condition, the AraBAD proteins are (or are not) induced.

	Initiator binding only		Enzyme induction
Growth medium	**AraC**	**CRP**	**AraBAD**
Lactose	☐	☐	☐
Arabinose	☐	☐	☐
Arabinose + glucose	☐	☐	☐

Answer According to the description of how the operon is regulated, AraC binds only the initiator when arabinose is present in the medium. The binding of AraC will induce transcription of the araBAD polycistronic mRNA, but AraC alone is not sufficient. As in the *lac* operon, transcription can occur only if the cAMP-CRP complex is also bound with the operator. The cAMP-CRP complex will form when cells have a high level of cyclic AMP, such as when growing in lactose or arabinose medium, but not when the level of cyclic AMP is low, such as under growth in glucose medium. The transcription of the araBAD mRNA results in the synthesis of the AraBAD enzymes. These considerations indicate how the table should be completed.

	Initiator binding only		Enzyme induction
Growth medium	**AraC**	**CRP**	**AraBAD**
Lactose	−	+	−
Arabinose	+	+	+
Arabinose + glucose	+	−	−

Problem 3 With regard to mating type in yeast, what phenotypes would each of the following haploid cells exhibit?

(a) a duplication of the mating-type gene giving the genotype *MAT***a** *MAT*α

(b) a deletion of the *HML*α cassette in a *MAT***a** cell

Answer

(a) The haploid cell expresses both **a** and α, so the **a**-specific genes, the α-specific genes, and the haploid-specific genes are all inactive. The phenotype is similar to the **a**α diploid. It will not mate, and if it attempts to sporulate, it will self-destruct.

(b) The phenotype is that of a normal **a** haploid, but mating-type switching to α cannot take place.

Analysis and Applications

12.1 The metabolic pathway for glycolysis is responsible for the degradation of glucose and is one of the fundamental energy-producing systems in living cells. Would you expect the enzymes in this pathway to be regulated? Why or why not?

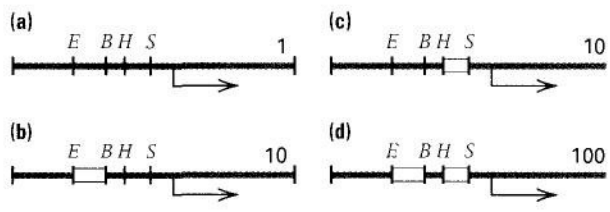
12.2 Among mammals, the reticulocyte cells in the bone marrow lose their nuclei in the process of differentiation into red blood cells. Yet the reticulocytes and red blood cells continue to synthesize hemoglobin. Suggest a mechanism by which hemoglobin synthesis can continue for a long period of time in the absence of the hemoglobin genes.

12.3 Several eukaryotes are known in which a single effector molecule regulates the synthesis of different proteins coded by distinct mRNA molecules—say, X and Y. At what level in the process of gene expression does regulation occur in each of the following situations?

(a) Neither nuclear nor cytoplasmic RNA can be found for either X or Y.
(b) Nuclear but not cytoplasmic RNA can be found for both X and Y.
(c) Both nuclear and cytoplasmic RNA can be found for both X and Y, but none of it is associated with polysomes.

12.4 Is it necessary for the gene that codes for the repressor of a bacterial operon to be near the structural genes? Why or why not?

12.5 In studies of the operator region of an inducible operon in *E. coli*, the four constructs shown below were examined for level of transcription *in vitro*. The number associated with each construct is the relative level of transcription observed in the presence of the repressor protein. The symbols *E*, *B*, *H*, and *S* stand for the restriction sites *Eco*RI, *Bam*HI, *Hin*dIII, and *Sac*I. Construct **(a)** is the wildtype operator region, and in parts **(b–d)** the open boxes indicate restriction fragments that were deleted. What hypothesis about repressor–operator interactions can explain these results? How could this hypothesis be tested?



12.6 Consider a eukaryotic transcriptional activator protein that binds to an enhancer sequence and promotes transcription. What change in regulation would you expect from a duplication in which several copies of the enhancer were present instead of just one?

12.7 Would you expect the regulation of a gene to be affected by an inversion of a promoter sequence? by an inversion of an enhancer sequence? Explain your answers.

12.8 The following questions pertain to the *lac* operon in *E. coli*.

(a) Which proteins are regulated by the repressor?
(b) How does binding of the *lac* repressor to the *lac* operator prevent transcription?
(c) Is production of the *lac* repressor constitutive or inducible?

12.9 The permease of *E. coli* that transports the α-galactoside melibiose can also transport lactose, but it is temperature-sensitive: Lactose can be transported into the cell at 30°C but not at 37°C. In a strain that produces the melibiose permease constitutively, what are the phenotypes of *lacZ*$^-$ and *lacY*$^-$ mutants at 30°C and 37°C?

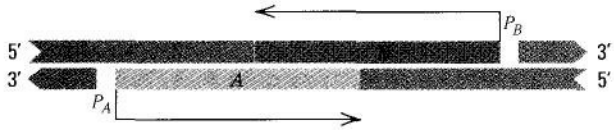
12.10 How do inducers enable transcription to occur in a bacterial operon under negative transcriptional control?

12.11 When glucose is present in an *E. coli* cell, is the concentration of cyclic AMP high or low? Can a mutant with either an inactive adenyl cyclase gene or an inactive *crp* gene synthesize β-galactosidase? Does the binding of cAMP–CRP to DNA affect the binding of a repressor?

12.12 The operon allows a type of coordinate regulation of gene activity in which a group of enzymes with related functions are synthesized from a single polycistronic mRNA. Does this imply that all proteins in the polycistronic mRNA are made in the same quantity? Explain.

12.13 Both repressors and aporepressors bind molecules that are substrates or products of the metabolic pathway encoded by the genes in an operon. Generally speaking, which binds the substrate of a metabolic pathway and which the product?

12.14 Genes *A* and *B* are partially overlapping in the manner shown in the diagram on the next page, in which P_A and P_B are the promoters and the arrows show the origin, direction, and extent of transcription. The overlap includes a protein-coding exon. For mutations that take



place in the protein-coding exon in the region of overlap, what are the consequences for reading frame B of a mutation that, in reading frame A, is:

(a) A missense mutation?
(b) A deletion?
(c) A frameshift mutation?

12.15 Is an attenuator a region of DNA that, like an operator, binds with a protein? Is RNA synthesis ever initiated at an attenuator?

12.16 In a bacterium related to *E. coli*, a biosynthetic operon containing genes for the synthesis of the amino acid proline is studied. This operon is found to have a regulatory system completely analogous to that of the *E. coli* tryptophan operon, with an aporepressor that binds to proline and a leader sequence in the mRNA that includes a sequence of five consecutive codons for proline.

(a) Under what conditions would transcription of the proline operon be initiated?
(b) Under what conditions would attenuation take place and transcription be halted?
(c) Under what conditions would transcription continue through to the end of the operon?

12.17 In order to study attenuation in the histidine operon, site-directed mutagenesis was used to convert the seven CAY codons for histidine in the *his* leader mRNA into UGG codons for tryptophan.

(a) Under what conditions would the histidine operon initiate transcription?
(b) Under what conditions would the attenuation take place, halting transcription?
(c) Under what conditions would transcription continue through to the end of the operon?

12.18 A mutant of *E. coli* is isolated that has a defective ribosomal protein such that translation stalls briefly whenever a codon for tryptophan is encountered, irrespective of the level of charged tryotophan tRNA. How would this mutation be expected to affect attenuation of the tryptophan operon?

12.19 Among amino acid biosynthetic operons that have attenuation, some leader mRNA sequences contain very few codons for the amino acid that is synthesized (for example, in the *trp* operon), whereas others contain many more codons for the amino acid (for example, in the *phe* operon). Explain why this might be expected.

12.20 In regard to mating-type switching in yeast, what phenotype would you expect from a type α cell that has a deletion of the *HMR***a** cassette?

12.21 What mating-type phenotype would you expect of a haploid cell of genotype *MAT***a**′, where the prime denotes a mutation that renders the **a**1 protein inactive? What mating type would you expect in the diploid *MAT***a**′/*MAT*α?

12.22 What mating-type phenotype would you expect from a diploid cell of genotype *MAT***a**/*MAT*α with a mutation in α in which the α2 gene product functions normally in turning off the **a**-specific genes but is unable to combine with the **a**1 product?

12.23 If a wildtype *E. coli* strain is grown in a medium without lactose or glucose, how many proteins are bound to the *lac* operon? How many are bound if the cells are grown in glucose?

12.24 A mutant strain of *E. coli* is found that produces both β-galactosidase and permease constitutively (that is, whether lactose is present or not).

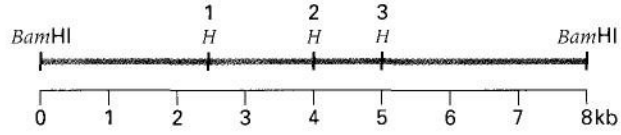
(a) What are two possible genotypes for this mutant?
(b) A second mutant is isolated that produces no active β-galactosidase at any time but produces permease if lactose is present in the medium. What is the genotype of this mutant?
(c) A partial diploid is created from the mutants in parts **(a)** and **(b)**: When lactose is absent, neither enzyme is made, and when lactose is present, both enzymes are made. What is the genotype of the mutant in part **(a)**?

12.25 A $lacI^+$ $lacO^+$ $lacZ^+$ $lacY^+$ Hfr strain is mated with an F^- $lacI^-$ $lacO^+$ $lacZ^-$ $lacY^-$ strain. In the absence of any inducer in the medium, β-galactosidase is made for a short time after the Hfr and F^- cells have been mixed. Explain why it is made and why only for a short time.

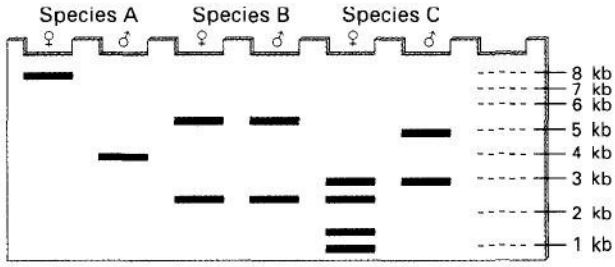
12.26 What amino acids can be inserted at the site of the UGA codon that is suppressed by a suppressor tRNA?

12.27 An *E. coli* mutant is isolated that is simultaneously unable to utilize a large number of sugars as sources of carbon. However, genetic analysis shows that the operons responsible for metabolism of each sugar are free of mutations. What genotypes of this mutant are possible?

12.28 The methylation-sensitive restriction enzyme *Hpa*II, which cleaves only unmethylated 5′-CCGG-3′ sites, is used to study methylation in three species of mammals in a region of a gene with the restriction map shown in the diagram below, where the *Hpa*II (*H*) sites are numbered.



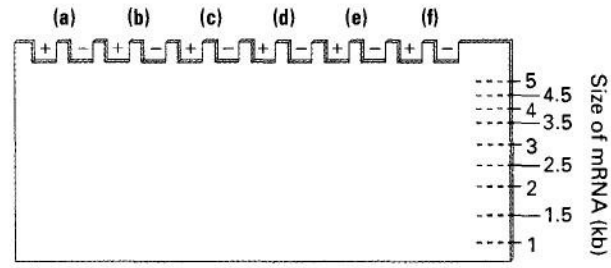
From each of the species, DNA is extracted from eggs and sperm, digested with *Bam*HI and *Hpa*II, and analyzed by Southern blotting using a probe that hybridizes with the entire region of interest. The results are shown in the gel diagram on the next page.



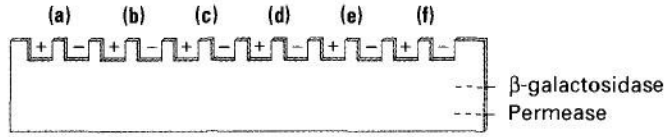
(a) In each type of gamete in each species, which *Hpa*II sites are methylated?
(b) Which species show differential methylation in the sexes?
(c) In which species do the females gametes show greater methylation than the male gametes?

12.29 In the *Northern blot* technique, mRNA molecules are separated on an electrophoresis gel, transferred to a membrane filter, and hybridized with DNA that has been labeled with a radioactive or light-sensitive moiety. RNA molecules that contain sequences homologous to the probe become visible as discrete bands. A Northern blot experiment is carried out with mRNA from *E. coli*, using a DNA probe corresponding to the *lacA* gene, which is adjacent to *lacY*. Wildtype *lac* mRNA is approximately 5 kb in length. Two kinds of *lacZ* and *lacY* mutations are studied. The $lacZ^-$ and $lacY^-$ mutations are simple nucleotide substitution mutations that encode inactive proteins. The *lacZ** and *lacY** mutations are deletions. The *lacZ** deletion is missing 2 kb of *lacZ* coding sequence, *lacY** is missing 0.5 kb of *lacY* coding sequence, and neither allele produces a polypeptide product. The mRNA from the following six genotypes is analyzed via Northern blot after growth in either the presence (+) or the absence (−) of inducer. The RNA molecules are separated under conditions yielding the size scale shown by the dashed lines in the accompanying gel diagram. For each genotype, under both inducing (+) and the noninducing (−) conditions, indicate in the gel diagram where an mRNA band would be expected.

(a) $lacI^+\ O^+\ Z^+\ Y^+$
(b) $lacI^-\ O^+\ Z^-\ Y^-$
(c) $lacI^-\ O^+\ Z^*\ Y^*$
(d) $lacI^+\ O^c\ Z^-\ Y^*$
(e) $F'\ lacI^+\ O^c\ Z^*\ Y^-/lacI^+\ O^+\ Z^-\ Y^*$
(f) $F'\ lacI^+\ O^+\ Z^*\ Y^+/lacI^+\ O^c\ Z^-\ Y^*$



12.30 In a *Western blot*, proteins are separated on an electrophoresis gel, transferred to a membrane filter, and mixed with a labeled antibody that combines specifically with one of the proteins of interest. The six strains of *E. coli* in the previous problem are analyzed using a Westen blot stained with a mixture of two antibodies, one for β-galactosidase and the other for the permease. The gel diagram below shows the positions to which these proteins would migrate, if they were present. Indicate in the gel diagram which bands, if any, would be found in each of the six strains grown under inducing (+) and under noninducing (−) conditions.



Problem 12.30

Challenge Problems

12.31 In a newly discovered organism the histidine operon is negatively regulated. It contains the structural genes for the enzymes needed to synthesize histidine, and the repressor protein is also coded within the operon—that is, in the polycistronic mRNA molecule that codes for the other proteins. Synthesis of this mRNA is controlled by a single operator regulating the activity of a single promoter. The co-repressor of this operon is $tRNA^{His}$, to which histidine is attached. This tRNA is not coded by the operon itself. A collection of mutants with the following defects is isolated. Determine whether the histidine enzymes would be synthesized by each of the mutants and whether each mutant would be dominant, *cis*-dominant only, or recessive to its wildtype allele in a partial diploid.

(a) The promoter cannot bind with RNA polymerase.
(b) The operator cannot bind the repressor protein.
(c) The repressor protein cannot bind with DNA.
(d) The repressor protein cannot bind histidyl-$tRNA^{His}$.
(e) The uncharged $tRNA^{His}$ (that is, without histidine attached) can bind to the repressor protein.

12.32 The attenuator of the histidine operon contains seven consecutive histidines. The relevant part of the attentuator coding sequence is

5'-AAACACCACCATCATCACCATCATCCTGAC-3'

A mutation occurs in which an additional A nucleotide is inserted immediately after the red A. What amino acid

GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by-chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- The goal of studying **gene regulation** is to be able to understand the DNA–protein and protein–protein interactions at the same level of detail as that at which computer scientists understand the connections in microcircuits. Among the seminal ideas in regulation is that DNA-binding proteins bind to DNA with positive cooperativity, which means that as each subunit of a complex is bound with DNA, it facilitates the binding of the next subunit. Cooperative binding creates efficient "molecular switches" to control transcription. This keyword site includes an excellent nontechnical summary of some of the crucial advances in studies of gene regulation.

- This keyword database of **transcription factors** gives easy access to sequences and structural diagrams showing the key motifs of each of the numerous types and subtypes of DNA-binding proteins involved in transcriptional regulation. View examples of each of the four major categories of transcription factors: those having basic domains (includes leucine zipper factors and helix–loop–helix factors), those having zinc-coordinating DNA-binding domains (includes all zinc-finger factors), those having helix–turn–helix motifs (includes all homeo domains), and those having beta-scaffold factors with minor groove contacts (includes the MADS box factor that figures prominently in plant development).

- More about the genes that control **mating type** in *Saccharomyces cerevisiae* can be retrieved at this site by examining the map of chromosome 3 and clicking on *HML*, *MAT*, and *HMR*. See whether you can follow the links to the Entrez Protein database to find the amino acid sequence of each of the regulatory proteins **a**1, α1, and α2.

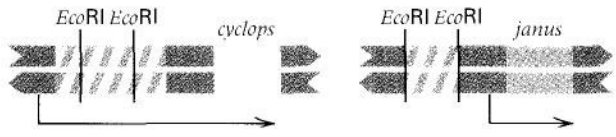
- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 12 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 12.



sequences are coded by the wildtype and mutant attenuators? What phenotype would you expect of the mutant?

12.33 In a species of insect, the genes *cyclops* and *janus* have different tissue-specific patterns of expression; *cyclops* is expressed only in photoreceptor cells of the eye and *janus* only in the testes. Near its 5′ end *cyclops* contains an intron, indicated in the diagram by the hatching, within which there are two *Eco*RI restriction sites. When the *Eco*RI fragment is removed and inserted at an *Eco*RI site immediately upstream of the *janus* promoter, *janus* is transcribed in photoreceptor cells as well as in the testes. The same result occurs irrespective of the orientation in which the *Eco*RI fragment is inserted. What hypothesis can explain this result? How could the hypothesis be tested?



Further Reading

Baylin, S. B., and J. G. Herman. 2000. DNA hypermethylation in tumorigenesis: Epigenetics joins genetics. *Trends in Genetics* 16: 168.

Cairns, B. R. 1998. Chromatin remodeling machines: Similar motors, ulterior motives. *Trends in Biochemical Sciences* 23: 20.

Carlson, M. 1997. Genetics of transcriptional regulation in yeast: Connections to the RNA polymerase II CTD. *Annual Review of Cell Developmental Biology* 13: 1–23.

Czaplinski K., M. J. Ruis-Echevarria, C. I. Gonzalez, and S. M. Peltz. 1999. Should we kill the messenger? The role of the surveillance complex in translation termination and mRNA turnover. *Bioessays* 21: 685.

Engel, J. D., and K. Tanimoto. 2000. Looping, linking, and chromatin activity: New insights into beta-globin locus regulation. *Cell* 100: 499.

Feil, R., and S. Khosla. 1999. Genomic imprinting in mammals: An interplay between chromatin and DNA methylation? *Trends in Genetics* 15: 431.

Fire, A. 1999. RNA-triggered gene silencing. *Trends in Genetics* 15: 358.

Fugmann, S. D., A. I. Lee, P. E. Shockett, I. J. Villey, and D. G. Schatz. 2000. The RAG proteins and V(D)J recombination: Complexes, ends, and transposition. *Annual Review of Immunology* 18: 495.

Gellert, M. 1992. V(D)J recombination gets a break. *Trends in Genetics* 8: 408.

Gralla, J. D. 1996. Activation and repression of *E. coli* promoters. *Current Opinion in Genetics & Development* 6: 526.

Grunberg-Manago, M. 1999. Messenger RNA stability and its role in control of gene expression in bacteria and phages. *Annual Review of Genetics* 33: 193.

Haber, J. E. 1998. Mating-type gene switching in *Saccharomyces cerevisiae*. *Annual Review of Genetics* 32: 561.

Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiology and Molecular Biology Reviews* 62: 465.

Jacob, F., and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318.

Kornberg, R. D. 1999. Eukaryotic transcriptional control. *Trends in Biochemical Sciences* 24: M46.

Laird, P. W., and R. Jaenisch. 1996. The role of DNA methylation in cancer genetics and epigenetics. *Annual Review of Genetics* 30: 441.

Marmorstein, R., M. Carey, M. Ptashne, and S. C. Harrison. 1992. DNA recognition by GAL4: Structure of a protein–DNA complex. *Nature* 356: 408.

Miller, J., and W. Reznikoff, eds. 1978. *The Operon*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Neidhardt, F. C., R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger, eds. 1996. Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (2 volumes). 2d ed. Washington, DC: American Society for Microbiology.

Novina, C. D., and A. L. Roy. 1996. Core promoters and transcriptional control. *Trends in Genetics* 12: 351.

Ptashne, M. 1992. *A Genetic Switch*. 2d ed. Cambridge, MA: Cell Press.

Ptashne, M., and A. Gann. 1997. Transcriptional activation by recruitment. *Nature* 386: 569.

Staudt, L. M., and P. O. Brown. 2000. Genomic views of the immune system. *Annual Review of Immunology* 18: 829.

Struhl, K. 1995. Yeast transcriptional regulatory mechanisms. *Annual Review of Genetics* 29: 651.

Struhl, K. 1996. Chromatin structure and RNA polymerase II connection: Implications for transcription. *Cell* 84: 179.

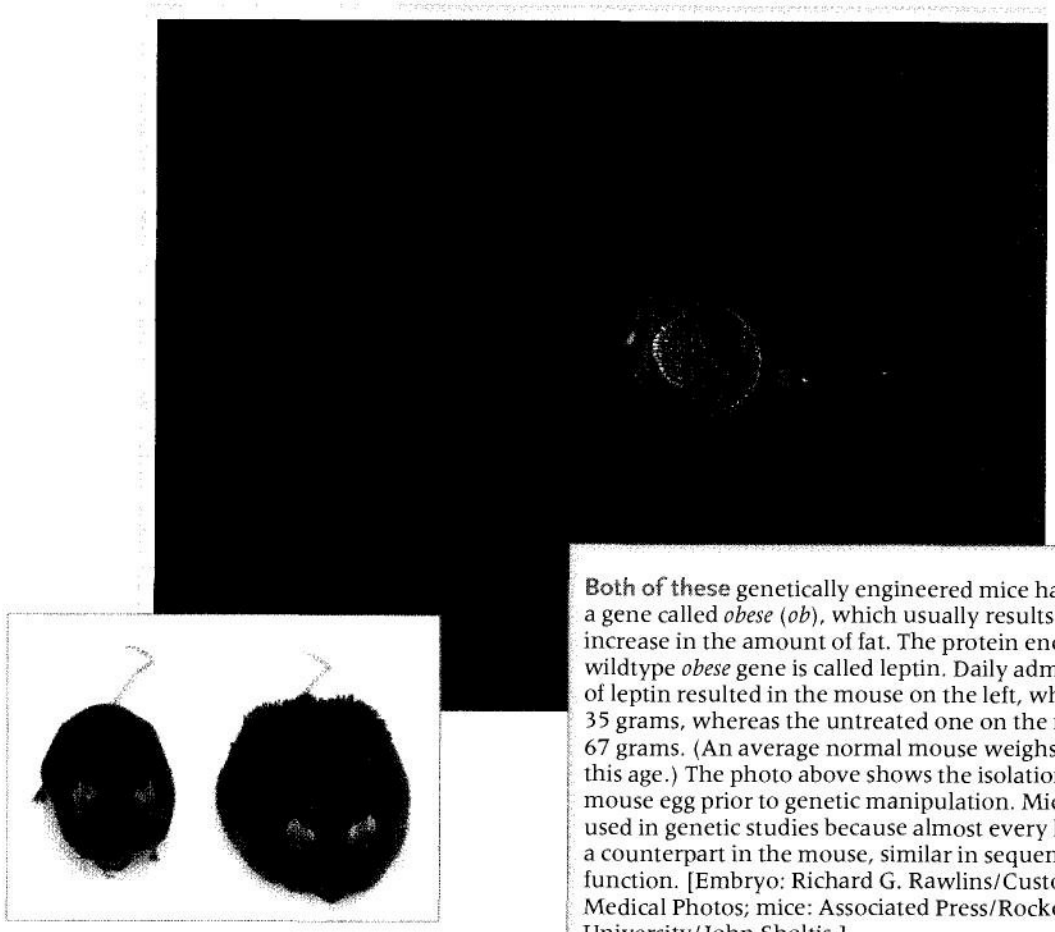
Tijan, R. 1995. Molecular machines that control genes. *Scientific American*, February.

Ullman, A., and A. Danchin. 1980. Role of cyclic AMP in regulatory mechanisms of bacteria. *Trends in Biochemical Sciences* 5: 95.

Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. *Nature* 289: 751.

CHAPTER 13

Genetic Engineering and Genomics



Both of these genetically engineered mice have a defect in a gene called *obese* (*ob*), which usually results in a marked increase in the amount of fat. The protein encoded in the wildtype *obese* gene is called leptin. Daily administration of leptin resulted in the mouse on the left, which weighs 35 grams, whereas the untreated one on the right weighs 67 grams. (An average normal mouse weighs 24 grams at this age.) The photo above shows the isolation of a fertilized mouse egg prior to genetic manipulation. Mice are often used in genetic studies because almost every human gene has a counterpart in the mouse, similar in sequences and basic function. [Embryo: Richard G. Rawlins/Custom Stock Medical Photos; mice: Associated Press/Rockefeller University/John Sholtis.]

CHAPTER OUTLINE

13.1 Restriction Enzymes and Vectors
- Production of Defined DNA Fragments
- Recombinant DNA Molecules
- Plasmid, Lambda, and Cosmid Vectors

13.2 Cloning Strategies
- Joining DNA Fragments
- Insertion of a Particular DNA Molecule Into a Vector
- The Use of Reverse Transcriptase: cDNA and RT-PCR

13.3 Dectection of Recombinant Molecules
- Gene Inactivation in the Vector Molecule
- Screening for Particular Recombinants
- Positional Cloning

13.4 Transgenic Organisms
- Germ-Line Transformation in Animals
- Genetic Engineering in Plants
- Transformation Rescue

13.5 Site-Directed Mutagenesis and Knockout Mutations

13.6 Some Applications of Genetic Engineering
- Giant Salmon with Engineered Growth Hormone
- Nutritionally Engineered Rice
- Production of Useful Proteins
- Genetic Engineering with Animal Viruses

13.7 Analysis of Complex Genomes
- Electrophoretic Separation of Large DNA Fragments
- Cloning of Large DNA Fragments
- Physical Mapping
- High-Throughput Genomic Sequencing and the Human Genome Project
- What Can We Learn from Genomic Sequencing?

13.8 Functional Genomics
- DNA Chips

PRINCIPLES

- In recombinant DNA (gene cloning), DNA fragments are isolated, inserted into suitable vector molecules, and introduced into bacteria or yeast cells where they are replicated.
- Transgenic organisms carry DNA sequences that have been introduced by germ-line transformation or other methods.
- Recombinant DNA is widely used in research, medical diagnostics, and the manufacture of drugs and other commercial products.
- Specialized methods for manipulating and cloning large fragments of DNA have resulted in integrated physical and genetic maps of the DNA in complex genomes.
- High-throughput automated DNA sequencing has resulted in the complete sequence of the genomes of many species of bacteria, several archaeons, *S. cerevisiae*, *C. elegans*, and *D. melanogaster*, as well as a "rough draft" of the genomic sequence of *Homo sapiens* with some chromosomes sequenced to completion. Genomic sequencing of many other organisms is well under way.
- Functional genomics using DNA chips enables the level of gene expression of all genes in the genome to be assayed simultaneously, which allows global patterns and coordinated regulation of gene expression to be investigated.

CONNECTIONS

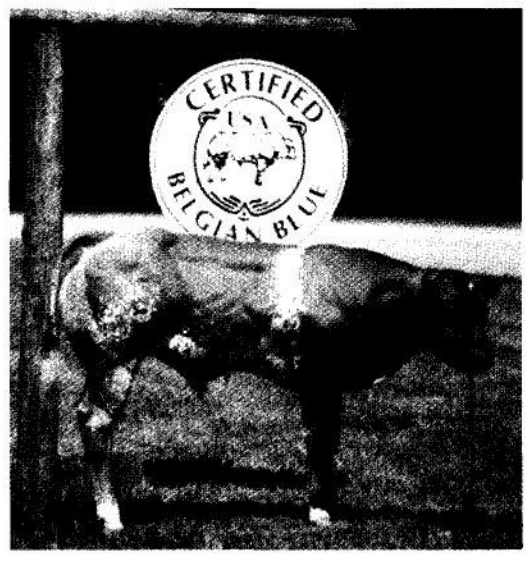
Hello, Dolly!
Ian Wilmut, Angelika E. Schnieke, Jim McWhir, Alex J. Kind, and Keith H. S. Campbell 1997
Viable Offspring Derived from Fetal and Adult Mammalian Cells

Catch-22
Ian Dunham, Nobuyoshi Shimizu, Bruce A. Roe, Stephanie L. Chissoe, and 213 other investigators
The DNA Sequence of Human Chromosome 22

Key techniques used in the isolation and characterization of particular DNA fragments were examined in Chapter 2 in the context of identifying DNA markers for genetic analysis. These same techniques also play an important role in the deliberate manipulation of genes to create genetically engineered organisms with novel genotypes and phenotypes. This approach is called **genetic engineering, gene cloning,** or **recombinant DNA.** The basic technique is quite simple: DNA is isolated and cut into fragments by one or more restriction enzymes; then the fragments are joined together in a new combination and introduced back into a cell or organism to change its genotype in a directed, predetermined way. Such genetically engineered organisms are called *transgenic organisms*. They may be created for experimental studies or for economic reasons, such as to improve varieties of domesticated animals and crop plants or to create organisms that produce pharmaceuticals for use in the treatment of human disease. Specific examples are considered later in this chapter.

13.1 Restriction Enzymes and Vectors

The technology of recombinant DNA relies heavily on *restriction enzymes*, which were examined in Section 2.3 with reference to their site-specific DNA cleavage in the generation of DNA fragments with defined ends. In recombinant DNA, after the DNA has been cleaved by a restriction enzyme, the restriction fragments are joined with a **vector** molecule, which is a (usually circular) DNA molecule into which a DNA fragment can be introduced and which can replicate in a suitable host organism (Figure 13.1). The bacterium *Escherichia coli* and the budding yeast *Saccharomyces cerevisiae* are typical host organisms for recombinant DNA. In this section, we take a closer look at the special features of restriction enzymes that make them so useful in DNA cloning. Some of the most popular types of vectors used in recombinant DNA are also examined.



A single mutant gene results in doubled muscles in the Belgian Blue breed of cattle, shown here. The overdeveloped muscles are tender and low in fat. Introduction of this mutant gene into other breeds could improve their meat quality. [Courtesy Shiloh Creek Farm, Fresno, Ohio.]

Production of Defined DNA Fragments

A restriction enzyme is a nuclease that cleaves duplex DNA wherever the DNA molecule contains a particular short sequence of nucleotides matching the *restriction site* of the enzyme (Section 2.3). The backbone of each strand is cleaved, creating a free 3' hydroxyl group (3'-OH) and a free 5' phosphate group (5'-P). Most restriction sites consist of four or six nucleotides. Several hundred restriction enzymes, each with a different restriction site, have been isolated from microorganisms. Most restriction sites are symmetrical in the sense that the recognition sequence is identical in both strands of the DNA duplex. For example, the restriction enzyme *Eco*RI, isolated from *Escherichia coli*, has the restriction site 5'-GAATTC-3'. The sequence of the other strand is 3'-CTTAAG-5', which is identical but is written with the 3' end at the left. This kind of symmetry in a short sequence of DNA is called a *palindrome*. In double-stranded DNA, *Eco*RI cleaves each strand of the palindrome between the G and the A.

Observations with the electron microscope, made soon after restriction enzymes were discovered, indicated that the fragments produced by many restriction enzymes could spontaneously form circles. The circles could be made linear again by heating; however, if after circularization they were treated with the enzyme **DNA ligase,** which joins 3'-OH and 5'-P groups,

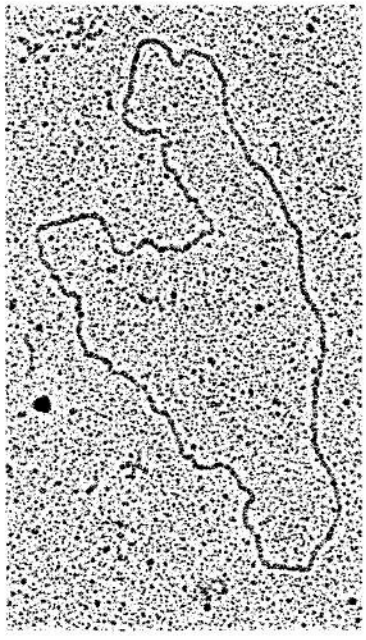


Figure 13.1 Electron micrograph of a circular plasmid used as a vector for cloning in *E. coli*.

then the ends became covalently joined. This observation was the first evidence for three important features of restriction enzymes:

- Restriction enzymes make breaks in symmetrical sequences (palindromes).
- The breaks need not be directly opposite one another in the two DNA strands.
- Enzymes that cleave the DNA strands asymmetrically generate DNA fragments with single-stranded ends that have complementary base sequences.

These properties are illustrated for *Eco*RI in **Figure 13.2**.

Most restriction enzymes are like *Eco*RI in that the cuts in the DNA strands are staggered; the resulting single-stranded ends are called **sticky ends**, and they can adhere to each other because they have complementary nucleotide sequences. Some restriction enzymes, including *Eco*RI, leave a single-stranded overhang at the 5′ end

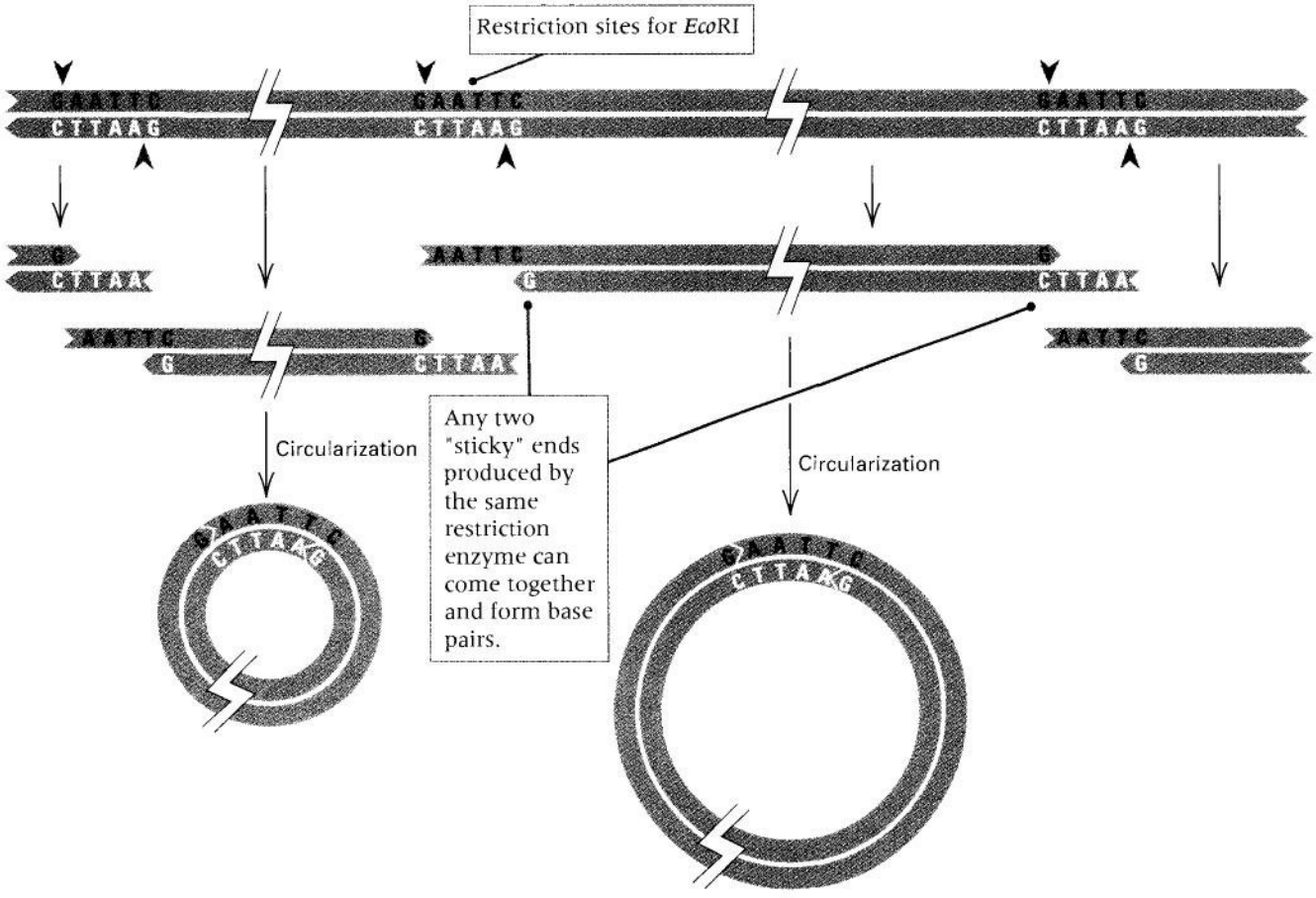


Figure 13.2 Circularization of DNA fragments produced by a restriction enzyme. The red arrowheads indicate the *Eco*RI cleavage sites.

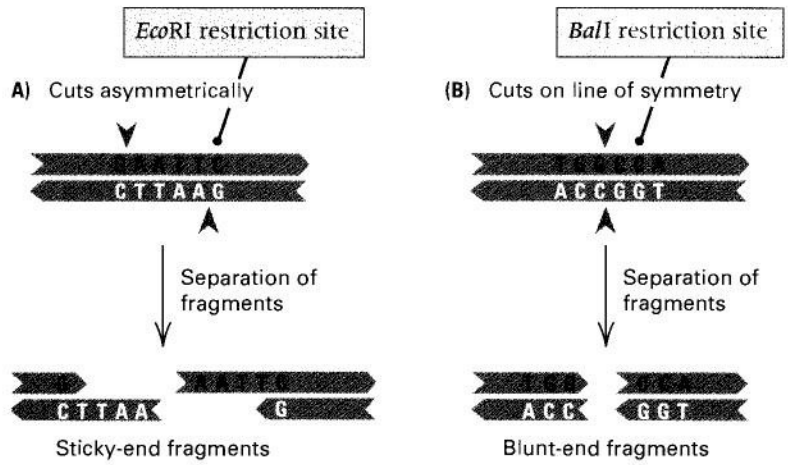


Figure 13.3 Two types of cuts made by restriction enzymes. The red arrowheads indicate the cleavage sites. (A) Cuts made in each strand at an equal distance from the center of symmetry of the restriction site. In this example with the enzyme *Eco*RI, the resulting molecules have complementary 5' overhangs. Other restriction enzymes produce fragments with complementary 3' overhangs. (B) Cuts made in each strand at the center of symmetry of the restriction site. The products have blunt ends. The specific enzyme in this example is *Bal*I.

(**Figure 13.3**, part A); others leave a 3' overhang. A number of restriction enzymes cleave both DNA strands at the center of symmetry, forming **blunt ends.** Figure 13.3B shows the blunt ends produced by the enzyme *Bal*I. Blunt ends also can be joined via a ligase obtained from bacteriophage T4 (*T4 ligase*), which differs from most other ligases in that it can join blunt ends as well as single-stranded nicks in the backbone of duplex DNA. However, whereas ligation of sticky ends recreates the original restriction site, any blunt end can join with any other blunt end and not necessarily create a restriction site.

Most restriction enzymes recognize their restriction sequence without regard to the source of the DNA:

> DNA fragments obtained from one organism will have the same sticky ends as the fragments from another organism if the fragments are produced by the same restriction enzyme.

This principle is one of the foundations of recombinant DNA technology.

Because restriction enzymes cleave only at their restriction site, the number of cuts made in the DNA of an organism by a particular enzyme is limited. For example, an *E. coli* DNA molecule contains 4.6×10^6 base pairs, and any enzyme that cleaves a six-base restriction site will cut the molecule into about a thousand fragments; the reason is that, with equal and random frequencies of each of the four nucleotides, a six-base restriction site is expected to occur, on average, every $4^6 = 4096$ base pairs. Human DNA, with a genome size of 3×10^9 base pairs, would be cut into about a million fragments. These numbers, though large, are still small compared with the number of fragments that would be produced by random shearing of the DNA. Of special interest are cleavage products from smaller DNA molecules that may have only a few cleavage sites (or possibly none) for any particular enzyme, such as viral DNA or plasmid DNA. A **plasmid** is a circular DNA molecule present in a cell, usually a bacterial cell, that is capable of replicating independently of the bacterial chromosome. Different plasmids range in size from about 1 to100 kb. As we shall see shortly, small plasmids containing a single cleavage site for a particular restriction enzyme are especially valuable as vectors in DNA cloning.

Recombinant DNA Molecules

In recombinant DNA, a particular DNA segment of interest is joined to a vector DNA molecule. This recombinant molecule is introduced into a cell by means of a *transformation* procedure very similar to that used by Avery, MacLeod, and McCarty in proving that DNA is the genetic material

(Chapter 1). After transformation, as the cellular DNA replicates, so does the recombinant DNA (**Figure 13.4**). When a stable transformant has been isolated, the DNA sequences linked to the vector are said to be **cloned.** In the following section, several types of vectors are described.

Plasmid, Lambda, and Cosmid Vectors

The most generally useful vectors have three properties:

- The vector DNA can be introduced into a host cell.
- The vector and any DNA it contains can be replicated inside the host cell.
- Cells containing the vector can be identified in a straightforward manner, most conveniently through a novel phenotype conferred on the host by DNA sequences present in the vector (for example, antibiotic resistance).

At present, the most commonly used vectors are *E. coli* plasmids and derivatives of the *E. coli* bacteriophages lambda (λ) and M13. Other plasmids and viruses also have

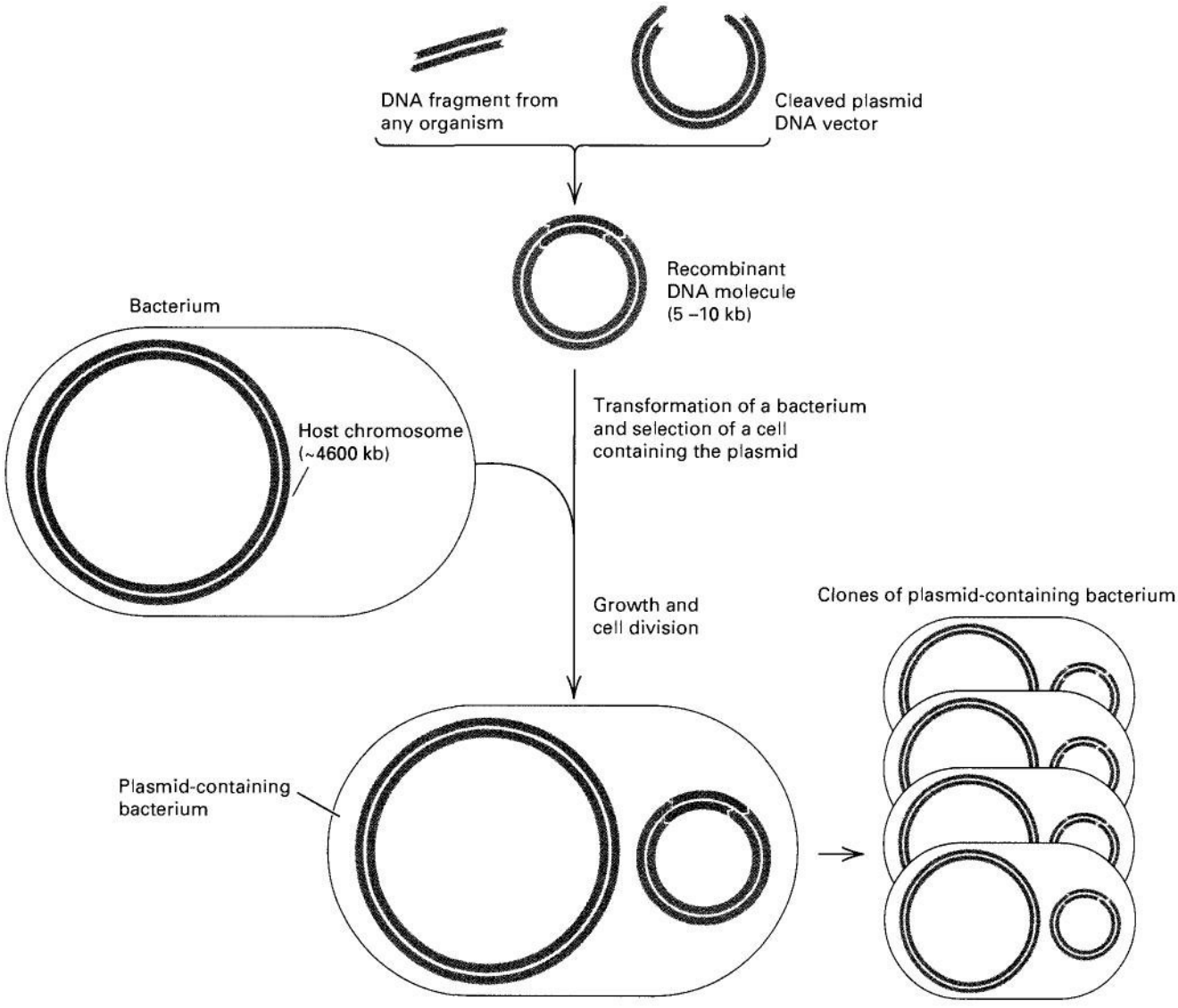


Figure 13.4 An example of cloning. A fragment of DNA from any organism is joined to a cleaved plasmid. The recombinant plasmid is then used to transform a bacterial cell, where the recombinant plasmid is replicated and transmitted to the progeny bacteria. The bacterial host chromosome is not drawn to scale. It is typically about 1000 times larger than the plasmid.

been developed for cloning into cells of animals, plants, and bacteria. Plasmid and phage DNA can be introduced into cells by a transformation procedure in which cells are enabled to take up free DNA by exposure to a $CaCl_2$ solution. Recombinant DNA can also be introduced into cells by a kind of electrophoretic procedure called **electroporation**. After introduction of the DNA, the cells containing the recombinant DNA are detected by means of vector-encoded functions that become manifest during bacterial growth. For example, *E. coli* cells that have been successfully transformed with a vector that encodes resistance to an antibiotic, such as tetracycline, will grow on medium containing tetracycline, whereas untransformed cells will not grow. Variants of this procedure are used to transform animal or plant cells with suitable vectors (Section 13.4), but the technical details differ considerably.

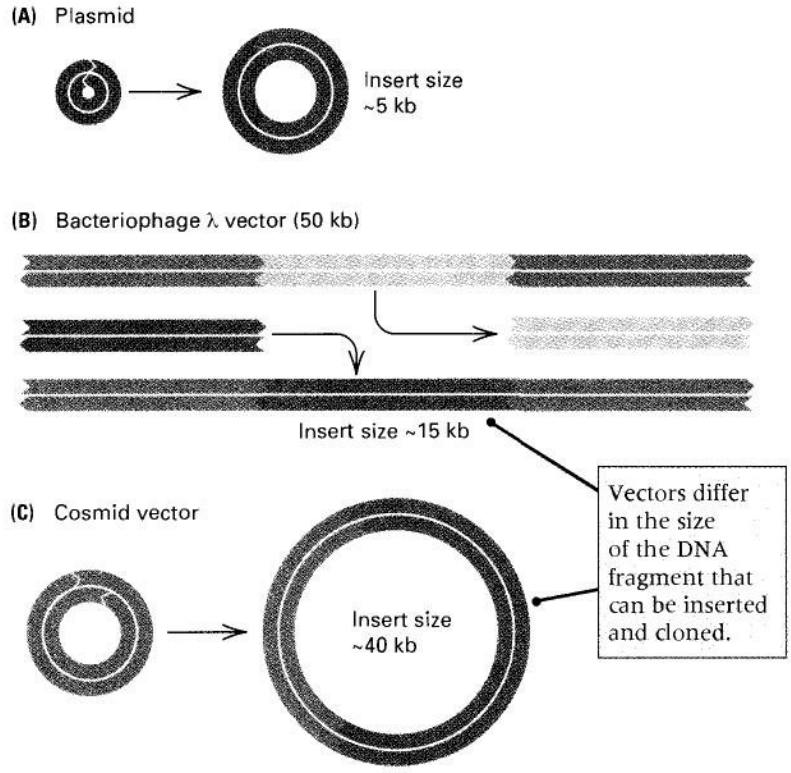


Figure 13.5 Common cloning vectors for use with *E. coli*, not drawn to scale. (A) Plasmid vectors are ideal for cloning relatively small fragments of DNA. (B) Bacteriophage λ vectors contain convenient restriction sites for removing the middle section of the phage and replacing it with the DNA of interest. (C) Cosmid vectors are useful for cloning DNA fragments up to about 40 kb; they can replicate as plasmids but contain the cohesive ends of phage λ and so can be packaged in phage particles.

Three types of vectors commonly used for cloning into *E. coli* are illustrated in **Figure 13.5**. Plasmids (Figure 13.5A) are most convenient for cloning relatively small DNA fragments (5–10 kb). Somewhat larger fragments can be cloned with bacteriophage λ (Figure 13.5B). The normal phage λ genome is approximately 50 kb in length, but because the central portion of the genome is not essential for infection and phage propagation, this portion can be removed and replaced with donor DNA. After the donor DNA has been ligated in place, the recombinant DNA is packaged into mature phage *in vitro*, and the recombinant phage are used to infect bacterial cells. However, to be packaged into a phage head, the recombinant DNA must be neither too large nor too small, which means that the donor DNA must be roughly the same size as the portion of the λ genome that was removed. Most λ cloning vectors accept inserts ranging in size from 12 to 20 kb. Still larger DNA fragments can be inserted into cosmid vectors (Figure 13.5C). These vectors can exist as plasmids, but they also contain the phage single-stranded sticky ends of 12 bp known as *cohesive ends* (Section 10.7). Cosmid clones can be packaged into mature phage particles by virtue of the cohesive ends, which are cleaved to enable the phage DNA to be packaged into the phage heads. The size limitation on cosmid inserts usually ranges from 40 to 45 kb. Vectors for cloning even larger fragments of DNA are referred to as *artificial chromosomes* and are discussed in Section 13.7.

13.2 Cloning Strategies

In genetic engineering, the immediate goal of an experiment is usually to insert a *particular* fragment of chromosomal DNA into a plasmid or a viral DNA molecule. Any strategy for achieving this goal must include several steps:

1. Purification of donor DNA and vector DNA
2. Cleavage with one or more restriction enzymes
3. Joining donor DNA with vector DNA
4. Identification and isolation of desired recombinant clones

Steps 1 and 2 were considered in Section 13.1. Here we examine steps 3 and 4.

Joining DNA Fragments

The circularization of restriction fragments having terminal single-stranded, "sticky" ends with complementary bases was described in Section 13.1. Because a particular restriction enzyme produces fragments with *identical* sticky ends, without regard for the source of the DNA, fragments from DNA molecules isolated from two different organisms can be joined, as shown in **Figure 13.6.** In this example, the restriction enzyme *Eco*RI is used to digest DNA from any organism of interest and to cleave a bacterial plasmid that contains only one *Eco*RI restriction site. The donor DNA is cleaved into many fragments (one of which is shown), and the plasmid is cleaved into a single linear fragment. When the donor fragment and the linearized plasmid are mixed, recombinant molecules can form by base pairing between the complementary single-stranded ends. At this point, the DNA is treated with DNA ligase to seal the joints, and the donor fragment becomes covalently joined to the vector. The ability to join a donor DNA fragment of interest to a vector is the basis of the recombinant DNA technology.

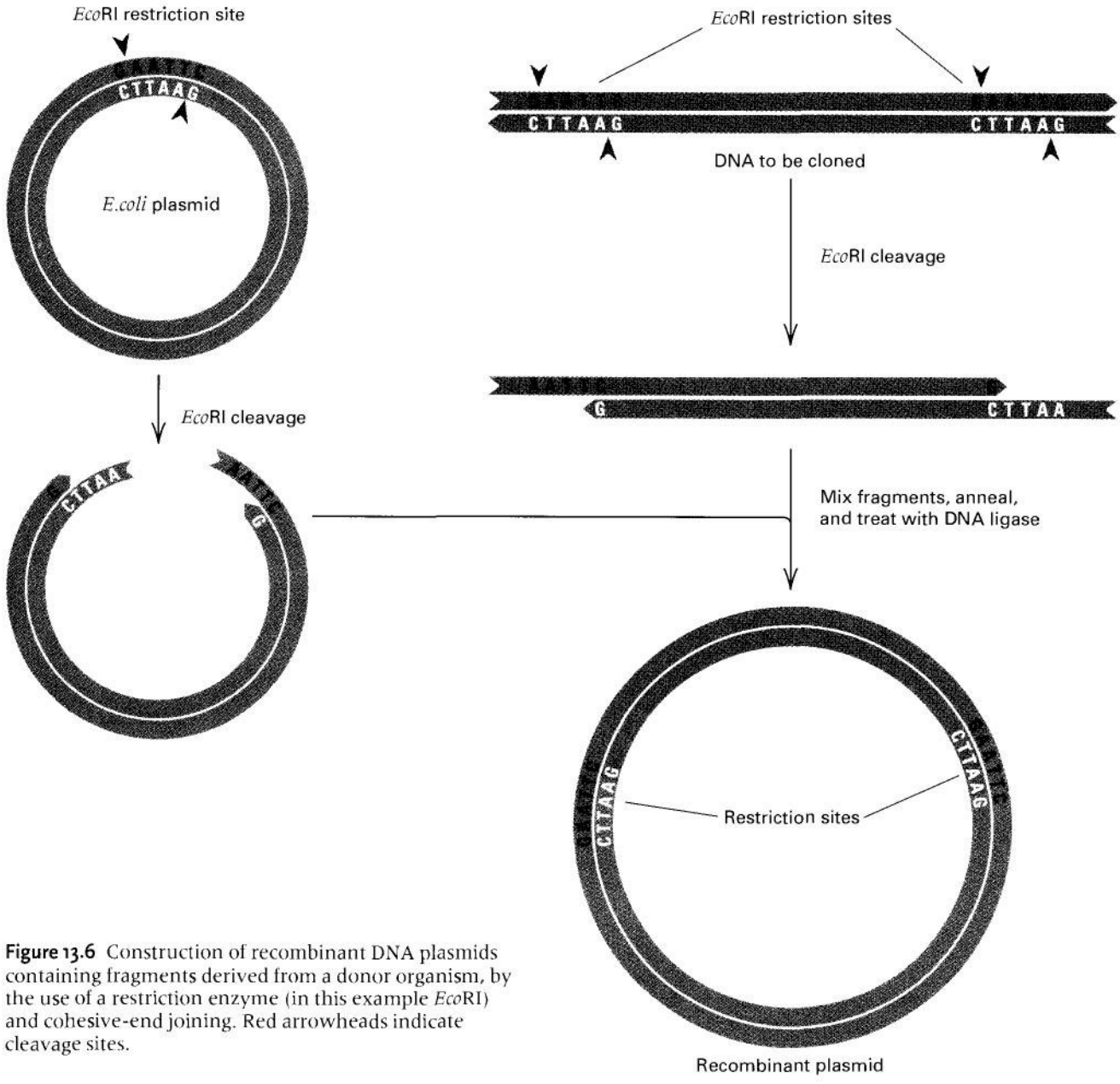


Figure 13.6 Construction of recombinant DNA plasmids containing fragments derived from a donor organism, by the use of a restriction enzyme (in this example *Eco*RI) and cohesive-end joining. Red arrowheads indicate cleavage sites.

Because any sticky end can base-pair with any other complementary sticky end, the joining of restriction fragments at random can produce a variety of products, including many that are useless. For example, consider a linear DNA molecule that is cleaved into four fragments—A, B, C, and D—present in the original molecule in the order A B C D (**Figure 13.7**). Reassembly of the fragments can occasionally yield the original molecule, but if B and C have the same pair of sticky ends, then molecules with different arrangements of the fragments are also formed, including arrangements in which one or more of the restriction fragments are inverted in orientation (indicated in Figure 13.7 by the reversed letters). Ligation of restriction fragments with blunt ends produces even more combinations, because any blunt end can be joined with any other blunt end.

Restriction fragments from the vector, if there are more than one, can also join together in the wrong order, but this potential

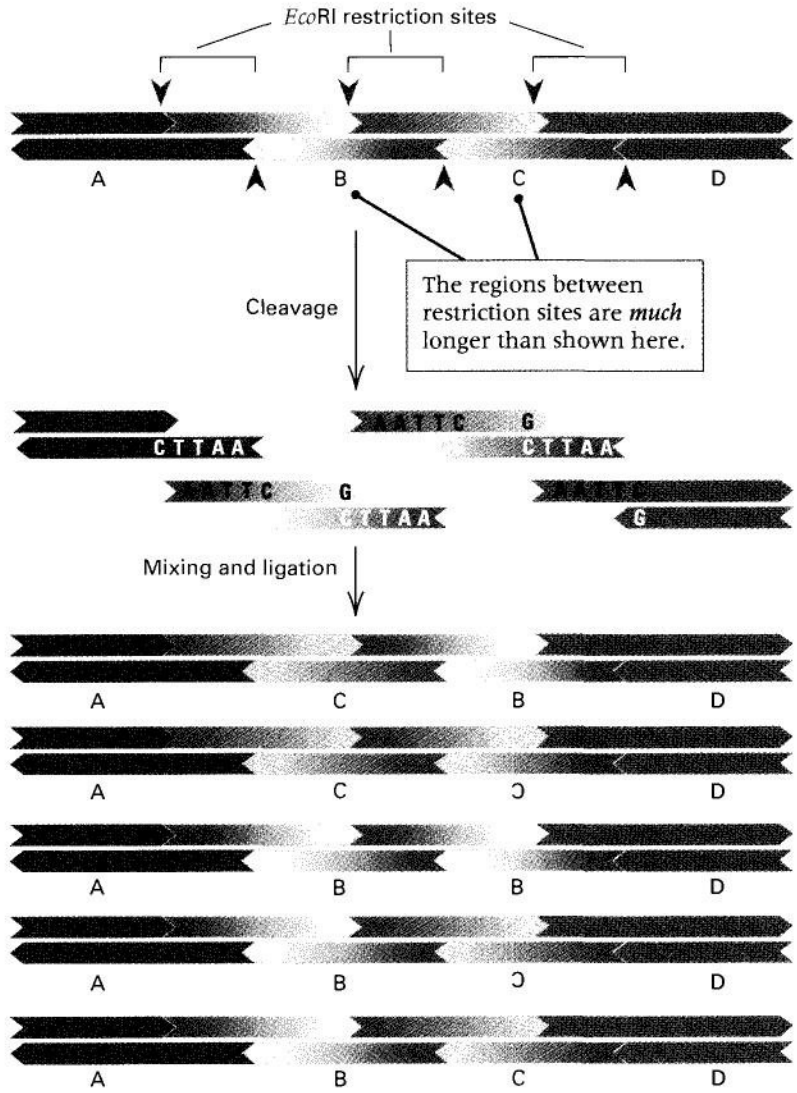


Figure 13.7 Fragments produced by a restriction enzyme can be rejoined in arbitrary ways because the ends of the cleavage sites are compatible. In this example, a DNA molecule is cut into four different fragments (A, B, C, and D) by a restriction enzyme. A few examples of how these can be rejoined are shown. Even among the rejoined molecules that have A and D at the ends and two fragments in the middle, many have one fragment represented twice and the other not at all, or they have a fragment present in the reverse orientation. Only occasionally does a rejoining recreate the original order A–B–C–D.

problem can be eliminated by the use of a vector that has only one cleavage site for a particular restriction enzyme. When a circular molecule has only one cleavage site for a restriction enzyme, as in Figure 13.6, cleavage occurs only at this one site, and any other DNA fragment of interest, if it has complementary ends, can be inserted into the gap. Many plasmids with single cleavage sites are available (most have been created by genetic engineering). Many vectors contain unique sites for several different restriction enzymes, but generally only one or two enzymes are used at a time.

Insertion of a Particular DNA Molecule Into a Vector

In the cloning procedure described so far, a collection of fragments obtained by digestion with a restriction enzyme can be joined with a cleaved vector molecule, yielding a large number of recombinant molecules containing different fragments of donor DNA. If a particular cloned segment of DNA is desired, then the recombinant molecule containing that particular segment must be isolated from among what is often a large background of other recombinant molecules. The simplest method is direct selection of the desired clone on the basis of a phenotypic attribute that it confers.

As an example of direct selection for the recovery of recombinant molecules containing a particular gene, suppose that we wished to clone a gene used by *E. coli* for the biosynthesis of the amino acid leucine. Direct selection would make use of a *leu*$^-$ mutant strain having a mutation in the gene of interest. The *leu*$^-$ cells would be unable to form colonies on solid medium unless the growth medium contained leucine. To clone the *leu*$^+$ gene from a wildtype strain able to grow without leucine, DNA from the *leu*$^+$ strain is isolated and cleaved with a restriction enzyme, and the resulting fragments are ligated into a plasmid vector that is then used to transform mutant cells of genotype *leu*$^-$. The transformed cells are spread onto solid medium lacking leucine. Any untransformed cells, or cells transformed with DNA not containing the *leu*$^+$ gene, will be unable to grow and form colonies. The medium selects for *leu*$^+$ transformants because the few cells that do contain the *leu*$^+$ gene inserted into the vector will form colonies. This experiment will not work if the *leu*$^+$ gene contains a restriction site for the enzyme used, because cleavage within the gene will destroy its ability to function. For this reason, cloning experiments of this type are usually carried out under conditions of **partial digestion,** which means that the cleavage reaction is terminated when only a fraction of the restriction sites have been cleaved. In this way, even though the *leu*$^+$ gene may contain one or more restriction sites for the enzyme, some of the *leu*$^+$ genes will, by chance, still remain intact after partial digestion.

Direct selection for recombinant clones, such as in the *leu*$^+$ example, is usually not possible. If the clone of interest is either very rare or very difficult to detect, it is often preferable to purify the DNA fragment that contains the gene of interest before joining it with the vector. Two methods of fragment isolation are in common use:

- Amplification by means of the *polymerase chain reaction (PCR),* discussed in Section 2.4, in which short oligonucleotide primers homologous to the ends of the fragment to be isolated are used to replicate the fragment of interest. After a sufficient number of rounds of replication, the resulting solution contains predominantly replicas of the amplified DNA fragment.
- Purification of a restriction fragment known to contain the gene of interest. For example, if a gene is known to be contained within a particular restriction fragment, this fragment can be isolated by purifying the corresponding DNA band from a gel after electrophoresis (Section 2.3) and then joined to an appropriate vector.

Some genes in higher eukaryotes, including many human genes, are very long. They extend for hundreds of kilobases. Any DNA fragment containing the complete gene is too long to be amplified by PCR and is likely to contain multiple cleavage sites for any restriction enzyme. However, such large genes usually produce mRNA molecules of modest size, typically less than 3 kb. If the purpose of cloning is to produce the protein product of the gene, then all the necessary information is contained in the mRNA. How to clone a DNA molecule whose sequence corresponds to a particular mRNA is described next.

The Use of Reverse Transcriptase: cDNA and RT-PCR

Some specialized animal cells contain a high abundance of a specific type of mRNA. For example, cells in the chicken oviduct contain so much mRNA for ovalbumin, an egg-white protein, that mRNA extracted from oviducts is predominantly ovalbumin mRNA. This purified mRNA can serve as the starting point for isolating recombinant clones that contain the coding sequence of ovalbumin. More generally, in the cloning of any gene whose mRNA is highly abundant in a particular type of tissue, purified mRNA from this tissue can serve as a starting point for creating a collection of recombinant plasmids, many of which contain coding sequences that correspond to the abundant mRNA.

Cloning from mRNA molecules makes use of an unusual polymerase, **reverse transcriptase,** which can bind with a single-stranded RNA molecule and use it as a template for the synthesis of a complementary strand of DNA. Reverse trancriptase was originally discovered in RNA tumor viruses, in which single-stranded RNA is the genetic material, where its function is to make double-stranded DNA complementary to the viral RNA that can be integrated into the DNA of the host cell; reverse transcriptase converts the viral genetic information into DNA for replication and transmission to the daughter cells when the host cell divides. Like other DNA polymerases, reverse transcriptase requires a primer. When the enzyme is used in genetic engineering, the poly-A tail usually found at the 3' end of eukaryotic mRNA serves as a convenient priming site, because the primer can be an oligonucleotide consisting of poly-T (**Figure 13.8**). Like any other single-stranded DNA molecule, the single strand of DNA produced from the RNA template can fold back upon itself at the ex-

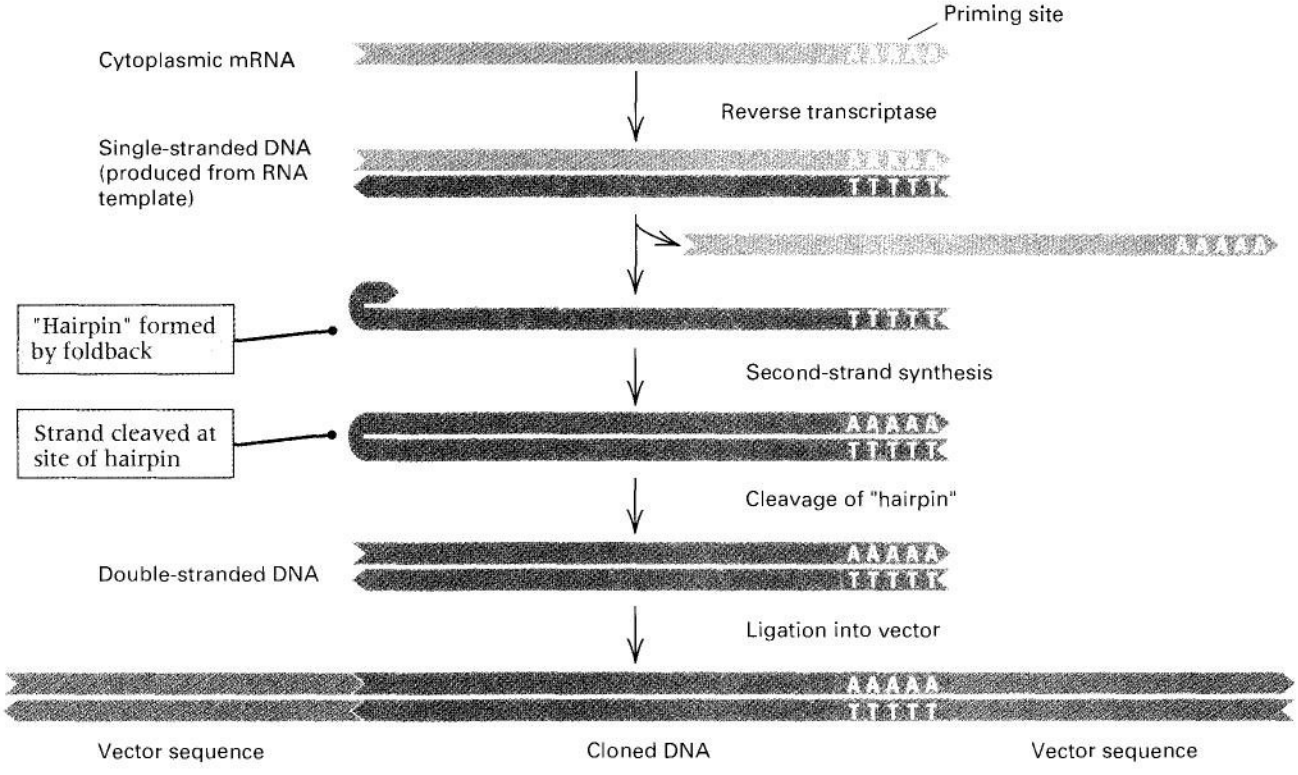


Figure 13.8 Reverse transcriptase produces a single-stranded DNA complementary in sequence to a template RNA. In this example, a cytoplasmic mRNA is copied. As indicated here, most eukaryotic mRNA molecules have a tract of consecutive A nucleotides at the 3' end, which serves as a convenient priming site. After the single-stranded DNA is produced, a foldback at the 3' end forms a hairpin that serves as a primer for second-strand synthesis. After the hairpin is cleaved, the resulting double-stranded DNA can be ligated into an appropriate vector either immediately or after PCR amplification.

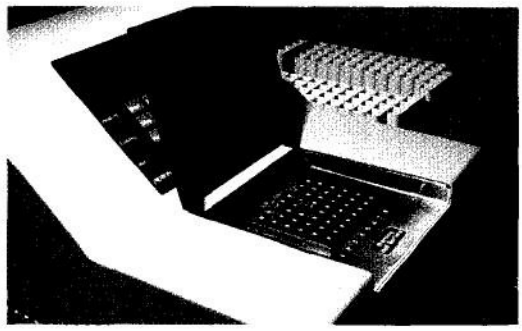
treme 3' end to form a "hairpin" structure that includes a short double-stranded region consisting of a few base pairs. The 3' end of the hairpin serves as a primer for synthesis of the second strand of DNA, forming a duplex DNA molecule. The second strand can be synthesized either by DNA polymerase or by reverse transcriptase. Reverse transcriptase itself produces the second strand in RNA viruses that encode a reverse transcriptase, such as the human immunodeficieny virus (HIV). Conversion into a conventional double-stranded DNA molecule is achieved by cleavage of the hairpin by a nuclease. The DNA complement of an RNA molecule is called **complementary DNA,** or **cDNA.**

In the reverse transcription of an mRNA molecule, the resulting full-length cDNA contains an uninterrupted coding sequence for the polypeptide of interest. Because many eukaryotic genes contain **introns** that are present in the primary transcript but removed in the production of the mature mRNA, the cDNA sequence is not identical to the genomic DNA. However, if the purpose of creating the recombinant DNA molecule is to identify the coding sequence or to synthesize the gene product in a genetically engineered organism, then cDNA formed from processed mRNA is the material of choice. The joining of cDNA to a vector can be accomplished by available procedures for joining blunt-ended molecules (Figure 13.8).

In a typical eukaryotic cell, the levels of mRNA abundance can be separated, very roughly, into three categories:

- About 1 percent of the expressed genes have an *abundant* mRNA, ranging from hundreds of copies to a few thousand copies per cell.
- About 10 percent of the expressed genes have a *moderately abundant* mRNA, ranging from about 10 to100 copies per cell.
- The remaining almost 90 percent of the expressed genes have a *rare* mRNA, ranging from one copy to a few copies per cell.

The cDNA molecules produced from such rare RNAs will also be rare. The efficiency of cloning rare cDNA molecules can be markedly increased by PCR amplification prior to ligation into the vector. The only limitation on the procedure is the requirement that enough DNA sequence be known at both ends of the cDNA for appropriate oligonucleotide primers to be designed. PCR amplification of the cDNA produced by reverse transcriptase is called **reverse transcriptase PCR (RT-PCR).** The resulting amplified molecules contain the coding sequence of the gene of interest with very little contaminating DNA.



A typical machine for carrying out the polymerase chain reaction (PCR) automatically. All of the multiple cycles of temperature shifting are programmable. [T. Bannor/Custom Medical Stock Photo.]

13.3 Detection of Recombinant Molecules

When genomic restriction fragments produced by a restriction enzyme are mixed with a vector cleaved by the same enzyme, many types of recombinant molecules can result—including such examples as a self-joined circular vector that has not acquired any fragments, a vector containing one or more fragments, and a molecule consisting only of many joined fragments. To facilitate the isolation of a vector containing a particular DNA fragment, some means is needed to ensure (1) that the vector does indeed possess an inserted DNA fragment, and (2) that the fragment is in fact the DNA segment of interest. This section describes several useful procedures for detecting recombinant DNA molecules that have the desired characteristics.

Gene Inactivation in the Vector Molecule

When transformation is used to introduce recombinant plasmids into bacterial cells, the initial goal is to isolate bacteria that contain the plasmid from among the resulting mixture of plasmid-free and

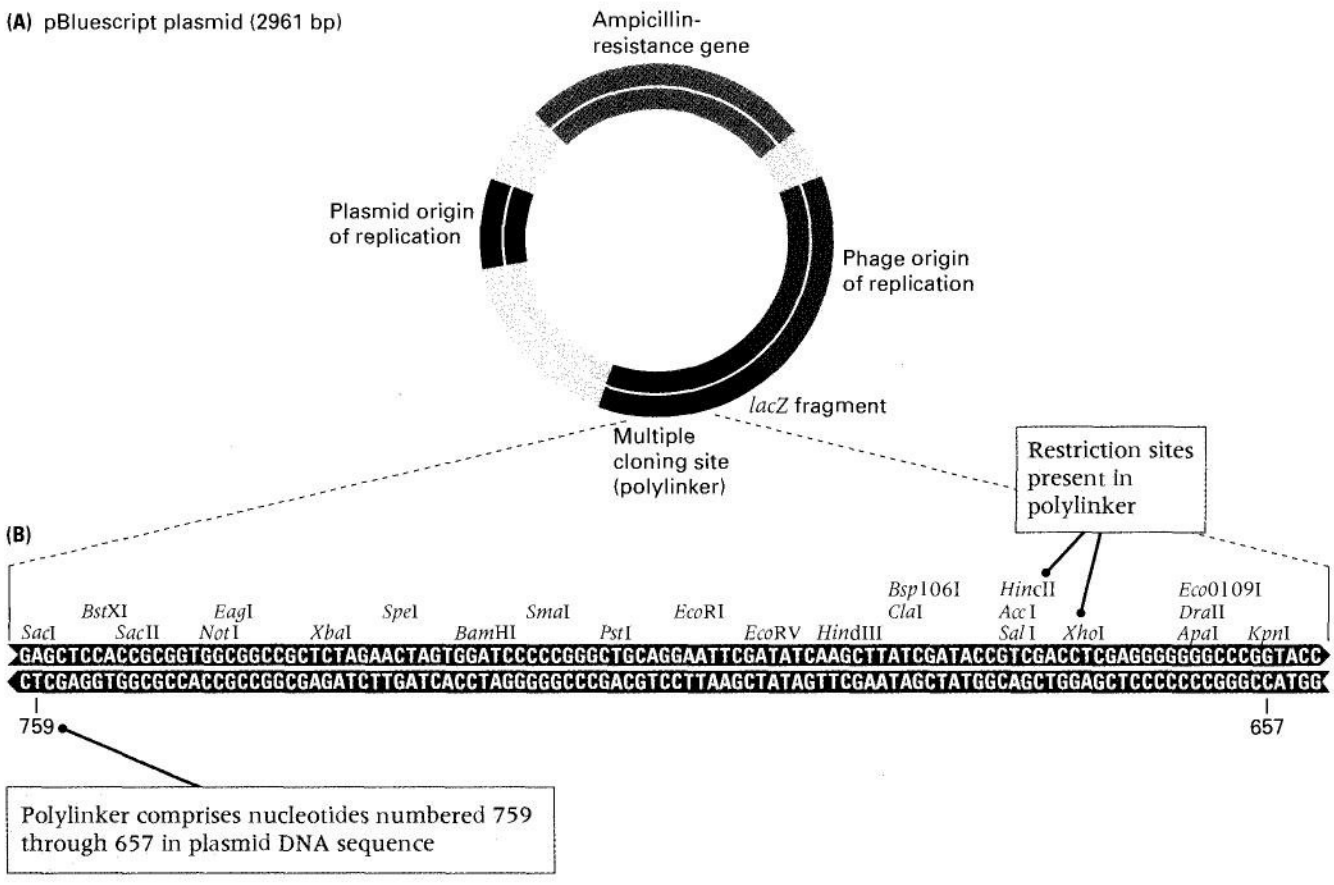


Figure 13.9 (A) Diagram of the cloning vector pBluescript II. It contains a plasmid origin of replication, an ampicillin-resistance gene, a multiple cloning site (polylinker) within a fragment of the *lacZ* gene from *E. coli*, and a bacteriophage origin of replication. (B) Sequence of the multiple cloning site showing the unique restriction sites at which the vector can be opened for the insertion of DNA fragments. The numbers 657 and 759 refer to the position of the base pairs in the complete sequence of pBluescript. [Courtesy of Stratagene Cloning Systems, La Jolla, CA.]

plasmid-containing cells. A common procedure is to use a plasmid that includes an antibiotic-resistance marker and to select the transformed bacteria on a medium containing the antibiotic. Only cells with the plasmid will be able to form colonies. An example of a state-of-the-art cloning vector is the pBluescript plasmid illustrated in **Figure 13.9**A. The entire plasmid is 2961 base pairs. Different regions contribute to its utility as a cloning vector:

- The plasmid *origin of replication* (the position at which DNA replication begins) is derived from the *E. coli* plasmid ColE1. The ColE1 is a high-copy-number plasmid, and its origin of replication enables pBluescript and its recombinant derivatives to exist in approximately 300 copies per cell.
- The ampicillin-resistance gene allows for selection of transformed cells in medium containing ampicillin.
- The cloning site, called a **multiple cloning site (MCS)** or **polylinker,** contains unique cleavage sites for many different restriction enzymes; it allows various types of restriction fragments to be inserted. In pBluescript, the MCS is a 108-bp sequence that contains cloning sites for 23 different restriction enzymes (Figure 13.9B).
- The detection of recombinant plasmids is achieved by means of a region containing the *lacZ* (β-galactosidase) gene from *E. coli*, shown in blue in Figure 13.9A. The basis of the selection is illustrated in **Figure 13.10**. When the *lacZ* region is interrupted by a fragment of DNA inserted into the MCS, the recombinant plasmids yield cells unable to produce the enzyme. Nonrecombinant plasmids lacking a

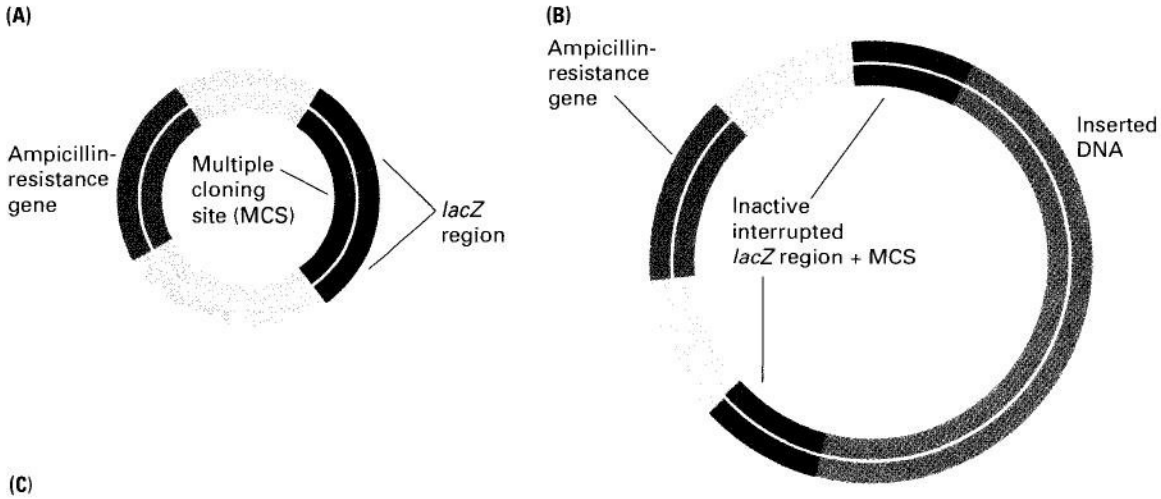


Figure 13.10 Detection of recombinant plasmids through insertional inactivation of a fragment of the *lacZ* gene from *E. coli*. (A) Nonrecombinant plasmid containing an uninterrupted *lacZ* region. The multiple cloning site (MCS) within the region (not drawn to scale) is sufficiently small that the plasmid still confers β-galactosidase activity. (B) Recombinant plasmid with donor DNA inserted into the multiple cloning site. This plasmid confers ampicillin resistance but not β-galactosidase activity, because the donor DNA interrupting the *lacZ* region is large enough to render the region nonfunctional. (C) Transformed bacterial colonies. Cells in the white colonies contain plasmids with inserts that disrupt the lacZ region; those in the blue colonies do not. [C courtesy of Elena R. Lozovsky.]

DNA fragment in the MCS yield cells that do produce the enzyme. The two types of cells can be distinguished by color when the growth medium contains a special β-galactoside compound called X-gal, which releases a deep blue dye when cleaved. On medium containing X-gal, as shown in Figure 13.10C, the colonies whose cells contain nonrecombinant plasmids produce β-galactosidase and turn a deep blue color, whereas colonies whose cells contain recombinant plasmids produce no β-galactosidase and remain the normal white color.

- The bacteriophage origin of replication derives from the single-stranded DNA phage f1. When cells containing a recombinant plasmid are infected with an f1 helper phage, the f1 origin enables a single strand of the inserted fragment, starting with *lacZ*, to be packaged in progeny phage. This feature is very convenient because it yields single-stranded DNA suitable as a template for DNA sequencing (Section 6.7). The plasmid shown in Figure 13.9A is the SK(+) variety. There is also an SK(−) variety in which the f1 origin is in the opposite orientation and packages the complementary DNA strand.

All good cloning vectors have an efficient origin of replication, at least one unique cloning site for the insertion of a DNA fragment, and a second gene whose interruption by inserted DNA yields a phenotype indicative of a recombinant plasmid. Once a **library,** or large set of clones, has been obtained in a particular vector, the next problem is how to identify the particular recombinant clones that contain the gene of interest.

Screening for Particular Recombinants

The procedure of **colony hybridization** makes it possible to detect the presence of any gene for which DNA or RNA labeled with radioactivity or some other means is available (**Figure 13.11**). Colonies to be tested are transferred (*lifted*) from a solid medium onto a nitrocellulose or nylon filter by gently pressing the filter onto the surface. A part of each colony remains on the agar medium, which constitutes the reference plate. The filter is treated with sodium hydroxide (NaOH), which simultaneously breaks open the cells and denatures the DNA. The filter is then saturated with a solution containing a *probe* consisting of labeled DNA or RNA, complementary to the gene being sought, and the cellular DNA is renatured. After washing to remove unbound probe, the positions of the bound probe identify the desired colonies. For example, with radioactively labeled probe, the desired colonies are located by means of autoradiography. A similar assay is done with bacterial cells that have been infected with phage vectors.

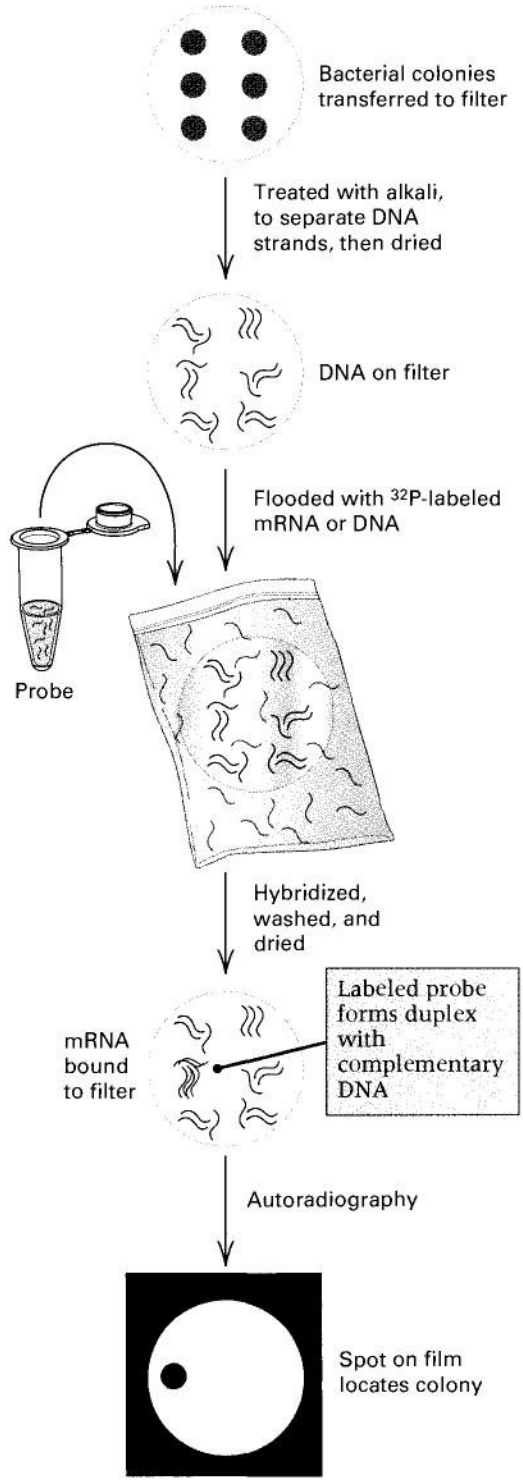


Figure 13.11 Colony hybridization.

If transformed cells can synthesize the protein product of a cloned gene or cDNA, then immunological techniques may allow the protein-producing colony to be identified. In one method, the colonies are transferred as in colony hybridization, and the transferred copies are exposed to a labeled antibody directed against the particular protein. Colonies to which the antibody adheres are those that contain the gene of interest.

Positional Cloning

Sometimes it is desired to clone a gene—for example, a genetic risk factor for an inherited disease—when nothing is known about the gene other than its position in the genetic map relative to other genetic markers, such as DNA markers of the type discussed in Chapter 2. The positional information contained in the genetic map can be of great utility in cloning, because in many organisms the genetic map is very densely populated with genetic markers. For example, in the current genetic map of the human genome, adjacent markers are spaced at an average of approximately 100 kb. The use of information from the genetic map in order to clone a gene is called **positional cloning** or **map-based cloning.**

The procedure for positional cloning has four steps:

1. Determine the genetic map position of the mutant gene as precisely as possible, using molecular markers linked to it in the same chromosome. This is done from pedigree data by using methods explained in Chapter 5.
2. For molecular markers (typically DNA polymorphisms) flanking the mutant gene, isolate clones that contain the marker sequences.
3. Verify that the DNA inserts in the isolated clones are mutually overlapping in such a way as to cover the entire region between the flanking markers without any gaps. Such a mutually overlapping set of clones is called a **contig** because adjacent clones are contiguous ("touching"). If the initial set of clones does include one or more gaps, then complete the contig by screening the clone library with additional probes derived from the ends of the clones flanking each gap.
4. Identify the clone or clones in the contig that contain the gene of interest. This usually entails determining the nucleotide sequence of all or part of the contig in wildtype and mutant chromosomes to identify the site of the mutation.

Note that the cloning procedure outlined in these steps requires no advance knowledge of the function of the gene of interest. All the method requires is that the gene be mapped genetically with respect to molecular markers, which in turn are used to isolate a contig of clones covering the relevant region. Positional cloning is one of the most important approaches to cloning genes in species that have a dense genetic map, and it is particularly important in human molecular genetics.

In some organisms, contigs of clones are already available that cover the entire genome. These organisms include bacteria, such as *E. coli*; fungi, including *S. cerevisiae;* the nematode *C. elegans;* the fruit fly *D. melanogaster;* plants, such as *A. thaliana;* the mouse *Mus musculus;* human beings, *Homo sapiens;* and many others. The utility of such contigs is illustrated for a 650-kb region of the *Drosophila* genome in **Figure 13.12**. The blue bars depict clones, each containing a large insert of approximately 85 kb of *Drosophila* DNA. The red lines represent the positions of molecular markers in the region, spaced at an average distance of about 30 kb, and the points of intersection with the clones show the presence of each marker within each clone. Seven cloned genes in the region are indicated at the left. In the cloning of any new mutant gene that maps within the region, the location of the mutation with respect to the markers immediately identifies the clone or clones in which the gene is located.

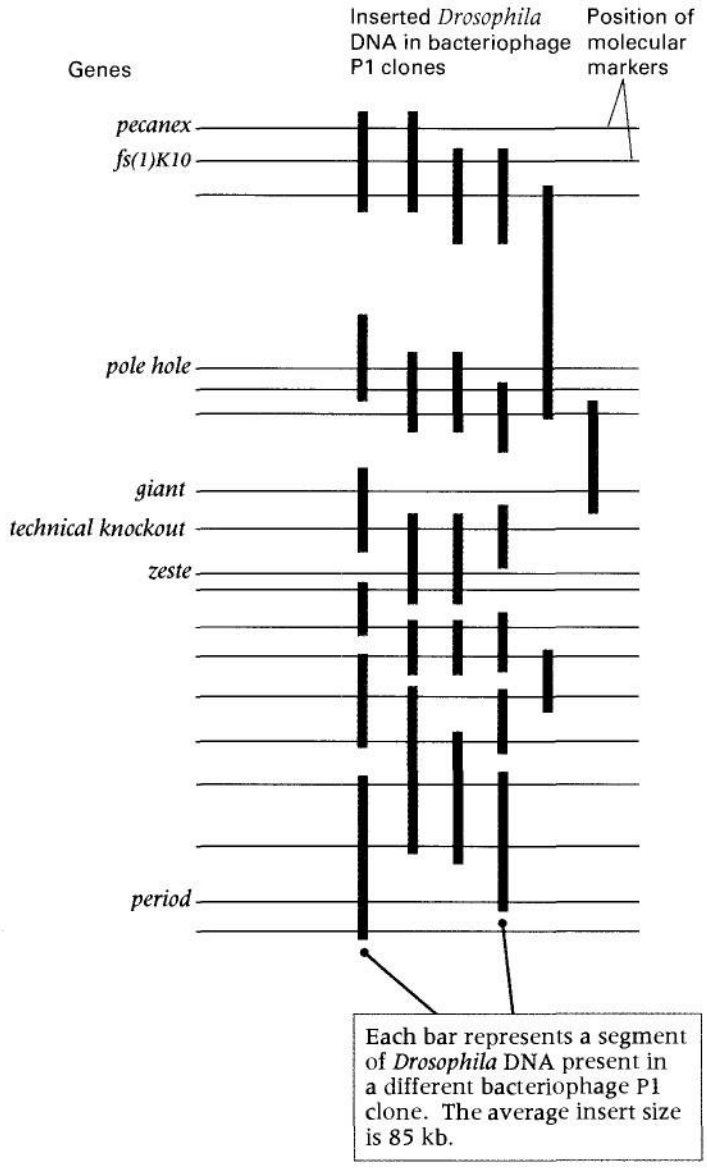


Figure 13.12 Positional cloning in *Drosophila* is made possible by a set of P1 clones that covers virtually the entire genome. Illustrated is a set of P1 clones covering a small region of the X chromosome containing molecular markers (red lines) that have known locations on the genetic map. Any new mutation in the region is cloned by first determining its locaton with respect to flanking molecular markers and then identifying the gene within the P1 clone that contains it. [After D. L. Hartl, D. I. Nurminsky, R. W. Jones, and E. R. Lozovskaya. 1994. *Proc. Natl. Acad. Sci. USA* 91: 6824.]

13.4 Transgenic Organisms

One motivation for cloning a gene is to determine its function by analysis of the DNA sequence and other methods. Another important application is to use a gene cloned from one species of organism to change the genotype of a different species of organism. The result of such genetic engineering is called a **transgenic organism.** To take a real example, the bacterium *Bacillus thuringiensis* produces a variety of insecticidal proteins called *delta-endotoxins* that form crystals in the bacterial spores. These proteins are toxic to insects of the orders *Lepidoptera* (moths and butterflies), *Diptera* (flies with two pairs of wings), and *Coleoptera* (beetles). When the bacterial spores are ingested by an insect, the crystals dissolve and are activated by proteases in the gut, where they bind to specific receptors in the intestinal epithelium and create "leakage channels" that ultimately kill the insect. Genetically engineered corn, cotton, and other crop plants have been produced whose genome contains a delta-endotoxin–coding region regulated by sequences taken from normal plant genes. Hence the engineered plants produce the delta-endotoxin protein in their tissues, making them lethal when ingested by insects such as the European corn borer, a lepidopteran pest that was introduced into the United States in the early 1900s and that currently causes annual crop losses of field corn, popcorn, seed corn, and sweet corn in excess of $1 billion.

In this section we shall examine some of the genetic procedures by which transgenic organisms are produced. We have already seen how recombinant DNA can be introduced into bacterial cells by means of transformation. Similar procedures work with yeast and many other unicellular microorganisms. Transformation in metazoan organisms, including domesticated animals and plants, requires different techniques.



Protests erupted when researchers reported that pollen from a variety of genetically engineered corn, when dusted onto milkweed plants, could kill caterpillars of the monarch butterfly. The engineered corn contains a gene from the bacterium *Bacillus thuringiensis* encoding a protein that kills insects and was originally intended to control damage from the European corn borer. The protesters feared that windborne pollen could dust many other plants and potentially harm beneficial insects such as butterflies and bees. [Associated Press/J. Scott Applewhite].

Germ-Line Transformation in Animals

In the nematode *Caenorhabditis elegans*, the transformation procedure is based on the fact that DNA injected directly into the reproductive organs is sometimes spontaneously incorporated into the genome of germ-line cells. Therefore, if an animal is injected with recombinant DNA that includes a genetic marker along with any sequence of interest, transformation is manifested by presence of the marker phenotype in the progeny of the injected animals.

A somewhat more elaborate procedure is used for germ-line transformation in *Drosophila*. The usual method employs a transposable element called the ***P* element,** which is a DNA sequence 2.9 kb in length that includes a central region, encoding the transposase, flanked by a 31-base-pair sequence that is repeated in inverted orientation at each end (**Figure 13.13**, part A). The terminal inverted repeats are essential for recognition by the transposase, which transposes the *P* element by excision and insertion. (Details of cut-and-paste transposition are examined in Section 7.3). For germ-line transformation, the vector is a plasmid containing a *P* element in which the transposase-coding region is replaced by the DNA sequence of interest, along with a suitable genetic marker (usually one affecting eye color). By itself, this *P* element cannot transpose because it encodes no

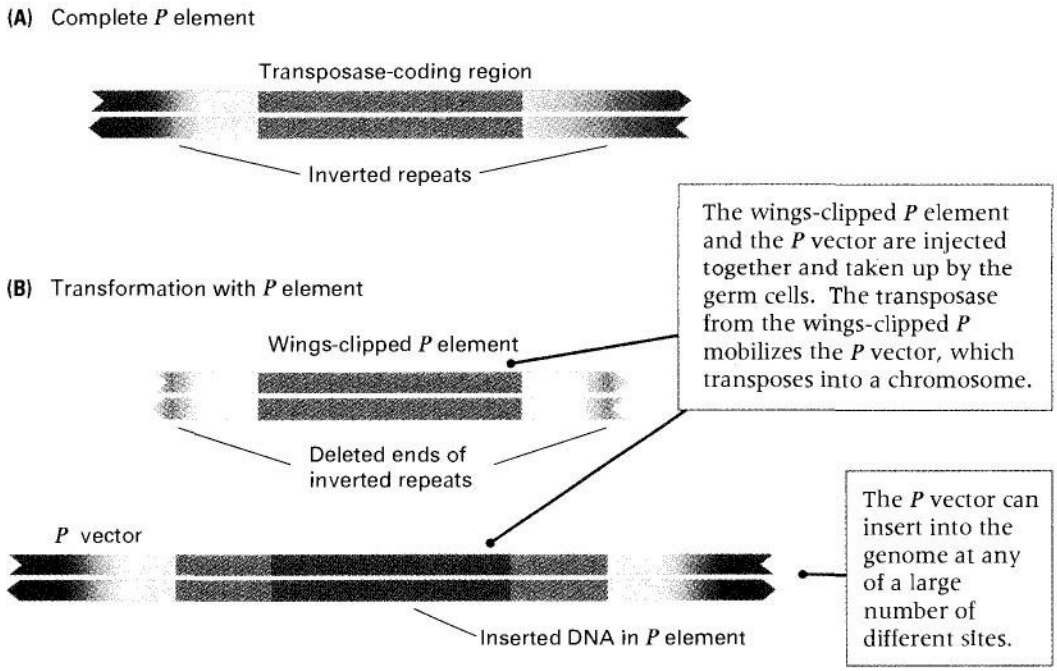


Figure 13.13 Transformation in *Drosophila* mediated by the transposable element *P*. (A) Complete *P* element containing inverted repeats at the ends and an internal transposase-coding region. (B) Two-component transformation system. The vector component contains the DNA of interest flanked by the recognition sequences needed for transposition. The wings-clipped component is a modified *P* element that codes for transposase but cannot transpose itself because critical recognition sequences are deleted.

transposase, but because it retains the inverted repeats it can be mobilized by transposase from another source. The source of transposase is usually a genetically engineered derivative of the *P* element, called *wings clipped*, that encodes functional transposase but cannot itself transpose because the ends of the inverted repeats are missing (Figure 13.13B).

In *Drosophila* transformation, any DNA sequence of interest can be placed between the ends of the deleted *P* element. The resulting vector, along with a different plasmid containing the wings-clipped element, are injected into the region of the early embryo containing the germ cells. The DNA is taken up by the germ cells, and the wings-clipped element makes possible the production of functional transposase (Figure 13.13B). This mobilizes the engineered *P* vector and results in its transposition into an essentially random location in the genome. Transformants are detected among the progeny of the injected flies because of the eye color or other genetic marker included in the *P* vector. Integration into the germ line is typically very efficient (10 to 20 percent of the injected embryos that survive and are fertile yield one or more transformed progeny). However, the efficiency decreases with the size of the DNA fragment in the *P* element, and the effective upper limit is approximately 20 kb.

In mammals, transformation of the germ line can be carried out in several ways. The most direct is the injection of vector DNA into the nucleus of fertilized eggs, which are then transferred to the uterus of foster mothers for development. The vector is usually a modified RNA virus, which uses a reverse transcriptase to convert the RNA genome into double-stranded DNA that becomes inserted into the genome in infected cells. Genetically engineered viruses containing inserted segments of RNA undergo the same integration process.

Another method of transforming mammals uses **embryonic stem cells** obtained

CoNnection

Hello, Dolly!

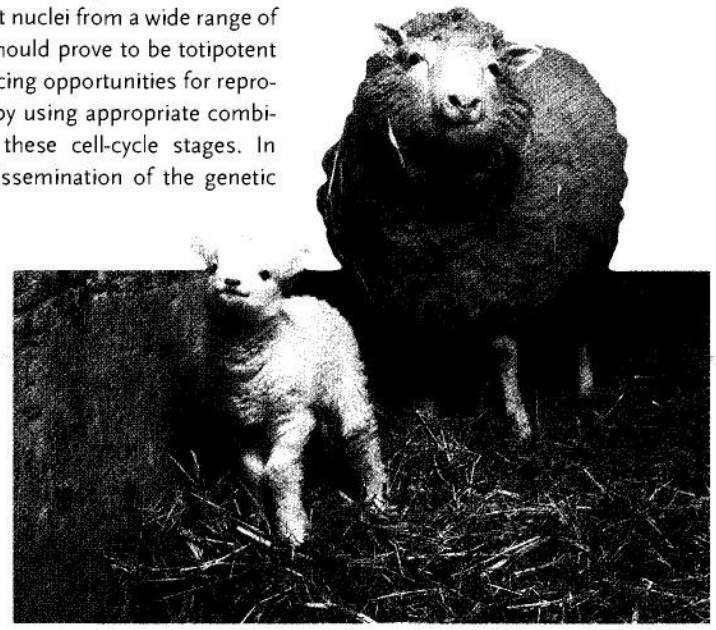
Ian Wilmut, Angelika E. Schnieke, Jim McWhir, Alex J. Kind, and Keith H. S. Campbell 1997
Roslin Institute, Roslin, Midlothian, Scotland
Viable Offspring Derived from Fetal and Adult Mammalian Cells

The Scottish Finn Dorset ewe known as "Dolly" is the first mammal to have been cloned from the nucleus of a cell taken from an adult mammal. The experiment created a press sensation. The President of the United States said that he would be against cloning people. A professor of public health said, "I don't think reasonable, rational people would want to clone themselves, but an eccentric millionaire might want to leave his money to a clone. There's no way of stopping the super-rich from going offshore to clone themselves." Some people were concerned for the animals. A spokesperson for People for the Ethical Treatment of Animals was quoted as saying, "It is time that society learned to respect our fellows, not exploit them for every fool thing." (Note in the excerpt that follows that the experiment was carried out with the prior approval of the appropriate Animal Welfare Committee.) Many important ethical issues are raised by the possibility of cloning humans. There are also important issues concerning the status of cloned animals. Should the clone of a racehorse that won the Triple Crown compete under the same rules as other racehorses? Should the clones of champion show dogs and cats compete under the same rules? Who should decide?

Transfer of a single nucleus at a specific stage of development, to an enucleated unfertilized egg, provided an opportunity to investigate whether cellular differentiation to that stage involved irreversible genetic modification. The first offspring to develop from a differentiated cell were born after nuclear transfer from an embryo-derived cell that had been induced to become quiescent. Using the same procedure, we now report the birth of live lambs from adult mammary gland, fetus, and embryo. The fact that a lamb was derived from an adult cell confirms that differentiation of that cell did not involve the irreversible modification of genetic material required for development to term. . . . If the recipient cytoplasm is prepared by enucleation of an oocyte at metaphase II, it is only possible to avoid chromosomal damage and maintain normal ploidy by transfer of diploid nuclei. Our studies with cultured cells suggest that there is an advantage if cells are quiescent. . . . Together our results indicate that nuclei from a wide range of cell types should prove to be totipotent after enhancing opportunities for reprogramming by using appropriate combinations of these cell-cycle stages. In turn, the dissemination of the genetic improvement obtained within elite selection herds will be enhanced by limited replication of animals with proven performance by nuclear transfer of cells derived from adult animals. . . . The lamb born after nuclear transfer from a mammary gland cell is, to our knowledge, the first mammal to develop from a cell derived from an adult tissue. . . . This is consistent with the generally accepted view that mammalian differentiation is almost all achieved by systematic, sequential changes in gene expression brought about by interactions between the nucleus and the changing cytoplasmic environment. . . . These experiments were conducted under the Animals (Scientific Procedure) Act 1986 [United Kingdom] and with the approval of the Roslin Institute Animal Welfare Experiments Committee.

The fact that a lamb was derived from an adult cell confirms that differentiation of that cell did not involve the irreversible modification of genetic material required for development to term.

Source: Nature 385: 810–813



"I love ewe, Mom!" Dolly and her natural-born lamb, whose fleece is white as snow. Who's Dolly? See the Connection on this page. [Roslin Institute/PA Photos.]

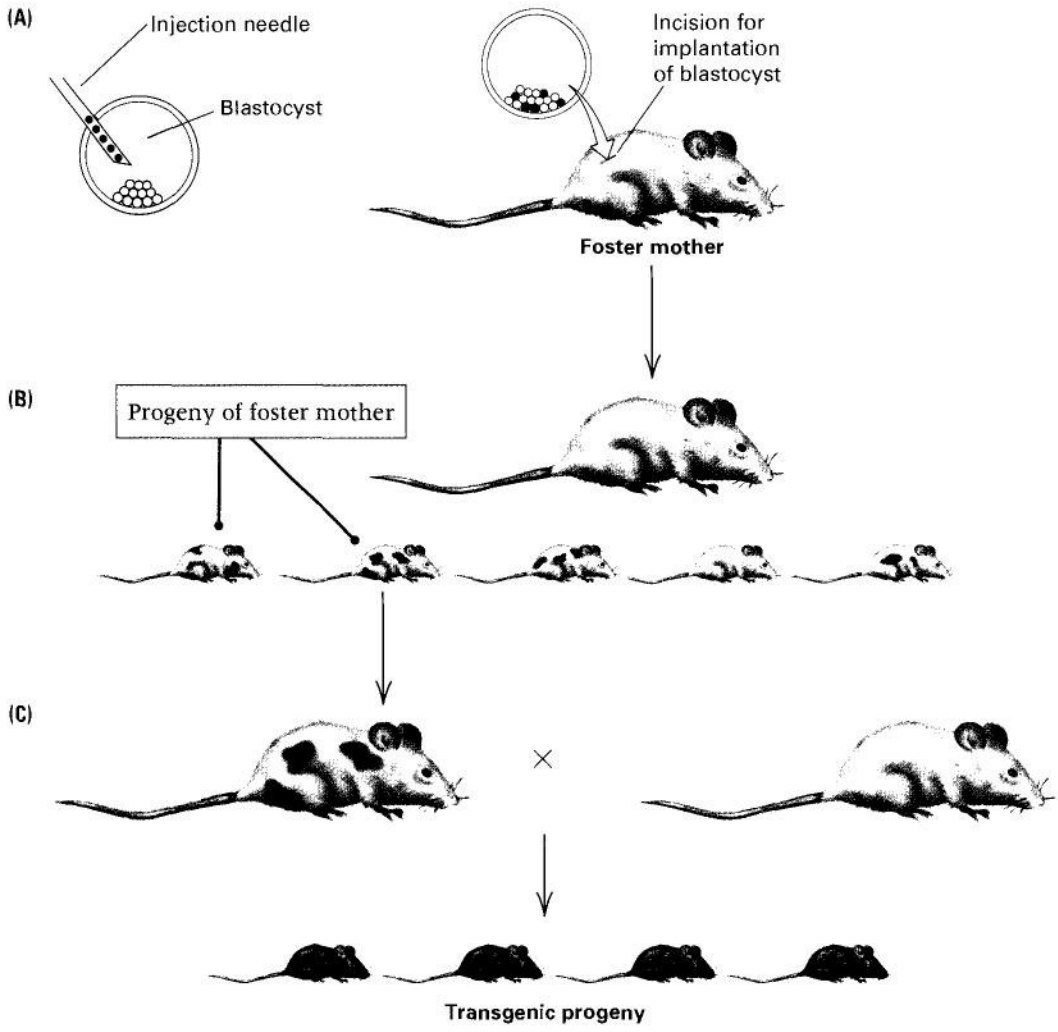


Figure 13.14 Transformation of the germ line in the mouse via embryonic stem cells. (A) Stem cells obtained from an embryo of a black strain are isolated and, after genetic manipulation in culture, injected into the embryo of a white strain, which is then introduced into the uterus of a foster mother. (B) The resulting offspring are often mosaics that contain cells from both the black and the white strains. (C) If cells from the black strain colonize the germ line, then the offspring of the mosaic animal will be black. [After M. R. Capecchi. 1989. *Trends Genet.* 5: 70.]

from embryos a few days after fertilization (**Figure 13.14**). Although embryonic stem cells are not very hardy, they can be isolated and then grown in culture and transformed with recombinant DNA vectors. The genetically altered stem cells are transferred into a developing embryo that is transplanted into the uterus of a foster mother (Figure 13.14A), where the stem cells become incorporated into various tissues of the embryo and often participate in forming the germ line. If the embryonic stem cells carry a genetic marker, such as a gene for black coat color, then mosaic animals can be identified by their spotted coats (B). Some of these animals, when mated, produce black offspring (C), which indicates that the embryonic stem cells had become incorporated into the germ line. In this way, genetic changes introduced into the embryonic stem cells while they were in culture may become incorporated into the germ line of living animals. The method shown in Figure 13.14 has been used to create strains of mice with mutations in genes associated with such human genetic diseases as cystic fibrosis. These strains serve as mouse models for studying the disease and for testing new drugs and therapeutic methods.

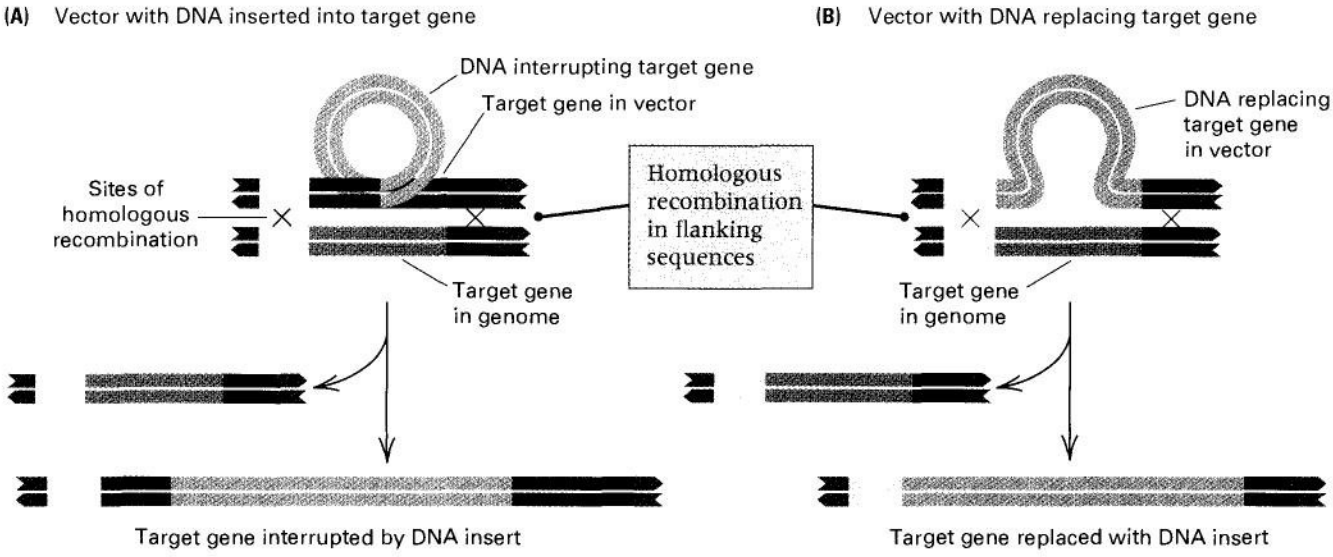


Figure 13.15 Gene targeting in embryonic stem cells. (A) The vector (top) contains the target sequence (brick red) interrupted by an insertion. Homologous recombination introduces the insertion into the genome. (B) In this case the vector contains DNA sequences flanking the targeted gene. Homologous recombination results in replacement of the targeted gene with an unrelated DNA sequence. [After M. R. Capecchi. 1989. *Trends Genet.* 5: 70.]

The procedure for introducing changes into specific genes is called **gene targeting.** The specificity of gene targeting comes from DNA sequence homology between the introduced recombinant DNA and normal sequences already present in the chromosome. When the DNA sequences are similar enough, they can undergo pairing as the result of an exchange of complementary strands followed by a process of breakage and reunion. Two examples are illustrated in **Figure 13.15**, where the DNA sequences present in gene-targeting vectors are shown as looped configurations paired with homologous regions in the chromosome prior to recombination. The targeted gene is shown in pink. In Figure 13.15A, the vector contains the targeted gene interrupted by an insertion of a novel DNA sequence, and homologous recombination results in the novel sequence becoming inserted into the targeted gene in the chromosome. In Figure 13.15B, the vector contains only flanking sequences, not the targeted gene itself, so homologous recombination results in replacement of the targeted gene with an unrelated DNA sequence. In both cases, cells with targeted gene mutations can be selected by including an antibiotic-resistance gene, or other selectable genetic marker, in the sequences that are incorporated into the genome through homologous recombination. The molecular mechanism of homologous recombination is discussed in Section 6.8.

Genetic Engineering in Plants

A procedure for the transformation of plant cells makes use of a plasmid found in the soil bacterium *Agrobacterium tumefaciens* and related species. Infection of susceptible plants with this bacterium results in the growth of tumors known as *crown gall tumors* at the entry site, usually a wound. Susceptible plants include about 160,000 species of flowering plants in the class *Dicotyledoneae*, which includes the great majority of common flowering plants.

The *Agrobacterium* contains a large plasmid of approximately 200 kb called the ***Ti* plasmid,** which includes a smaller (about 25-kb) region known as the **T DNA** terminated by 25-bp direct repeats (sequences repeated in the same orientation) at the extreme ends (**Figure 13.16**, part A). Infection with *Agrobacterium* causes a profound change in the metabolism of infected cells, because the T DNA encodes proteins that stimulate cell division, which causes

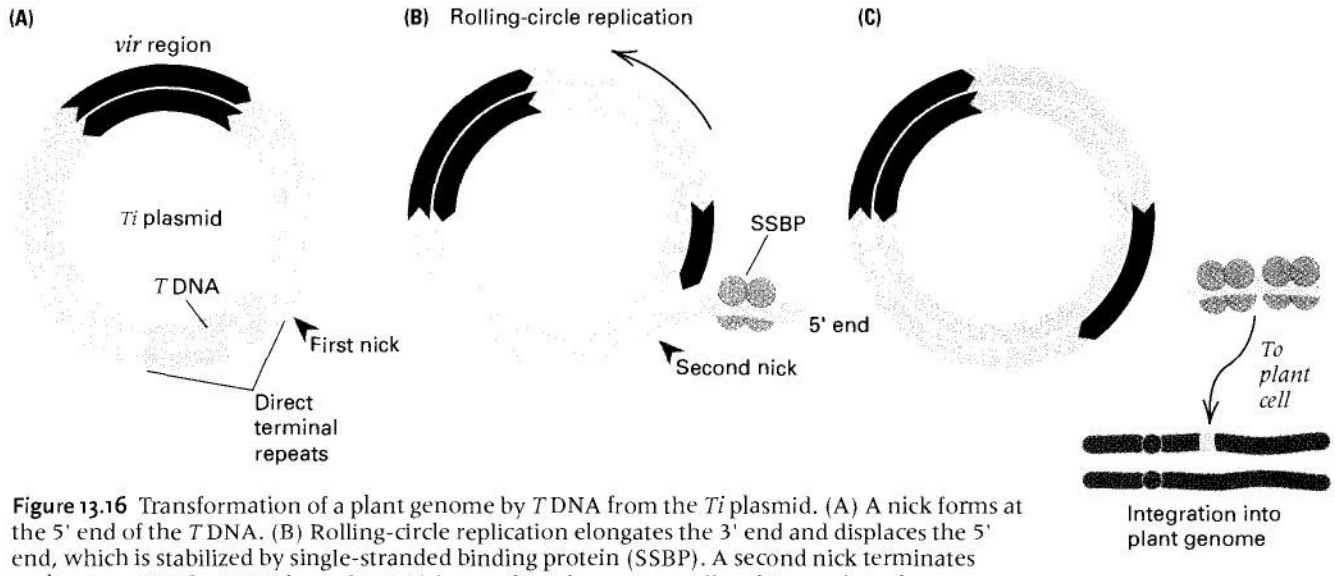


Figure 13.16 Transformation of a plant genome by *T* DNA from the *Ti* plasmid. (A) A nick forms at the 5' end of the *T* DNA. (B) Rolling-circle replication elongates the 3' end and displaces the 5' end, which is stabilized by single-stranded binding protein (SSBP). A second nick terminates replication. (C) The SSBP-bound *T* DNA is transferred to a plant cell and inserts into the genome.

the tumor, and also encodes enzymes that convert the amino acid arginine into an unusual derivative (generally *nopaline* or *octopine*, depending on the particular type of *Ti* plasmid) that the bacterium needs in order to grow.

The change in plant cell metabolism is brought about by physical transfer and incorporation of the T DNA into the plant genome by the mechanism outlined in Figure 13.16. The transfer process is mediated by the products of six genes present in a 40-kb region of the *Ti* plasmid called the *virulence* (*vir*) region. The transfer of T DNA begins with the formation of a single-stranded nick (Figure 13.16A), from which the 5' end of the T DNA peels off the plasmid; the resulting gap is closed by newly replicated DNA added onto the 3' end of the nick, using the unnicked circle as the template strand (Figure 13.16B). The region of the plasmid that is transferred is delimited by a second nick at the other end of the T DNA, but the position of this nick is somewhat variable. The resulting single-stranded T DNA is bound with molecules of a single-stranded binding protein (SSBP) and is transferred into the plant cell and incorporated into the nucleus. There it is integrated into the chromosomal DNA by a mechanism that is still unclear (Figure 13.16C). Although the SSBP has certain similarities in amino acid sequence to the *recA* protein from *E. coli*, which plays a key role in homologous recombination, the integration of T DNA does not require sequence homology.

The use of T DNA in plant transformation is made possible by engineered plasmids in which the sequences normally present in the T DNA are removed and replaced with a selectable marker along with the DNA sequences to be incorporated into the plant genome. A second plasmid containing the *vir* genes permits mobilization of the engineered T DNA. In infected tissues, the *vir* functions mobilize the T DNA for transfer into the host cells and integration into the chromosome. Transformed cells are selected in plant cell culture by the use of the selectable marker and then grown into mature plants.

Transformation Rescue

One of the important applications of germline transformation is to define experimentally the limits of any particular gene along the DNA. Every gene includes upstream and downstream sequences that are necessary for its correct expression, as well as the coding sequences. As we saw in Chapter 12, these regulatory sequences control the time in development, the cell types, and the level at which transcription occurs. In defining the upstream and downstream

limits of a gene, even knowing the complete nucleotide sequence of the coding region and the flanking DNA may be insufficient. The main reason is that there is no general method by which to identify regulatory sequences. Regulatory sequences are often composites of short, seemingly nondescript sequences, which are in fact the critical binding sites for regulatory proteins that control transcription.

To see how germ-line transformation is used to define the limits of a functional gene, consider the example of the *Drosophila* gene *white*, which when mutated results in flies with white eyes instead of red. The genetic organization of *white* is illustrated in **Figure 13.17**. The primary RNA transcript (B) is a little more than 6 kb in length and includes five introns that are excised and degraded during RNA processing. The resulting mRNA (C) is translated into a protein of 720 amino acids (D), which is a member of a large family of related ATP-dependent transmembrane proteins known as *ABC transporters* that regulate the traffic of their target molecules or ions across the cell membrane. The White transporter is located in the pigment-producing cells in the eye, and its target molecule is one of the key precursors of the eye-color pigments; hence flies with a mutant White transporter have white eyes.

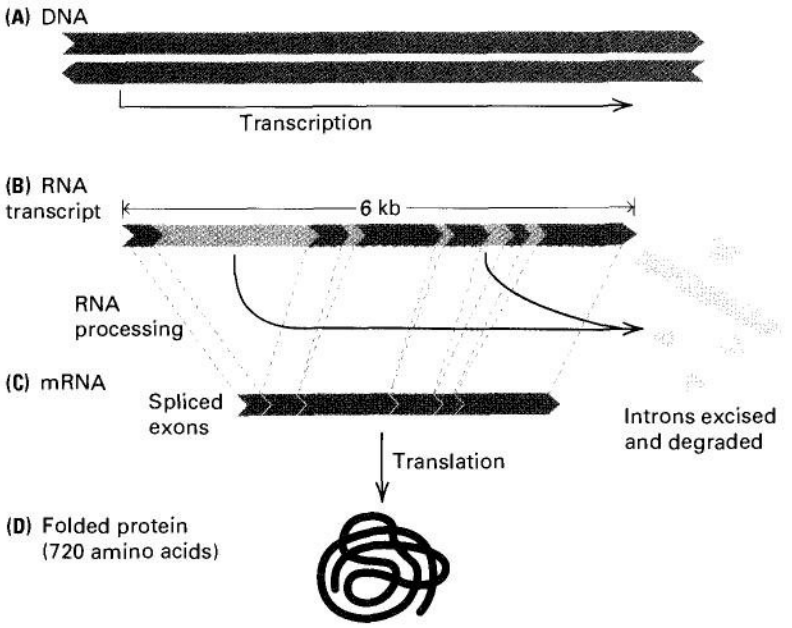


Figure 13.17 Genetic organization of the *Drosophila white* gene.

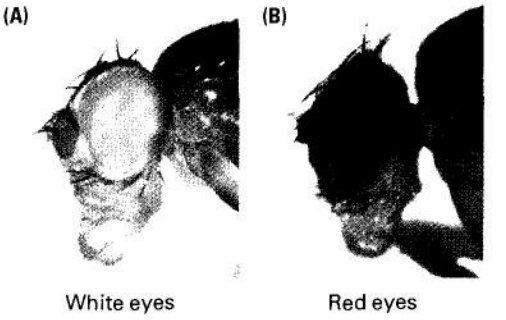


Figure 13.18 Mutant white-eyed and wildtype red-eyed males of *Drosophila melanogaster*. [Courtesy of E.R. Lozovsky.]

The question addressed by germ-line transformation is this: How much DNA upstream and downstream of *white* is necessary for the fly to produce the wildtype transporter protein in the pigment cells? The experimental approach is first to clone a large fragment of DNA that includes the coding sequence for the wildtype White protein, and then to use germ-line transformation to introduce this fragment into the genome of a fly that contains a *white* mutation. If the introduced DNA includes all the 5' and 3' regulatory sequences necessary for correct gene expression, then the phenotype of the resulting flies will be wildtype, because the wildtype gene is dominant to the *white* mutation. The ability of an introduced DNA fragment to correct a genetic defect in a mutant organism is called **transformation rescue** (**Figure 13.18**), and it means that the fragment contains all the essential regulatory sequences. The fragment may also include some sequences that are nonessential, so finding the *minimal* 5' and 3' flanking regions requires that the smaller pieces of the original fragment also be assayed for transformation rescue. In the case of *white*, the minimal DNA fragment is about 8.5 kb in length, starting about 2 kb upstream from the transcription start site.

13.5 Site-Directed Mutagenesis and Knockout Mutations

Once a gene has been cloned and sequenced, specific mutations can be introduced into the gene and transformed into the germ line in order to assay the resulting

phenotype and make it possible to ascertain the function or functions of the gene. This approach is known as **reverse genetics** because it reverses the traditional flow of genetic analysis. Instead of starting with a mutant phenotype and trying to identify the mutant gene, reverse genetics starts with a known gene and determines how mutations affect the phenotype.

Several methods are available for creating a defined mutation at a specific site in cloned DNA. One method of **site-directed mutagenesis,** based on the polymerase chain reaction (PCR), is outlined in **Figure 13.19**. It starts with a unique restriction fragment isolated from a cloned wildtype gene, which in this example is assumed to be an *Eco*RI–*Bam*HI fragment. Two different PCR reactions are carried out. One PCR uses a primer pair designated 1 and 2, in which primer 1 extends beyond and overlaps the *Eco*RI site, allowing this site to be reconstituted upon amplification, and primer 2 is a *mutagenic primer* containing a single-base mismatch (in this example a G–A mismatch). The other PCR uses primer pairs 3 and 4, in which the 5' end of primer 3 is adjacent to the 5' end of primer 2, and primer 4 extends beyond and overlaps the *Bam*HI site, allowing this site to be reconstituted upon amplification. The amplification products are shown in part C. When these are ligated and then digested with *Eco*RI and *Bam*HI, the original *Eco*RI–*Bam*HI fragment is reconstituted, except that the new fragment has a specific A–T → G–C base-pair substitution at the site indicated. When the *Eco*RI–*Bam*HI fragment is substituted back into the original wildtype gene, the reconstituted gene has the A–T → G–C base-pair mutation that can be introduced into an organism by germ-line transformation.

Similar methods can be used to create *deletion mutations*, in which part of a gene is missing, or mutations in which part of a gene is replaced with a completely unrelated sequence. When such constructs are used in gene-targeting experiments to replace the wildtype gene in a genome by homologous recombination (see Figure 13.15), the result is a **knockout mutation** in which the wildtype gene has been rendered completely nonfunctional ("knocked out"). Two important applications of gene knockouts can be highlighted. The first is the generation of mouse strains with

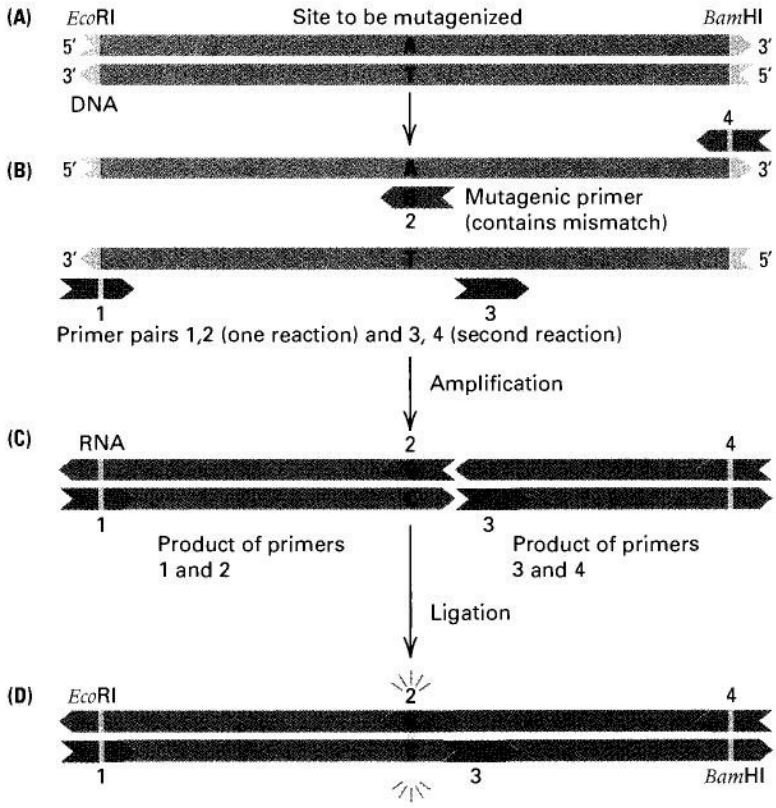


Figure 13.19 Site-directed mutagenesis using PCR. (A) An A–T nucleotide pair is to be replaced with a G–C nucleotide pair. (B) The original fragment is amplified in two parts, using a primer (number 2 in this example) that contains a G instead of an A. (C) The amplified product on the left contains a G–C nucleotide pair at the targeted site. (D) Ligation of the amplified fragments reconstitutes a DNA duplex that contains the site-directed mutation.

knockouts of the homologs of human disease genes in order to serve as models for the human disease. For example, the effectiveness of medication for different genetic causes of high blood pressure can be examined in mouse strains that have knockouts of genes whose products are known to be important in regulating blood pressure. A second important application of knockout mutagenesis is in determining gene function. For example, among the approximately 6000 genes identified from complete genomic sequencing of the genome of budding yeast (*Saccharomyces cerevisiae*), only about 25 percent are associated with genes previously identified through the isolation of mutants. Approximately another 35 percent have functions that may be inferred from homologies to genes that have been identified in other organisms. But the remaining approximately 40 percent have completely unknown functions. The functions of many of these mystery genes can be revealed by examining the phenotypes associated with knockout mutations under a variety of growth conditions.

13.6 Some Applications of Genetic Engineering

Recombinant DNA technology has revolutionized modern biology not only by opening up new approaches in basic research but also by making possible the creation of organisms with novel genotypes for practical use in agriculture and industry. In this section we examine a few of many applications of recombinant DNA.

Giant Salmon with Engineered Growth Hormone

In many animals, the rate of growth is controlled by the amount of growth hormone produced. Transgenic animals with a growth-hormone gene under the control of a highly active promoter to drive transcription often grow larger than their normal counterparts. An example of a highly active promoter is found in the gene for **metallothionein.** The metallothioneins are proteins that bind heavy metals. They are ubiquitous in eukaryotic organisms and are encoded by members of a family of related genes. Human beings, for example, have more than ten metallothionein genes that can be separated into two major groups according to their sequences. The promoter region of a metallothionein gene drives transcription of any gene to which it is attached, in response to heavy metals or steroid hormones. For example, when DNA constructs consisting of a rat growth-hormone gene under metallothionein control are used to produce transgenic mice, the resulting animals grow about twice as large as normal mice.

The effect of another growth-hormone construct is shown in **Figure 13.20.** The fish are coho salmon at 14 months of age. Those

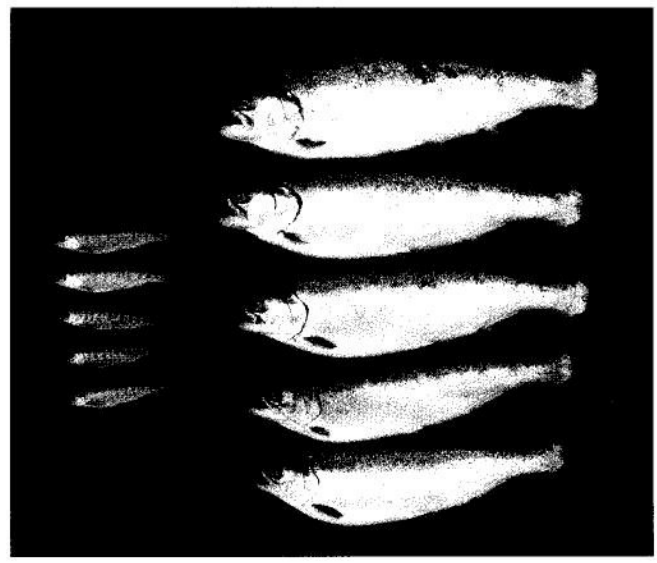


Figure 13.20 Normal coho salmon (left) and genetically engineered coho salmon (right) containing a sockeye salmon growth-hormone gene driven by the regulatory region from a metallothionein gene. The transgenic salmon average 11 times the weight of the nontransgenic fish. The smallest fish on the left is about 4 inches long. [Courtesy of R. H. Devlin. Reprinted by permission from *Nature* 371: 209, R. H. Devlin, T. Y. Yesaki, C. A. Biagi, E. M. Donaldson, P. Swanson, and W. K. Chan. Copyright 1994 Macmillan Magazines Ltd.]

on the left are normal, whereas those on the right are transgenic animals that contain a salmon growth-hormone gene driven by a metallothionein regulatory region. Both the growth-hormone gene and the metallothionein gene were cloned from the sockeye salmon. As an indicator of size, the largest transgenic fish on the right has a length of about 42 cm. On average, the transgenic fish are 11 times heavier than their normal counterparts; the largest transgenic fish was 37 times the average weight of the nontransgenic animals. And not only do the transgenic salmon grow faster and become larger than normal salmon; they also mature faster.

Nutritionally Engineered Rice

Although thus far we have focused on genetic engineering through the manipulation of single genes, it is also possible to create transgenic organisms that have entire new metabolic pathways introduced. A remarkable example is in the creation of a genetically engineered rice that contains an introduced biochemical pathway for the synthesis of β-carotene, a precursor of vitamin A found primarily in yellow vegetables and greens. (Deficiency of vitamin A affects some 400 million people throughout the world, predisposing them to skin disorders and night blindness.) The β-carotene pathway includes four enzymes, which in the engineered rice are encoded in genes from different organisms (**Figure 13.21**). Two of the genes come from the common daffodil (*Narcissus pseudonarcissus*), whereas the other two come from the bacterium *Erwinia uredovora*. Each pair of genes was cloned into T DNA and transformed into rice using *Agrobacterium tumefaciens* by the mechanism outlined in Figure 13.16. Transgenic plants were then crossed to produce progeny containing all four enzymes. The engineered rice seeds contain enough β-carotene to provide the daily requirement of vitamin A in 300 grams of cooked rice; they even have a yellow tinge (Figure 13.21B).

People on high-rice diets are also prone to iron deficiency because rice contains a small phosphorus-storage molecule called *phytate* (myo-inositol hexakis dihydrogen phosphate),which binds with iron and interferes with its absorption through the intestine. The transgenic β-carotene rice

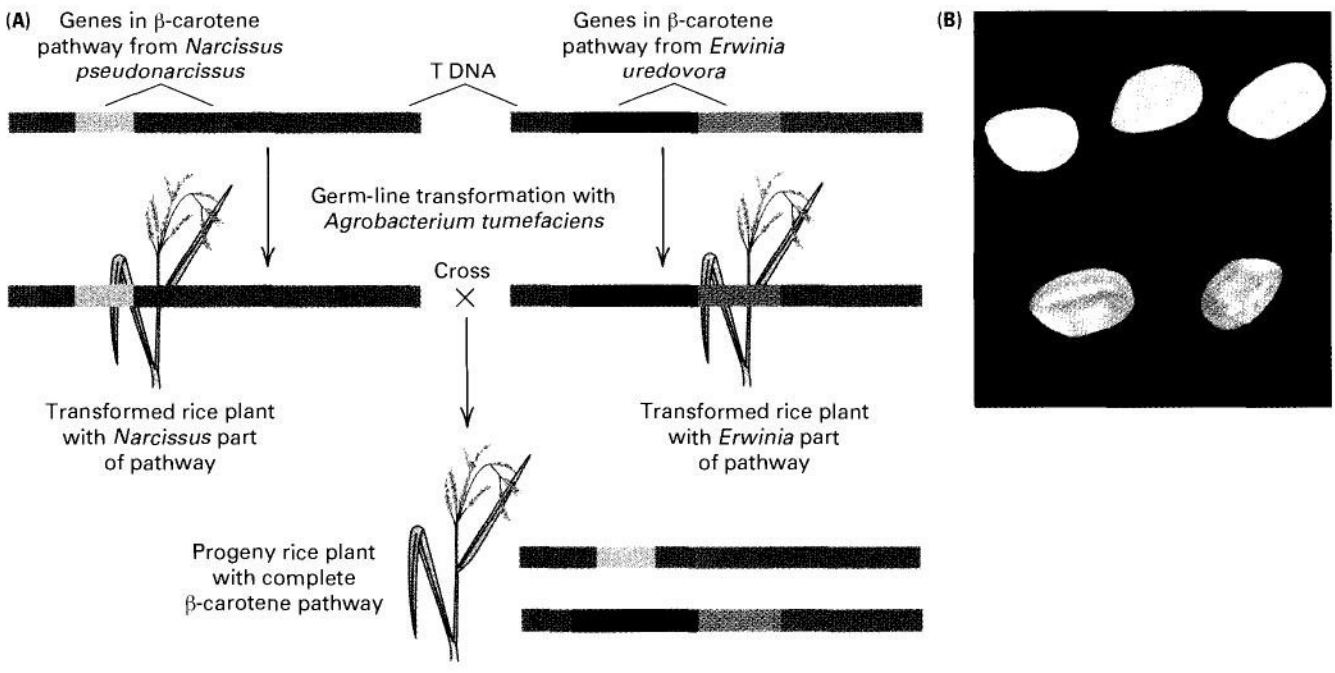


Figure 13.21 Genetically engineered rice containing a biosynthetic pathway for β-carotene. (A) Enzymes in the pathway derive from genes in two different species. (B) Rice plants with both parts of the pathway produce grains with a yellowish cast (top) due to the β-carotene they contain, in contrast to the pure white grains (bottom) of normal plants. [B courtesy of Ingo Potrykus.]

was also engineered to minimize this problem by introducing the fungal enzyme from *Aspergillus ficuum* that breaks down phytate, along with a gene encoding the iron-storage protein ferritin from the French bean, *Phaseolus vulgaris*, plus yet another gene from basmati rice that encodes a metallothionein-like gene that facilitates iron absorption in the human gut. Altogether, then, the transgenic rice strain rich in β-carotene and available iron contains six new genes taken from four unrelated species plus one gene from a totally different strain of rice!

Production of Useful Proteins

Among the most important applications of genetic engineering is the production of large quantities of particular proteins that are otherwise difficult to obtain (for example, proteins that are present in only a few molecules per cell or that are produced in only a small number of cells or only in human cells). The method is simple in principle. A DNA sequence coding for the desired protein is cloned in a vector adjacent to an appropriate regulatory sequence. This step is usually done with cDNA, because cDNA has all the coding sequences spliced together in the right order. Using a vector with a high copy number ensures that many copies of the coding sequence will be present in each bacterial cell, which can result in synthesis of the gene product at concentrations ranging from 1 to 5 percent of the total cellular protein. In practice, the production of large quantities of a protein in bacterial cells is straightforward, but there are often problems that must be overcome, because in the bacterial cell, which is a prokaryotic cell, the eukaryotic protein may be unstable, may not fold properly, or may fail to undergo necessary chemical modification. Many important proteins are currently produced in bacterial cells, including human growth hormone, blood-clotting factors, and insulin. Patent offices in Europe and the United States have already issued 12,000 patents for the clinical use of the products of genetically engineered human genes. **Figure 13.22** gives a breakdown of the numbers of patents issued relative to clinical application.

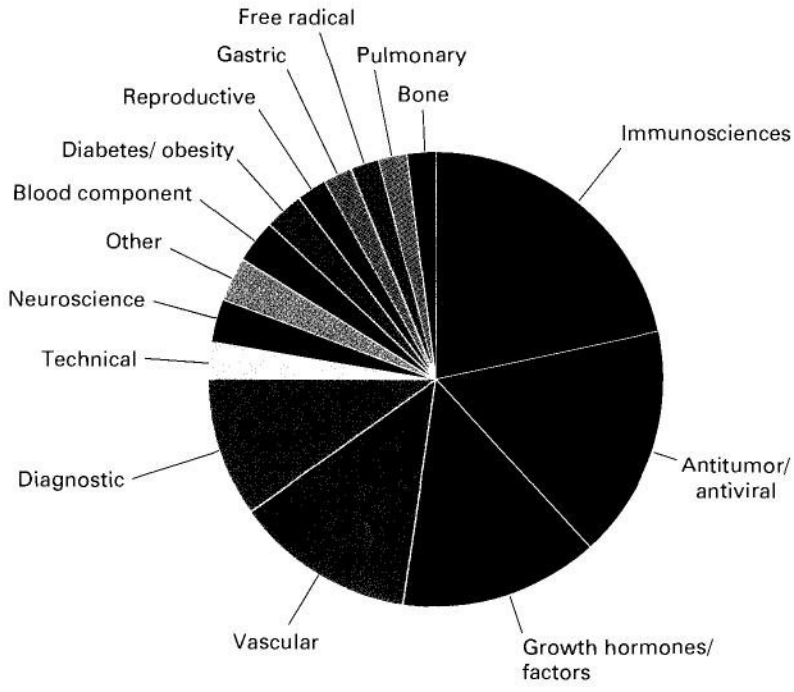


Figure 13.22 Relative number of patents issued for various clinical applications of the products of genetically engineered human genes. [Data from *Nature* 1996, 380: 388.]

Genetic Engineering with Animal Viruses

Genetic engineering in animal cells often exploits RNA viruses called **retroviruses** that use reverse transcriptase to make a double-stranded DNA copy of their RNA genome. The DNA copy then becomes inserted into the chromosomes of the cell. Ordinary transcription of DNA to RNA occurs only after the DNA copy is inserted. The infected host cell survives the infection, retaining the DNA copy of the retroviral RNA in its genome. These features of retroviruses make them convenient vectors for the genetic manipulation of animal cells, including those of birds, rodents, monkeys, and human beings.

Genetic engineering with retroviruses allows the possibility of altering the genotypes of animal cells. Because a wide variety of retroviruses are known, including many that infect human cells, genetic defects may be corrected by these procedures in the future. The recombinant DNA procedure employed with retroviruses consists of the *in vitro* synthesis of double-stranded DNA from the viral RNA by means of reverse

transcriptase. The DNA is then cleaved with a restriction enzyme and, using techniques already described, any DNA fragment of interest is inserted. Transformation yields cells with the recombinant retroviral DNA permanently inserted into the genome. However, many retroviruses contain genes that result in uncontrolled proliferation of the infected cell, thereby causing a tumor. When retrovirus vectors are used for genetic engineering, the cancer-causing genes are first deleted. The deletion also provides the space needed for incorporation of the desired DNA fragment.

Attempts are currently under way to assess the potential use of retroviral vectors in **gene therapy**, or the correction of genetic defects in somatic cells by genetic engineering. Examples might include correcting immunological deficiencies in patients with various kinds of inherited disorders. However, a number of major problems stand in the way of gene therapy becoming widely used. At this time, there is no completely reliable way to ensure that a gene will be inserted only into the appropriate target cell or target tissue, and there is no completely reliable means of regulating the expression of the inserted genes.

A major breakthrough in disease prevention would come through the development of synthetic vaccines produced by recombinant DNA. Production of natural vaccines is often unacceptable because of the extreme hazards of working with large quantities of the active virus—for example, the human immunodeficiency virus (HIV-1) that causes acquired immune deficiency syndrome (AIDS). The danger can be minimized by cloning and producing viral antigens in a nonpathogenic organism. Vaccinia virus, the agent used in smallpox vaccination, has attracted much attention as a candidate for this application. Viral antigens are often on the surface of virus particles, and some of these antigens can be engineered into the coat of vaccinia. For example, engineered vaccinia virus with certain surface antigens of hepatitis B virus, influenza virus, and vesicular stomatitis virus (which kills cattle, horses, and pigs) has already proved effective in animal tests. One of the great challenges is malaria, which affects approximately 300 million people in Africa, Asia, and Latin America, causing approximately 1–1.5 million deaths per year. Control of the disease by drugs is increasingly compromised by the spread of resistance mutations. Successful vaccination of mice against malaria has been achieved using recombinant vaccinia virus expressing certain surface antigens of *Plasmodium berghei*. This result encourages hope that a vaccine against the human agent, *Plasmodium falciparum*, will eventually be developed.

13.7 Analysis of Complex Genomes

Gene structure and function are understood in considerable detail because of the manipulative capability of recombinant DNA technology teamed with genetic analysis. Most recombinant DNA techniques are limited to the manipulation of DNA fragments smaller than about 40 kb. Although this range includes most cDNAs and many genes, some genes are much larger than 50 kb and so must be analyzed in smaller pieces. Special methods are also available to manipulate and clone DNA molecules that are very large. These methods are suitable not only for the analysis of very large genes, but also for the study of entire genomes.

Electrophoretic Separation of Large DNA Fragments

Section 2.3 included an examination of conventional DNA electrophoresis, a procedure by which DNA fragments up to about 20 kb can be separated. But a eukaryotic chromosome contains a single DNA molecule much longer than this. For example, the largest chromosome in the *D. melanogaster* genome has a DNA content of about 65 Mb (6.5×10^7 nucleotide pairs), which is equivalent to a continuous linear duplex about 22 mm in length (almost an inch). Such long molecules usually break into pieces during isolation.

In organisms such as budding yeast, *Saccharomyces cerevisiae*, which have small genomes, the DNA molecules present in the chromosomes are sufficiently small that they can be separated by modifications of conventional electrophoresis. The most common modification is to change the spatial orientation of the electric field at periodic intervals during the course of the

electrophoretic separation. Some electrophoretic apparatuses alternate the electric field between two sets of electrodes oriented at right angles or among three sets of electrodes forming a hexagon. In one type of large-fragment electrophoresis, called **pulsed-field gel electrophoresis (PFGE),** the apparatus is similar to that for conventional electrophoresis, but the orientation of the electric field is changed periodically. The improved separation of large DNA molecules apparently results from the additional time it takes for large molecules to reorient themselves when the orientation of the electric field is changed. Part A of **Figure 13.23** illustrates the separation of yeast chromosomes by means of PFGE, and Figure 13.23B is a diagram showing the relative sizes of the 16 chromosomes that make up the yeast genome. The chromosomes range in size from approximately 230 kb to 1.5 Mb. Electrophoretic separation yields a visual demonstration that each chromosome contains a single DNA molecule that runs continuously throughout its length.

In only a few organisms are any chromosomes as small as several megabases, however. In *Drosophila*, the smallest wild-type chromosome is about 6 Mb. Most DNA molecules in chromosomes of higher eukaryotes are very much longer than this. The smallest human chromosome is chromosome 21 (the one exception to the "number by size" rule), and its long arm alone is approximately 30 Mb. Therefore, even with the ability to separate large DNA fragments by electrophoresis, it is necessary to be able to cut the DNA in the genome into fragments of manageable size. This can be done with a class of restriction enzymes that cleaves at a restriction site consisting of eight bases rather than the usual six or four.

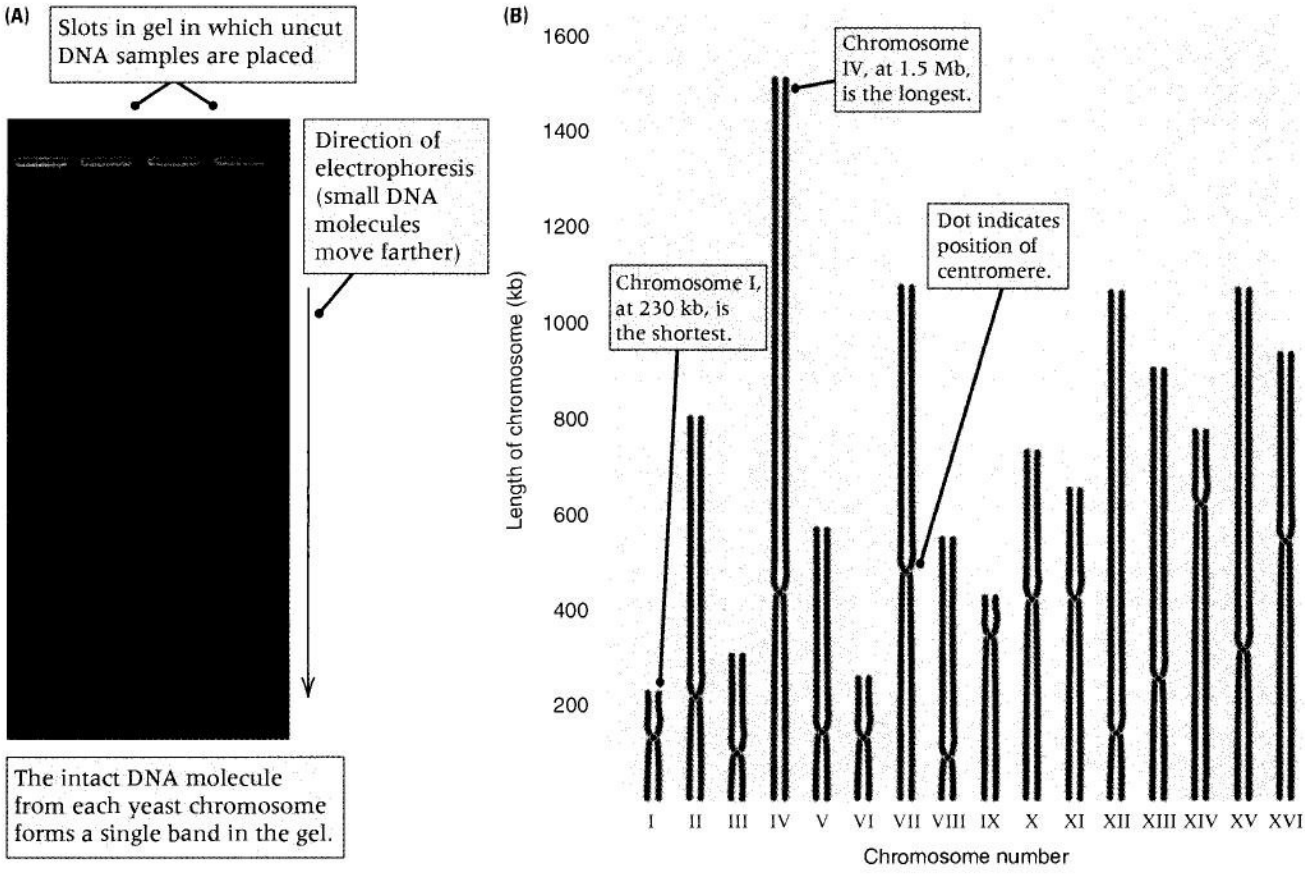


Figure 13.23 (A) Separation of chromosomes of the yeast *Saccharomyces cerevisiae* by pulsed-field gel electrophoresis, in which there is regular change in the orientation of the electric field. (B) Histogram of sizes of the 16 yeast chromosomes. [©1988 BioRad Laboratories, Inc.; permission to use this image has been granted by Bio-Rad Laboratories Inc.]

For example, the restriction sites of the eight-cutter enzymes *Not*I and *Sfi*I are

*Not*I 5'-GC ↓ GGCCGC-3'
*Sfi*I 5'-GGCCNNNN ↓ NGGCC-3'

These enzymes cleave double-stranded DNA at the positions of their restriction sites, and the downward arrows denote the position where the backbone is cut in each DNA strand. (The N's in the *Sfi*I restriction site mean that any nucleotide can be present at this site.) In a genome with equal proportions of the four nucleotides and random nucleotide sequences, the average size of a *Not*I or *Sfi*I fragments is 4^8 = 65,536 nucleotide pairs, or about 66 kb. Many genomes are relatively A + T-rich, and in vertebrate genomes there is a bias against long runs of G's and C's, so many *Not*I and *Sfi*I fragments are considerably larger than 66 kb. In any case, the use of eight-cutter restriction enzymes allows complex DNA molecules to be cleaved into a relatively small number of large fragments that can be separated, cloned, and analyzed individually.

Cloning of Large DNA Fragments

Large DNA molecules can be cloned intact in bacterial cells with the use of specialized vectors that can accept large inserts. The vectors that can accept large DNA fragments are called artificial chromosomes. They fall into three classes: **P1 artificial chromosomes (PACs), bacterial artificial chromosomes (BACs),** and **yeast artificial chromosomes (YACs)** (**Figure 13.24**). PACs and BACs are basically large circular plasmids maintained in *E. coli*,

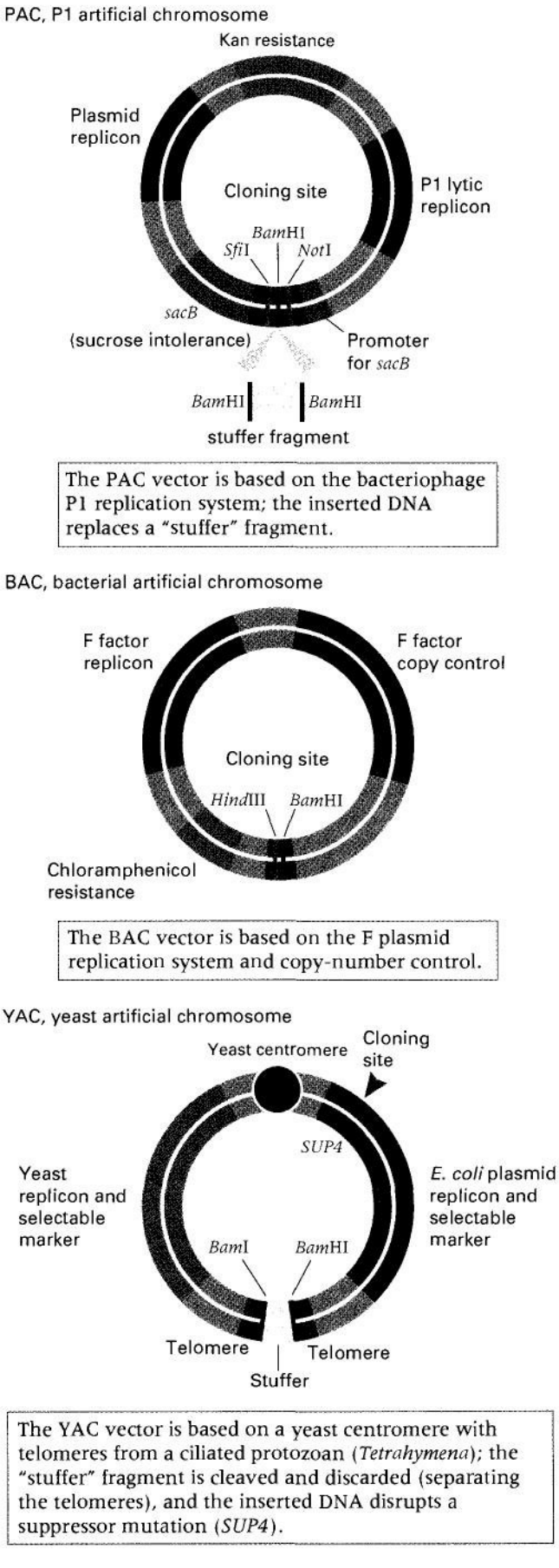


Figure 13.24 Artificial chromosome vectors. (A) P1 artificial chromosomes contain a plasmid replicon that results in 1–2 copies per cell, but there is also an inducible lytic replicon by which the number of copies can be amplified. Donor DNA is selected by insertional inactivation of the *sacB* gene, which allows survival in medium containing sucrose. (B) Bacterial artificial chromosomes are based on a vector that contains the F factor replicon as well as genes for regulating copy number. (C) Yeast artificial chromosomes are based on a vector that can be replicated in *E. coli* but that also contains a yeast centromere, replication origins, and selectable marker, as well as telomeres from *Tetrahymena*.

whereas the parent YAC replicates as a circular plasmid in bacteria but, after ligation with the inserted DNA, as a linear chromosome in yeast.

The genome of the wildtype P1 phage is approximately 100 kb, but a smaller region is sufficient to support replication and persistence as a plasmid in *E. coli*. One type of PAC vector is illustrated in Figure 13.24A. The essential elements of the vector are contained in 17 kb of DNA, which is sufficient for replication of the vector as a plasmid, for amplification of the vector via the lytic replicon, and for selection of the plasmid via kanamycin resistance. Selection for DNA inserts occurs through interruption of the *sacB* gene. When the *sacB* gene is expressed, the cells die if grown in medium containing sucrose (ordinary table sugar). Insertion of DNA at the cloning site separates the *sacB* gene from its promoter, and the gene cannot be expressed, so vectors with inserts in the cloning site allow growth in sucrose. PAC libraries often have an average insert size of 145 kb, and the largest inserts are about 300 kb.

The BAC vector (Figure 13.24B) is based on the F factor of *E. coli*, which was discussed in Chapter 10 in the context of its role in conjugation. The essential functions included in the 6.8-kb vector are genes for replication (*repE* and *oriS*), for regulating copy number (*parA* and *parB*), and for chloramphenicol resistance. BAC vectors with inserts greater than 300 kb can be maintained.

The YAC vector (Figure 13.24C) incorporates essential features of linear chromosomes and includes four types of genetic elements: (1) a cloning site; (2) a yeast centromere, replication origins, and genetic markers that are selectable in yeast; (3) an *E. coli* origin of replication and genetic markers that are selectable in *E. coli*; and (4) a pair of telomere sequences from the ciliated protozoan *Tetrahymena*. A YAC vector is therefore a *shuttle vector* that can replicate and be selected both in yeast and in *E. coli*. YAC chromosomes with inserts as large as 1 Mb can be recovered intact in yeast cells.

DNA fragments in the appropriate size range can be produced by breaking larger molecules into fragments of the desired size by physical means, by treatment with restriction enzymes that have infrequent cleavage sites (for example, enzymes such as *Not*I and *Sfi*I), or by treatment with ordinary restriction enzymes under conditions in which only a fraction of the restriction sites are cleaved (*partial digestion*). Cloning the large molecules consists of mixing the large fragments of source DNA with the vector; ligation with DNA ligase; introduction of the recombinant molecules into bacterial or yeast cells, depending on the vector; and selection for the clones of interest. These methods are generally similar to those described in Section 13.2 for the production of recombinant molecules containing small inserts of cloned DNA.

Use of the YAC vector in cloning is illustrated in **Figure 13.25**. The circular vector is isolated after growth in *E. coli* and cleaved with two different restriction enzymes: one that cuts only at the cloning site (denoted A) and one that cuts near the tip of each of the telomeres (denoted B). Discarding the segment between the telomeres results in the two telomere-bearing fragments shown, which form the arms of the yeast artificial chromosomes. Ligation of a mixture containing source DNA and YAC vector arms results in a number of possible products. However, transformation of yeast cells and selection for the genetic markers in the YAC arms yields only the ligation product shown in Figure 13.25, in which a fragment of source DNA is sandwiched between the left and right arms of the YAC vector. This product is recovered because it is the only true chromosome that possesses a single centromere and two telomeres. Products with two YAC left arms are dicentric, those with two YAC right arms are acentric, and both of these types of products are genetically unstable. YACs that have donor DNA inserted at the cloning site can be identified because the inserted DNA interrupts a yeast gene present at the cloning site and renders it nonfunctional. Yeast cells containing YACs with inserts of particular sequences of donor DNA can be identified in a number of ways, including colony hybridization (Figure 13.11), and these cells can be grown and manipulated in order to isolate and study the YAC insert. **Figure 13.26** shows a region of the *Drosophila* salivary gland chromosomes that hybridizes with a YAC containing an insert of 300 kb. The entire genome of *Drosophila* could be contained in only 550 YAC clones of this size.

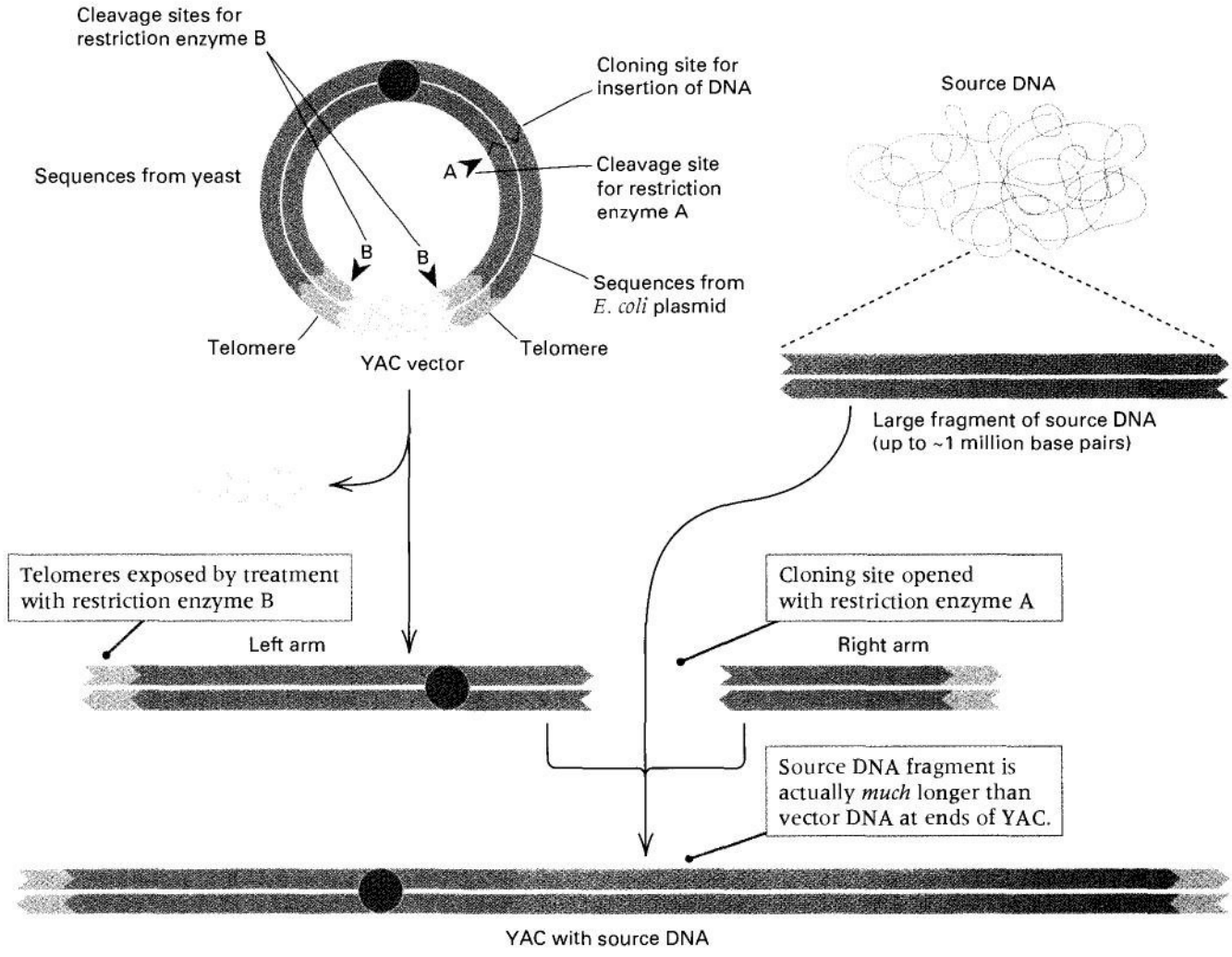


Figure 13.25 Cloning large DNA fragments in yeast artificial chromosomes (YACs). The vector (upper left) contains sequences that allow replication and selection in both *E. coli* and yeast, a yeast centromere, and telomeres from *Tetrahymena*. In producing the YAC clones, the vector is cut by two restriction enzymes (A and B) to free the chromosome arms. These arms are ligated to the ends of large fragments of source DNA, and yeast cells are transformed. Many ligation products are possible, but only those that consist of source DNA flanked by the left and right vector arms form stable artificial chromosomes in yeast.



Figure 13.26 Hybridization *in situ* between *Drosophila* DNA cloned into a yeast artificial chromosome and the giant salivary gland chromosomes.

CoNnection

Catch-22

Ian Dunham,[1] Nobuyoshi Shimizu,[2] Bruce A. Roe,[3] Stephanie L. Chissoe,[4] and 213 other investigators

1. The Sanger Centre, Hinxton, United Kingdom; 2. Keio University, Tokyo, Japan; 3. The University of Oklahoma, Norman, Oklahoma; 4. Washington University, St. Louis, Missouri; and 5 other research institutions

The DNA Sequence of Human Chromosome 22

Less than a century passed between the rediscovery of Mendel's work in 1900 and the DNA sequencing of the first human chromosome. Few fields in the history of science have had such a run. For those of us who participated in, or witnessed, even a part of it, it has been exhilarating—ingenious technical advances, stunning new interpretations, dazzling discoveries, all happening in rapid succession. We geneticists, like Alice in Wonderland's Red Queen, have had to learn as fast as we could just to stay in the same place!

Here we report the first sequencing landmark of the human genome project. . . . Clones representing parts of chromosome 22 were identified by screening BAC and PAC libraries . . . using sequence-tagged site (STS) markers. . . . Overlapping clone contigs were assembled . . . and ordered relative to each other. . . . A set of minimally overlapping clones (the "tile path") was chosen from the physical map and sequenced using a combination of random shotgun assembly followed by directed sequencing to close gaps. . . . The completed sequence covers 33.4 Mb of 22q with 11 gaps and has been estimated to be accurate to less than 1 error in 50,000 bases. . . . The sequence is complete to 97% coverage. . . . The total length of the sequence occupied by the annotated genes, including their introns, is 13.0 Mb (39% of the total sequence). About 3% of the total sequence is occupied by the exons. This contrasts sharply with the 41.9% of the sequence that represents tandem and dispersed repeat sequences. . . . The size of individual genes encoded on this chromosome varies over a wide range. The smallest complete genes are only on the order of 1 kb in length, whereas the largest single gene stretches over 583 kb. The mean genomic size of the genes is 19.2 kb (median 3.7 kb). Some complete gene structures appear to contain only a single exon, whereas the largest number of exons in a gene is 54. The mean exon number is 5.4 (median 3). The mean exon size is 266 bp (median 135 bp). . . . Two genes occur within the introns of other expressed genes. . . . We also observe pseudogenes frequently lying within the introns of other functional genes. . . . Nineteen percent of the coding sequences identified were identified as pseudogenes because they had significant similarity to known genes or proteins but had disrupted protein-coding reading frames. . . . Given that the sequence of 33.4 Mb represents 1.1% of the genome and encodes 679 genes [including 134 pseudogenes] . . . the minimum number of genes in the entire human genome would be at least 61,000. . . . Comparison of genetic distance against physical distance for all the [relevant genetic markers] shows a mean value of 1.87 cM [centimorgans] per Mb. . . . Examination of the mouse chromosomal locations mapped onto the human chromosome 22 sequence confirms that conserved synteny groups corresponding to human chromosome 22 are on mouse chromosomes 6, 16, 10, 5, 11, 8, and 15.

The first sequencing landmark of the human genome project . . . covers 33.4 Mb of 22q.

Source: Nature 402: 489–495

Physical Mapping

The first stage in the analysis of complex genomes is usually the production of a **physical map,** which is a diagram of the genome depicting the physical locations of various landmarks along the DNA. The landmarks in a physical map usually consist of the locations of particular DNA sequences, such as coding regions or sequences present in particular cloned DNA fragments. If the landmarks are the locations of the cleavage sites for restriction enzymes, then the physical map is also a restriction map (Section 2.3). More useful landmarks are the positions of molecular markers with known positions in the genetic map, because genetically mapped markers serve to integrate the genetic map and the physical map of an organism. The conceptual utility of a physical map is that it affords us a single framework for organizing and integrating diverse types of genetic information, including the positions of chromosome bands, chromosome breakpoints, mutant genes, transcribed regions, and DNA sequences. The practical utility of a physical map is that it contains the in-

formation needed for positional cloning (Section 13.3).

A completely integrated physical map is illustrated in **Figure 13.27**. The map covers a small part of human chromosome 16, and it illustrates how the physical map is used to organize and integrate several different levels of genetic information. A map of the metaphase banding pattern of chromosome 16 is shown across the top. (The entire chromosome contains about 95 Mb of DNA.) Beneath the cytogenetic map, the somatic cell hybrid map shows the locations of chromosome breakpoints observed in

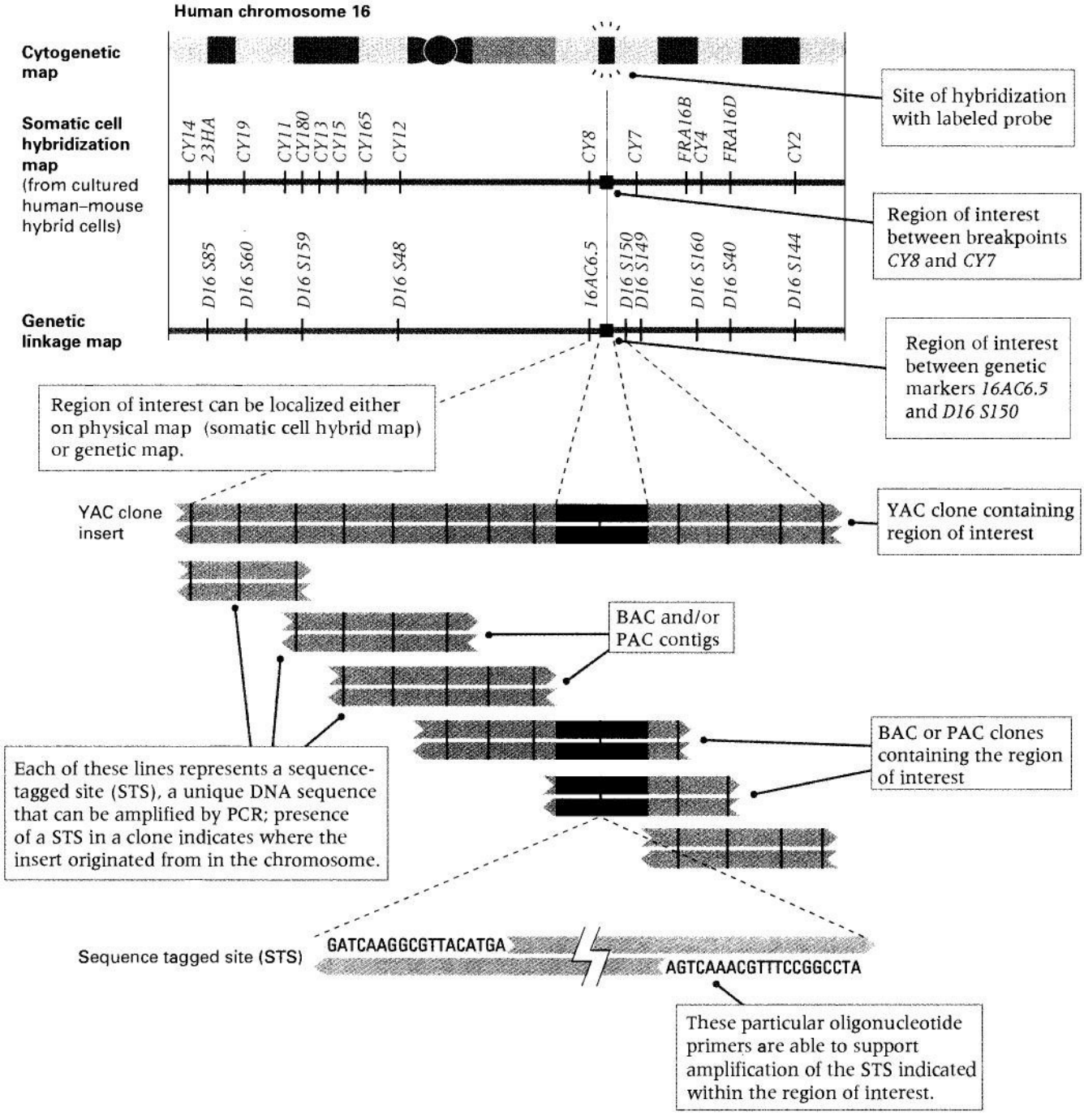


Figure 13.27 Integrated physical map of a small part of human chromosome 16. The various levels of the map are integrated by sequence-tagged sites(STSs). [Marker data from *Human Genome: 1991–92 Program Report,* United States Department of Energy.]

cultured hybrid cells that contain only a part of chromosome 16. The genetic linkage map depicts the locations of various genetic markers studied in pedigrees; most of the genetic markers are molecular markers. One region of the genetic linkage map shows the location of a YAC clone, which has been assigned a physical location in the cytogenetic map by *in situ* hybridization. The large DNA insert in the YAC clone is also represented in a contig of overlapping PAC or BAC clones.

The various levels of the physical map in Figure 13.27 are connected by a special type of genetic marker shown at the bottom: a **sequence-tagged site (STS).** An STS marker is a DNA sequence, present once per haploid genome, that can be amplified with a suitable pair of oligonucleotide primers by means of the polymerase chain reaction (PCR). Hence an STS marker defines a unique site in the genome whose presence in a cloned DNA fragment can be detected by PCR amplification. In Figure 13.27, for example, the STS marker located within the region of interest is present in two clones of the PAC/BAC contig, as well as in the YAC clone. Furthermore, the STS can be positioned on the somatic cell hybrid map by carrying out the PCR reaction with DNA from hybrid cells that contain rearranged or deleted chromosomes, and the amplified PCR product (or a clone containing the STS) can be used with *in situ* chromosome hybridization to localize the STS on the cytogenetic map. In this manner, STS markers can be used to integrate different types of information in a physical map. At present, more than 95 percent of the human genome is covered by a set of YAC clones interconnected by 30,000 STS markers. There is an average spacing between adjacent STS markers of 100 kb in the 3000-Mb human genome. Nearly 24,000 STS markers have been assigned to particular chromosomes, and a large fraction of these have been assigned to particular contigs within the chromosome.

Although an STS may consist of any type of sequence present once per haploid genome, some genome projects rely extensively on STS markers derived from cDNAs, because these markers represent coding regions that are likely to be of greater long-term interest than single-copy sequences that are noncoding. Sequences identified from cDNAs are called **expressed sequence tags (ESTs).** Known ESTs include a substantial proportion of the estimated 80,000 or so genes in the human genome. The most ambitious EST project to date included sequencing about 300,000 partial cDNAs, among them cDNAs obtained from cDNA libraries prepared from 37 distinct human organs and tissues. The total DNA sequence obtained was 83 million base pairs. Computer matching among the cDNA sequences revealed 87,983 distinct sequences, many of which could be assigned a function on the basis of similarity to known genes from human beings or from other organisms. **Figure 13.28** gives a breakdown of the ESTs by type of function. Approximately 40 percent of human genes are implicated in basic energy metabolism, cell structure, homeostasis, or cell division; a further 22 percent are concerned with RNA and protein synthesis and processing; and 12 percent are associated with signaling and communication between cells. **Figure 13.29** summarizes the results of examining the tissue-specific cDNA libraries. For each organ or tissue type, the first number is the total number of cDNA clones sequenced, and the number in parentheses is the number of distinct cDNA sequences found among the total from that organ or tissue type.

High-Throughput Genomic Sequencing and the Human Genome Project

The development of methods for isolating and cloning large DNA fragments has stimulated major efforts to map and sequence the human genome, which is the principal goal of an effort termed the **Human Genome Project.** The project also included sequencing the genomes of a number of model genetic organisms, including *E. coli*, yeast, *Arabidopsis thaliana*, the nematode *C. elegans*, *Drosophila*, and the laboratory mouse. To this end there have been set up, throughout the world, a number of specialized sequencing centers equipped with the robotic, imaging, and computational infrastructure to carry out high-throughput genomic sequencing on the order of many megabases of sequence per year. These and more conventional facilities have already sequenced numerous mitochondrial and

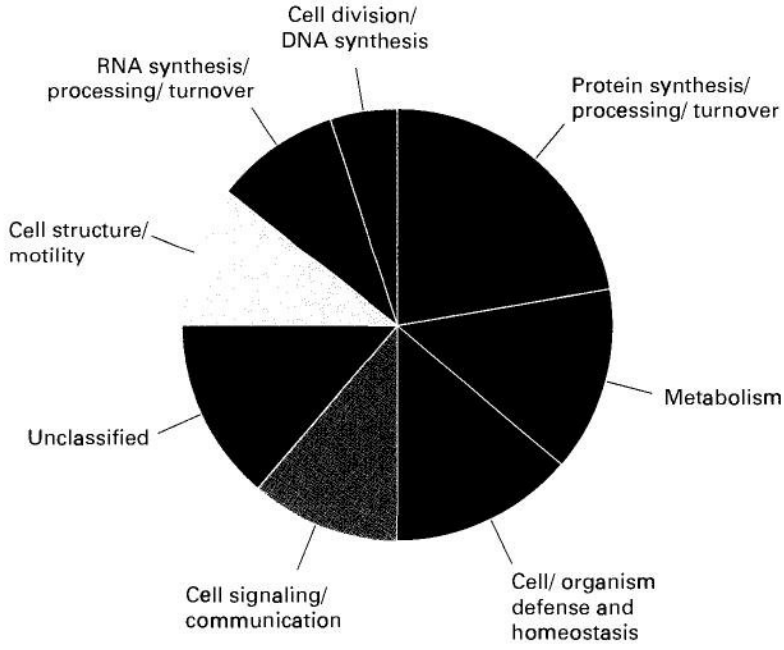
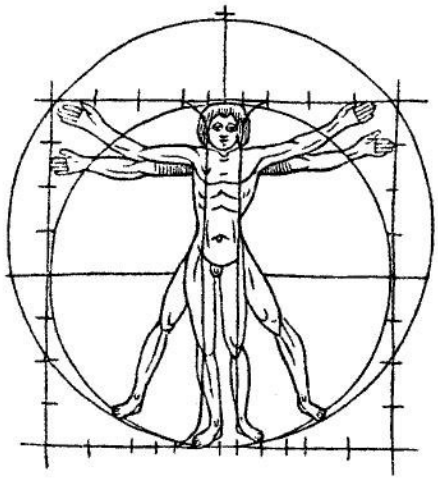


Figure 13.28 Classification of cDNA sequences by function. The pie chart is based on over 13,000 distinct, randomly selected human cDNA sequences. [Data courtesy of Craig Venter and the Institute for Genomic Research.]



Source of cDNA	Sequences	Source of cDNA	Sequences
Skin	3,043 (629)	Brain	67,679 (3,195)
Blood cells	23,505 (2,194)	Eye	1,932 (574)
Bone	5,736 (904)	Salivary gland	186 (17)
Thyroid	2,381 (584)	Esophagus	194 (76)
Parathyroid	197 (46)	Adipose tissue	2,412 (581)
Endothelial cells	5,736 (1,031)	Heart	9,400 (1,195)
Liver	37,807 (2,091)	Peritoneum	884 (163)
Gall bladder	3,754 (768)	Spleen	7,924 (924)
Colon	4,832 (879)	Adrenal gland	3,427 (658)
Smooth muscle	297 (127)	Pancreas	5,534 (1,094)
Small intestine	1,009 (297)	Kidney	3,213 (712)
Prostate	7,971 (1,203)	Epididymus	1,716 (370)
Testis	7,117 (1,232)	Skeletal muscle	4,693 (735)
Ovary	3,848 (504)	Synovial membrane	3,889 (813)
Uterus	6,392 (1,059)	Embryo	19,291 (1,989)
Placenta	12,148 (1,290)	Thymus	2,412 (261)

Figure 13.29 Classification of cDNA sequences by organ or tissue type. In each category, the initial number is the total number of cDNA clones examined. The number in parentheses is the number of distinct sequences found per organ or tissue type. [Data from M. D. Adams and 84 other authors. 1995. *Nature* 377 (Suppl.): 3.]

chloroplast genomes, more than 30 bacterial and 10 archaeal genomes, 3 complete eukaryotic genomes (*S. cerevisiae*, *C. elegans*, and *D. melanogaster*), and a number of human chromosomes. Some of these genomes are shown in **Table 13.1**. Genomic sequencing is partially completed for many organisms not shown in the table, including the agents that cause bubonic plague, gonorrhea, leprosy, and necrotizing fasciitis (caused by the Group A *Streptococcus*, "flesh-eating" bacterium), as well as rice, two fish species, and the mouse. There are clear biases in the genomes chosen for sequencing: they tend to be disease-causing organisms, model organisms, or archaeans.

What Can We Learn from Genomic Sequencing?

In terms of practical usefulness, a knowledge of the genomes of pathogens may help us devise better ways to make vaccines and hence to limit the death and destruction that these organisms cause. In a wider view, knowledge of the genomes of model organisms and that of human beings may reveal important new information about disease, development, neurobiology, aging, and many other biological processes. In a still wider view, genomic sequencing will inevitably lead to a greater understanding of ourselves, our world, and the evolutionary history of organisms.

Genomic sequencing can also raise new questions. For example, comparisons of the genomes of archaeans, bacteria, and eukaryotes indicate that the machinery for DNA replication and transcription in archaeans resembles that of eukaryotes, whereas the metabolism and the genes for metabolism strongly resemble those of bacteria. About half of the genes are unique to the Kingdom Archaea. However, neither the archaeal nor the bacterial genomes contain genes that encode proteins related to the actin-based cytoskeleton. This finding, together with evolutionary trees based on rDNA sequences, suggests that a still unidentified group of organisms may have participated in the origin of the eukaryotic cell.

With the recent completion of the sequence of DNA from the euchromatin of *Drosophila melanogaster*, it has become possible to compare the complete genomic DNA sequence of three eukaryotes (the yeast *Saccharomyces cerevisiae*, the nematode worm *Caeorhabditis elegans*, and the fruit fly *Drosophila melanogaster*) with each other as well as with a host of prokaryotic genomes and with genes and ESTs from two mammals (the mouse *Mus musculus* and human *Homo sapiens*). In *Haemophilus*, a prokaryote, there are 1709 protein-coding sequences. These sequences encode 1247 *unique proteins*, which have no more similarity to other proteins in the organism than would be expected by chance, and 178 *families of proteins*, each family consisting of two or more proteins that are significantly more similar than would be expected by chance. The individual members of each protein family are known as **paralogs**; their coding sequences all derive from a common ancestor through gene duplication and subsequent divergence. The **core proteome** of an organism consists of all unique proteins plus the number of families of paralogs encoded in the genome. In the case of *Haemophilus*, the core proteome consists of 1247 unique sequences + 178 families of paralogs = 1425 proteins. Eukaryotic proteomes are substantially larger. The yeast genome encodes 6241 predicted proteins, of which 4383 constitutes the core proteome. Worms have 18,424 and flies 13,601 predicted proteins, and their core proteomes number 9453 and 8065 proteins, respectively. Surprisingly, the relatively "complex" metazoan *Drosophila melanogaster*—which has a nervous system, an immune system, tissues and organs, and a complex development—has a core proteome only about twice that of yeast.

The genomes of the yeast, worm, and fly have also been analyzed to identify proteins that are **orthologs;** this term refers to highly similar proteins from different organisms that derive from a common ancestral coding sequence through speciation rather than gene duplication. (In comparing gene families, it is often difficult to distinguish paralogs from orthologs). About 41% of all yeast proteins have orthologs in flies, and 38% have orthologs in worms. On the other hand, whereas 28% of fly proteins have orthologs in yeast, 48% of fly proteins have orthologs in worms. In comparison with mammals, mammalian orthologs are found for 37% of yeast proteins, 34% of worm proteins, and 39% of fly proteins. In some instances, the

Table 13.1 Some examples of genomic sequencing

Domain	Organism	%A + T	Size (Mbp)	Comments/disease associations
Eubacteria				
Gamma (Enterobacteriaceae)	*Escherichia coli*	49	4.64	Normal intestinal flora
Gamma (Pasteurellaceae)	*Haemophilus influenzae*	62	1.83	Meningitis
Gamma (Pseudomonas group)	*Pseudomonas aeruginosa*	33	6.26	Respiratory infections
Gamma (Vibrionaceae)	*Vibrio cholerae*	53	4.0	Cholera
Epsilon subdivision	*Helicobacter pylori*	61	1.67	Duodenal ulcers
	Campylobacter jejuni	69	1.64	Gastroenteritis (food poisoning)
Alpha subdivision	*Rickettsia prowazekii*	71	1.11	Brill's disease and European epidemic typhus
Beta subdivision	*Neisseria meningitidis*	48	2.18	Meningitis
Bacillus/Clostridium group	*Bacillus subtilis*	56	4.20	Normal soil bacterium
	Mycoplasma genitalium	68	0.58	Smallest cellular genome
	Mycoplasma pneumoniae	60	0.82	Lung infections
	Ureaplasma urealyticum	75	0.75	Urethral infections
	Clostridium acetobutylicum	69	3.94	Food poisoning
	Streptococcus pyrogenes	61	1.85	Post-partum infections (puerperal fever)
Actinobacteria	*Mycobacterium tuberculosis*	34	4.41	Tuberculosis
Spirochaetales	*Treponema pallidum*	47	1.14	Syphilis
	Borrelia burgdorferi	71	0.91	Lyme disease
Chlamydia group	*Chlamydia pneumoniae*	59	1.23	Mild form of pneumonia
	Chlamydia trachomatis	59	1.04	Sexually transmitted
Aquificales	*Aquifex aeolicus*	57	1.55	Heat-tolerant, oxidizes hydrogen
Cyanobacteria	*Synechocystis* sp.	52	3.57	"Blue-green algae"
Deinococcus/Thermus	*Deinococcus radiodurans*	34	3.61	Radiation-resistant
Thermatogales	*Thermatoga maritima*	54	1.86	Anaerobic, hot marine mud
Archaea				
Euryarchaeota	*Methanococcus jannaschii*	69	1.66	Marine hydrothermal vents
	Archaeoglobus fulgidus	51	2.18	Marine hydrothermal sediment
	Pyrococcus abyssi	55	1.77	Extreme thermophile
	Pyrococcus furiosus	59	1.75	Extreme thermophile
	Pyrococcus horikoshii	58	1.74	Extreme thermophile
Crenarchaeota	*Aeropyrum pernix*	44	1.67	Marine sulfur vents
	Sulfolobus solfataricus	64	2.25	Yellowstone, USA, sulfur hot springs
Eukaryota				
Fungi	*Saccharomyces cerevisiae*	62	12.07	Budding yeast
Protozoa	*Leishmania major* (1 chsm)	52	0.27	Leishmaniasis
	Plasmodium falciparum (2 chsms)	80	2.01	Malaria
Nematode	*Caenorhabditis elegans*	64	100	Soil organism
Insect	*Drosophila melanogaster*	57	165	Fermenting fruits
Primate	*Homo sapiens* (chsm 22)	52	32	Ubiquitous; first chromosome sequenced

Source: Data courtesy of David Ussery, Peder Worning, Hans Henrik Stærfeldt, and Lars Juhl Jensen, Center for Biological Sequence Analysis, Department of Biotechnology, The Technical University of Denmark.

similarities in sequence are known to reflect conservation of function, because more that 70 cases have been described in which a DNA sequence encoding a human protein will substitute for a yeast gene and rescue a mutant phenotype. These include genes encoding DNA ligase, translation initiation factors, proteins involved in signal transduction, cyclins that control the cell cycle, and RAS proto-oncogenes that are associated with cancer.

Studies in model organisms have contributed to our understanding of genes and pathways that are implicated in human disease. Among a set of 289 genes that are known to be mutated or amplified in human diseases, 230 of the encoded proteins are similar to proteins found in *D. melanogaster*, 212 to ones in *C. elegans*, and 121 to ones in *S. cerevisiae*. These include proteins involved in DNA synthesis, DNA repair, general metabolism, and the cytoskeleton. Among the human genes associated with cancer, 68 percent have an ortholog in *Drosophila*; somewhat fewer have an ortholog in *Caenorhabditis*, and the similarity with worm genes is not usually as great as that found in fly genes. For some of the cancer genes, even yeast has an ortholog, particularly genes involved in DNA metabolism.

In summary, comparison of the genomes of three eukaryotic model organisms—a yeast, a nematode worm, and a fruit fly—has yielded interesting and surprising conclusions. One finding is that the core proteomes of the worm and the fly are similar in size and only about twice that of yeast. Thus, the complexity of metazoan organisms does not arise primarily through an increased number of genes and proteins. However, the proteins encoded in the metazoan genomes are, on average, larger and more complex (have more domains) than those encoded in the yeast genome. These complex proteins include, for example, proteins with extracellular domains that allow cell–cell and cell–substrate interactions. The main message from genomic analysis is therefore an encouraging one. The evolutionary conservation of similarity of protein sequence and function in the "lower" eukaryotes and in humans is a clear indication that what we learn from model organisms will continue to serve as a reliable guide to understanding human biology so long as we proceed cautiously and intelligently.

13.8 Functional Genomics

Genomic sequencing has made possible a new approach to genetics called **functional genomics,** which focuses on genome-wide patterns of gene expression and the mechanisms by which gene expression is coordinated. As changes take place in the cellular environment—for example, through changes in the external conditions, development, or aging—the patterns of gene expression change also. But genes are usually deployed in sets, not individually. As the level of expression of one coordinated set is decreased, the level of expression of a different coordinated set may be increased. How can one study tens of thousands of genes all at the same time?

DNA Chips

The study of genome-wide patterns of gene expression became feasible with the development of the **DNA chip** (or *microarray*), a flat surface about the size of a postage stamp on which 10,000–100,000 distinct spots are present, each containing a different immobilized DNA sequence suitable for hybridization with DNA or RNA isolated from cells growing under different conditions, from cells not exposed or cells exposed to a drug or toxic chemical, from different stages of development, or from different types or stages of a disease such as cancer. Two types of DNA chips are presently in use:

- A chip arrayed with oligonucleotides synthesized directly on the chip, one nucleotide at a time, by automated procedures; these chips typically have hundreds of thousands of spots per array.
- A chip arrayed with denatured, double-stranded DNA sequences of 500–5000 bp, in which the spots, each about a millionth of a drop in volume, are deposited by capillary action from miniaturized fountain-pen-like devices mounted on the moveable head of a flatbed robotic workstation; these chips typically have tens of thousands of spots per array.

Figure 13.30 shows one method by which DNA chips are used to assay the global levels of gene expression in an experimental sample relative to a control. At the upper right are shown six adjacent spots in the microarray, each of which contains a DNA sequence

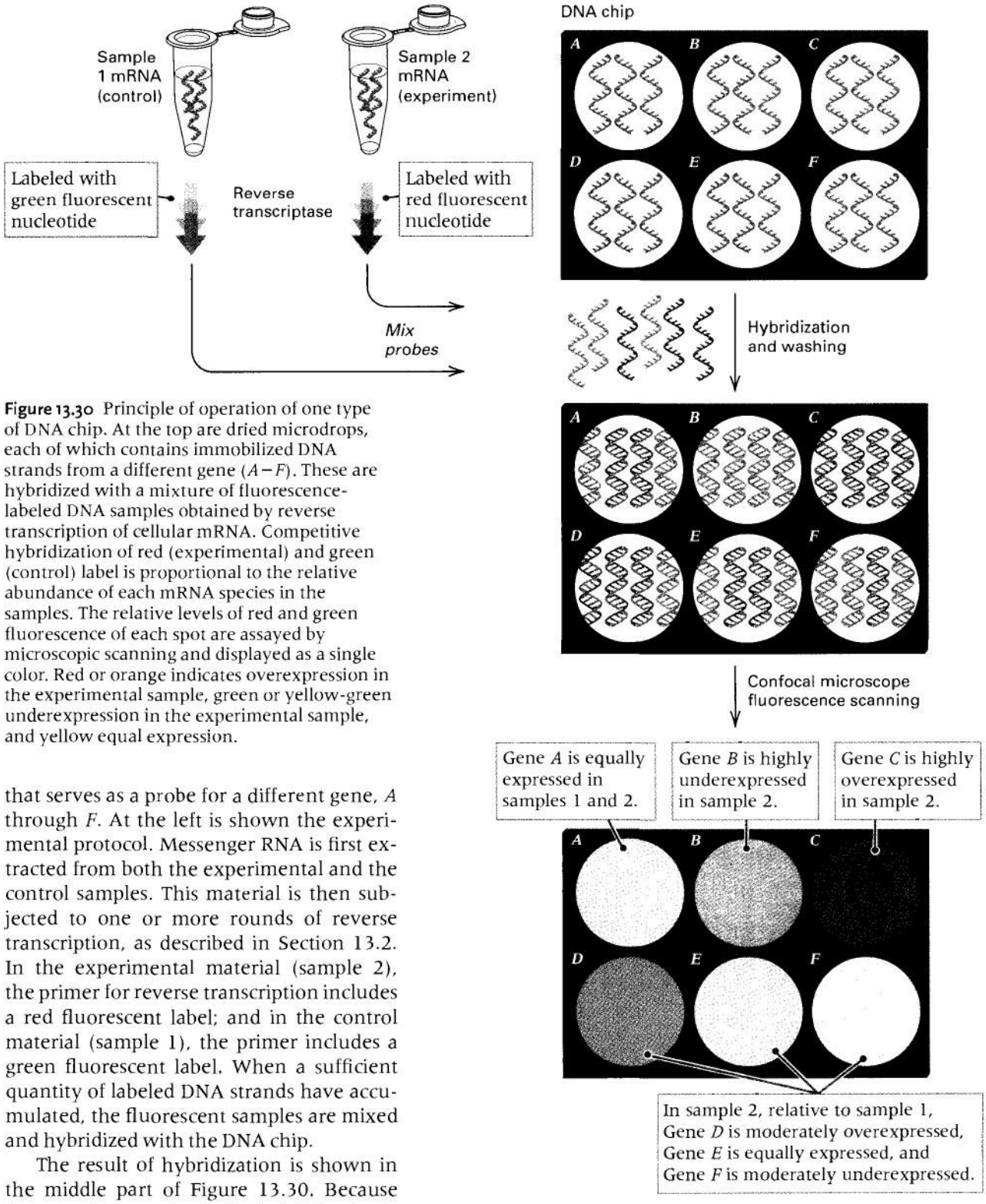
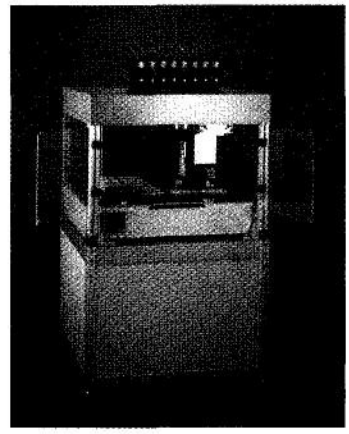


Figure 13.30 Principle of operation of one type of DNA chip. At the top are dried microdrops, each of which contains immobilized DNA strands from a different gene ($A-F$). These are hybridized with a mixture of fluorescence-labeled DNA samples obtained by reverse transcription of cellular mRNA. Competitive hybridization of red (experimental) and green (control) label is proportional to the relative abundance of each mRNA species in the samples. The relative levels of red and green fluorescence of each spot are assayed by microscopic scanning and displayed as a single color. Red or orange indicates overexpression in the experimental sample, green or yellow-green underexpression in the experimental sample, and yellow equal expression.

that serves as a probe for a different gene, *A* through *F*. At the left is shown the experimental protocol. Messenger RNA is first extracted from both the experimental and the control samples. This material is then subjected to one or more rounds of reverse transcription, as described in Section 13.2. In the experimental material (sample 2), the primer for reverse transcription includes a red fluorescent label; and in the control material (sample 1), the primer includes a green fluorescent label. When a sufficient quantity of labeled DNA strands have accumulated, the fluorescent samples are mixed and hybridized with the DNA chip.

The result of hybridization is shown in the middle part of Figure 13.30. Because the samples are mixed, the hybridization is competitive, and therefore the density of red or green strands bound to the DNA chip is proportional to the concentration of red or green molecules in the mixture. Genes that are overexpressed in sample 2 relative to sample 1 will have more red strands hybridized to the spot, whereas those that are underexpressed in sample 2 relative to sample 1 will have more green strands hybridized to the spot.

After hybridization, the DNA chip is placed in a confocal fluorescence scanner



A commercially available microarrayer used to make the types of DNA microarrays illustrated in Figure 13.30. The interior of the chamber must be kept scrupulously clean. [Courtesy: Cartesian Technologies, Inc.]

that scans each pixel (the smallest discrete element in a visual image) first to record the intensity of one fluorescent label and then again to record the intensity of the other fluorescent label. These signals are synthesized to produce the signal value for each spot in the microarray. A great deal of automated computation is necessary at this stage. The density of the DNA is not necessarily uniform from one spot to the next across the chip, and it can even vary according to location within a particular spot. There are also differences in background fluorescence as a result of inhomogeneities in the material of the chip itself. These are compensated as follows:

- The average background fluorescence surrounding each spot is measured for each wavelength and subtracted from the fluorescence of each pixel within the spot.
- After correction for local background, the signal for each pixel is obtained by combining the intensities of the red and green signals.
- The final signal for each spot is obtained by averaging across all pixels in the spot. (In the type of DNA chip illustrated here, there are approximately 400 pixels per spot.)

The final signals indicate the relative levels of gene expression by color, as shown at the bottom of Figure 13.30. A spot that is red or orange indicates high or moderate overexpression of the gene in the experimental sample, a spot that is green or yellow-green indicates high or moderate underexpression of the gene in the experimental sample, and a spot that is perfectly

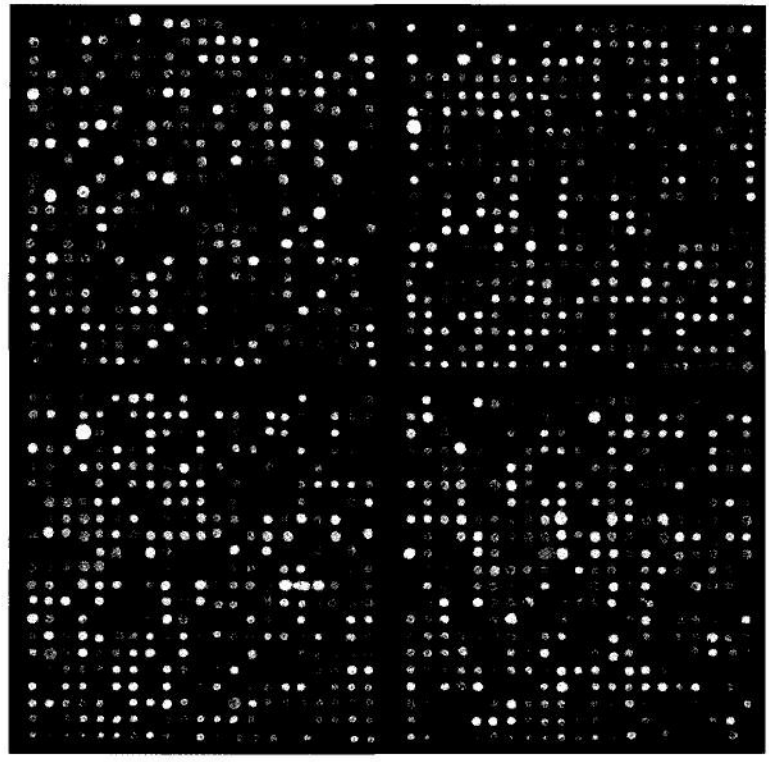


Figure 13.31 Small part of a yeast DNA chip showing 1764 spots, each specific for hybridization with a different mRNA sequence. The color of each spot indicates the relative level of gene expression in experimental and control samples. The complete chip for all yeast open reading frames includes over 6200 spots. [Courtesy of Jeffrey P. Townsend, Duccio Cavalieri, and the Harvard Center for Genomics Research.]

yellow indicates equal levels of gene expression in the samples. In this manner, DNA chips can assay the relative levels of any mRNA species whose abundance in the sample is more than one molecule per 10^5, and differences in expression as small as approximately twofold can be detected. **Figure 13.31** shows an image of part of a yeast DNA chip; each spot corresponds to a different gene in the organism. An important component of functional genomics is bioinformatics: the challenge to acquire, store, retrieve, and analyze such immense amounts of data as represented by simultaneous patterns of expression of thousands of genes. Take the yeast chip, part of which is shown in Figure 13.31, for example. If each gene were compared only with each other gene, the total number of pairwise combinations would already be more than 20×10^6.

Chapter Summary

Recombinant DNA technology makes it possible to modify the genotype of an organism in a directed, predetermined way by enabling different DNA molecules to be joined into novel genetic units, altered as desired, and reintroduced into the germ line. Restriction enzymes play a key role in the technique, because they can cleave DNA molecules within particular base sequences. Many restriction enzymes generate DNA fragments with complementary single-stranded ends, which can anneal and be ligated together with similar fragments from other DNA molecules. The carrier DNA molecule used to propagate a desired DNA fragment is called a vector. The most common vectors are plasmids, phages, viruses, and yeast artificial chromosomes (YACs). Transformation is an essential step in the propagation of recombinant molecules because it enables the recombinant DNA molecules to enter host cells, such as those of bacteria, yeast, or mammals. If the recombinant molecule has its own replication system or can use the host replication system, then it can replicate. Plasmid vectors become permanently established in the host cell; phage can multiply and produce a stable population of phage carrying source DNA; retroviruses can be used to establish a gene in an animal cell; and YACs include source DNA within an artificial chromosome that contains a functional centromere and telomeres.

Recombinant DNA can also be used to transform the germ line of animals or to genetically engineer plants. These techniques form the basis of reverse genetics, in which genes are deliberately mutated in specified ways and introduced back into the organism to determine the effects on phenotype. Reverse genetics is routine in genetic analysis in bacteria, yeast, nematodes, *Drosophila*, the mouse, and other organisms. In *Drosophila*, transformation employs a system of two vectors based on the transposable *P* element. One vector contains sequences that produce the *P* transposase; the other contains the DNA of interest between the inverted repeats of *P* and other sequences needed for mobilization by transposase and insertion into the genome. Germ-line transformation in the mouse makes use of retrovirus vectors or embryonic stem cells. Dicotyledonous plants are transformed with T DNA derived from the *Ti* plasmid found in species of *Agrobacterium*, whose virulence genes promote a conjugation-like transfer of T DNA into the host plant cell, where it is integrated into the chromosomal DNA.

Practical applications of recombinant DNA technology include the efficient production of useful proteins, the creation of novel genotypes for the synthesis of economically important molecules, the generation of DNA and RNA sequences for use in medical diagnosis, the manipulation of the genotype of domesticated animals and plants, the development of new types of vaccines, and the potential correction of genetic defects (gene therapy). Production of eukaryotic proteins in bacterial cells is sometimes hampered by protein instability, inability to fold properly, or failure to undergo necessary chemical modification. These problems are often eliminated by production of the protein in eukaryotic cells.

Specialized cloning vectors (PACs, BACs, and YACs) allow large DNA molecules to be isolated and manipulated. They have made possible the construction of detailed physical maps that integrate many levels of genetic information, such as the positions of sequence-tagged sites in the genetic (recombination) map or the positions and lengths of cloned contigs. Complete DNA sequences have been determined for numerous mitochondria and chloroplasts, as well as for the genomes of many bacteria, several archaeans, and a few eukaryotes with relatively small genomes. The sequence of the yeast genome has revealed a high density of genetic information and evidence of an ancient genome duplication. Many coding regions in yeast have previously unknown functions and yield no recognizable phenotype when mutated. The sequencing of two metazoan genomes has revealed that the proteins encoded are on average larger and more complex than those of a unicellular eukaryote. Many human disease genes have counterparts in other metazoan and even the yeast genome. Beyond analysis of the DNA sequences themselves, and comparisons among organisms, genomics has spawned the field of functional genomics to describe and understand the genome-wide patterns of gene expression in cells under different conditions, including normal development and disease.

Key Terms

bacterial artificial chromosome (BAC)
blunt end
cloned DNA
colony hybridization
complementary DNA (cDNA)
contig
core proteome
DNA chip
DNA ligase
electroporation
embryonic stem cell
expressed sequence tag (EST)
functional genomics
gene cloning
gene targeting
gene therapy
genetic engineering
Human Genome Project
knockout mutation
library
map-based cloning
metallothionein
multiple cloning site
ortholog
P1 artificial chromosome (PAC)
paralog
partial digestion
physical map
plasmid
polylinker
positional cloning
pulsed-field gel electrophoresis (PFGE)
P element
recombinant DNA
retrovirus
reverse genetics
reverse transcriptase
reverse transcriptase PCR (RT-PCR)
sequence-tagged site (STS)
site-directed mutagenesis
sticky end
T DNA
Ti plasmid
transformation rescue
transgenic organism
vector
yeast artificial chromosome (YAC)

Review the Basics

- What is meant by the term "recombinant DNA"? What are some of the practical uses of recombinant DNA?
- Why are restriction endonucleases essential in most types of recombinant DNA research?
- What features are essential in a bacterial cloning vector? How can a vector have more than one cloning site? What is a multiple cloning site (polylinker)?
- What is the reaction catalyzed by the enzyme reverse transcriptase? How is this enzyme used in recombinant DNA technology?
- Distinguish between a physical map and a genetic map. If recombination took place at uniform frequencies throughout a genome, how would the genetic map compare with the physical map?
- Explain how the genetic map of a chromosome is used in positional cloning.
- What is a transgenic organism? What are some of the practical uses of transgenic organisms?
- Explain the role of recombination in gene targeting in embryonic stem cells. How can gene targeting be used to create a "knockout" mutation?
- Where does T DNA come from, and how is it used in making transgenic plants?
- In the context of genome analysis, what is a BAC? an STS? a contig?
- What are the major goals of the Human Genome Project? Which of them have been achieved?
- Explain the term *functional genomics*. What are DNA chips and how are they used in functional genomics? In the type of functional genomics experiment described in the text, how would you interpret a spot on a DNA microarray that is red or orange? yellow? green or yellow-green?

Guide to Problem Solving

Problem 1 What is the average distance between restriction sites for each of the following restriction enzymes? Assume that the DNA substrate has a random sequence with equal amounts of each base. The symbol N stands for any nucleotide, R for any purine (A or G), and Y for any pyrimidine (T or C).

(a) TCGA *Taq*I
(b) GGTACC *Kpn*I
(c) GTNAC *Mae*III
(d) GGNNCC *Nla*IV
(e) GRCGYC *Acy*I

Answer The average distance between restriction sites equals the reciprocal of the probability of occurrence of the restriction site. You must therefore calculate the probability of occurrence of each restriction site in a random DNA sequence.

(a) The probability of the sequence TCGA is

$$1/4 \times 1/4 \times 1/4 \times 1/4 = (1/4)^4 = 1/256$$

so 256 bases is the average distance between *Taq*I sites.

(b) By the same reasoning, the probability of a *Kpn*I site is $(1/4)^6 = 4096$, so 4096 bases is the average distance between *Kpn*I sites.

(c) The probability of N (any nucleotide at a site) is 1, so the probability of the sequence GTNAC equals

$$1/4 \times 1/4 \times 1 \times 1/4 \times 1/4 = (1/4)^4 = 1/256$$

therefore, 256 is the average distance between *Mae*III sites.

(d) The same reasoning yields the average distance between *Nla*IV sites as

$$(1/4 \times 1/4 \times 1 \times 1 \times 1/4 \times 1/4)^{-1} = 256 \text{ bases.}$$

(e) The probability of an R (A or G) at a site is 1/2, and the probability of a Y (T or C) at a site is 1/2. Hence the probability of the sequence GRCGYC is

$$1/4 \times 1/2 \times 1/4 \times 1/4 \times 1/2 \times 1/4 = 1/1024$$

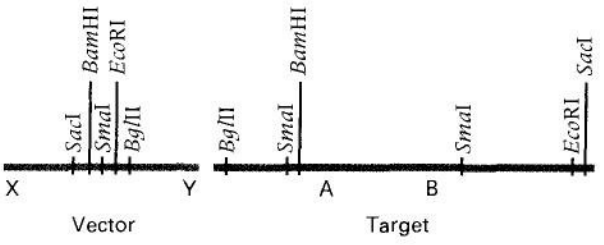
so there is an average of 1024 bases between *Acy*I sites.

Problem 2 Many engineered vectors have a polylinker, or multiple cloning site (MCS), instead of a single restriction site available for insertion of DNA. A polylinker is a short DNA sequence containing multiple different restriction sites, each unique to the vector, any combination of which can be cleaved with the appropriate restriction enzymes to yield a linear vector molecule with defined restriction sites at the ends. The accompanying diagram illustrates five restriction sites in the polylinker of a vector; the vector sequences adjacent to the polylinker are denoted X and Y. Also shown is the restriction map of a genomic target sequence whose ends are denoted A and B. The restriction enzymes *Sac*I, *Bam*HI, *Bgl*II, and *Eco*RI all produce sticky ends; *Sma*I produces blunt ends.

(a) If the vector and target were both digested with *Sma*I and the resulting fragments ligated in such a way that each vector molecule was ligated with one and only one target fragment, in what orientation would the A–B fragment be ligated into the polylinker?

(b) What restriction enzyme (or enzymes) would you use to digest the vector and target so that after mixing of the fragments and ligation, the cloned DNA would have the sequence X–B–A–Y?

(c) What restriction enzyme (or enzymes) would you use to digest the vector and target so that after mixing of the fragments and ligation, the cloned DNA would have the sequence X–A–B–Y?



Answer **(a)** *Sma*I produces blunt ends, and any blunt end can be ligated onto any other. Hence the *Sma*I fragment A–B can be ligated into the polylinker in either of two orientations. The resulting clones are expected to be X–A–B–Y and X–B–A–Y in equal frequency. **(b)** A restriction enzyme produces fragments whose ends are identical, so either end can be ligated onto a complementary sticky end. To force the orientation X–A–B–Y, one needs to cleave the vector and the target with two restriction enzymes that produce different sticky ends. The site nearest X in the vector must match the site nearest A in the target, and the site nearest Y in the vector must match the site nearest B in the vector. In this case, if the vector and target are both cleaved with *Bam*HI and *Eco*RI, the resulting clone is expected to have the orientation X–A–B–Y. **(c)** Following the logic of part **(b)**, digestion of the vector and target with *Bgl*II and *Sac*I will yield clones with the orientation X–B–A–Y.

Problem 3 The accompanying table shows the presence or absence of seven sequence-tagged sites (STSs), numbered 1 through 7, in each of six BAC clones, designated A through F. A + in the table means that PCR primers specific to the STS are able to amplify the sequence from the clone, − means that the STS cannot be amplified from the clone. Use these data to make a map of the order of the BAC clones and of the STS markers located within the clones. In orienting the map, put BAC clone A at the top.

	STS marker						
BAC clone	1	2	3	4	5	6	7
A	−	−	+	+	−	−	−
B	−	+	−	−	−	−	+
C	+	−	−	−	+	−	−
D	+	−	−	−	−	+	−
E	−	−	−	+	−	+	−
F	−	+	−	−	+	−	−

GeNETics on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

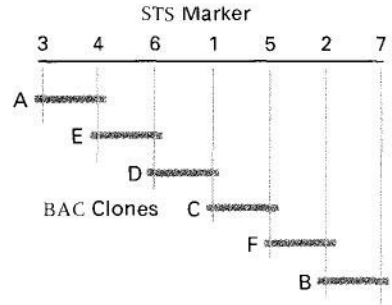
For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by-chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- Genomic sequencing has already been completed or is nearing completion for 20 eukaryotic genomes, including the human genome, and 65 prokaryotic genomes—not to mention the genomes of organelles of numerous species as well as numerous viruses, bacteriophage, and plasmids. An extensive list of **genome sequencing projects** is maintained at this keyword site. It is useful to have a biological dictionary or encyclopedia close at hand because the organisms are listed according to their official binomial Latin names (for example, *Oryza sativa*), rather than their common names (in this case, rice).

- The use of **human gene therapy** has been controversial as a consequence of several well-publicized failures, but recent experience has been more promising. One example is in the treatment of X-linked severe combined immunodeficiency disease at Hospital Necker in Paris. This disease is caused by a mutation in a cytokine receptor that blocks the development of T cells and NK (natural killer) lymphocytes. Because affected boys lack a functional immune system, they usually die from infections such as chicken pox or pneumonia before two years of age. In this case, bone marrow transplants were performed with two patients, using their own cells that had been transformed with the wildtype X-linked gene. After recovery, T-cell and NK-cell counts were comparable to those of normal children. At this keyword site you can learn more about this and other examples of human gene therapy.

- What are the **social implications** of advances in genomics? Some groups are worried because the application of genetic technologies raises ethical and legal issues of the foreknowledge of one's health, as well as issues of genetic privacy and insurability. Others are optimistic that the technologies will yield great benefits for medicine and society. This keyword

Answer The principle in ordering the BAC clones is that the DNA fragments present in any two (or more) clones that contain the same STS marker must overlap. The table contains no information bearing on the relative sizes of the BAC clones or on the distance between the STS markers. Thus each clone is drawn as though it had the same size, and the STS markers are spaced equally. Considering the STS markers shared between clones, and starting with clone A, we obtain the map shown at right. The mirror-image map with STS 3 on the right and STS 7 on the left is also a valid solution.



Analysis and Applications

13.1 The restriction enzyme *Nla*IV cleaves the recognition sequence GGNNCC between the two N's (N stands for any nucleotide). How many *Nla*IV sites are possible? How many of these are palindromes?

13.2 The restriction enzyme *Asp*I cleaves the restriction site GACN↓NNGTC and cleaves at the position of the arrow. How many *Asp*I sites are possible? If two randomly chosen *Asp*I sites are cleaved, what is the probability that their 5' overhangs are compatible?

13.3 The restriction enzyme *Sfi*I has the restriction site GGCCNNNNNGGCC, and the restriction enzyme *Eac*I has the restriction site YGGCCR, where Y is any pyrimidine and R is the complementary purine. What is the probability that an *Sfi*I site includes one *Eac*I site? two *Eac*I sites?

13.4 The enzyme *Ava*II cleaves the restriction site G↓GYCC at the position of the arrow. If two randomly chosen *Ava*II sites are cleaved, what is the probability that their "sticky ends" are compatible?

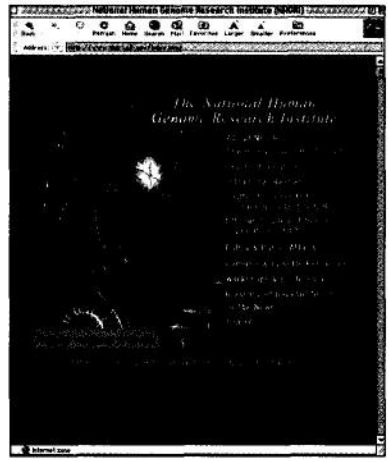
13.5 A circular plasmid has three restriction sites for the enzyme *Eco*RI. After digestion the fragments are ligated together, and a circular product is isolated that has each of the three fragments present and in the correct order.

site is devoted to the exploration of ethical, legal, and social issues related to sequencing the human genome, to the stimulation of public discussion, and to the formulation of public policies to prevent misuse of genetic information.

- The debate over the growing and marketing of **genetically engineered crops** has been fierce, especially in Europe. Opponents call them Frankenstein's plants. They argue that it is foolhardy to plant millions of acres of novel crop plants and let their products enter the food supply without evaluating the possible long-term effects on human health and the environment. Proponents argue that the maintenance of an affordable food supply depends on the increased productivity expected from genetically modified plants. This keyword site provides a balanced discussion of the issues and points out that while the debate continues, genetically modified crops already constitute a large portion of the food supply in the United States.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 13 and you will be linked automatically.

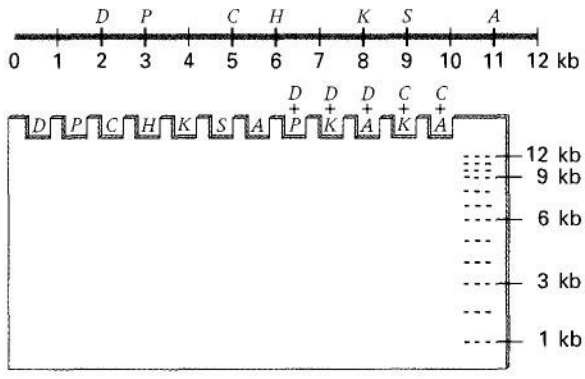
- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 13.



[Courtesy National Human Genome Research Institute/National Institutes of Health.]

Does this mean that the ligated plasmid is the same as the original? Explain.

13.6 Shown here is a restriction map of a 12-kb linear plasmid isolated from cells of *Borrelia burgdorferi*, a spirochete bacterium transmitted by the bite of Ixodes ticks that causes Lyme disease. The symbols *D, P, C, H, K, S,* and *A* represent cleavage sites for the restriction enzymes *Dde*I, *Pst*I, *Cla*I, *Hin*dIII, *Kpn*I, *Sac*I, and *Alu*I, respectively. In the accompanying gel diagram, show the positions at which bands would be found after digestion of the plasmid with the indicated restriction enzyme or enzymes.



13.7 A DNA chip is hybridized with fluorescently labeled reverse-transcribed DNA as described in the text, where the control mRNA (C) is labeled with a green fluor and the experimental mRNA (E) with a red fluor. Indicate what you can conclude about the relative levels of expression of a spot in the microarray that fluoresces:

(a) Red.
(b) Green.
(c) Yellow.
(d) Orange.
(e) Lime green.

13.8 A functional genomics experiment is carried out using a DNA chip to assay levels of gene expression in a species of bacteria. What genes would you expect to find overexpressed in cells grown in mimimal medium compared to cells grown in complete medium?

13.9 The euchromatic part of the *Drosophila* genome that is highly replicated in the banded salivary gland chromosomes is approximately 120 Mb (million base pairs) in size. The salivary gland chromosomes contain approximately 5000 bands. For ease of reference, the salivary chromosomes are divided into about 100 approximately equal, numbered sections (1–100), each of which consists of six lettered subdivisions (A–F). On average, how much DNA is there in a salivary gland band? in a lettered subdivision? in a numbered section? How do these compare with the size of the DNA insert in a 200-kb (kilobase pair) YAC? with the size of the DNA insert in an 80-kb P1 clone?

13.10 Restriction enzymes generate one of three possible types of ends on the DNA molecules that they cleave. What are the three possibilities?

13.11 Are the ends of different restriction fragments produced by a particular restriction enzyme always the same? Must opposite ends of each restriction fragment be the same? Why?

13.12 Will the sequences 5'-GGCC-3' and 3'-GGCC-5' in a double-stranded DNA molecule be cut by the same restriction enzyme?

13.13 In cloning into bacterial vectors, why is it useful to insert DNA fragments to be cloned into a restriction site inside an antibiotic-resistance gene? Why is another gene for resistance to a second antibiotic also required?

13.14 How frequently would the restriction enzymes *Taq*I (restriction site TCGA) and *Mae*III (restriction site GTNAC, in which N is any nucleotide) cleave double-stranded DNA molecules containing random sequences of

(a) 1/6 A, 1/6 T, 1/3 G, and 1/3 C?

(b) 1/3 A, 1/3 T, 1/6 G, and 1/6 C?

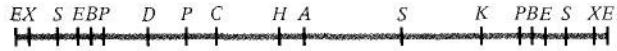
13.15 If the genomic and cDNA sequences of a gene are compared, what information does the cDNA sequence give you that is not obvious from the genomic sequence? What information does the genomic sequence contain that is not in the cDNA?

13.16 What might prevent a cloned eukaryotic gene from yielding a functional mRNA in a bacterial host? Assuming that these problems are overcome, why might the desired protein still not be produced?

13.17 You want to introduce the human insulin gene into a bacterial host in hopes of producing a large amount of human insulin. Should you use the genomic DNA or the cDNA? Explain your reasoning.

13.18 When DNA isolated from phage J2 is treated with the enzyme *Sal*I, eight fragments are produced with sizes of 1.3, 2.8, 3.6, 5.3, 7.4, 7.6, 8.1, and 11.4 kilobase pairs. However, if J2 DNA is isolated from infected cells, only seven fragments are found, with sizes of 1.3, 2.8, 7.4, 7.6, 8.1, 8.9, and 11.4 kb. What form of the intracellular DNA can account for these results?

13.19 A mutant allele is found to express the wildtype gene product, but at only about 20 percent of the wildtype level. The mutation is traced to an intron whose size has increased by 3.1 kb because of the presence of a DNA fragment with the restriction map shown here. The symbols *A*, *B*, *C*, *D*, *E*, *H*, *K*, *P*, *S*, and *X* represent cleavage sites for the restriction enzymes *Alu*I, *Bam*HI, *Cla*I, *Dde*I, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Sac*I, and *Xho*II, respectively. Does the restriction map of the insertion give any clues to what it is?



13.20 Phage X82 DNA is cleaved into six fragments by the enzyme *Bgl*I. A mutant is isolated with plaques that look quite different from the wildtype plaque. DNA isolated from the mutant is cleaved into only five fragments. What possible genetic changes can account for the difference in the number of restriction fragments?

13.21 The wildtype allele of a bacterial gene is easily selected by growth of cells on a special medium. Repeated attempts to clone the gene by digestion of cellular DNA with the enzyme *Eco*RI are unsuccessful. However, if the enzyme *Hin*dIII is used, then clones are easily found. Explain.

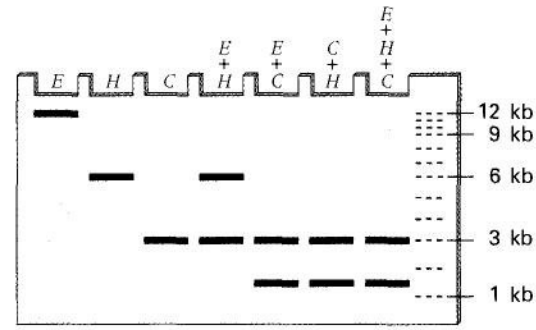
13.22 A *kan-r tet-r* plasmid is treated with the restriction enzyme *Bgl*I, which cleaves in the *kan* (kanamycin) gene. The DNA is annealed with a *Bgl*I digest of *Neurospora* DNA and then used to transform *E. coli*.

(a) What antibiotic would you put in the growth medium to ensure that each colony has the plasmid?

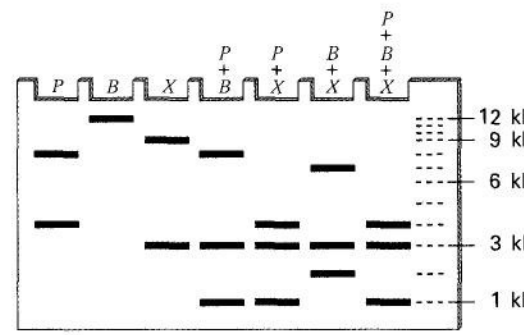
(b) What antibiotic-resistance phenotypes will be found among the resulting colonies?

(c) Which phenotype will contain *Neurospora* DNA inserts?

13.23 A circular plasmid gives the illustrated pattern of bands when digested singly or in combination with three restriction enzymes, *Eco*RI (*E*), *Hin*dIII (*H*), and *Cla*I (*C*). Draw a restriction map of the plasmid. (*Hint*: Some bands may conceal multiple fragments, because the total size of all fragments produced must be the same across all lanes in the gel.)

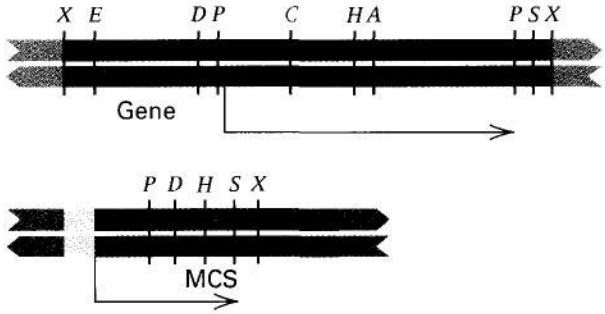


13.24 A circular plasmid gives the illustrated pattern of bands when digested singly or in combination with three restriction enzymes, *Pst*I (*P*), *Bam*HI (*B*), and *Xho*II (*X*). Draw a restriction map of the plasmid. (*Hint*: Some bands may conceal multiple fragments, because the total size of all fragments produced must be the same across all lanes in the gel.)



13.25 A lac^+ *tet-r* plasmid is cleaved in the *lac* gene with a restriction enzyme. The enzyme has a four-base restriction site and generates fragments with a two-base single-stranded overhang. The cutting site in the *lac* gene is in the codon for the second amino acid in the chain, a site that can tolerate any amino acid without loss of function. After cleavage, the single-stranded ends are converted to blunt ends with DNApolymerase I, and then the ends are joined by blunt-end ligation to recreate a circle. A lac^- *tet-s* bacterial strain is transformed with the DNA, and tetracycline-resistant bacteria are selected. What is the Lac phenotype of the colonies?

13.26 Shown here is a gene with its transcription start site and direction of transcription indicated, along with the multiple cloning site (MCS) in a plasmid vector. The symbols *A, C, D, E, H, P, S,* and *X* represent cleavage sites for the restriction enzymes *Alu*I, *Cla*I, *Dde*I, *Eco*RI, *Hin*dIII, *Pst*I, *Sac*I, and *Xho*II, respectively.



A geneticist wishes to place the coding sequence of the gene under the control of a promoter located to the left of the MCS, as shown in the diagram. To do this, she isolates the *Pst*I fragment containing the coding sequence, adds it to a solution containing a plasmid that has been cleaved with *Pst*I, ligates the sticky ends together, and transforms bacterial cells. She chooses a single transformant containing a plasmid that has the *Pst*I fragment inserted in the right place but finds, to her disappointment, that the gene is not transcribed. Suggest an explanation.

13.27 The nonfunctional recombinant DNA clone described in the previous problem is analyzed to deduce a restriction map in the vicinity of the MCS. The result is shown in the accompanying illustration. Does this finding help explain the unexpected result?



13.28 Analysis of additional clones from Problem 13.26 reveals that they fall into two classes with respect to the restriction map in the vicinity of the MCS. One of the restriction maps is shown in the previous problem. What does the other look like? Would this type of clone be expected to transcribe the gene of interest? Why or why not?

13.29 Another geneticist examines the restriction maps in Problem 13.26 and suggests that *directional cloning* would be a more fruitful approach to making sure that the coding sequence is transcribed in the recombinant plasmid. In this approach, the gene of interest and the plasmid are both cleaved with the same two restriction enzymes, which are chosen so that their sticky ends will ensure that the fragment of interest is inserted in the correct orientation in the plasmid. For directional cloning of the fragment in Problem 13.26, which two restriction enzymes should be used? What is the restriction map of the resulting plasmid in the vicinity of the MCS?

13.30 Plasmid pBR607 DNA is a double-stranded circle of 4 kilobase pairs. This plasmid carries two genes whose protein products confer resistance to tetracycline (Tet-r) and ampicillin (Amp-r) in host bacteria. The DNA has a single site for each of the following restriction enzymes: *Eco*RI, *Bam*HI, *Hin*dIII, *Pst*I, and *Sal*I. Cloning DNA into the *Eco*RI site does not affect resistance to either drug. Cloning DNA into the *Bam*HI, *Hin*dIII, and *Sal*I sites abolishes tetracycline resistance. Cloning into the *Pst*I site abolishes ampicillin resistance. Digestion with the following mixtures of restriction enzymes yields fragments with the sizes listed below. Indicate the positions of the *Pst*I, *Bam*HI, *Hin*dIII, and *Sal*I cleavage sites on a restriction map, relative to the *Eco*RI cleavage site.

Enzyme mixture	Fragment size (kb)
*Eco*RI + *Pst*I	0.70, 3.30
*Eco*RI + *Bam*HI	0.30, 3.70
*Eco*RI + *Hin*dIII	0.08, 3.92
*Eco*RI + *Sal*I	0.85, 3.15
*Eco*RI + *Bam*HI + *Pst*I	0.30, 0.70, 3.00

Challenge Problems

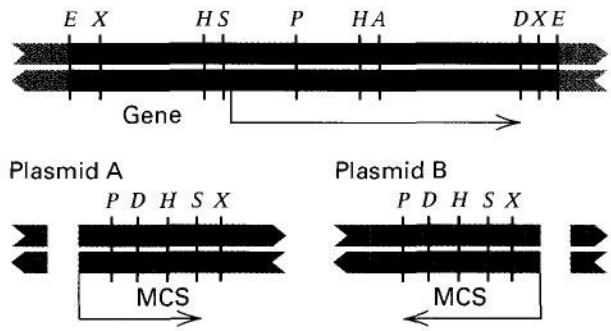
13.31 A geneticist decides to use yeast DNA chips (microarrays) to study the effects of deletion mutations in the yeast mating-type locus *MAT*. To perform these experiments, RNA is extracted from two strains of yeast to be compared, and labeled cDNA probe is prepared from each strain by reverse transcription of the RNA in the presence of either green (Cy3) or red (Cy5) fluorescent label. The differently labeled cDNA solutions are mixed and applied to a microarray with spots of DNA sequence for each of the types of genes shown in the accompanying table.

Red (Cy5)	Wildtype α cells	Wildtype α cells	Wildtype **a** cells	$\Delta MATa1$ cells	$\Delta MAT\alpha1$ cells	$\frac{\Delta MAT(\alpha2\alpha1)}{\Delta MATa1}$
Red (Cy5)	Wildtype **a** cells	Wildtype **a**/α cells	Wildtype **a**/α cells	Wildtype **a** cells	Wildtype α cells	Wildtype *a* cells
MATα1						
MATα2						
MATa1						
αsg						
asg						
hsg						

Problem 13.31

MATα1, *MATα2*, and *MATa1* are DNAs for the possible coding sequences present at the *MAT* locus. *αsg* and *asg* spots contain DNA of genes expressed specifically in α and **a** cells, respectively, and *hsg* spots contain DNA of genes expressed specifically in haploid cells. For each of the hybridizations in the accompanying table, indicate the expected color of the hybridization (red, green, yellow, or none). Yellow indicates equal amounts of Cy3 and Cy5, and "none" indicates lack of expression in both strains. The symbol Δ means a deletion mutation. (Before tackling this problem, you may wish to review the genetic control of yeast mating type in Section 12.9.)

13.32 The accompanying illustration shows a gene with its transcription start site and direction of transcription, along with the multiple cloning site (MCS) in two plasmid vectors, A and B, which differ according to the position of a promoter adjacent to the MCS. The symbols *A*, *D*, *E*, *H*, *P*, *S*, and *X* represent cleavage sites for the restriction enzymes *Alu*I, *Dde*I, *Eco*RI, *Hin*dIII, *Pst*I, *Sac*I, and *Xho*II, respectively.



In directional cloning, a gene of interest and a plasmid are both cleaved with the same two restriction enzymes, which are chosen so that their sticky ends will ensure that the fragment of interest is inserted in the correct orientation in the plasmid. A geneticist plans to use directional cloning to place the coding region of the gene under the control of the promoter in each of the two plasmids.

(a) Which restriction enzymes should be used for directional cloning in plasmid A?

(b) Which restriction enzymes should be used for directional cloning in plasmid B?

(c) Draw restriction maps of the resulting plasmids in the vicinity of the MCS.

13.33 The directional cloning described in the previous problem was successful in both cases, but unfortunately, the labels on the tubes became smudged and unreadable. Do all of the plasmids have to be restriction mapped in full, or is there a simpler tell-tale digest that could be done to distinguish between them?

Further Reading

Azpirozleehan, R., and K. A. Feldmann. 1997. T-DNA insertion mutagenesis in *Arabidopsis*: Going back and forth. *Trends in Genetics* 13: 152.

Blaese, R. M. 1997. Gene therapy for cancer. *Scientific American*, June.

Botstein, D., A. Chervitz, and J. M. Cherry. 1997. Yeast as a model organism. *Science* 277: 1259.

Capecchi, M. R. 1994. Targeted gene replacement. *Scientific American*, March.

Carlson, M. 2000. The awesome power of yeast biochemical genomics. *Trends in Genetics* 16: 49.

Clark, M. S. 1999. Comparative genomics: The key to understanding the human genome project. *Bioessays* 21: 121.

Cooke, H. 1987. Cloning in yeast: An appropriate scale for mammalian genomes. *Trends in Genetics* 3: 173.

Eisenberg, D., E. M. Marcotte, I. Xenarios, and T. O. Yeates. 2000. Protein function in the post-genomic era. *Nature* 405: 823.

Felgner, P. L. 1997. Nonviral strategies for gene therapy. *Scientific American*, June.

Friedmann, T. 1997. Overcoming the obstacles to gene therapy. *Scientific American*, June.

Gasser, C. S., and R. T. Fraley. 1992. Transgenic crops. *Scientific American*, June.

Gossen, J., and J. Vigg. 1993. Transgenic mice as model systems for studying gene mutations in vivo. *Trends in Genetics* 9: 27.

Granjeaud, S., F. Bertucci, and B. R. Jordan. 1999. Expression profiling: DNA arrays in many guises. *Bioessays* 21: 781.

Havukkala, I. J. 1996. Cereal genome analysis using rice as a model. *Current Opinion in Genetics & Development* 6: 711.

Houdebine, L. M., ed. 1997. *Transgenic Animals: Generation and Use*. New York: Gordon and Breach.

Lockhart, D. J., and E. A. Winzeler. 2000. Genomics, gene expression and DNA arrays. *Nature* 405: 827.

Nettelbeck, D. M., V. Jerome, and R. Muller. 2000. Gene therapy: Designer promoters for tumour targeting. *Trends in Genetics* 16: 174.

Pandey, A., and M. Mann. 2000. Proteomics to study genes and genomes. *Nature* 405: 837,

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual*. 2d ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Vukmirovic, O. G., and S. M. Tilghman. 2000. Exploring genome space. *Nature* 405: 820.

Watson, J. D. 1999. Interviews on genomics: Interview with Dr. James D. Watson. *Bioessays* 21: 175–178.

Wilson, R. K. 1999. How the worm was won: The *C. elegans* genome sequencing project. *Trends in Genetics* 15: 51.

CHAPTER 14

Genetic Control of Development



Human fertilization and an embryo at 4 months.

A million million spermatozoa
All of them alive:
Out of their cataclysm but one poor Noah
Dare hope to survive.
And among the billion minus one
Might have chanced to be
Shakespeare, another Newton, a new Donne—
But the One was Me.

[Photos: Lennart Nilsson/Albert Bonniers Publishing Company, *A Child Is Born*, Dell Publishing Co.; verse: Aldous Huxley, from the *Fifth Philosopher's Song*.]

CHAPTER OUTLINE

14.1 **Genetic Determinants of Development**

14.2 **Early Embryonic Development in Animals**
Autonomous Development and Intercellular Signaling
Early Development and Activation of the Zygotic Genome
Composition and Organization of Oocytes

14.3 **Genetic Analysis of Development in the Nematode**
Analysis of Cell Lineages
Mutations Affecting Cell Lineages
Types of Lineage Mutations
Loss-of-Function and Gain-of-Function Alleles

14.4 **Genetic Control of Development in *Drosophila***
Maternal-Effect Genes and Zygotic Genes
Genetic Basis of Pattern Formation in Early Development
Coordinate Genes
Gap Genes
Pair-Rule Genes
Segment-Polarity Genes
Metamorphosis of the Adult Fly
Homeotic Genes
HOX Genes in Evolution

14.5 **Genetic Control of Development in Higher Plants**
Flower Development in *Arabidopsis*
Combinatorial Determination of the Floral Organs

PRINCIPLES

- In animal cells, maternal gene products in the oocyte control the earliest stages of development, including the establishment of the main body axes.
- Developmental genes are often controlled by gradients of gene products, either within cells or across parts of the embryo.
- Regulation of developmental genes is hierarchical—genes expressed early in development regulate the activities of genes expressed later.
- Regulation of developmental genes is combinatorial—each gene is controlled by a combination of other genes.
- Many of the fundamental processes of pattern formation appear to be similar in animals and plants.

CONNECTIONS

Distinguished Lineages
John E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson 1983
The Embryonic Cell Lineage of the Nematode Caenorhabditis elegans

Embryo Genesis
Christiane Nüsslein-Volhard and Eric Wieschaus 1980
Mutations Affecting Segment Number and Polarity in Drosophila

Understanding gene regulation is central to understanding how an organism as complex as a human embryo develops from a fertilized egg. In development, genes are expressed according to a prescribed program to ensure that the fertilized egg divides repeatedly and that the resulting cells become specialized in an orderly way to give rise to the fully differentiated organism. The genotype determines not only the events that take place in development, but also the temporal order in which the events unfold.

Genetic approaches to the study of development often make use of mutations that alter developmental patterns. These mutations interrupt developmental processes and make it possible to identify factors that control development and to study the interactions among them. This chapter demonstrates how genetics is used in the study of development. Many of the examples come from the nematode *Caenorhabditis elegans* and the fruit fly *Drosophila melanogaster,* because these organisms have been studied intensively from the standpoint of developmental genetics.

The key process in development is **pattern formation,** which means the emergence, from cell division and differentiation of the fertilized egg, of spatially organized and specialized cells that form the embryo. One might draw an analogy between pattern formation in development and a pattern formed by fitting the pieces of a jigsaw puzzle together, but the analogy is not particularly apt. In a jigsaw puzzle, the pattern exists independently of the shape and position of the pieces; the cuts in the pattern are superimposed on a preexisting picture that is merely reassembled. In biological development, the emergence of the image on each piece is *caused by* the shape and position of the piece. It is remarkable also that the picture changes throughout life as the organism undergoes growth and aging.

14.1 Genetic Determinants of Development

Conceptually, the relationship between genotype and development is straightforward:

> The genotype contains a developmental program that unfolds and results in the expression of different sets of genes in different types of cells.

In other words, development consists of a program that results in the specific expression of some genes in one cell type and not in another. As development unfolds, cell types progressively appear that differ qualitatively in the genes that are expressed. Whether a particular gene is expressed may depend on the presence or absence of a particular transcription factor, a change in chromatin structure, or the synthesis of a particular receptor molecule. These and other molecular mechanisms define the pattern of interactions by which one gene controls another. Collectively, these regulatory interactions ultimately determine the fate of the cell. Moreover, many interactions that are important in development make use of more general regulatory elements. For example, developmental events often depend on regional differences in the concentration of molecules within a cell or within an embryo, and the activity of enhancers may be modulated by the local concentration of these substances.

The formation of an embryo is also affected by the environment in which development takes place. The genotype determines the developmental potential of the organism. Genes determine whether a developing embryo will become a nematode, a fruit fly, a chicken, or a mouse. However, the expression of the genetically determined developmental potential is also influenced by the environment—in some cases, very dramatically. Fetal alcohol syndrome is one example in which an environmental agent, chronic alcohol poisoning, affects various aspects of fetal growth and neurological development.

Development includes many examples in which genes are selectively turned on or off by the action of regulatory proteins that respond to environmental signals. The identification of genetic regulatory interactions that operate during development is an important theme in developmental genetics. The implementation of different regulatory interactions means that genetically identical cells can become qualitatively different in the genes that are expressed.

Several common mechanisms by which genes become activated or repressed at particular stages in development or in particular tissues are illustrated in the generalized example in **Figure 14.1**. In this example, the developmentally regulated genes are denoted *r* (red), *g* (green), *b* (blue), and *p* (purple). The initial event is transcription of the *r* gene in the oocyte (Figure 14.1A). Unequal partitioning of the *r* gene product into one region of the egg establishes a polarity, or regionalization, of the egg with respect to presence of the *r* gene product (Figure 14.1B). When cleavage takes place (Figure 14.1C), the polarized cell produces daughter cells either with (L, left) or without (R, right) the *r* gene product. The *r* gene product is a transcriptional activator of the *g* gene, so the *g* gene product is expressed in cell L but not in cell R. Hence, even at the two-cell stage, the daughter cells may differ in gene expression as a result of the initial polarization of the egg. Also in the two-cell stage, the *p* gene is expressed in both daughter cells. This gene codes for a *transmembrane receptor* containing amino acid sequences that span the cell membrane.

The presence of receptor molecules in the cell membrane provides a key mechanism for intercellular signaling, as illustrated in this example with the regulation of the *b* gene in Figure 14.1D. In the absence of the transmembrane receptor, the *b* gene would be expressed in all four cells.

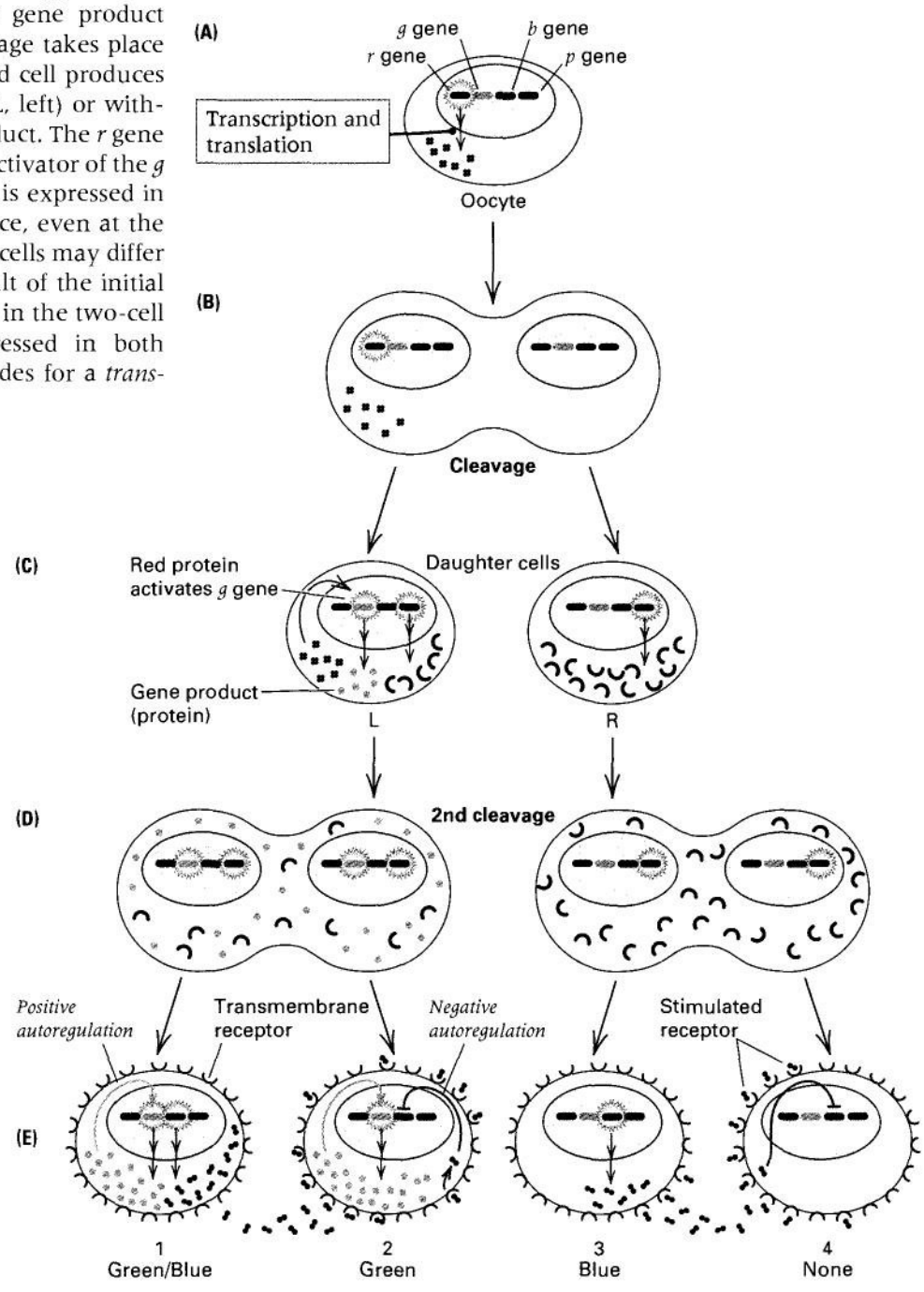


Figure 14.1 Generalized example illustrating some regulatory mechanisms that result in differences in gene expression in development. (A) The *r* gene is expressed in the oocyte. (B) Polarized presence of the *r* gene product in the egg. (C) Presence of the *r* gene product in cell L stimulates transcription of the *g* gene. The *p* gene codes for a transmembrane receptor expressed in both L and R cells. (D) The *g* gene has positive autoregulation and so continues to be expressed in cells 1 and 2. (E) Expression of the *b* gene in one cell represses its expression in the neighboring cell by stimulation of the *p* gene transmembrane receptor. The result of these mechanisms is that cells 1 through 4 differ in their combinations of *b* gene and *g* gene activity.

However, expression of the *b* gene in one cell inhibits its expression in the neighboring cell because of the transmembrane receptor. The mechanism is that the product of the *b* gene stimulates the receptor of neighboring cells, and stimulation of the receptor represses transcription of *b*. The process in which gene expression in one cell inhibits expression of the same or differing genes in neighboring cells is known as *lateral inhibition*.

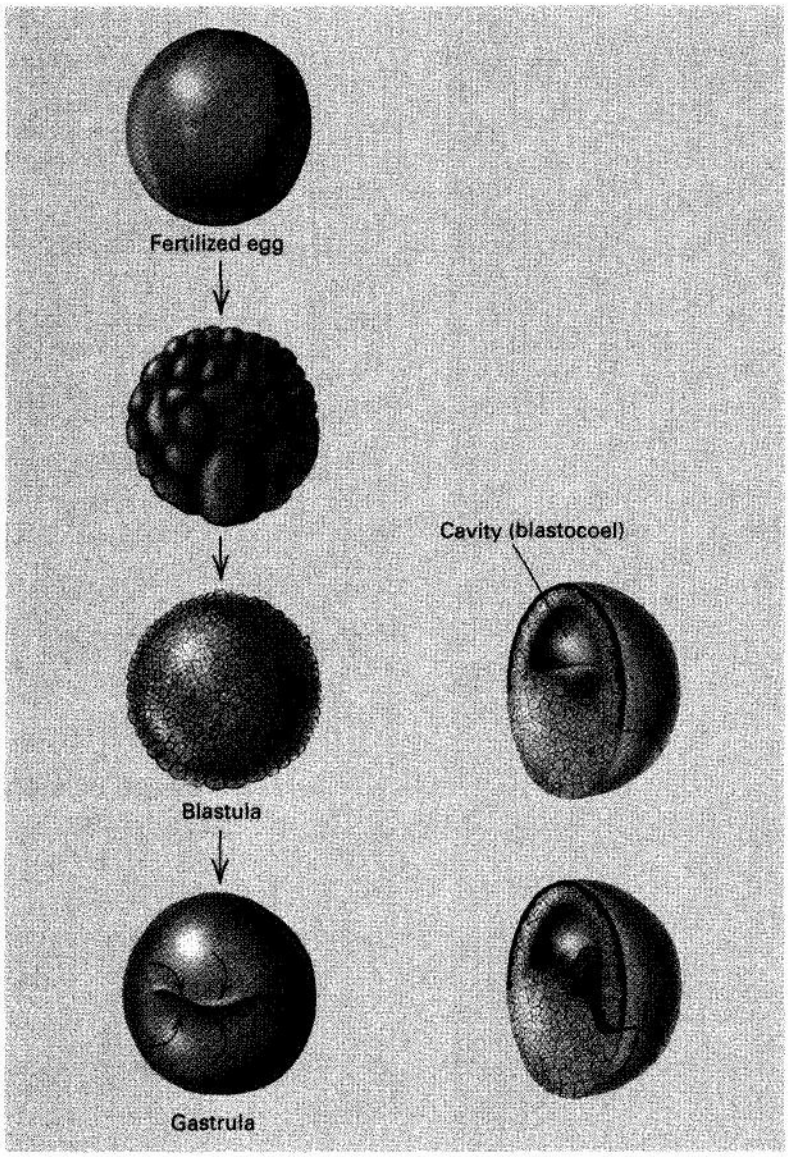


Figure 14.2 Early development of the animal embryo. The cleavage divisions of the fertilized egg result in first a clump of cells and then a hollow ball of cells (the blastula). Extensive cell migrations form the gastrula and establish the basic body plan of the embryo.

At the same time that *b* gene expression is regulated by lateral inhibition, the initial activation of the *g* gene in cell L is retained in daughter cells 1 and 2, because the *g* gene product has a positive autoregulatory activity and stimulates its own transcription. Positive autoregulation is one way in which cells can amplify weak or transient regulatory signals into permanent changes in gene expression. The result of lateral inhibition of *b* expression and *g* autoregulation is that cells 1 through 4 in Figure 14.1E, though genetically identical, are developmentally different because they express different combinations of the *g* and *b* genes. Specifically, cell 1 has *g* and *b* activity, cell 2 has *g* activity only, cell 3 has *b* activity only, and cell 4 has neither.

Among the most intensively studied developmental genes are those that act early in development, before tissue differentiation, because these genes establish the overall pattern of development.

14.2 Early Embryonic Development in Animals

The early development of the animal embryo establishes the basic developmental plan for the whole organism. Fertilization initiates a series of mitotic **cleavage divisions** in which the embryo becomes multicellular. There is little or no increase in overall size compared with the egg, because the cleavage divisions are accompanied by little growth; they merely partition the fertilized egg into progressively smaller cells. The cleavage divisions form the **blastula,** which is essentially a ball of about 10^4 cells containing a cavity (**Figure 14.2**). Completion of the cleavage divisions is followed by the formation of the **gastrula** through an infolding of part of the blastula wall and extensive cellular migration. In the reorganization of cells in the gastrula, the cells become arranged in several distinct layers. These layers establish the basic body plan of an animal. In higher plants, as we shall see later, the developmental processes differ substantially from those in animals.

A fertilized animal egg has full developmental potential because it contains the

genetic program in its nucleus and macromolecules present in the oocyte cytoplasm that are necessary for giving rise to an entire organism. Full developmental potential is maintained in cells produced by the early cleavage divisions. However, the developmental potential is progressively channeled and limited in early embryonic development. Cells within the blastula usually become committed to particular developmental outcomes, or **cell fates,** that limit the differentiated states possible among descendant cells.

Autonomous Development and Intercellular Signaling

Two principal mechanisms progressively restrict the developmental potential of cells within a lineage:

- Developmental restriction may be **autonomous,** which means that it is determined by genetically programmed changes in the cells themselves.
- Cells may respond to **positional information,** which means that developmental restrictions are imposed by the position of cells within the embryo. Positional information may be mediated by signaling interactions between neighboring cells or by gradients in the concentration of morphogens. A **morphogen** is a molecule that participates directly in the control of growth and form during development.

Restriction of developmental fate is usually studied by transplanting cells of the embryo to new locations to determine whether they can substitute for the cells that they displace. An example showing the distinction between autonomous development and response to positional information is illustrated in **Figure 14.3**. In the normal embryo (Figure 14.3A), cells 1 and 2 have different fates either because of autonomous developmental programs or because they respond to positional information near the anterior and posterior ends. When the cells are transplanted (Figure 14.3B), autonomous development is indicated if the developmental fate of cells 1 and 2 is unchanged in spite of their new locations; in this case, the embryo has anterior and posterior reversed. When development depends on positional

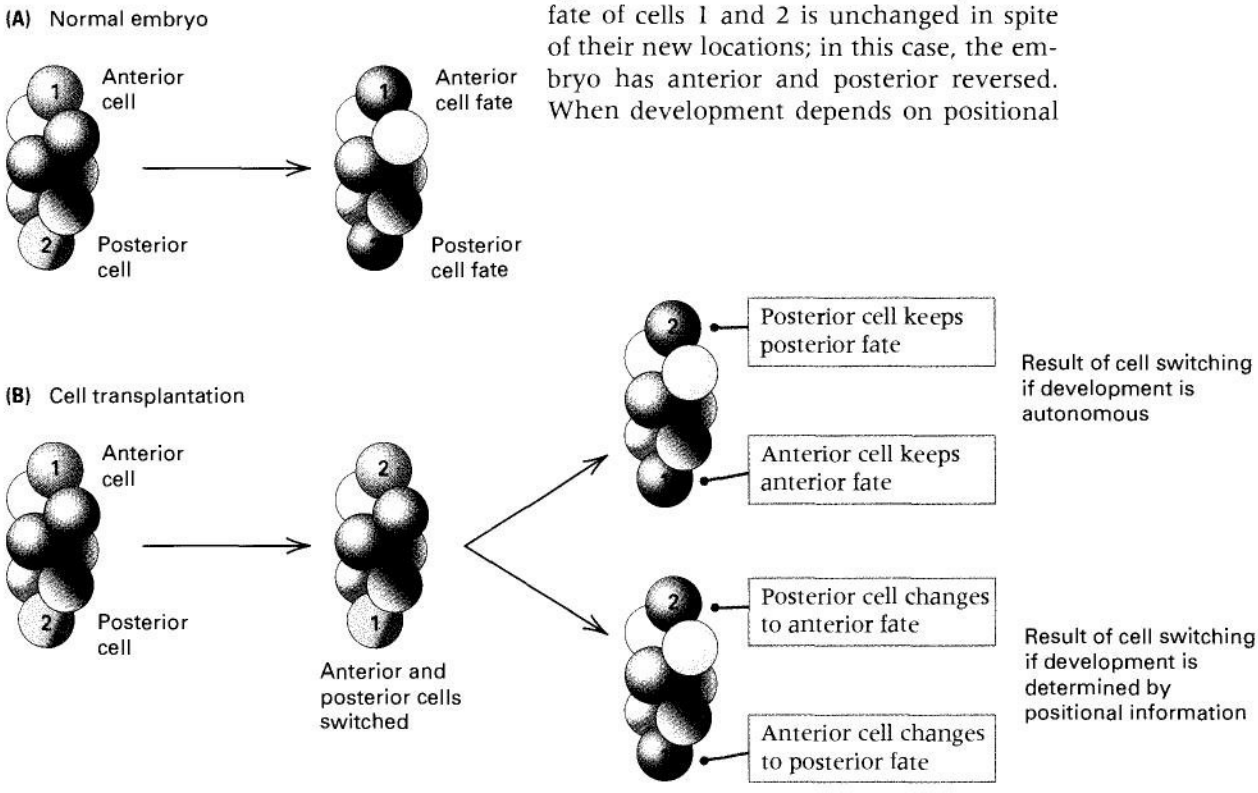


Figure 14.3 Distinction between autonomous determination and positional information. (A) Cells 1 and 2 differentiate normally as shown. (B) Transplantation of the cells to reciprocal locations. If the transplanted cells differentiate autonomously, then they differentiate as they would in the normal embryo, but in their new locations. If they differentiate in response to positional information (signals from neighboring cells), then their new positions determine their fate.

information, the transplanted cells respond to their new locations, and the embryo is normal.

In some eukaryotes, such as the soil nematode *Caenorhabditis elegans*, many lineages develop autonomously according to genetic programs that are induced by interactions with neighboring cells very early in embryonic development. Subsequent cell interactions also are important, and each stage in development is set in motion by the successful completion of the preceding stage. **Figure 14.4** illustrates the first three cell divisions in the development of *C. elegans*, which result in eight embryonic cells that differ in their genetic activity and developmental fate. The determination of cell fate in these early divisions is in part autonomous and in part a result of interactions between cells. Figure 14.4B shows the lineage relationships between the cells. Cell-autonomous mechanisms are illustrated by the transmission of cytoplasmic particles called *polar granules* from the cells P0 to P1 to P2 to P3. Normal segregation of the polar granules is a function of microfilaments in the cytoskeleton. Cell-signaling mechanisms are illustrated by the effects of P2 on EMS and on ABp. The EMS fate is determined by the activity of the *mom-2* gene in P2. The P2 cell also produces a signaling molecule, APX-1, which determines the fate of ABp through the cell-surface receptor GLP-1. (Note that the specific mechanisms that determine early cell fate in *C. elegans* strongly resemble some of the general mechanisms outlined in Figure 14.1.) In contrast to *C. elegans*, in which many developmental decisions are cell-autonomous, in *Drosophila* and *Mus* (the mouse), regulation by intercellular signaling is more the rule than the exception. The use of intercellular signaling to regulate development provides a sort of insurance that helps to overcome the death of individual cells that might happen by chance during development.

One special case of intercellular signaling is **embryonic induction,** in which the development of a major embryonic structure is determined by a signal sent from neighboring cells. An example is found in early development of the sea urchin, *Strongylocentrotus purpuratus*, in which, after

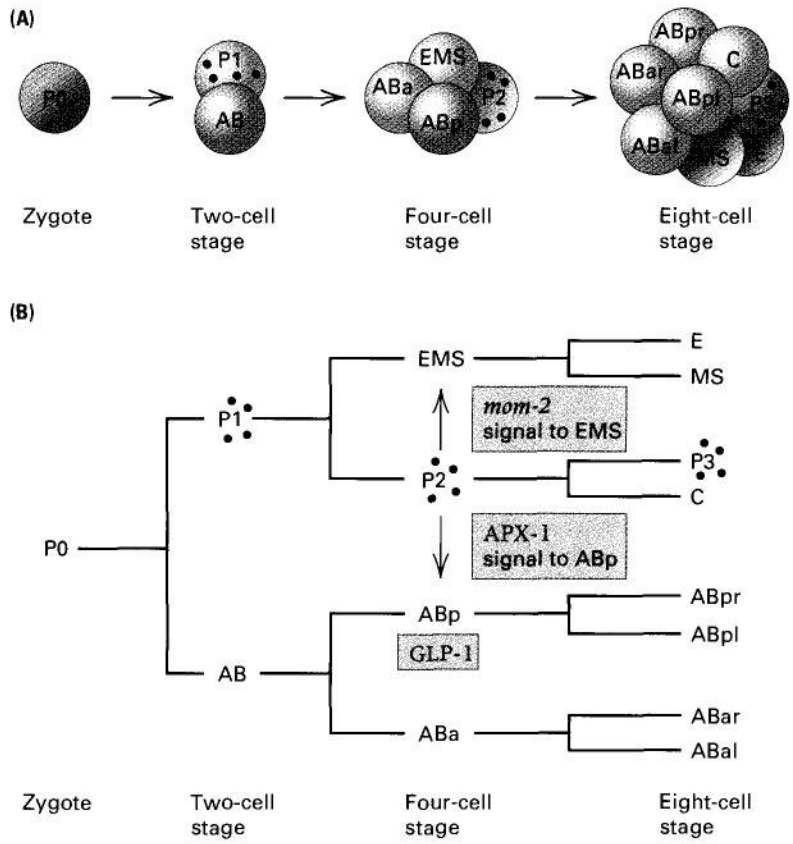


Figure 14.4 Early cell divisions in *C. elegans* development. (A) Spatial organization of cells. (B) Lineage relationships of the cells. The transmission of the polar granules illustrates cell-autonomous development. The arrows denote cell-to-cell signaling mechanisms that determine developmental fate. (After Wade Roush. 1996. *Science* 272: 1871.)

the fourth cleavage division, specialized cells called *micromeres* are produced at the ventral pole of the embryo (**Figure 14**, parts A and B). After additional cell proliferation, the region of the embryo immediately above the micromeres folds inward to form a tube, the *archenteron*, that eventually develops into the stomach, intestine, and related structures. If micromere cells from a 16-cell embryo are removed and transplanted into the dorsal region of another embryo at the 8-cell stage (Figure 14.5C), then the next cleavage results in an embryo with two sets of micromeres (Figure 14.5D). In these embryos, as development proceeds, an archenteron forms in the normal position from cells lying above the ventral micromeres. But a second archenteron also forms beneath the transplanted micromeres (Figure 14.5E), which indicates that the transplanted micromeres are capable of inducing archenteron development in adjacent cells.

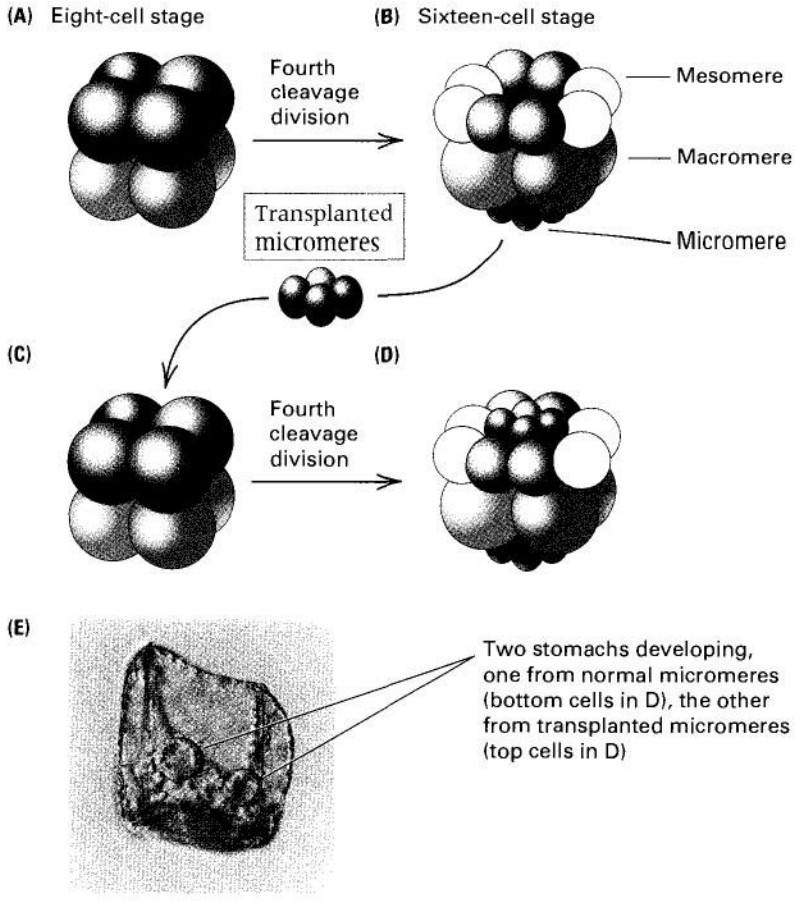


Figure 14.5 Embryonic induction of the archenteron by micromere cells in the sea urchin. (A) Normal embryo at the 8-cell stage. (B) Normal embryo at the 16-cell stage, showing micromeres on the ventral side. (C) Transplanted 8-cell embryo with micromeres from another embryo placed at the dorsal end. (D) Transplanted embryo after the fourth cleavage division with two sets of micromeres. (E) The result of the transplant is an embryo with two archenterons. The doughnut-shaped objects in this embryo are the developing stomachs. [Photomicrograph, courtesy of Eric H. Davidson and Andrew Ransick. Reprinted with permission from E. H. Davidson and A. Ransick, *Science* 259: 1134. Copyright 1993 American Association for the Advancement of Science.]

Early Development and Activation of the Zygotic Genome

Fertilized frog eggs with the nucleus removed are still able to carry out the cleavage divisions and form rudimentary blastulas. Similarly, when gene transcription in sea urchin or amphibian zygotes is blocked, there is no effect on the cleavage divisions or on blastula formation, but gastrulation does not take place. These observations imply that early development through blastula formation does not depend on genetic information in the zygotic cell nucleus. All the necessary macromolecules are in the oocyte cytoplasm.

Following this early period when zygotic genes are not needed, the embryo becomes dependent on the activity of its own genes. In mammals, the zygotic genes are needed much earlier than in lower vertebrates. Inhibitors of transcription stop development of the mouse embryo when they are applied at any time after the first cleavage division. This result implies that the shift from control by the maternal genome to control by the zygotic genome begins in the two-cell stage. This is also the stage when proteins coded by the zygotic nucleus are first detected. However, even in mammals, the earliest stages of development are greatly influenced by the cytoplasm of the oocyte, which determines the planes of the initial cleavage divisions and other events that ultimately affect cell fate.

Early activation of the zygotic nucleus in mammals may be necessary because, in gamete formation, certain genes undergo the process of epigenetic inactivation known as *imprinting* (discussed in Chapter 12). Not many genes are subject to imprinting. Among these are the gene for an insulin-like growth factor (*Igf2*), which is imprinted during oogenesis, and the gene for the *Igf2* transmembrane receptor (*Igf2-r*), which is imprinted during spermatogenesis. Therefore, expression of both *Igf2* and *Igf2-r* in the

embryo requires activation of the sperm nucleus for *Igf2* and of the egg nucleus for *Igf2-r*. Imprinting also affects genes in the human genome and has been implicated in conditions such as the fragile-X syndrome (Chapter 7) and the Prader-Willi and Angelman syndromes (Chapter 12).

Composition and Organization of Oocytes

The oocyte is a diploid cell during most of oogenesis. The cytoplasm of the oocyte is extensively organized and regionally differentiated (**Figure 14.6**). This spatial differentiation ultimately determines the different developmental fates of cells in the blastula. The cytoplasm of the animal egg has two essential functions:

1. Storage of the molecules needed to support the cleavage divisions and the rapid RNA and protein synthesis that take place in early embryogenesis.
2. Organization of the molecules in the cytoplasm to provide the positional information that results in differences between cells in the early embryo.

To establish the proper composition and organization of the oocyte requires the participation of many genes within the oocyte itself and gene products supplied by adjacent helper cells of various types (**Figure 14.7**). Numerous maternal genes are transcribed in oocyte development, and the mature oocyte typically contains an abundance of transcripts that code for proteins needed in the early embryo. Some of the maternal mRNA transcripts are masked in special ribonucleoprotein particles and cannot be translated, and the release of the mRNA, enabling it to be translated, does not take place until after fertilization.

Developmental instructions in oocytes are determined in part by the presence of distinct types of molecules at different positions within the cell and in part by gradients of morphogens that differ in concentration from one position to the next. Although the oocyte contains the products of many genes, only a small number of genes are expressed exclusively in the formation of the oocyte. Most genes expressed in oocyte formation are also important in the development of other tissues or at later times in development. Therefore, it is not only gene products but also their spatial organization within the cell that give the oocyte its unique developmental potential.

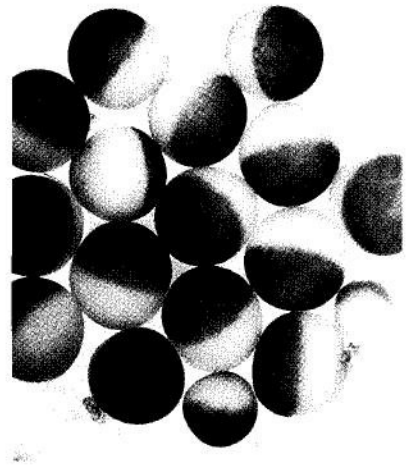
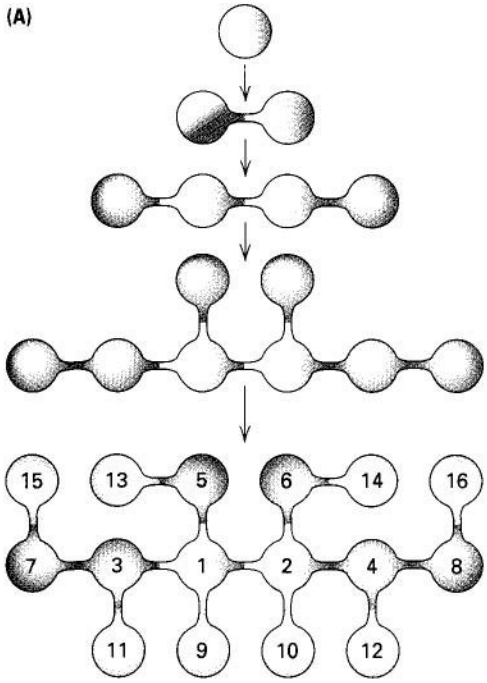


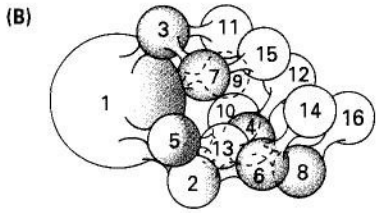
Figure 14.6 The animal oocyte is highly organized, which is revealed by the visible differences between the dorsal (dark) and ventral (light) parts of these *Xenopus* eggs. The regional organization of the oocyte determines many of the critical events in early development. [Courtesy of Michael W. Klymkowsky.]

14.3 Genetic Analysis of Development in the Nematode

The soil nematode *Caenorhabditis elegans* (**Figure 14.8**) is popular for genetic studies because it is small, is easy to culture, and has a short generation time with a large number of offspring. The worms are grown on agar surfaces in petri dishes and feed on bacterial cells such as *E. coli*. Because they are microscopic in size, as many as 10^5 animals can be contained in a single petri dish. Sexually mature adults of *C. elegans* are capable of laying more than 300 eggs within a few days. At 20°C, it requires about 60 hours for the eggs to hatch, undergo four larval molts, and become sexually mature adults.

Nematodes are diploid organisms with two sexes. In *C. elegans*, the two sexes are the hermaphrodite and the male. The hermaphrodite contains two X chromosomes (XX), produces both functional eggs and





Spatial organization of the oocyte and nurse cells. In this case, cell 1 becomes the oocyte and all other cells are nurse cells.

Figure 14.7 Pattern of cell divisions in the development of the *Drosophila* oocyte. (A) The cells surrounding the oocyte are connected to it and to each other by cytoplasmic bridges. These cells synthesize products transported into the oocyte and contribute to its regional organization. (B) Geometrical organization of the cells. [From R. C. King. 1965. *Genetics,* Oxford University Press.]

functional sperm, and is capable of self-fertilization. The male produces only sperm and fertilizes the hermaphrodites. The sex-chromosome constitution of *C. elegans* consists of a single X chromosome; there is no Y chromosome, and the male karyotype is XO.

Analysis of Cell Lineages

The transparent body wall of the worm has made possible the study of the division, migration, and differentiation or death of all cells present in the course of development. Nematode development is unusual in that the pattern of cell division and differentiation is highly stereotyped, with almost no variation from one individual to the next. The hermaphrodite contains exactly 959 somatic cells, the male exactly 1031. The complete developmental history of each somatic cell is known at the cellular and ultrastructural levels. The pedigree relationships among a group of cells is known as a cell **lineage.** A cell lineage can be illustrated with a **lineage diagram** that shows the ancestor–descendant relationships and the terminal differentiated state of each

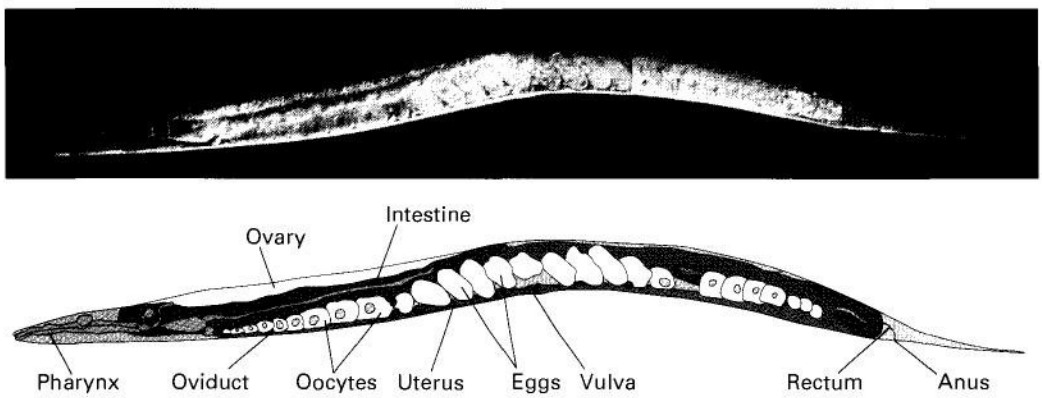


Figure 14.8 The soil nematode *Caenorhabditis elegans.* [Photograph courtesy of Tim Schedl.]

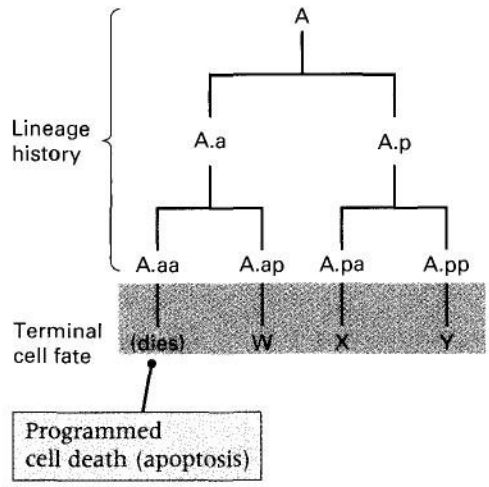


Figure 14.9 Generalized cell-lineage diagram. Different terminally differentiated cell fates are denoted W, X, and Y. One cell in the lineage (A.aa) undergoes programmed cell death. The lowercase letters "*a*" and "*p*" denote anterior and posterior daughter cells. For example, cell A.ap is the posterior daughter of the anterior daughter of cell A.

cell. **Figure 14.9** shows a generalized lineage diagram of a cell A whose descendants undergo either programmed cell death or terminal differentiation into cell type W, X, or Y. The letter symbols are the kind normally used for cells in nematodes, in which the name denotes the cell lineage according to ancestry and position in the embryo. For example, the cells A.a and A.p are the anterior and posterior daughters of cell A, respectively; and A.aa and A.ap are the anterior and posterior daughters of cell A.a.

Mutations Affecting Cell Lineages

Many mutations that affect cell lineages have been studied in nematodes. They reveal several general features by which genes control development:

- The division pattern and fate of a cell are generally affected by more than one gene and can be disrupted by mutations in any of them.
- Most genes that affect development are active in more than one type of cell.
- Complex cell lineages often include simpler, genetically determined lineages within them; these components are called *sublineages* because they are expressed as an integrated pattern of cell division and terminal differentiation.
- The lineage of a cell may be triggered autonomously within the cell itself or by signaling interactions with other cells.
- Regulation of development is controlled by genes that determine the different sublineages that cells can undergo and the individual steps within each sublineage.

The next section deals with some specific types of mutations that affect cell lineages and development.

Types of Lineage Mutations

Mutations can affect cell lineages in two major ways. One is through nonspecific metabolic blocks (for example, in DNA replication) that prevent the cells from undergoing division or differentiation. The other is through specific molecular defects that result in patterns of division or differentiation that are normally found elsewhere in the embryo or at a different time in development. From the standpoint of genetic analysis of development, the latter class of mutants is the more informative, because the mutant genes must be involved in developmental processes rather than in general "housekeeping" functions found in all cells.

Mutations that cause cells to undergo developmental fates characteristic of other types of cells are called **transformation mutations.** For example, in the wildtype cell lineage shown in **Figure 14.10**A, the cell designated A.ap differentiates into cell-type W. In the transformation mutant (Figure 14.10B), cell A.ap differentiates into cell-type Y. Transformation mutations have the consequence that one or more differentiated cell types are missing from the embryo and are replaced with extra, supernumerary copies of other cell types. In this example, the missing cell type is W, and the extra cell type is Y. In the nematode, the mutation *unc-86* (*unc* = *uncoordinated*) provides an example in which the supernumerary cells become neurons. Supernumerary cells can impair development by interfering with the normal developmental process in nearby cells. In Section 14.4, we shall see that transformation mutations in nematodes are analogous to a class of mutations in *Drosophila* called *homeotic* mutations, in

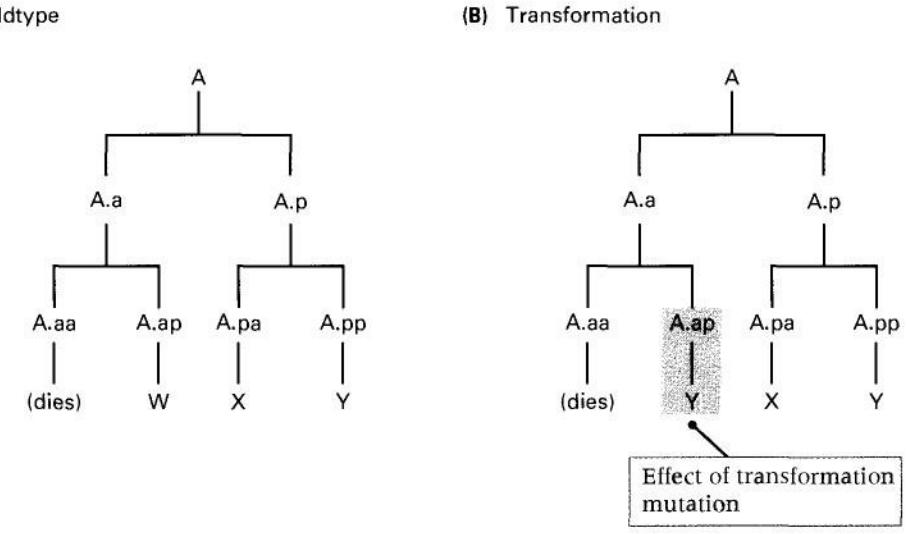


Figure 14.10 Transformation mutations cause cells to undergo abnormal terminal differentiation. (A) The wildtype lineage. (B) A mutant lineage in which the cell A.ap differentiates into a Y-type cell rather than the normal W-type cell.

which certain cells undergo developmental fates that are normally characteristic of other cells. Indeed, the wildtype *unc-86* gene codes for a transcription factor containing a DNA-binding domain similar to that found in proteins encoded by the *Drosophila* homeotic genes.

Developmental programs often require sister cells or parent and offspring cells to adopt different fates. Mutations that cause sister cells or parent and offspring cells to fail to become differentiated from each other are called **segregation mutations** because the factors that govern cell fate fail to segregate (become separated) in the daughter cells. The *lin-17* (*lin* = lineage) mutation in *C. elegans* provides an example of a segregation mutation that affects sister cells (**Figure 14.11**). In males, a gonadal precursor cell, Z1, produces daughter cells Z1.a and Z1.p that have different lineages and fates (Figure 14.11A). In *lin-17* mutants, the developmental determinants do not segregate, and the sister cells Z1.a and Z1.p both have the same lineage and fate as the wildtype Z1.p (Figure 14.11B).

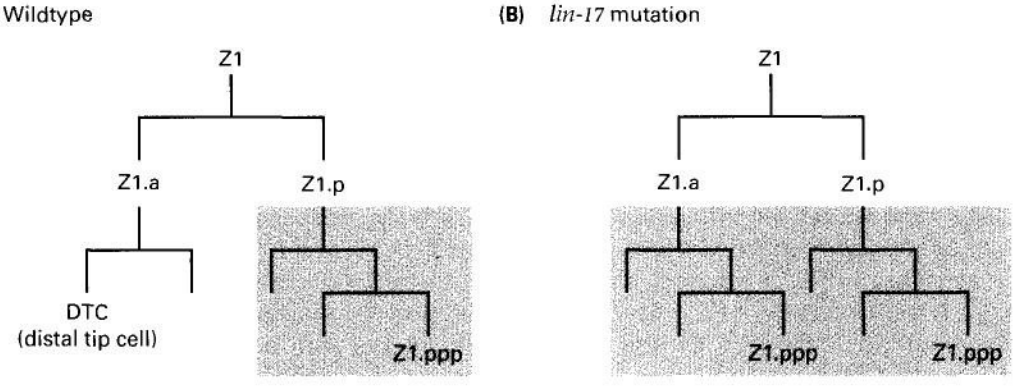


Figure 14.11 Aberrant sister-cell segregation caused by the *lin-17* mutation in *C. elegans*. (A) Z1.a and Z1.p normally undergo different fates, and the Z1.a lineage includes a distal tip cell (DTC). (B) In the *lin-17* segregation mutant, Z1.a and Z1.p have the same fate, and the distal tip cell is absent.

Connection

Distinguished Lineages

John E. Sulston,[1] E. Schierenberg,[2] J. G. White,[1] and J. N. Thomson[1] 1983
[1]Medical Research Council Laboratory for Molecular Biology, Cambridge, England
[2]Max Planck Institute for Experimental Medicine, Gottingen, Germany
The Embryonic Cell Lineage of the Nematode Caenorhabditis elegans

The data produced in this landmark study form the basis for interpreting developmental mutants in the nematode worm. This long paper offers voluminous data, and it is presently available through the Internet. During embryogenesis, 671 cells are generated; 113 of these, or 17%, undergo programmed cell death. What is the reason for such a high proportion of programmed cell deaths? The embryonic lineage is highly invariant—the same from one organism to the next. Why is there not more developmental flexibility, as is found in most other organisms? These issues are addressed in this excerpt, in which the emphasis is on the historical background and motivation of the study, the big picture of development, and interpretation of the lineage in terms of the evolution of the nematode. The technique of Nomarski microscopy is a modern invention also called differential interference contrast microscopy. When light passes through living material, it changes phase according to the refractive index of the material. Adjacent parts of a cell or organism that differ in refractive index cause different changes in phase. When two sets of waves combine after passing through an object, the difference in phase creates an interference pattern that yields an image of the object. The major advantage of Nomarski microscopy is that it can be used to observe living tissue.

This report marks the completion of a project begun over one hundred years ago—namely the determination of the entire cell lineage of a nematode. Nematode embryos were attractive to nineteenth-century biologists because of their simplicity and the reproducibility of their development, and considerable progress was made in determining their lineages by the use of fixed specimens. By the technique of Nomarski microscopy, which is nondestructive and yet produces high resolution, cells can be followed in living larvae. The use of living material lends a previously unattainable continuity and certainty to the observations, and has permitted the origin and fate of every cell in one nematode species [*Caenorhabditis elegans*] to be determined. Thus, not only are the broad relationships between tissues now known unambiguously but also the detailed pattern of cell fates is clearly revealed. . . . The lineage is of significance both for what it can tell us immediately about relationships between cells and also as a framework into which future observations can be fitted. . . . The embryonic cell lineage is essentially invariant. The patterns of division, programmed cell death, and terminal differentiation are constant

When mutant cells are unable to execute their normal developmental fates, the mutation is an **execution mutation.** The *lin-11* gene in *C. elegans* provides an example of an execution mutation (**Figure 14.12**). In the wildtype development of the vulva in the hermaphrodite, a lineage designated the 2° lineage gives rise to four cells in the spatial pattern N-T-L-L (Figure 14.12A), in which N indicates a cell that does not divide, T a cell that divides in a transverse plane relative to the orientation of the larva, and L a cell that divides in a longitudinal plane. In *lin-11* mutants, the 2° lineage is not executed, and the four cells divide in the abnormal pattern, L-L-L-L (Figure 14.12B).

The process of **programmed cell death,** technically known as **apoptosis,** is an important feature of normal development in many organisms. Apoptosis is a process in which, at the appropriate time in development, a cell commits suicide. In many cases, the signaling molecules that determine this fate have been identified (a number are known to be transcription factors). Failure of programmed cell death often results in specific developmental abnormalities. When apoptosis fails and the surviving cells differentiate into recognizable cell types, the result is the presence of supernumerary cells of that type. For example, with mutations in the *ced-3* gene (*ced = cell death abnormal*) in *C. elegans*, a particular cell that normally undergoes programmed cell death survives and often differentiates into a supernumerary neuron. There are exactly 113 programmed cell deaths in the development of the *C. elegans* hermaphrodite. None of these deaths is essential. Mutants that cannot execute programmed cell death are viable and fertile

from one individual to another, and no great differences are seen in timing. . . . We shall use the term "sublineage" as an abbreviation for the more descriptive, but cumbersome, phrase "intrinsically determined sublineage"—namely, a fragment of the lineage which is thought to be generated by a program within its precursor cell. . . . Two of the available criteria for postulating the existence of a sublineage are: (1) the generation of the same lineage, giving rise to the same cell fates, from a series of precursors of diverse origin and position; (2) evidence for cell autonomy within the lineage, obtained from laser ablation experiments or the study of mutants. . . . The large number of programmed cell deaths, and their reproducibility, are evident from the lineage. The most likely reason for the occurrence of most of them is that, because of the existence of sublineages, unneeded cells are frequently generated along with needed ones. . . . Perhaps the most striking findings are firstly the complexity and secondly the cell autonomy of the lineages. . . . With hindsight, we can rationalize both this complexity and this rigidity. The nematode belongs to an ancient phylum, and its cell lineage is a piece of frozen evolution. In the course of time, new cell types were generated from precursors selected not so much for their intrinsic properties as for the accident of their position in the embryo. . . . Cell-cell interactions that were initially necessary for developmental decisions may have been gradually supplanted by autonomous programs that were fast, economical, and reliable, the loss of flexibility being outweighed by the gain in efficiency. On this view, the perverse assignments, the cell deaths, the long-range migration—all the features which could, it seems, be eliminated from a more efficient design—are so many developmental fossils. These are the places to look for clues both to the course of evolution and to the mechanisms by which the lineage is controlled today.

The nematode belongs to an ancient phylum, and its cell lineage is a piece of frozen evolution. In the course of time, new cell types were generated from precursors selected not so much for their intrinsic properties as for the accident of their position in the embryo

Source: Developmental Biology 100: 64–119

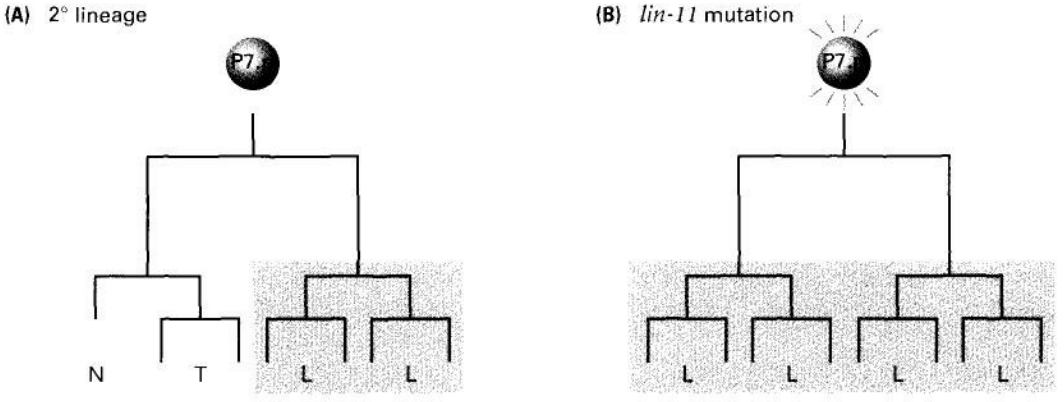


Figure 14.12 Example of an execution mutation. (A) Wildtype development of the 2° lineage in the vulva of *C. elegans*. The letter N denotes no further cell division, and T and L denote cell division in either a transverse (T) or a longitudinal (L) plane with respect to the embryo. (B) The *lin-11* mutation results in failure to execute the N and T sublineages. All cells divide in a longitudinal plane.

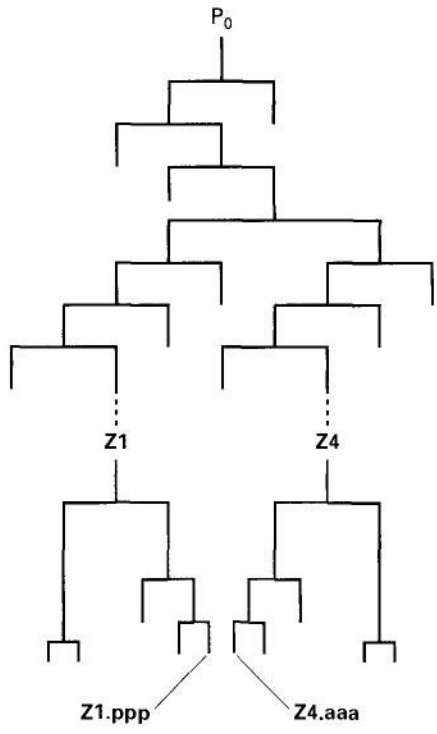


Figure 14.13 Complete lineage of Z1.ppp and Z4.aaa in *C. elegans*. P_0 represents the zygote, and the dashed lines indicate three cell divisions not shown. In the normal development of the vulva, Z1.ppp and Z4.aaa are equally likely to differentiate into the anchor cell. Whichever cell remains differentiates into a ventral uterine precursor cell.

but are slightly impaired in development and in some sensory capabilities. On the other hand, mutants in *Drosophila* that fail to execute apoptosis are lethal, and in mammals, including human beings, failure of programmed cell death results in severe developmental abnormalities or, in some instances, leukemia or other forms of cancer (Chapter 15).

Loss-of-Function and Gain-of-Function Alleles

Genes that cause cells to differ in developmental fate are not always easy to recognize. For example, an execution mutation may identify a gene that is *necessary* for the expression of a particular developmental fate, but the gene may not be *sufficient* to determine that developmental fate in the cells in which it is expressed. This possibility complicates the search for genes that control major developmental decisions.

Genes that control decisions about cell fate can sometimes be identified by the unusual characteristic that dominant and recessive mutations have opposite effects. The reasoning is that, if different mutations in the same gene can alter cell fate in opposite directions, then the product of the gene must be both necessary and sufficient for

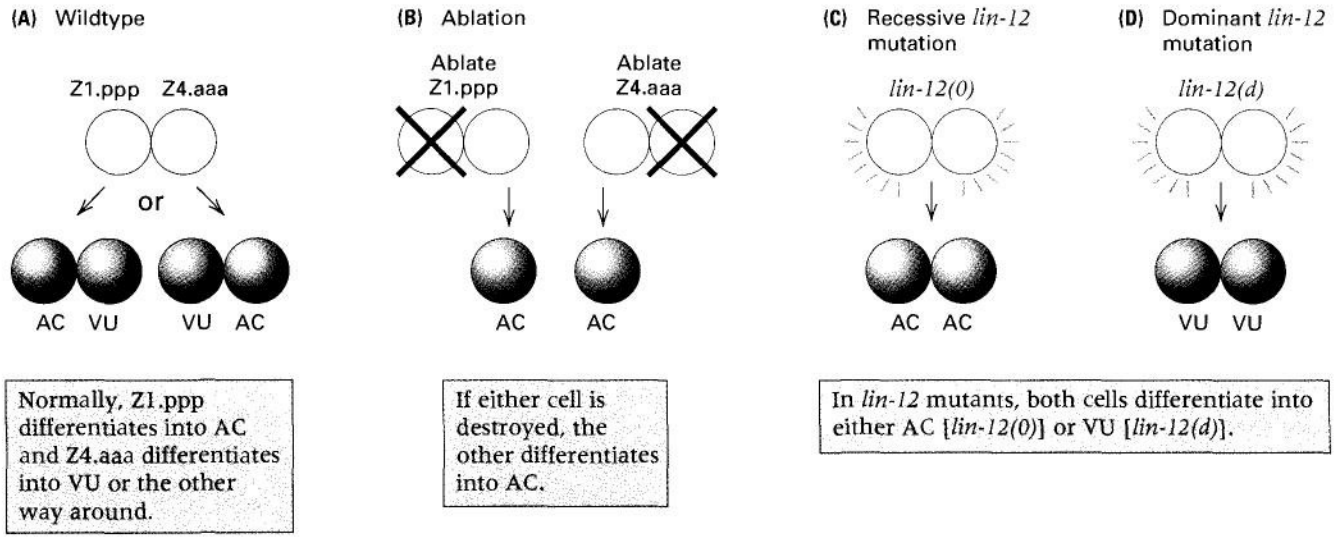


Figure 14.14 Control of the fates of Z1.ppp and Z4.aaa in vulval development. (A) In wildtype cells, each cell has an equal chance of becoming the anchor cell (AC); the other becomes a ventral uterine precursor cell (VU). (B) If either cell is destroyed (ablated) by a laser beam, then the other differentiates into the anchor cell. (C) Genetic control of cell fate by the *lin-12* gene. With recessive loss-of-function mutations [*lin-12(0)*], both cells become anchor cells. (D) With dominant gain-of-function mutations [*lin-12(d)*], both cells become ventral uterine precursor cells.

expression of the fate. Recessive alleles in genes controlling development are often *loss-of-function mutations*, such as deletions, nonsense mutations, or inactivating amino acid replacements. Dominant alleles in developmental-control genes are often *gain-of-function mutations* in that the gene is overexpressed or is expressed ectopically in the wrong cells or at the wrong time.

In *C. elegans*, only a small number of genes have dominant and recessive alleles that affect cell fate in opposite ways. Among them is the *lin-12* gene, which controls developmental decisions in a number of cells. One example concerns the cells denoted Z1.ppp and Z4.aaa in **Figure 14.13**. These cells lie side by side in the embryo, but they derive from cell P_0 (the zygote) via completely different lineages. Normally, as shown in **Figure 14.14**A, one of the cells differentiates into an *anchor cell* (AC), which participates in development of the vulva, and the other differentiates into a *ventral uterine precursor cell* (VU). Z1.ppp and Z4.aaa are equally likely to become the anchor cell.

Direct cell–cell interaction between Z1.ppp and Z4.aaa controls the AC-VU decision. If either cell is ablated (burned away) by a laser microbeam, then the remaining cell differentiates into an anchor cell (Figure 14.14B). This result implies that the preprogrammed fate of both Z1.ppp and Z4.aaa is that of an anchor cell. In the normal situation, commitment of either Z1.ppp or Z4.aaa to the anchor-cell fate also commits the other cell to the ventral uterine precursor fate. As noted, recessive and dominant mutations of *lin-12* have opposite effects. Recessive mutations in which *lin-12* activity is lacking or is greatly reduced are denoted *lin-12(0)*, and in the mutants both Z1.aaa and Z4.aaa become anchor cells (Figure 14.14C). In contrast, dominant mutations in which *lin-12* activity is overexpressed are denoted *lin-12(d)*, and in these mutants both Z1.aaa and Z4.ppp become ventral uterine precursor cells (Figure 14.14D).

Molecular analysis of *lin-12* revealed a gene encoding a typical transmembrane receptor protein, a function consistent with the opposite effects of *lin-12* mutations. Too little receptor renders cells insensitive to signal; too much receptor renders them super sensitive. The transmembrane region of the LIN-12 protein (**Figure 14.15**) divides the molecule into an extracellular part (the amino end) and an intracellular part (the carboxyl end). The extracellular part contains 13 repeats of a domain found in a mammalian peptide hormone, epidermal growth factor (EGF), as well as in the product of the *Notch* gene in *Drosophila*, which controls the decision between epidermal and neural cell fates. Nearer the transmembrane region, the amino end contains three

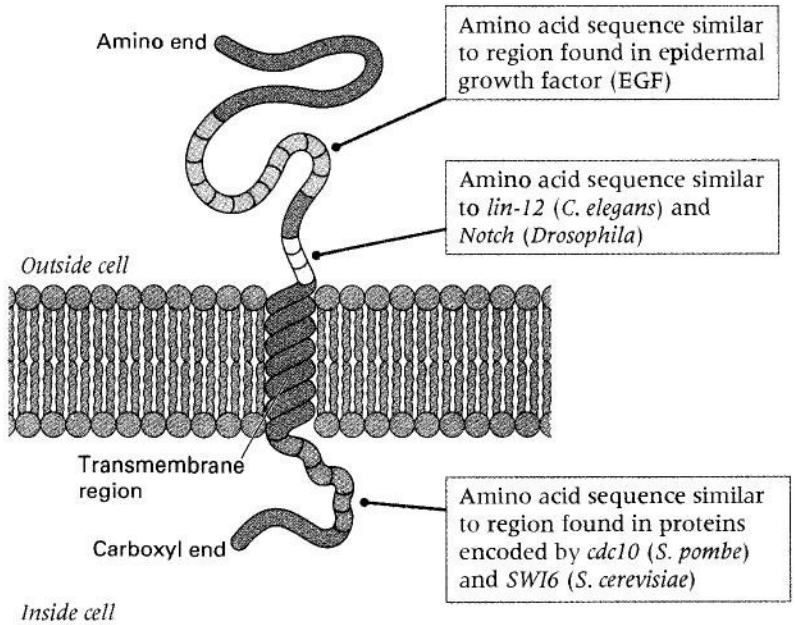


Figure 14.15 The structure of the LIN-12 protein is that of a receptor protein containing a transmembrane region and various types of repeated units that resemble those in epidermal growth factor and other developmental-control genes.

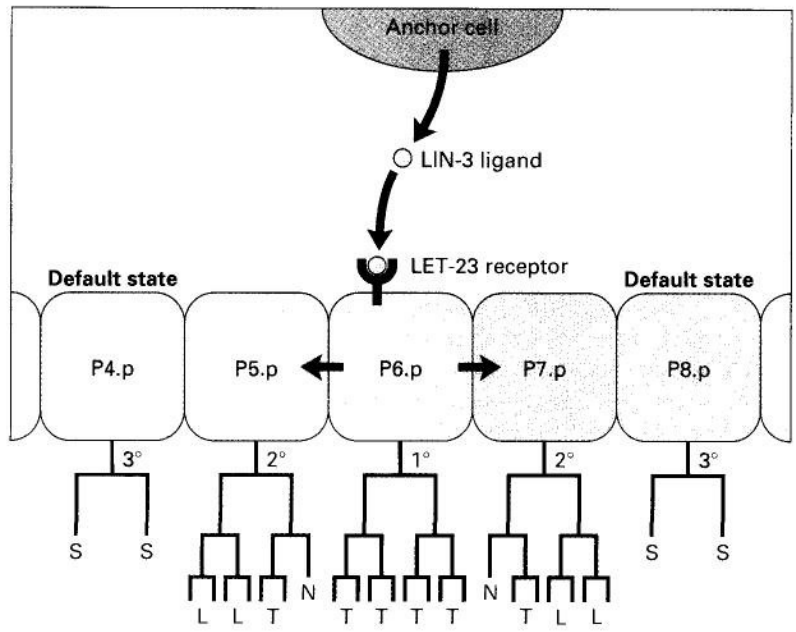


Figure 14.16 Determination of vulval differentiation by means of intercellular signaling. Cells P4.p through P8.p in the hermaphrodite give rise to lineages in the development of the vulva. The three types of lineages are designated 1°, 2°, and 3°. The 1° lineage is induced in P6.p by the ligand LIN-3 produced in the anchor cell (AC), which stimulates the LET-23 receptor tyrosine kinase in P6.p. The P6.p cell, in turn, produces a ligand that stimulates receptors in P5.p and P7.p to induce the 2° fate. The 3° fate is the default condition, which P4.p and P8.p adopt normally and all cells adopt in the absence of AC.

repeats of a cysteine-rich domain also found in the *Notch* protein. Inside the cell, the carboxyl part of the LIN-12 protein contains six repeats of a domain also found in the genes *SWI6* and *cdc10*, which control cell division in budding yeast and fission yeast.

Figure 14.16 illustrates another example in which loss-of-function and gain-of-function alleles have opposite effects on phenotype. The anchor cell expresses a signaling gene, called *lin-3*, whose product controls the fate of target cells in the development of the vulva. Figure 14.16 illustrates five precursor cells, P4.p through P8.p, that participate in the development of the vulva. Each precursor cell has the capability of differentiating into one of three fates, called the 1°, 2°, and 3° lineages, which differ in whether descendant cells remain in a syncytium (S) or divide longitudinally (L), transversely (T), or not at all (N). The precursor cells normally differentiate as shown, yielding five lineages in the order 3°–2°–1°–2°–3°. The vulva itself is formed from the 1° and 2° cell lineages. The spatial arrangement of some of the key cells is shown in the photograph in **Figure 14.17**. The black arrow indicates the anchor cell, and the white lines show the pedigrees of 12 cells. The four cells in the middle derive from P6.p, and the four on each side derive from P5.p, and P7.p.

The important role of the *lin-3* gene product (LIN-3) is suggested by the opposite phenotypes of loss-of-function and gain-of-function alleles. Loss of LIN-3 results in the complete absence of vulval development. Conversely, overexpression of LIN-3 results in excess vulval induction.

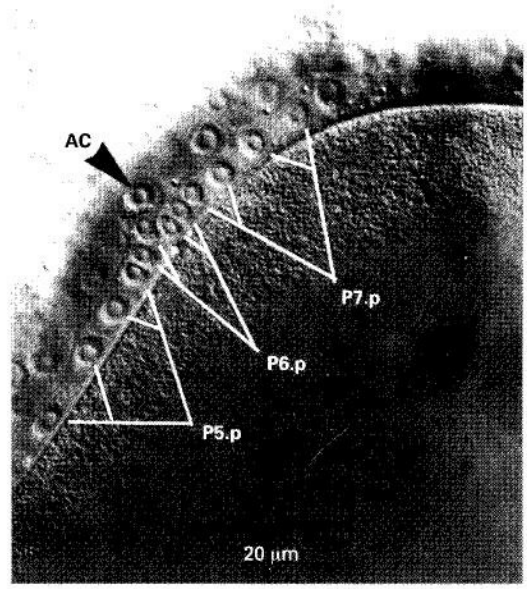


Figure 14.17 Spatial organization of cells in the vulva, including the anchor cell (black arrowhead) and the daughter cells produced by the first two divisions of P5.p through P7.p (white tree diagrams). The length of the scale bar equals 20 micrometers. [Courtesy of G. D. Jongeward, T. R. Clandinin, and P. W. Sternberg. 1995. *Genetics* 139: 1553.]

LIN-3 is a typical example of an interacting molecule, or **ligand,** that interacts with an EGF-type transmembrane receptor. In this case, the receptor is located in cell P6.p and is the product of the gene *let-23*. The LET-23 protein is a tyrosine–kinase receptor that, when stimulated by the LIN-3 ligand, stimulates a series of intracellular signaling events that ultimately results in the synthesis of transcription factors that determine the 1° fate. Among the genes that are induced is a gene for yet another ligand, which stimulates receptors on the cells P5.p and P7.p, causing these cells to adopt the 2° fate (blue horizontal arrows in Figure 14.16). The evidence for horizontal signaling is found in genetic mosaic worms in which the LET-23 receptor is missing in some or all of P5.p through P7.p. If the receptor is missing in all three cell types, none of the cells adopts its normal fate. However, if the receptor is present in P6.p but absent in P5.p and P7.p, then all three cell types differentiate as they should, which implies that receipt of the LIN-3 signal is necessary for 1° determination and that a stimulated P6.p is necessary for 2° determination.

In vulval development, the adoption of the 3° lineages by the P4.p and P8.p cells is determined not by a positive signal but by the lack of a signal because in the absence of the anchor cell, all of the cells P4.p through P8.p express the 3° lineage. Thus development of the 3° lineage is the default, or uninduced, state, which means that the 3° fate is preprogrammed into the cell and must be overridden by another signal if the cell's fate is to be altered.

14.4 Genetic Control of Development in *Drosophila*

Many important insights into developmental processes have been gained from genetic analysis in *Drosophila melanogaster*. **Figure 14.18** summarizes the developmental cycle through egg, larval, pupal, and adult stages. Early development includes a series of cell divisions, migrations, and infoldings that result in the gastrula. About 24 hours after fertilization, the first-stage larva, composed of about 10^4 cells, emerges from the egg.

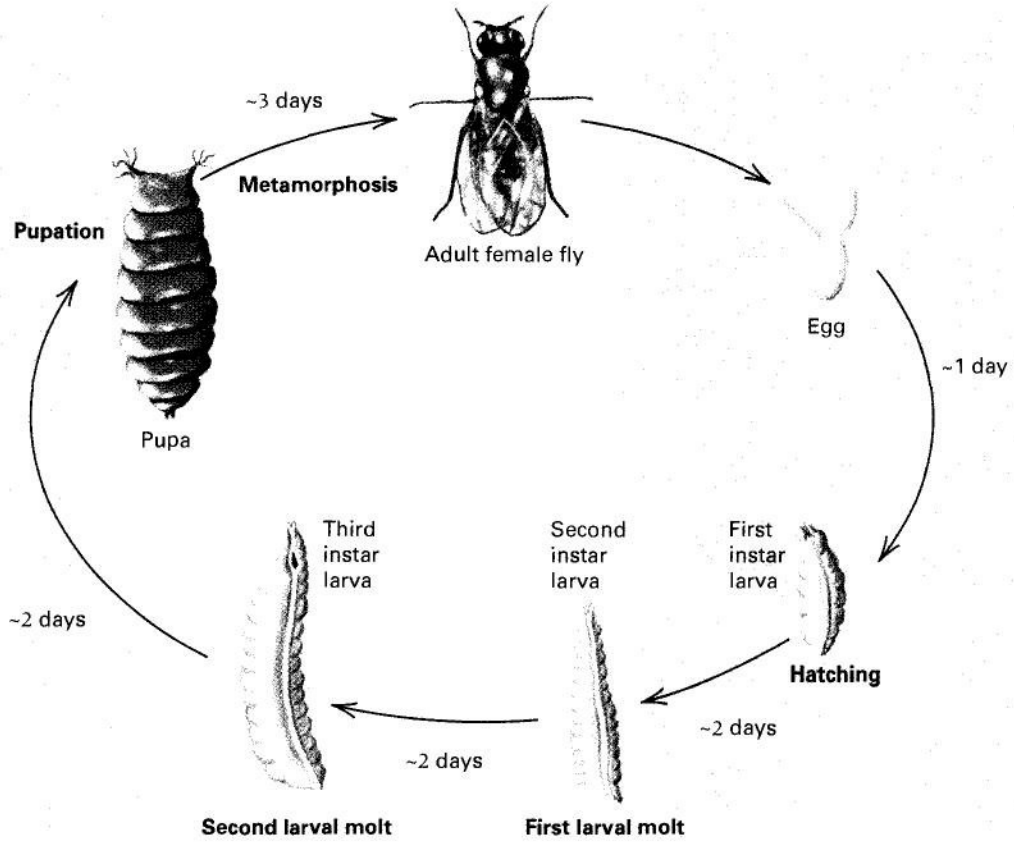


Figure 14.18 Developmental program of *Drosophila melanogaster*. The durations of the stages are at 25°C.

Each larval stage is called an *instar*. Two successive larval molts that give rise to the second and third instar larvae are followed by pupation and a complex metamorphosis that gives rise to the adult fly composed of more than 10^6 cells. In wildtype strains reared at 25°C, development requires from 10 to 12 days.

Early development in *Drosophila* takes place within the egg case (**Figure 14.19**A). The first nine mitotic divisions occur in rapid succession, without division of the cytoplasm, and produce a cluster of nuclei within the egg (Figure 14.19B). The nuclei migrate to the periphery, and the germ line is formed from about 10 **pole cells** set off at the posterior end (Figure 14.19C); the pole cells undergo two additional divisions and are reincorporated into the embryo by invagination. The nuclei within the embryo undergo four more mitotic divisions without division of the cytoplasm, forming the **syncytial blastoderm**, which contains about 6000 nuclei (Figure 14.19D). Cellularization of the blastoderm takes place from about 150 to 180 minutes after fertilization by the synthesis of membranes that separate the nuclei. The **blastoderm** formed by cellularization (Figure 14.19E) is a flattened hollow ball of cells that corresponds to the blastula in higher animals.

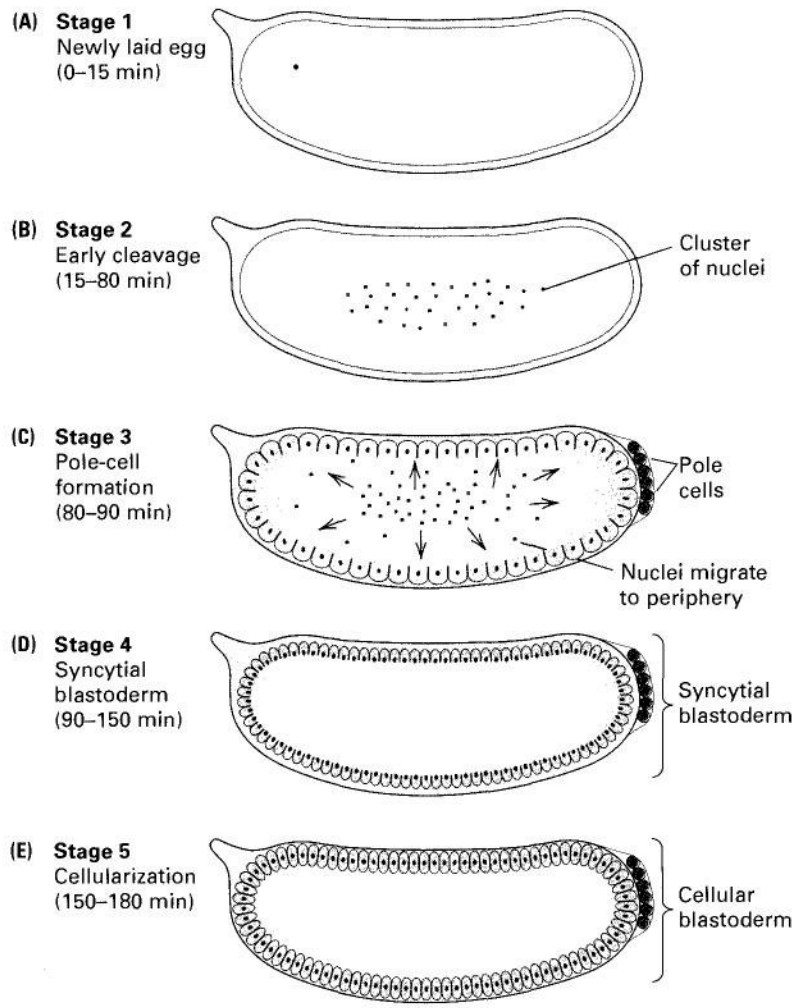
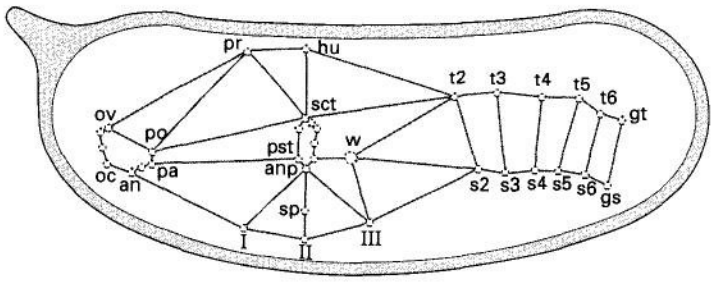


Figure 14.19 Early development in *Drosophila*. (A) The nucleus in the fertilized egg. (B) Mitotic divisions take place synchronously within a syncytium. (C) Some nuclei migrate to the periphery of the embryo, and at the posterior end, the pole cells (which form the germ line) become cellularized. (D) Additional mitotic divisions occur within the syncytial blastoderm. (E) Membranes are formed around the nuclei, giving rise to the cellular blastoderm.

The ablation of patches of cells within a *Drosophila* blastoderm results in localized defects in the larva and adult. The spatial correlation between the position of the cells destroyed and the type of defects results in a **fate map** of the blastoderm, which specifies the cells in the blastoderm that give rise to the various larval and adult structures. Use of genetic markers in the blastoderm has made possible further refinement of the fate map (**Figure 14.20**). Cell lineages can be genetically marked during development by inducing recombination between homologous chromosomes in mitosis, which results in genetically different daughter cells (Chapter 5). Much like the cells in the early blastula of *Caenorhabditis*, cells in the blastoderm of *Drosophila* have predetermined developmental fates, with little ability to substitute in development for other, sometimes even adjacent, cells. Evidence for this conclusion comes from experiments in which cells from a genetically marked blastoderm are implanted into host blastoderms. Blastoderm cells implanted into the equivalent regions of the host become part of the normal adult structures. However, blastoderm cells implanted into different regions develop autonomously and are not integrated into host structures.

Because of the relatively high degree of determination in the blastoderm, genetic analysis of *Drosophila* development has tended to focus on the early stages of development, when the basic body plan of the embryo is established and key regulatory processes become activated. The following



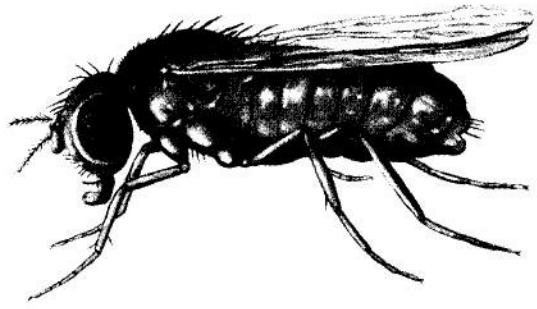


Figure 14.20 Fate map of the *Drosophila* blastoderm, which shows the adult structures that derive from various parts. The map was determined by correlating the expression of genetic markers in different adult structures in genetic mosaics. The abbreviations stand for various body parts in the adult fly. For example, ov and oc are head structures; w is the wing; I, II, and III are the first, second, and third legs; and gs and gt are genital structures. [After J. C. Hall, W. M. Gelbart, and D. R. Kankel. 1976. In *The Genetics and Biology of Drosophila*, vol. 1a, M. Ashburner and E. Novitski, eds. New York: Academic Press, pp. 265–314.]

sections summarize the genetic control of these early events.

Maternal-Effect Genes and Zygotic Genes

Early development in *Drosophila* requires translation of maternal mRNA molecules present in the oocyte. Blockage of protein synthesis during this period arrests the early cleavage divisions. Expression of the zygote genome is also required, but the timing varies. Blockage of transcription of the zygote genome at any time after the ninth cleavage division prevents formation of the blastoderm.

Because the earliest stages of *Drosophila* development are programmed in the oocyte, mutations that affect oocyte composition or structure can upset development of the embryo. Genes that function in the mother and are needed for development of the embryo are called **maternal-effect genes,** and developmental genes that function in the embryo are called **zygotic genes.** The interplay between the two types of genes is as follows:

> The zygotic genes interpret and respond to the positional information laid out in the egg by the maternal-effect genes.

Mutations in maternal-effect genes are easy to identify because homozygous females produce eggs that are unable to support normal embryonic development, whereas homozygous males produce normal sperm. Therefore, reciprocal crosses give dramatically different results. For example, a recessive maternal-effect mutation, *m*, will yield the following results in crosses:

$$m/m\ ♀ \times +/+\ ♂ \rightarrow +/m \text{ progeny}$$
(abnormal development)

$$+/+\ ♀ \times m/m\ ♂ \rightarrow +/m \text{ progeny}$$
(normal development)

The $+/m$ progeny of the reciprocal crosses are genetically identical, but development is upset when the mother is homozygous m/m.

The reason why maternal-effect genes are needed in the mother is that the maternal-effect genes establish the polarity of the *Drosophila* oocyte even before fertilization takes place. They are active during the earliest stages of embryonic development, and they determine the basic body plan of the embryo. Maternal-effect mutations provide a valuable tool for investigating the genetic control of pattern formation and for identifying the molecules that are important in morphogenesis.

Genetic Basis of Pattern Formation in Early Development

The *Drosophila* embryo features 14 superficially similar repeating units visible as a pattern of stripes along the main trunk (**Figure 14.21**). The stripes can be recognized externally by the bands of *denticles*, which are tiny, pigmented, tooth-like projections from the surface of the larva. The 14 stripes in the embryo correspond to the segments in the larva that forms from the embryo. Each **segment** is defined morphologically as the region between successive indentations formed by the sites of muscle attachment in the larval cuticle. The designations of the segments are indicated in Figure 14.21. There are three head segments (C1–C3), three thoracic segments (T1–T3), and eight abdominal segments (A1–A8). In addition to the segments, another type of repeating unit is also important in devel-

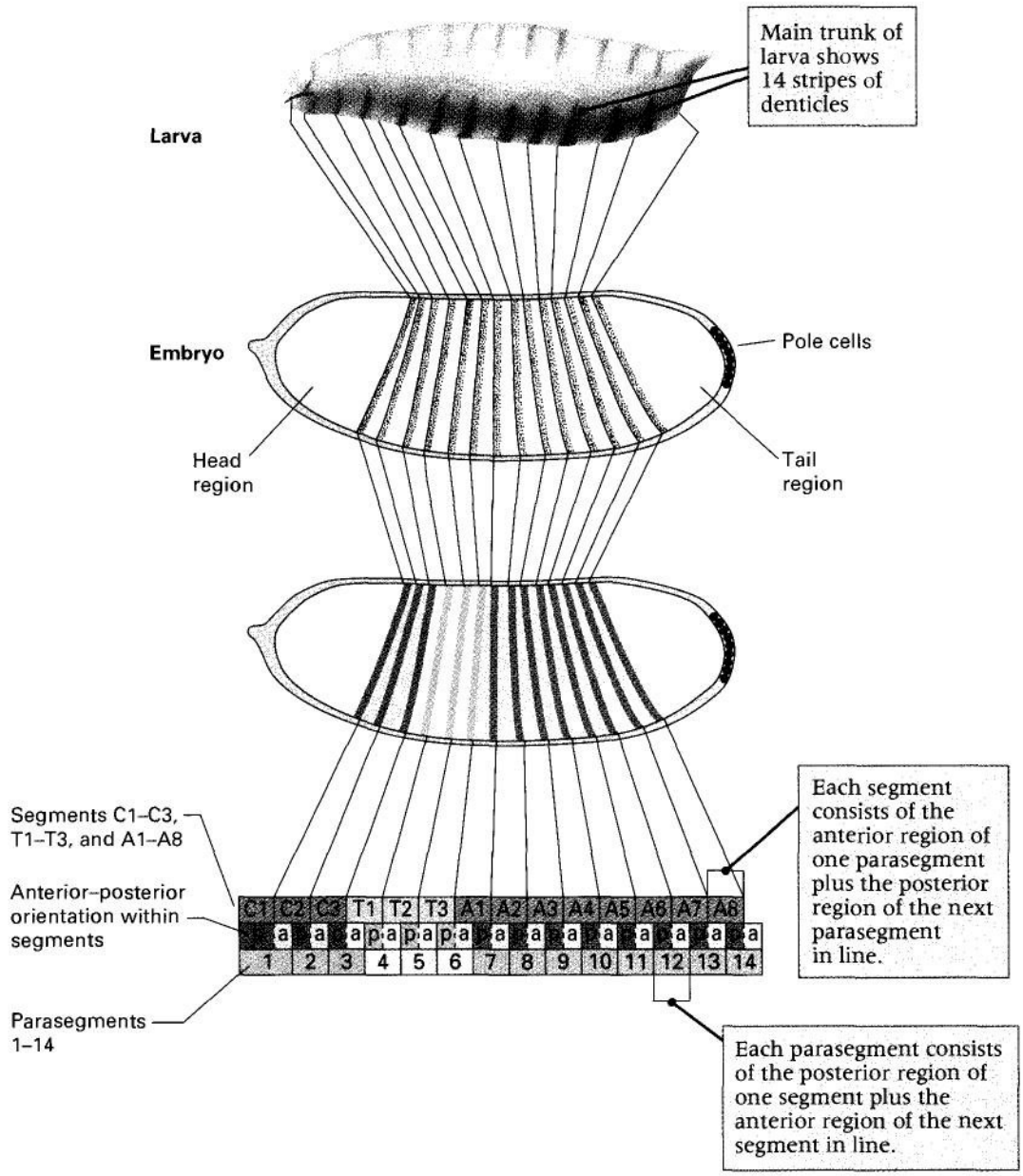


Figure 14.21 Segmental organization of the *Drosophila* embryo and larva. The segments are defined by successive indentations formed by the sites of muscle attachment in the larval cuticle. The parasegments are not apparent morphologically but include the anterior and posterior regions of adjacent segments. The patterns of expression of segmentation genes are more often correlated with the parasegment boundaries than with the segment boundaries.

opment. These repeating units are called **parasegments;** each parasegment consists of the posterior region of one segment and the anterior region of the adjacent segment. Parasegments have a transient existence in embryonic development. Although they are not visible morphologically, they are important in gene expression because the patterns of expression of many genes coincide with the boundaries of the parasegments rather than with the boundaries of the segments.

The early stages of pattern formation are determined by genes that are often called **segmentation genes** because they determine the origin and fate of the segments and parasegments. There are four classes of segmentation genes, which differ in their times and patterns of expression in the embryo.

1. The *coordinate genes* determine the principal coordinate axes of the embryo: the anterior–posterior axis, which defines the front and rear, and the dorsal–ventral axis, which defines the top and bottom.
2. The *gap genes* are expressed in contiguous groups of segments along the embryo (**Figure 14.22**A), and they establish the next level of spatial organization. Mutations in gap genes result in the absence of contiguous body segments, so gaps appear in the normal pattern of structures in the embryo.
3. The *pair-rule genes* determine the separation of the embryo into discrete segments (Figure 14.22B). Mutations in pair-rule genes result in missing pattern elements in alternate segments. The reason for the two-segment periodicity of pair-rule genes is that the genes are expressed in a zebra-stripe pattern along the embryo.
4. The *segment-polarity genes* determine the pattern of anterior–posterior development within each segment of the embryo (Figure 14.22C). Mutations in segment-polarity genes affect all segments or parasegments in which the normal gene is active. Many segment-polarity mutations have the normal number of segments, but part of each segment is deleted and the remainder is duplicated in mirror-image symmetry.

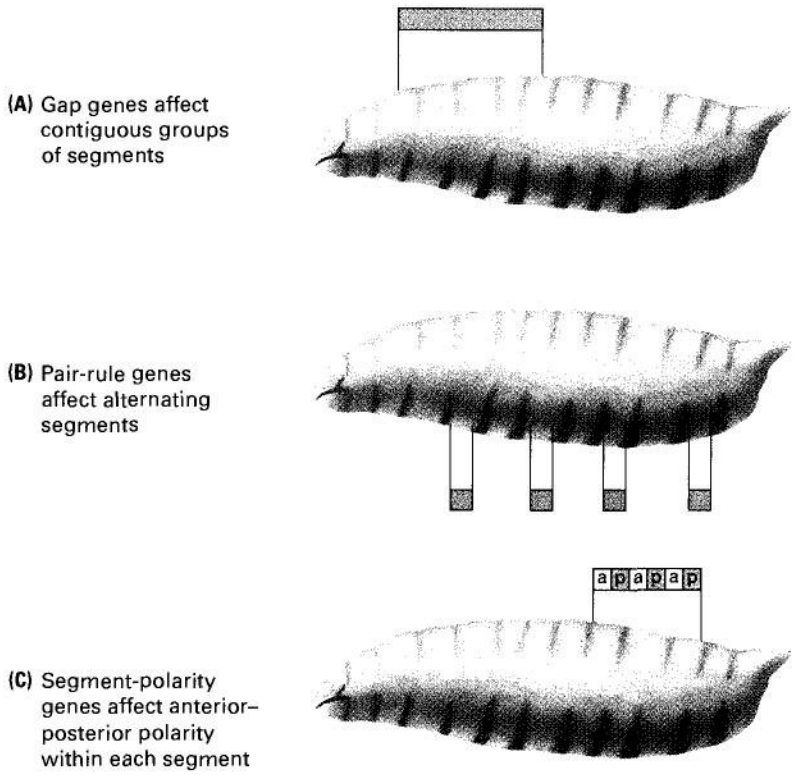


Figure 14.22 Patterns of expression of different types of segmentation genes. (A) The gap genes are expressed in a set of contiguous segments. (B) The pair-rule genes are expressed in alternating segments. (C) The segment-polarity genes are expressed in each segment and determine the anterior-posterior pattern of differentiation within each parasegment.

Evidence for the existence of these four classes of segmentation genes is presented in the following sections.

Coordinate Genes

The **coordinate genes** are maternal-effect genes that establish early polarity through the presence of their products at defined positions within the oocyte or through gradients of concentration of their products. As illustrated in Figure 14.23, the genes that determine the anterior–posterior axis can be classified into three groups—*anterior, posterior,* and *terminal*—according to the morphology of mutant phenotypes.

- The *anterior genes* affect the head and thorax. The key gene in this class is *bicoid.* Mutations in *bicoid* produce embryos that lack the head and thorax and occasionally have abdominal segments in reverse polarity duplicated at the anterior end. The *bicoid* phenotype resembles that produced by certain kinds of surgical manipulations. For example, when *Drosophila* eggs are punctured and small amounts of cytoplasm allowed to escape, loss of cytoplasm from the anterior end results in embryos in which some posterior structures develop in place of the head. Similarly, replacement of anterior cytoplasm with posterior cytoplasm by injection yields embryos with two mirror-image abdomens and no head.

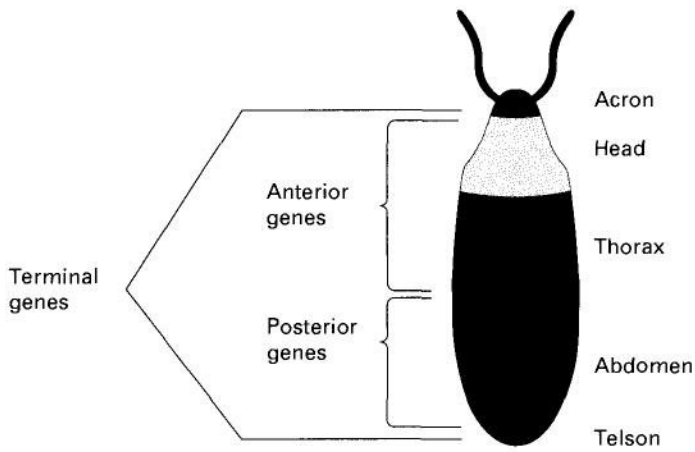


Figure 14.23 Regional differentiation of the early *Drosophila* embryo along the anterior-posterior axis. Mutations in any of the classes of genes shown result in elimination of the corresponding region of the embryo.

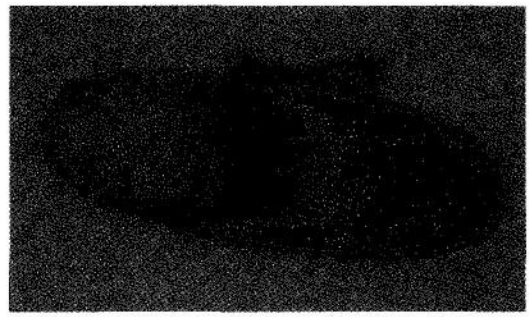


Figure 14.24 A gradient of gene expression resembling that of *bicoid* in the early *Drosophila* embryo. In this photograph, the intensity of the fluorescent signal has been pseudocolored so that the region of highest expression is pink and the region of lowest expression is green. [Courtesy of James Langeland, Stephen Paddock, and Sean Carroll.]

The *bicoid* gene product is a transcription factor for genes that determine anterior structures. Because the *bicoid* mRNA is localized in the anterior part of the early-cleavage embryo, these genes are activated primarily in the anterior region. The *bicoid* mRNA is produced in nurse cells (the cells surrounding the oocyte in Figure 14.7) and exported to a localized region at the anterior pole of the oocyte. The protein product is less localized and, during the syncytial cleavages, forms an anterior–posterior concentration gradient with the maximum at the anterior tip of the embryo (Figure 14.24). The bicoid protein is a principal morphogen in determining the blastoderm fate map. The protein is a transcriptional activator that contains a helix–turn–helix motif for DNA binding (Chapter 12). Genes affected by the bicoid protein contain multiple upstream binding domains that consist of nine nucleotides resembling the consensus sequence 5'-TCTAATCCC-3'. Binding sites that differ by as many as two base pairs from the consensus sequence can bind the bicoid protein with high affinity, and sites that contain four mismatches bind with low affinity. The combination of high- and low-affinity binding sites determines the concentration of bicoid protein needed for gene activation; genes with many high-affinity binding sites can be activated at low concentrations, but those with many low-affinity binding sites need higher concentrations. Such differences in binding affinity

CoNnection

Embryo Genesis

Christiane Nüsslein-Volhard and Eric Wieschaus 1980

European Molecular Biology Laboratory, Heidelberg, Germany

Mutations Affecting Segment Number and Polarity in Drosophila

Nüsslein-Volhard and Wieschaus were exceptionally bold in supposing that the molecular mechanisms governing a process as complex as early embryonic development could be understood by the genetic and molecular analysis of mutations. The phenotype of such mutants is superficially identical: The embryo dies. The Drosophila *genetic map was already littered with mutations classified collectively as "recessive lethals." These were generally considered as not amenable to further analysis because, in any particular case, the search for the specific defect was regarded as a needle-in-a-haystack problem. Nüsslein-Volhard and Wieschaus ignored most of the existing mutants. They set out to acquire systematically a new set of recessive-lethal mutants, each showing a specific and characteristic type of defect in the formation of organized patterns in the early embryo. Their first efforts, reported in this paper, yielded a number of mutations in each of three major classes of genes concerned with development. The paper sparked an enormous interest in* Drosophila *developmental genetics. Today, a typical Annual Drosophila Research Conference includes approximately 500 presentations (mainly posters) dealing with aspects of* Drosophila *development. Nüsslein-Volhard and Wieschaus were awarded a Nobel Prize in 1995. They shared it with Edward B. Lewis for his pioneering genetic studies of the homeotic genes.*

The construction of complex form from similar repeating units is a basic feature of spatial organization in all higher animals. Very little is known for any organism about the genes involved in this process. In *Drosophila* , the metameric [repeating] nature of the pattern is most obvious in the thoracic and abdominal segments of the larval epidermis and we are attempting to identify all loci required for the establishment of this pattern. . . . We have undertaken a systematic search for mutations that affect the segmental pattern. We describe here mutations at 15 loci which show one of three novel types of pattern alteration: pattern duplication (segment polarity mutants; six loci), pattern deletion in alternating segments (pair-rule mutants; six loci) and deletion of a group of adjacent segments (gap mutants; three loci). . . . Segment polarity mutants have the normal number of segments. However, in each segment a defined fraction of the normal pattern is deleted and the remainder is present as a mirror-image duplication. The duplicated part is posterior to the 'normal' part and has reversed polarity. . . . In pair-rule mutants homologous parts of the pattern are deleted in every other segment. Each of the six loci is characterized by its own pattern of deletions. . . . One of the striking features of the [segment-polarity and pair-rule] classes is that the alteration in the pattern is repeated at specific intervals along the antero-posterior axis of the embryo. No such repeated pattern is found in mutants of the gap class and instead a group of up to eight adjacent segments is deleted. . . . The lack of a repeated pattern suggests that the loci are involved in processes in which position along the antero-posterior axis of the embryo is defined by unique values. . . . The majority of mutants described here have been isolated in systematic searches for mutations affecting the segmentation pattern. These experiments are still incomplete. . . . In *Drosophila*, it would seem feasible to identify all genetic components involved in the complex process of embryonic pattern formation.

In *Drosophila*, it would seem feasible to identify all genetic components involved in the complex process of embryonic pattern formation

Source: Nature 287: 795–801

mean that the level of gene expression can differ from one regulated gene to the next along the bicoid concentration gradient. One of the important genes activated by *bicoid* is the gap gene *hunchback*. Other genes in the anterior class code for cellular components necessary for *bicoid* localization.

• The *posterior genes* affect the abdominal segments (Figure 14.23). Some of the mutants also lack pole cells. One of the posterior mutations, *nanos*, yields embryos with defective abdominal segmentation but normal pole cells, which resemble abnormalities produced by surgical removal of the posterior cytoplasm. The phenotype does not result merely from a generalized disruption of development at the posterior end, because the pole cells (as well as a posterior structure called the telson that

normally develops between the pole cells and the abdomen) are not affected in either *nanos* or the surgically manipulated embryos. The *nanos* mRNA is localized tightly to the posterior pole of the oocyte, and the gene product is a repressor of translation. Among the mRNAs not translated in the presence of nanos protein is the *hunchback* mRNA. Hence *hunchback* expression is controlled jointly by the bicoid and nanos proteins; bicoid protein activates transcription in an anterior–posterior gradient, and nanos protein represses translation in the posterior region.

- The *terminal genes* affect both the most anterior structure (the acron) and the most posterior structure (the telson) (Figure 14.23). The key gene in this class is *torso*, which codes for a transmembrane receptor that is uniformly distributed throughout the embryo in the early developmental stages. The torso receptor is activated by a signal released only at the poles of the egg by the nurse cells in that location (Figure 14.7). The torso receptor is a tyrosine kinase that initiates cellular differentiation by means of phosphorylation of specific tyrosine residues in one or more target proteins, among them a *Drosophila* homolog of the vertebrate cancer-associated gene *D-raf*.

Apart from the three sets of genes that determine the anterior–posterior axis of the embryo, a fourth set of genes determines the dorsal–ventral axis. The morphogen for dorsal–ventral determination is the product of the gene *dorsal*, which is present in a pronounced ventral-to-dorsal gradient in the late syncytial blastoderm. The dorsal protein is a transcription factor related to the avian cancer-associated gene *v-rel*. More than a dozen other genes are known to affect dorsal–ventral determination. Mutations in these genes eliminate ventral and lateral pattern elements. In many cases, the mutant embryos can be rescued by the injection of wildtype cytoplasm, no matter where the wildtype cytoplasm is taken from or where it is injected. Examples include the genes *snake*, *gastrulation-defective*, and *easter*. All three genes code for proteins called *serine proteases*. Serine proteases are synthesized as inactive precursors that require a specific cleavage for activation. They often act in a temporal sequence, which means that activation of one enzyme in the pathway is necessary for activation of the next enzyme in line. About half the clotting factors in human blood are serine proteases. The serial activation of the enzymes results in a *cascade effect* that greatly amplifies an initial signal. Each step in the cascade multiplies the signal produced in the preceding step, like a nuclear chain reaction in which each event triggers the occurrence of multiple new events at the next instant in time.

Gap Genes

The main role of the coordinate genes is to regulate the expression of a small group of approximately six gap genes along the anterior–posterior axis. The genes are called **gap genes** because mutations in them result in the absence of pattern elements derived from a group of contiguous segments (Figure 14.22A). Gap genes are zygotic genes. The gene *hunchback* serves as an example of the class because *hunchback* expression is controlled by offsetting effects of *bicoid* and *nanos*. Transcription of *hunchback* is stimulated in an anterior-to-posterior gradient by the bicoid transcription factor, but posterior *hunchback* expression is prevented by translational repression because of the posteriorly localized nanos protein. In the early *Drosophila* embryo in **Figure 14.25**, the gradient of *hunchback* expression is indicated by the green fluorescence of an antibody specific to the *hunchback* gene product. The superimposed red fluorescence results from antibody specific to the product of *Krüppel*, another gap gene. The region of overlapping gene expression appears in yellow. The products of both *hunchback* and *Krüppel* are transcription factors of the zinc finger type (Chapter 12). Other gap genes also encode transcription factors. Together, the gap genes have a pattern of regional specificity and partly overlapping domains of expression that enable them to act in combinatorial fashion to control the next set of genes in the segmentation hierarchy, the pair-rule genes.

Pair-Rule Genes

The coordinate and gap genes determine the polarity of the embryo and establish broad regions within which subsequent

development takes place. As development proceeds, the progressively more refined organization of the embryo is correlated with the patterns of expression of the segmentation genes. Among these are the **pair-rule genes,** in which the mutant phenotype has alternating segments absent or malformed (Figure 14.22B). Approximately eight pair-rule genes have been identified. For example, mutations of the pair-rule gene *even-skipped* affect even-numbered segments, and those of another pair-rule gene, *odd-skipped,* affect odd-numbered segments. The function of the pair-rule genes is to give the early larva a segmented body pattern with both repetitiveness and individuality of segments. For example, there are eight abdominal segments that are repetitive in that they are regularly spaced and share several common features, but they differ in the details of their differentiation.

One of the earliest pair-rule genes expressed is *hairy,* whose pattern of expression is under both positive and negative regulation by the products of *hunchback, Krüppel,* and other gap genes. Expression of *hairy* yields seven stripes (**Figure 14.26**). The striped pattern of pair-rule gene expression is typical, but the stripes of expression of one gene are usually slightly out of register with those of another. Together with the continued regional expression of the gap genes, the combinatorial patterns of gene expression in the embryo are already complex and linearly differentiated. **Figure 14.27** shows an embryo stained for the products of three genes: *hairy* (green), *Krüppel* (red), and *giant* (blue). The regions of overlapping expression appear as color mixtures—orange, yellow, light green, or blue. Even at the early stage in Figure 14.27, there is a unique combinatorial pattern of gene expression in every segment and parasegment. The complexity of combinatorial control can be appreciated by considering that the expression of the *hairy* gene in stripe 7 depends on a promoter element smaller than 1.5 kb that contains a series of binding sites for the protein products of the genes *caudal, hunchback, knirps, Krüppel, tailless, huckbein, bicoid,* and perhaps still other proteins yet to be identified. The combinatorial patterns of gene expression of the pair-rule genes define the boundaries of expression of the

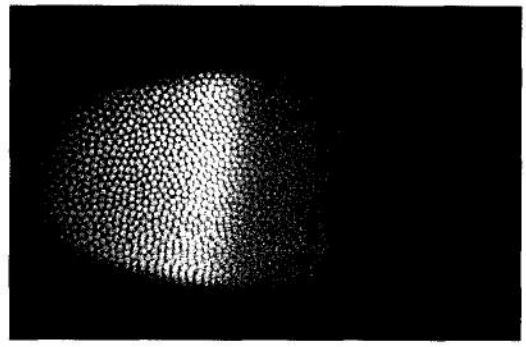


Figure 14.25 An embryo of *Drosophila,* approximately 2.5 hours after fertilization, showing the regional localization of the *hunchback* gene product (green), the *Krüppel* gene product (red), and their overlap (yellow). [Courtesy of James Langeland, Stephen Paddock, and Sean Carroll.]



Figure 14.26 Characteristic seven stripes of expression of the gene *hairy* in a *Drosophila* embryo approximately 3 hours after fertilization. [Courtesy of James Langeland, Stephen Paddock, and Sean Carroll.]

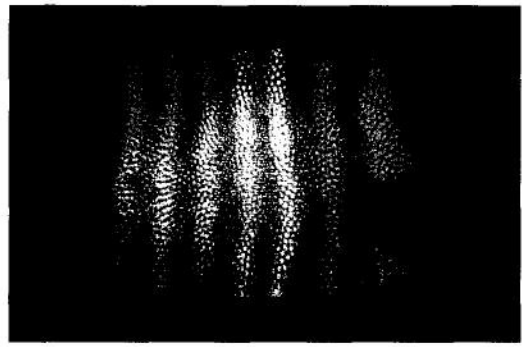


Figure 14.27 Combined patterns of expression of *hairy* (green), *Krüppel* (red), and *giant* (blue) in a *Drosophila* embryo approximately 3 hours after fertilization. Already, considerable linear differentiation is apparent in the patterns of gene expression. [Courtesy of James Langeland, Stephen Paddock, and Sean Carroll.]

segment-polarity genes, which function next in the hierarchy.

Segment-Polarity Genes

Whereas the pair-rule genes determine the body plan at the level of segments and parasegments, the **segment-polarity genes** create a spatial differentiation within each segment. Approximately 14 segment-polarity genes have been identified. The mutant phenotype has repetitive deletion of pattern along the embryo (Figure 14.22C) and usually a mirror-image duplication of the part that remains. Among the earliest segment-polarity genes expressed is *engrailed*, whose stripes of expression approximately coincide with the boundaries of the parasegments and so divide each segment into anterior and posterior domains (**Figure 14.28**).

The expression of the segment-polarity genes finally establishes the early polarity and linear differentiation of the embryo into segments and parasegments. The regulatory interactions within the hierarchy of segmentation genes are illustrated in **Figure 14.29**. These interactions govern the activities of the second set of developmental genes, the *homeotic genes*, which control the pathways of differentiation in each segment or parasegment.



Figure 14.28 Expression of the segment-polarity gene *engrailed* partitions the early *Drosophila* embryo into 14 regions. These eventually differentiate into three head segments, three thoracic segments, and eight abdominal segments. [Courtesy of James Langeland, Stephen Paddock, and Sean Carroll.]

Metamorphosis of the Adult Fly

As with most other insects, the larvae and adults of *Drosophila* have a segmented body plan consisting of a head, three thoracic segments, and eight abdominal segments (**Figure 14.30**). Metamorphosis makes use of about 20 structures called **imaginal disks** present inside the larvae (**Figure 14.31**). Formed early in development, the imaginal disks ultimately give rise to the principal

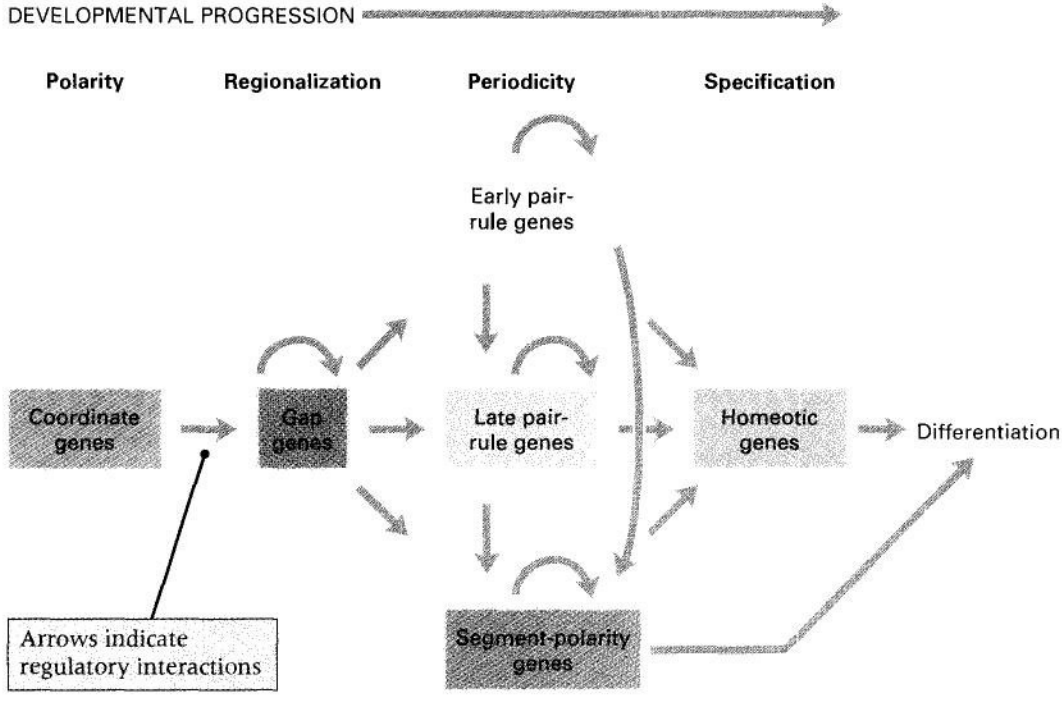


Figure 14.29 Hierarchy of regulatory interactions among genes controlling early development in *Drosophila*. Each gene is controlled by a unique combination of other genes. The terms *polarity*, *regionalization*, *periodicity*, and *specification* refer to the major developmental determinations that are made in each time interval.

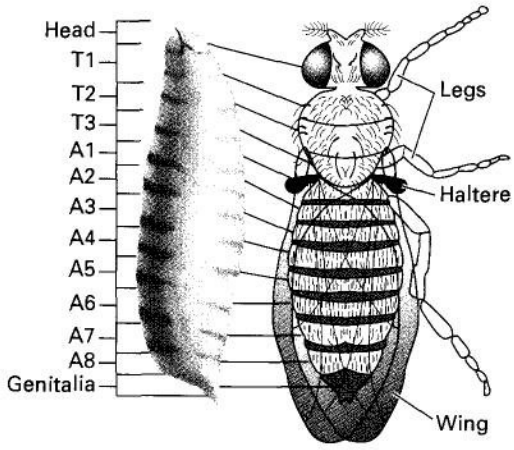


Figure 14.30 Relationship between larval and adult segmentation in *Drosophila*. Each of the three thoracic segments in the adult carries a pair of legs. The wings develop on the second thoracic segment (T2) and the halteres (flight balancers) on the third thoracic segment (T3).

structures and tissues in the adult organism. Examples of imaginal disks are the pair of wing disks (one on each side of the body), which give rise to the wings and related structures; the pair of eye-antenna disks, which give rise to the eyes, antennae, and related structures; and the genital disk, which gives rise to the reproductive apparatus. During the pupal stage, when many larval tissues and organs break down, the imaginal disks progressively unfold and differentiate into adult structures. The morphogenic events that take place in the pupa are initiated by the hormone *ecdysone*, secreted by the larval brain.

Cell determination in *Drosophila* also takes place within bounded units called **compartments.** Cells in the body segments and imaginal disks do not migrate across the boundaries between compartments. For example, the *Drosophila* wing disk includes five compartment boundaries, and most body segments include one boundary that divides the segment into anterior and posterior halves. The evidence for compartments comes from genetic marking of individual cells by means of mitotic recombination and observation of the positions of their descendants. Within each compartment, neighboring groups of cells not necessarily related by ancestry undergo developmental determination together.

As in the early embryo, overlapping patterns of gene expression and combinatorial control guide later events in *Drosophila*

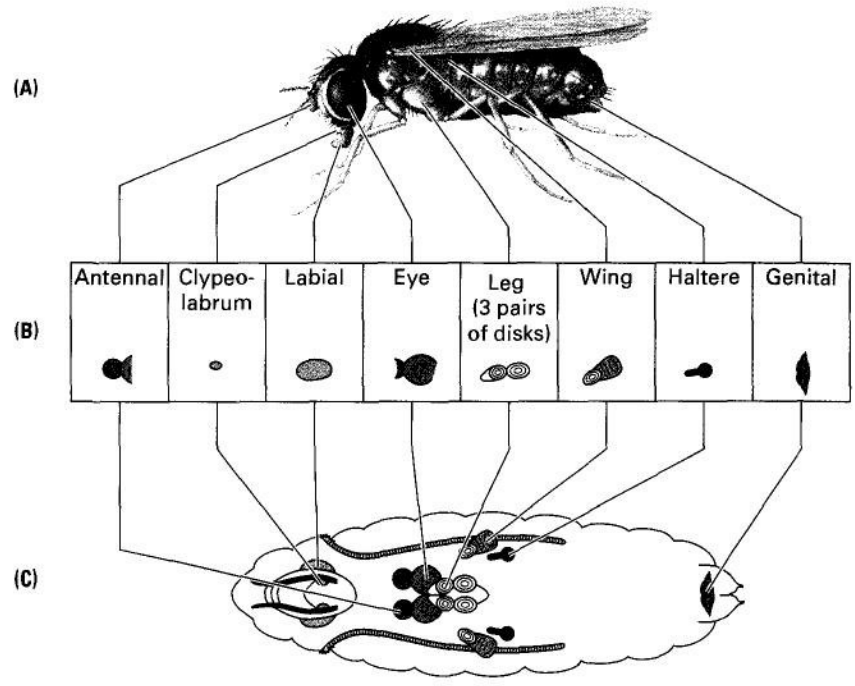


Figure 14.31 Adult *Drosophila* structures (A) and the imaginal disks (B) from which they derive. (C) General morphology and positions of the disks in the late third instar larva.

development. The pattern of expression of a key gene in wing development, *vestigial*, in a wing disk is shown in **Figure 14.32**A. The apparently uniform and approximately circular pattern of expression is actually the summation of *vestigial* response to two separate signaling pathways shown in **Figure 14.33**, which result in the cross-shaped and four-part patterns of expression shown at the bottom. Separate visualization of these patterns in the wing disk is shown in Figure 14.32B. The signaling pathway A in Figure 14.33 consists of the products of the genes *Apterous, fringe, Serrate,* and so forth; and pathway B consists of the products of the genes *engrailed, hedgehog,* and so forth. The Suppressor-of-Hairy protein binds to a *boundary enhancer* in the vestigial gene, which induces gene expression in the cross-shaped pattern. The Mad protein binds to a separate *quadrant enhancer,* which induces gene expression in the quadrant pattern. Such overlapping patterns of gene expression of *vestigial* and other genes in wing development ultimately yield the exquisitely fine level of cellular and morphological differentiation observed in the adult animal.

Homeotic Genes

Among the genes that transform the periodicity of the *Drosophila* embryo into a body plan with linear differentiation that includes the development of wings from segment T2 and halteres from segment T3 are two small sets of **homeotic (*HOX*) genes.** Homeotic genes are defined by mutations that result in the transformation of one body segment into another; the phenotype is that the affected segments differentiate into structures that are normally present elsewhere in the organism. The normal functions of homeotic genes are to regulate target genes concerned with such characteristics as rates of cell division, orientation of mitotic spindles, and the capacity to differentiate bristles, legs, and other features. Homeotic genes are also important in restricting the activities of groups of structural genes to definite spatial patterns. One class of homeotic mutation is illustrated by *bithorax,* which causes transformation of the anterior part of the third thoracic segment into the anterior part of the second thoracic segment, with the result that the halteres are transformed into an extra pair of wings (**Figure 14.34**). The other class of homeotic mutation is illustrated by *Antennapedia,* which results in transformation of the antennae into legs. The *HOX* genes represented by *bithorax* and *Antennapedia* are in fact gene clusters. The cluster containing *bithorax* is designated BX-C (stands for *bithorax*-complex), and that containing *Antennapedia* is called ANT-C (stands for *Antennapedia*-complex). Both

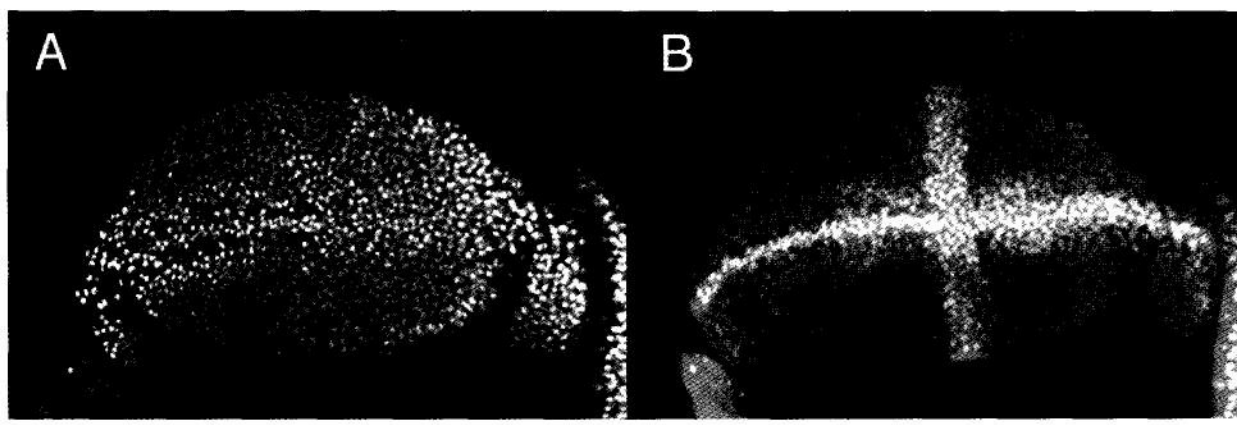


Figure 14.32 (A) Expression of the *vestigial* gene (green) in the developing wing imaginal disk. The approximately circular area of expression gives rise to the wing proper. (B) Visualization of the underlying boundary and quadrant patterns of *vestigial* expression in the same disk. [Courtesy of Stephen Paddock and Sean Carroll.]

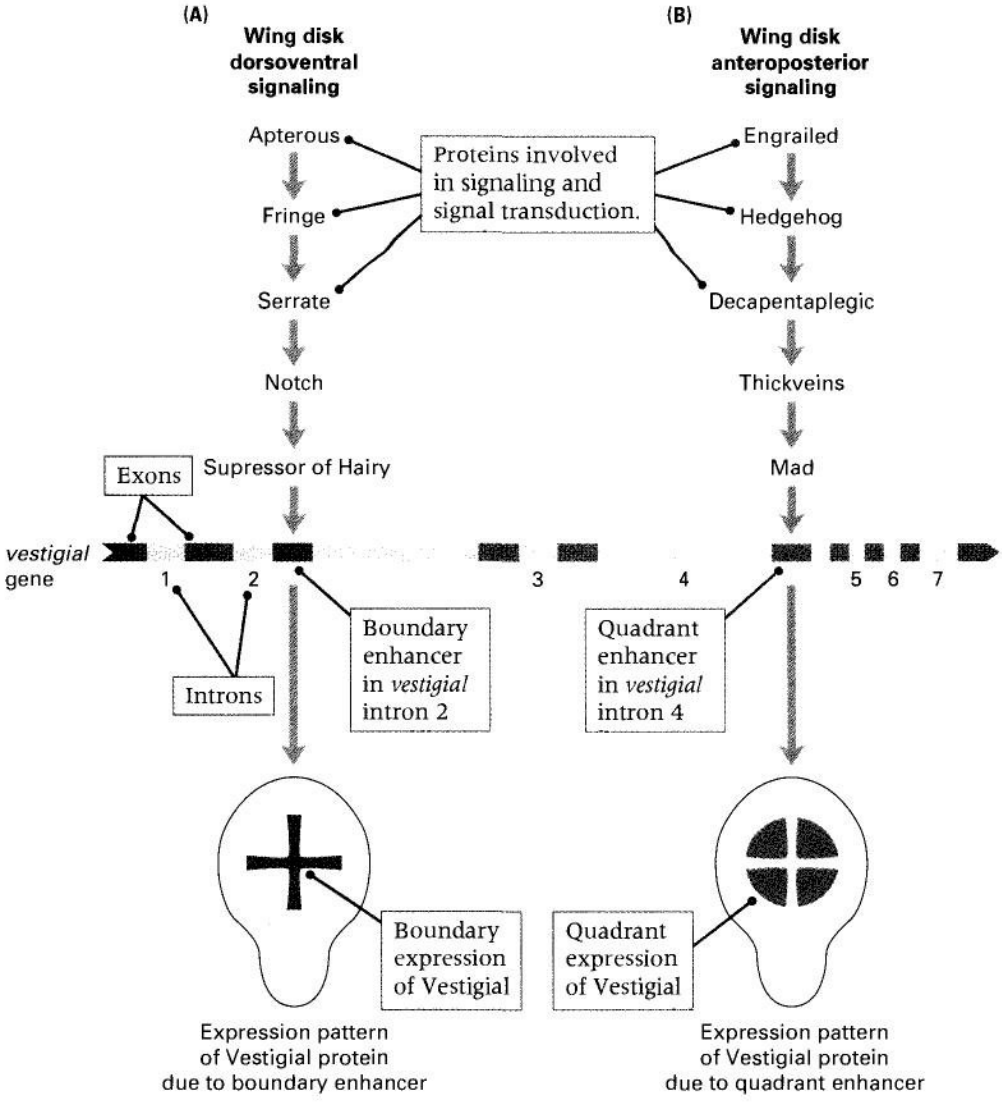


Figure 14.33 The uniform pattern of *vestigial* expression in the wing imaginal disk results from the superposition of two separate patterns. (A) The boundary expression pattern is determined by a dorsoventral signaling pathway. (B) The quadrant expression pattern is determined by the anteroposterior signaling pathway.

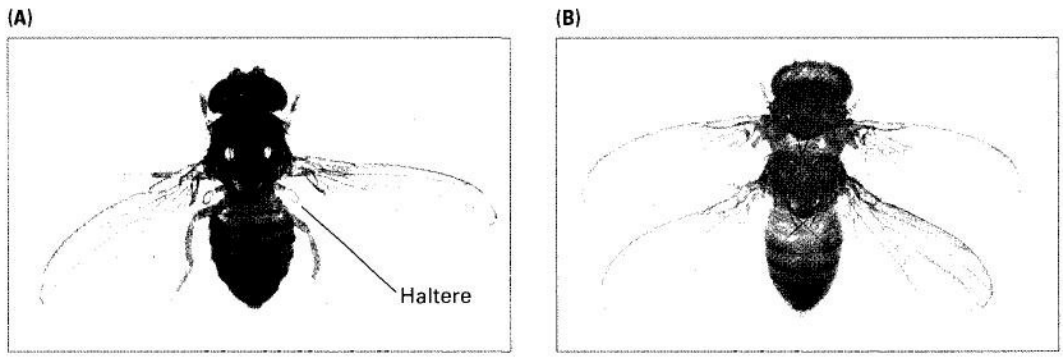


Figure 14.34 (A) Wildtype *Drosophila* showing wings and halteres (the pair of knob-like structures protruding posterior to the wings). (B) A fly with four wings produced by mutations in the *bithorax* complex. The mutations convert the third thoracic segment into the second thoracic segment, and the halteres that are normally present on the third thoracic segment become converted into the posterior pair of wings. [Courtesy of E. B. Lewis.]

gene clusters were initially discovered through their homeotic effects in adults. Later they were shown to affect the identity of larval segments. The BX-C is primarily concerned with the development of larval segments T3 through A8 (Figure 14.30), with principal effects in T3 and A1. The BX-C region extends across approximately 300 kb of DNA yet contains only three essential coding regions. The rest of the region appears to consist of a complex series of enhancers and other regulatory elements that function to specify segment identity by activating the different coding regions to different degrees in particular parasegments. The ANT-C is primarily concerned with the development of the head (H) and thoracic segments T1 and T2.

The homeotic genes encode transcriptional activator proteins. Most *HOX* genes contain one or more copies of a characteristic sequence of about 180 nucleotides called a **homeobox,** which is also found in key genes concerned with the development of embryonic segmentation in organisms as diverse as segmented worms, frogs, chickens, mice, and human beings. Homeobox sequences are present in exons and code for a protein-folding domain that includes a helix–turn–helix DNA-binding motif (Chapter 12).

HOX Genes in Evolution

In *Drosophila*, a *HOX* gene called *Ultrabithorax (Ubx)* is normally active in segment T3, and only in segment T3, which gives rise to the halteres. Reduced *Ubx* expression makes the haltere more wing-like, and in the absence of *Ubx* expression, the haltere disk forms a structurally normal wing (Figure 14.34). Given this information, it is tempting to say that *Ubx* "controls" the development of the haltere. But this inference is wrong, because it implies that *Ubx* expression is both necessary and sufficient for the development of the haltere. The fact is that *Ubx* homologs are expressed in T3 in probably all insects, but segment T3 carries halteres only in the dipteran insects (Figure 12.28), of which *Drosophila* is an example. The ancestral condition in insects is similar to that in today's dragonfly, in which the hind wings are virtually identical to the forewings; the dragonfly *Ubx* homolog clearly does not make a haltere. In the insect lineage leading to present-day lepidopterans, the hind wing became modified from the forewing, but not to such an extreme extent as in the dipterans; yet the lepidopteran *Ubx* homolog is expressed in the hind wing.

The resolution of the paradox of how *Ubx* can seemingly "control" the development of the haltere in dipterans without doing so in other insects is found in the gradual evolution of sets of genes that come under *Ubx* control. The situation is exemplified in **Figure 14.35**, in which the small spheres denote genes and the arrows gene activation in segment T3. The condition in part A is assumed to be the ancestral state, in which the gene *c1* has evolved in such a way that it is activated by the *Ubx* homolog. In this organism, the active genes in T3 will consist of the genes activated by the *Ubx* homolog (*c1*), plus genes further along the regulatory hierarchy activated by these genes (*c2* and *c3*), plus other genes (not shown) whose expression in T3 is independent of the *Ubx* pathway. The complete set of *Ubx*-induced genes, direct and indirect, is shown at the right of the regulatory hierarchy. These are the genes that account for the slight modifications of the hind wing as compared with the forewing. In the absence of the *Ubx* homolog, the hind wing would lose these slight modifications and become even more like a forewing.

As evolution proceeds, mutations can occur that, by chance, can either release a gene from *Ubx* control or bring a gene under *Ubx* control. Such mutations would typically include the loss or gain of *Ubx*-responsive enhancers. A newly evolved *Ubx* responsiveness, or loss of *Ubx* responsiveness, would not affect the expression of the same genes in other tissues, because *Ubx* is expressed specifically in T3. If the novel pattern of T3 gene expression is favored by natural selection, then the mutation will become incorporated into the species. After a long period of evolution, numerous genes will have evolved in this way, progressively modifying the hind-wing pattern in a manner favored by natural selection. In dipterans, for example, the selection was for extreme modification of size and structure to form the haltere. The result of the evolution is shown diagrammatically in

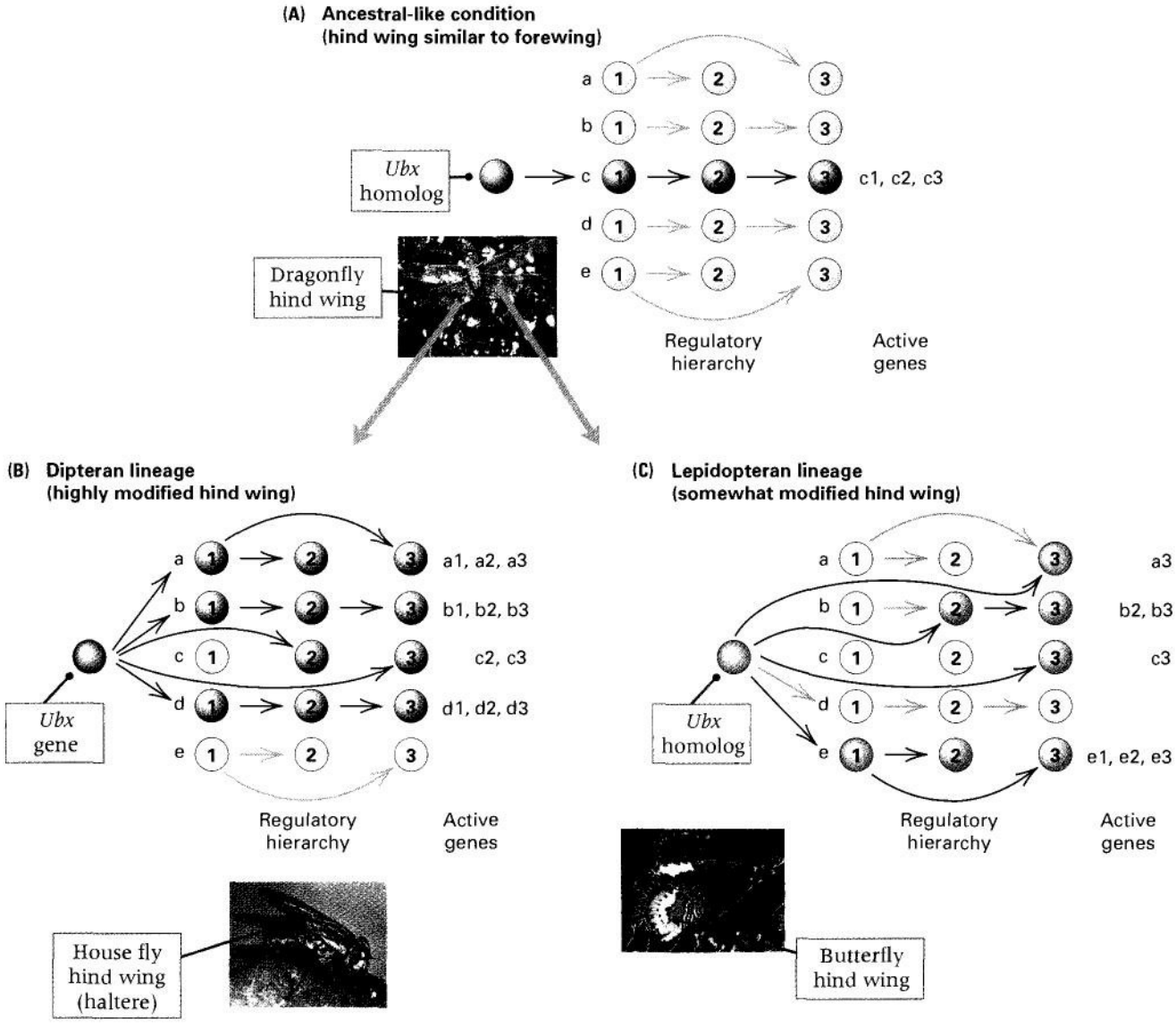


Figure 14.35 Evolutionary scheme by which the *Ubx* homolog expressed in segment T3 produces different developmental pathways for the hind wing in different insect lineages. As each lineage evolves, mutations gradually accumulate, through natural selection, that bring different sets of genes under the control of the *Ubx* homolog. [Based on ideas of Sean Carroll; photos © PhotoDisc.]

Figure 14.35B, where *c1* is depicted as having lost *Ubx* activation and *a1*, *b1*, *c2* and *c3*, and *d1* as having gained it. (Note that the *Ubx* gene can control genes anywhere in the regulatory hierarchy, not only at the beginning.) Similarly, in the lepidopteran lineage (part C), there was selection for less extreme modification, which is depicted as loss of *c1* activation but gain of *a3*, *b2*, *c3*, and *e1* activation. The point is that, independently in each lineage, the hind wing was modified gradually while a sequence of genes was brought under or released from *Ubx* control. The result is that in dipterans the *Ubx* gene "controls" haltere development, whereas in lepidopterans it "controls" a less extreme modification of the forewing. In each lineage, the *Ubx* homolog is critical in developmental control, but it regulates different genes. In each lineage also, mutations in *Ubx* will make the hind wing more similar to the forewing.

14.5 Genetic Control of Development in Higher Plants

Reproductive and developmental processes in plants differ significantly from those in other eukaryotes. For example, plants have an alternation of generations between the diploid sporophyte and the haploid gametophyte, and the plant germ line is not established in a discrete location during embryogenesis but rather at many locations in the adult organism. In a corn plant, for example, each ear contains germ-line cells that undergo meiosis to form the pollen and ovules.

In animals, as we have seen, most of the major developmental decisions are made early in life, in embryogenesis. In higher plants, differentiation takes place almost continuously throughout life in regions of actively dividing cells called **meristems** in both the vegetative organs (root, stem, and leaves) and the floral organs (sepal, petal, pistil, and stamen). The shoot and root meristems are formed during embryogenesis and consist of cells that divide in distinctive geometric planes and at different rates to produce the basic morphological pattern of each organ system. The floral meristems are established by a reorganization of the shoot meristem after embryogenesis and eventually differentiate into floral structures characteristic of each particular species. One important difference between animal and plant development is that:

> In higher plants, as groups of cells leave the proliferating region of the meristem and undergo further differentiation into vegetative or floral tissue, their developmental fate is determined almost entirely by their position relative to neighboring cells.

The critical role of positional information in the development of higher plants stands in contrast to animal development, in which cell lineage often plays a key role in determining cell fate.

The plastic or "indeterminate" growth patterns of higher plants are the result of continuous production of both vegetative and floral organ systems. These patterns are conditioned largely by day length and the quality and intensity of light. The plasticity of plant development gives plants a remarkable ability to adjust to environmental insults. Figure **14.36** shows a tree that, over time, adjusts to the presence of a nearby fence by engulfing it into the trunk. Higher plants can also adjust remarkably well to a variety of genetic aberrations. For example, transgenic plants of *Arabidopsis thaliana* (a member of the mustard family) have been created that either overexpress or underexpress cyclin B. Overexpression of cyclin B results in an accelerated rate of cell division; underexpression of cyclin B results in a decelerated rate. Plants with the faster rate of cell division contain more cells and are somewhat larger than their wildtype counterparts, but otherwise they look completely normal. Likewise, plants with the decreased rate of cell division have less than half the normal number of cells, but they grow at almost the same rate and reach almost the same size as wildtype plants, because as the number of cells decreases, each individual cell gets larger. One of the important consequences of plants being able to adjust to abnormal growth conditions is that plant cells rarely undergo a transformation into proliferative cancer



Mosses are among the groups of plants in which the haploid gametophyte is the dominant phase of the life cycle, well differentiated into stems and leaves. The gametophytes produce the gametes by mitosis, which fuse to form the diploid sporophyte. Shown here is the hairy-cap moss *Polytrichum*. The green stuctures at the base are the gametophytes, the stalks above the sporophytes. [John D. Cunningham/Visuals Unlimited.]



Figure 14.36 The ability of plant development to adjust to perturbations is illustrated by this tree. Encountering a fence, it eventually incorporates the fence into the trunk. [Courtesy of Robert Pruitt.]

cells, as often happens in animals (Chapter 15). The common plant tumors are produced only as a result of complex interactions with pathogens such as *Agrobacterium* (Chapter 13).

Flower Development in *Arabidopsis*

Genetic analysis of *Arabidopsis* has revealed some important principles in the genetic determination of floral structures. As is typical of flowering plants, the flowers of *Arabidopsis* are composed of four types of organs arranged in concentric rings, or whorls. **Figure 14.37** illustrates the geometry, looking down at a flower from the top. From outermost to innermost, the whorls are designated 1, 2, 3, and 4 (Figure 14.37A). In the development of the flower, each whorl gives rise to a different floral organ (Figure 14.37B). Whorl 1 yields the sepals (the green, outermost floral leaves), whorl 2 the petals (the white, inner floral leaves), whorl 3 the stamens (the male organs, which form pollen), and whorl 4 the carpels (which fuse to form the ovary).

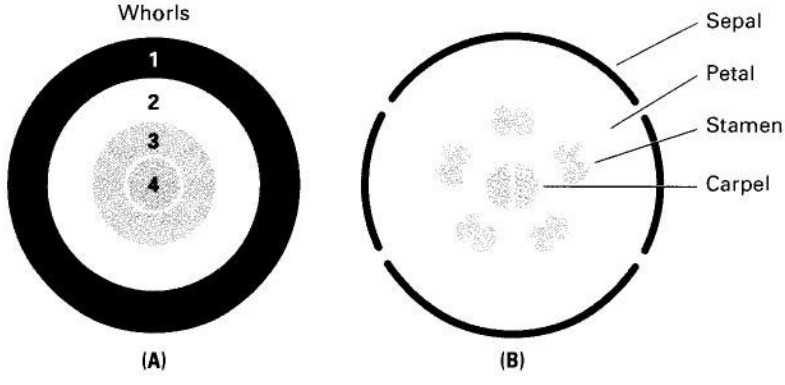
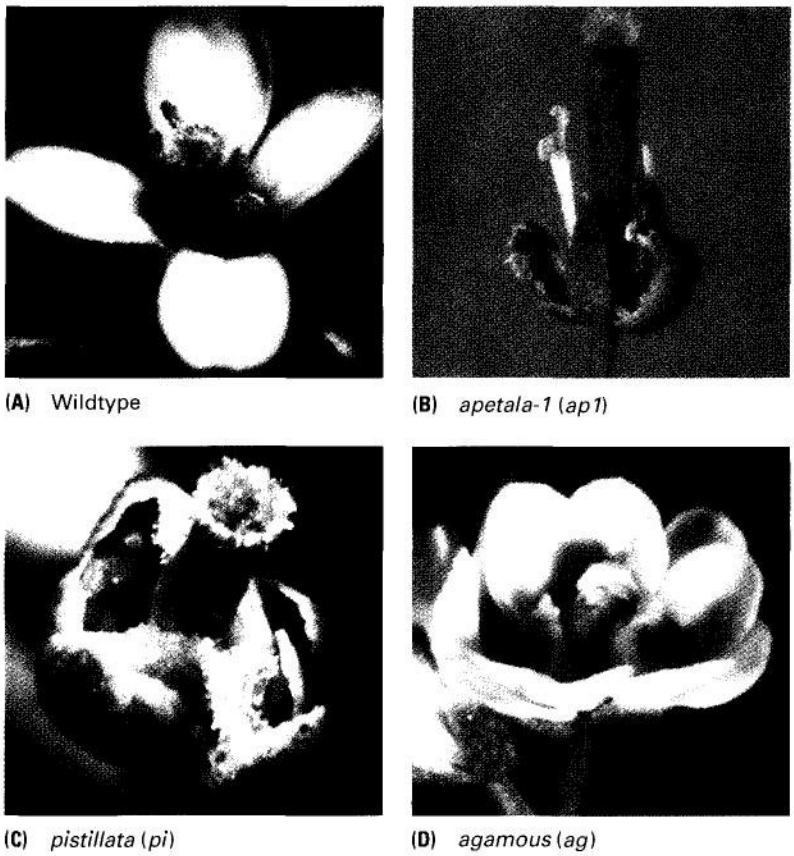


Figure 14.37 (A) The organs of a flower are arranged in four concentric rings, or whorls. (B) Whorls 1, 2, 3, and 4 give rise to sepals, petals, stamens, and carpels, respectively.



(A) Wildtype

(B) *apetala-1 (ap1)*

(C) *pistillata (pi)*

(D) *agamous (ag)*

Figure 14.38 Phenotypes of the major classes of floral mutations in *Arabidopsis*. (A) The wildtype floral pattern consists of concentric whorls of sepals, petals, stamens, and carpels. (B) The homozygous mutation *ap1* (*apetala-1*) results in flowers missing sepals and petals. (C) Genotypes that are homozygous for either *ap3* (*apetala-3*) or *pi* (*pistillala*) yield flowers that have sepals and carpels but lack petals and stamens. (D) The homozygous mutation *ag* (*agamous*) yields flowers that have sepals and petals but lack stamens and carpels. [Courtesy of Elliot M. Meyerowitz and John Bowman. Part B from Elliot M. Meyerowitz. 1994. *Scientific American* 271: 56.]

Table 14.1 Floral development in mutants of *Arabidopsis*

	Whorl			
Genotype	**1**	**2**	**3**	**4**
wildtype	sepals	petals	stamens	carpels
ap1/ap1	carpels	stamens	stamens	carpels
ap3/ap3	sepals	sepals	carpels	carpels
pi/pi	sepals	sepals	carpels	carpels
ag/ag	sepals	petals	petals	sepals

Mutations that affect floral development fall into three major classes, each with a characteristic phenotype (**Figure 14.38**). Compared with the wildtype flower (panel A), one class lacks sepals and petals (panel B), another class lacks petals and stamens (panel C), and the third class lacks stamens and carpels (panel D). On the basis of crosses between homozygous mutant organisms, these classes of mutants can be assigned to different complementation groups, each of which defines a different gene (Chapter 3). Each gene and the phenotype of a plant homozygous for a recessive mutation in the gene are shown in **Table 14.1**. The phenotype lacking sepals and petals is caused by mutations in the gene *ap1 (apetala-1)*. The phenotype lacking stamens and petals is caused by a mutation in either of two genes, *ap3 (apetala-3)* or *pi (pistillata)*. The phenotype lacking stamens and carpels is caused by mutations in the gene *ag (agamous)*. Each of these genes has been cloned and sequenced. They all encode transcription factors. The transcription factors encoded by *ap3*, *pi*, and *ag* are members of what is called the *MADS box* family of transcription factors; each member of this family contains a sequence of 58 amino acids in which common features can be identified. MADS box transcription factors are very common in plants and are also found, less frequently, in animals.

Combinatorial Determination of the Floral Organs

The role of the *ap1*, *ap3*, *pi*, and *ag* transcription factors in the determination of floral organs can be inferred from the phenotypes of the mutations. The logic of the inference is based on the observation (see Table 14.1) that mutation in any of the genes eliminates two floral organs that arise from adjacent whorls. This pattern suggests that *ap1* is necessary for sepals and petals, *ap3* and *pi* are both necessary for petals and stamens, and *ag* is necessary for stamens and carpels. Because the mutant phenotypes are caused by loss-of-function alleles of the genes, it may be inferred that *ap1* is expressed in whorls 1 and 2, that *ap3* and *pi* are expressed in whorls 2 and 3, and that *ag* is expressed in whorls 3 and 4. The overlap-

ping patterns of expression are shown in **Table 14.2.**

The model of gene expression in Table 14.2 suggests that floral development is controlled in combinatorial fashion by the four genes. Sepals develop from tissue in which only *ap1* is active; petals are evoked by a combination of *ap1*, *ap3*, and *pi*; stamens are determined by a combination of *ap3*, *pi*, and *ag*; and carpels derive from tissue in which only *ag* is expressed. This model is illustrated graphically in **Figure 14.39.**

You may have noted already that the model in Table 14.2 does not account for all of the phenotypic features of the *ap1* and *ag* mutations in Table 14.1. In particular, according to the combinatorial model in Table 14.2, the development of carpels and stamens from whorls 1 and 2 in homozygous *ap1* plants would require expression of *ag* in whorls 1 and 2. Similarly, the development of petals and sepals from whorls 3 and 4 in homozygous *ag* plants would require expression of *ap1* in whorls 3 and 4. This discrepancy can be explained if it is assumed that *ap1* expression and *ag* expression are mutually exclusive: In the presence of the *ap1* transcription factor, *ag* is repressed; in the presence of the *ag* transcription factor, *ap1* is repressed. If this were the case, then in *ap1* mutants, *ag* expression would spread into whorls 1 and 2; in *ag* mutants, *ap1* expression would spread into whorls 3 and 4. This additional assumption enables us to explain the phenotypes of the single and even double mutants.

With the additional assumption we have made about *ap1* and *ag* interaction, the model in Table 14.2 fits the data, but is the model true? For these genes, the patterns of gene expression, assayed by *in situ* hybridization of RNA in floral cells with labeled probes for each of the genes, fit the patterns in Table 14.2. In particular, *ap1* is expressed in whorls 1 and 2, *ap3* and *pi* in whorls 2 and 3, and *ag* in whorls 3 and 4. Furthermore, the seemingly arbitrary assumption about *ap1* and *ag* expression being mutually exclusive turns out to be true. In *ap1* mutants, *ag* is expressed in whorls 1 and 2; reciprocally, in *ag* mutants, *ap1* is expressed in whorls 3 and 4. It is also known how *ap3* and *pi* work together. The active transcription factor that corresponds to these genes is a dimeric protein composed

Table 14.2 Domains of expression of genes determining floral development

Whorl	Genes expressed	Determination
1	*ap1*	sepal
2	*ap1* + *ap3* and *pi*	petal
3	*ap3* and *pi* + *ag*	stamen
4	*ag*	carpel

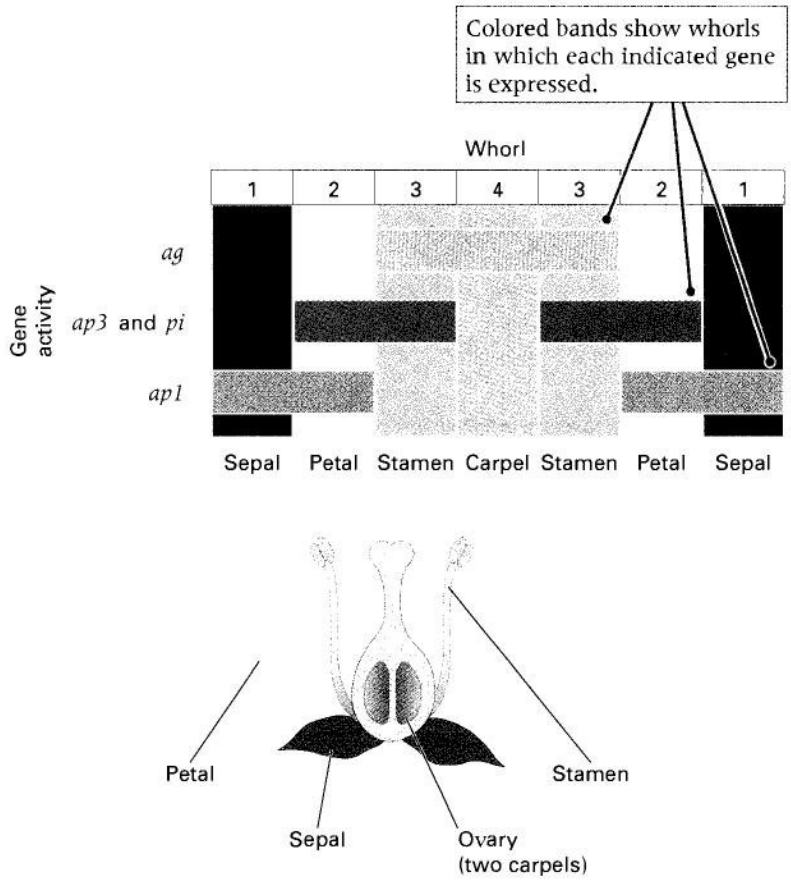


Figure 14.39 Control of floral development in *Arabidopsis* by the overlapping expression of four genes. The sepals, petals, stamens, and carpels are floral organ systems that form in concentric rings, or whorls. The developmental identity of each concentric ring is determined by the genes *ap1*, *ap3* and *pi*, and *ag*, each of which is expressed in two adjacent rings. Gene *ap1* is expressed in the outermost two rings, *ap3* and *pi* in the middle two, and *ag* in the inner two. Therefore, each whorl has a unique combination of active genes.

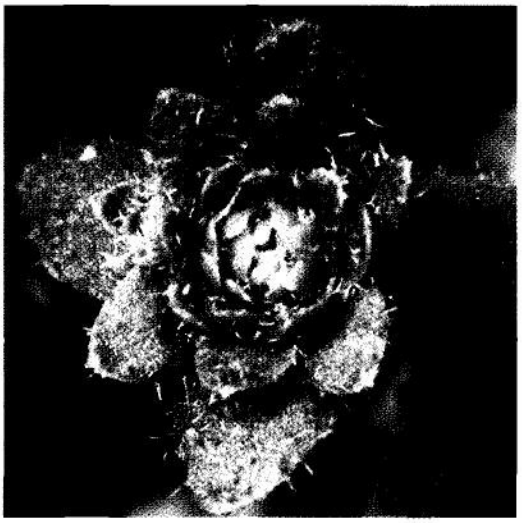


Figure 14.40 The homozygous triple mutant *ap1 pi ag* lacks all of the transcription factors needed for floral development. Hence the "flowers" lack all of the wildtype floral structures. They have no sepals, petals, stamens, or carpels. Without the floral genetic determinants, the flowers consist entirely of leaves arranged in concentric whorls. [Courtesy of Elliot M. Meyerowitz and John Bowman.]

of Ap3 and Pi polypeptides. Each component polypeptide, in the absence of the other, remains inactive in the cytoplasm. Together, they form an active dimeric transcription factor that migrates into the nucleus.

Given the critical role of the Ap1, Ap3/Pi, and Ag transcription factors in floral determination, it might be speculated that triple mutants lacking all three types of transcription factors would have very strange flowers. The phenotype of the *ap1 pi ag* triple mutant is shown in **Figure 14.40**. The flowers have none of the normal floral organs. They consist merely of leaves arranged in concentric whorls.

Chapter Summary

The genotype determines the developmental potential of the embryo. By means of a developmental program that results in different sets of genes being expressed in different types of cells, the genotype controls the developmental events that take place and their temporal order. Mutations that interrupt developmental processes identify genetic factors that control development.

The early development of the animal embryo establishes the basic developmental plan for the whole organism. The earliest events in embryonic development depend on the correct spatial organization of numerous constituents present in the oocyte. Developmental genes that are needed in the mother for proper oocyte formation and zygotic development are maternal-effect genes. Genes that are required in the zygote nucleus are zygotic genes. Fertilization of the oocyte initiates a series of mitotic cleavage divisions that form the "hollow ball" blastula, which rapidly undergoes a restructuring into the gastrula. Accompanying these early morphological events are a series of molecular events that determine the developmental fates that cells undergo. Execution of a developmental state may be autonomous (genetically programmed), or it may require positional information supplied by neighboring cells or the local concentration of one or more morphogens.

The soil nematode *Caenorhabditis elegans* is used widely in studies of cell lineages because many lineages in the organism undergo virtually autonomous development and the developmental program is identical from one organism to the next. Most lineages are affected by many genes, including genes that control the sublineages into which the lineage can differentiate. Mutations that affect cell lineages define several types of developmental mutants: (1) execution mutations, which prevent a developmental program from being carried out; (2) transformation mutations, in which an incorrect developmental program is executed; (3) segregation mutations, in which developmental determinants fail to segregate normally between sister cells or mother and daughter cells; and (4) apoptosis (programmed cell death) mutations, in which cells fail to undergo a normal developmental program that normally leads to their death.

Genes that control key points in development can often be identified by the unusual feature that loss-of-function mutations (usually recessive) and gain-of-function mutations (usually dominant) have opposite effects on phenotype. For example, if loss of function results in failure to execute a developmental program in a particular anatomical position, then gain of function should result in execution of the program in an abnormal location.

The *lin-12* gene in *C. elegans* is an important developmental control gene identified by these criteria. The *lin-12* gene, which encodes a transmembrane receptor, controls the developmental "decision" of a pair of cells whether to become anchor cells or ventral uterine precursor cells.

Early development in *Drosophila* includes the formation of a syncytial blastoderm by early cleavage divisions without cytoplasmic division, the setting apart of pole cells that form the germ line at the posterior of the embryo, the migration of most nuclei to the periphery of the syncytial blastoderm, cellularization to form the cellular blastoderm, and determination of the blastoderm fate map at or before the cellular blastoderm stage. Metamorphosis into the adult fly makes use of about 20 imaginal disks present in the larva that contain developmentally committed cells that divide and develop into the adult structures. Most imaginal disks include several discrete groups of cells or compartments separated by boundaries that progeny cells do not cross.

Early development in *Drosophila* to the level of segments and parasegments requires four classes of segmentation genes: (1) coordinate genes that establish the basic anterior–posterior and dorsal–ventral aspect of the embryo, (2) gap genes for longitudinal separation of the embryo into regions, (3) pair-rule genes that establish an alternating on–off striped pattern of gene expression along the embryo, and (4) segment-polarity genes that refine the patterns of gene expression within the stripes and determine the basic layout of segments and parasegments. For example, the *bicoid* gene is a maternal-effect gene in the coordinate class that controls anterior–posterior differentiation of the embryo and determines position on the blastoderm fate map. The *bicoid* gene product is a transcriptional activator protein that is present in a concentration gradient from anterior to posterior along the embryo. Genes that are regulated by *bicoid* have different sensitivities to its concentration, depending on the number and types of high-affinity and low-affinity binding sites that they contain. The segmentation genes can regulate themselves, other members of the same class, and genes of other classes farther along the hierarchy. Together, the segmentation genes control the homeotic genes that initiate the final stages of developmental specification.

Mutations in homeotic genes result in the transformation of one body segment into another. For example, *bithorax* mutations cause transformation of the anterior part of the third thoracic segment into the anterior part of the second thoracic segment, and *Antennapedia* mutations result in transformation of the dorsal part of the head into the second thoracic segment. Homeotic genes act within developmental compartments to control other genes concerned with such characteristics as rates of cell division, orientation of mitotic spindles, and the capacity to differentiate bristles, legs, and other features. Both the *bithorax* complex (BX-C) and the *Antennapedia* complex (ANT-C) are clusters of genes, each containing homeoboxes that are characteristic of developmental-control genes in many organisms and that code for DNA-binding and other protein domains.

Developmental processes in higher plants differ significantly from those in animals in that developmental decisions continue throughout life in the meristem regions of the vegetative organs (root, stem, and leaves) and the floral organs (sepal, petal, pistil, and stamen). However, genetic control of plant development is mediated by transcription factors analogous to those in animals. For example, control of floral development in *Arabidopsis* is through the combinatorial expression of genes in each of a series of four concentric rings, or whorls, of cells that eventually form the sepals, petals, stamens, and carpels.

Key Terms

apoptosis
autonomous determination
blastoderm
blastula
cell fate
cleavage division
compartment
coordinate gene
embryonic induction
execution mutation
fate map
gap gene
gastrula
homeobox
homeotic gene
imaginal disk
ligand
lineage
lineage diagram
maternal-effect gene
meristem
morphogen
pair-rule gene
parasegment
pattern formation
pole cell
positional information
programmed cell death
segment
segmentation gene
segment-polarity gene
segregation mutation
syncytial blastoderm
transformation mutation
zygotic gene

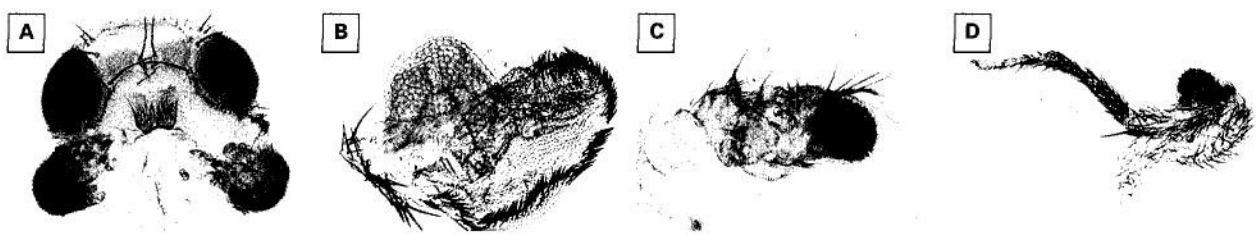
Review the Basics

- What is meant by the term *positional information* in regard to development? How can positional information affect cell fate?
- How does knowledge of the complete cell lineage of nematode development demonstrate the importance of programmed cell death (apoptosis) in development?
- If a gene is both necessary and sufficient for determining a developmental pathway, explain why loss-of-function mutants would be expected to have a different phenotype than gain-of-function mutants.
- What is a receptor? What is a ligand? What role do these types of molecules play in signaling between cells?
- What is meant by the term *polarity* in reference to the mature oocyte?
- How was the study of maternal-effect lethal genes a key to deciphering the genetic control of early embryogenesis in *Drosophila?*
- Among genes that control embryonic development in *Drosophila,* distinguish among coordinate genes, gap genes, pair-rule genes, and segment-polarity genes. Generally speaking, what is the temporal order of expression of these classes of genes?
- What is a homeotic mutation? Give an example from *Drosophila.* Do homeotic mutations occur in organisms other than *Drosophila?*
- Do plants have a germ line in the same sense as animals? What does the difference in germ-cell origin imply about the potential role of "somatic" mutations in the evolution of each type of organism?
- What is the genetic basis of the developmental determination of sepals, petals, stamens, and carpels in floral development in *Arabidopsis?*

Guide to Problem Solving

Problem 1 Flies of *Drosophila melanogaster* that are mutant for the gene *eyeless* are almost completely devoid of eye tissue in the head region normally occupied by the compound eyes. A deletion of *eyeless,* when homozygous, produces the eyeless phenotype. Transgenic flies that carry the wildtype *eyeless* gene under the control of a promoter that causes expression in abnormal body parts (called *ectopic expression*) develop misplaced eye tissue in these body parts. Some of the phenotypes are shown in the figure below. Part A shows the result of ectopic expression in the entire head; eyes develop in the normal place, but also eye tissue develops instead of the antennae. Other results of ectopic expression are shown in parts B through D. Part B shows a wing with eye tissue growing out from it, part C a single antenna in which most of the third segment consists of eye tissue, and part D a middle leg with an eye outgrowth at the base of the tibia. What do these results imply about the role of the *eyeless* gene in eye development?

Answer The key feature of this situation is that loss-of-function mutants and gain-of-function mutants have opposite phenotypes. A deletion of *eyeless* is a loss-of-function mutation, and normal eye development is completely absent. The transgenic flies with ectopic expression are the equivalent of gain-of-function mutants, and they develop eyes in the ectopic locations. The inference from these results is that *eyeless* is a key regulatory gene in eye development. The expression of *eyeless* is *necessary* for eye tissue development and also *sufficient* for eye tissue development. (Interestingly, the human genome contains a homolog of *eyeless,* which also plays a critical role in the specification of eye development.)



[Photos courtesy of G. Halder and W. J. Gehring. From G. Halder, P. Callaerts, and W. J. Gehring. 1995. *Science* 267: 1788.]

Problem 1

Problem 2: The *bicoid* mutation in *Drosophila* results in absence of head structures in the early embryo. A similar phenotype results when cytoplasm is removed from the anterior end of a *Drosophila* embryo. What developmental effect would you expect in each of the following circumstances?

(a) A *bicoid* embryo was injected in the anterior end with some cytoplasm taken from the anterior end of a wildtype embryo.

(b) A wildtype *Drosophila* embryo was injected in the middle with some cytoplasm taken from the anterior end of another wildtype embryo.

Answer The experimental results imply that the anterior cytoplasm, including the *bicoid* gene product, induces development of head structures.

(a) Anterior cytoplasm from a wildtype embryo should be able to rescue *bicoid*, so head structures should be formed.

(b) Anterior cytoplasm injected into the middle of the embryo should induce head structures in that location, so the embryo should have head structures at the anterior end and also in the middle.

Problem 3 Consider a hypothetical mutant protease that affects floral development in *Arabidopsis thaliana*. The protease has an altered substrate specificity that enables it to cleave and inactivate both Ap1 and Ag proteins (the products of *ap1* and *ag*, respectively). Considering that tissue containing the Ap3/Pi dimeric protein, but neither Ap1 nor Ag, alone develops into floral organs intermediate between petals and stamens, what floral phenotype would be expected in the protease mutant?

Answer In whorl 1, Ap1 activity is missing, so this region will develop as a whorl of leaves. Likewise, in whorl 4, Ag activity is missing, so this region will develop as a whorl of leaves. In whorls 2 and 3, only Ap3/Pi is present, so these will develop as whorls of tissue intermediate between petals and stamens. Therefore, the flower phenotype will be leaves in whorls in 1 and 4 and will be petal/stamens intermediates in whorls 2 and 3.

Analysis and Applications

14.1 Distinguish between a developmental fate determined by autonomous development and one determined by positional information. What two types of surgical manipulation are used to distinguish between the processes experimentally?

14.2 What are lineage mutations and how are they detected?

14.3 What is the *Drosophila* blastoderm, and what is the blastoderm fate map? Does the existence of a fate map imply that *Drosophila* developmental decisions are autonomous?

14.4 What is the result of a maternal-effect lethal allele in *Drosophila*? If an allele is a maternal-effect lethal, how can a fly be homozygous for it?

14.5 How would you determine whether two independently isolated maternal-effect lethal mutations were alleles of the same gene?

14.6 What is the principal consequence of a failure in programmed cell death?

14.7 With regard to its effects on cell lineages, what kind of mutation is a homeotic mutation in *Drosophila*?

14.8 Can the key genes involved in the specification of floral organ identity in *Arabidopsis* be regarded as homeotic genes? Explain your answer.

14.9 A cell, A, in *Caenorhabditis elegans* normally divides and the daughter cells differentiate into cell types B and C. What developmental pattern would result from a mutation in cell A that prevents sister-cell differentiation? From a mutation in cell A that prevents parent-offspring segregation?

14.10 A mutation is found in which the developmental pattern is normal but slow. Does this qualify as a heterochronic mutation? Explain.

14.11 Why is transcription of the zygote nucleus dispensable in *Drosophila* early development but not in early development of the mouse?

14.12 Distinguish between a loss-of-function mutation and a gain-of-function mutation. Can the same gene undergo both types of mutations? Can the same allele have both types of effects?

14.13 A particular gene is necessary, but not sufficient, for a certain developmental fate. What is the expected phenotype of a loss-of-function mutation in the gene? Is the allele expected to be dominant or recessive?

14.14 A particular gene is sufficient for a certain developmental fate. What is the expected phenotype of a gain-of-function mutation in the gene? Is the allele expected to be dominant or recessive?

14.15 What is the phenotype of a *Drosophila* mutation in the *gap* class? of a mutation in the *pair-rule* class?

GENETICS on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by-chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- A very effective way to appreciate the complex, interactive hierarchy of gene activation and repression in *Drosophila* development is to explore the **interactive fly** at this keyword site. Created by Tom Brody and Judith Brody, this resource enables you to learn more about the stages and time course of developmental processes—and much more about the genes that are involved—all richly illustrated with drawings and photographs.

- The nematode worm *C. elegans* was the first multicellular organism whose complete developmental program of cell lineages was described; it was also the first multicellular organism whose genome was completely sequenced. You can explore the **lineage diagram** at this keyword site and the **worm proteome** at this keyword site. The challenge for developmental genetics is to be able to understand in detail the relationships between the constituents of the proteome and the patterns in the lineage diagram.

- Check out this site to learn about other genes involved in **flower development** in *Arabodopsis* and to see pictures of the mutant phenotypes. You can also learn more about the family of over 30 *Arabidopsis* genes known as MADS-box genes, which encode transcription factors, many of which are important in flower development.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 14 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 14.



14.16 What floral phenotype would be expected of a mutation resulting in a loss-of-function of *ap3* and *pi*? a gain-of-function mutation in which both *ap3* and *pi* were expressed in whorls 1 and 4 as well as whorls 2 and 3? What do these results imply about Ap3/Pi being necessary or sufficient for the developmental fate of the four whorls?

14.17 What floral phenotype would be expected of a mutation in *Arabidopsis* whose function was necessary for Ap3/Pi and Ag to exert their effects on development?

14.18 What phenotype would be expected of a gain-of-function mutation in *Arabidopsis* that resulted in expression of Ap3 and Pi in whorl 1?

14.19 The drug actinomycin D prevents RNA transcription but has little direct effect on protein synthesis. When fertilized sea urchin eggs are immersed in a solution of the drug, development proceeds to the blastula stage, but gastrulation does not take place. How would you interpret this finding?

14.20 A mutation in the axolotl designated *o* is a maternal-effect lethal because embryos from *oo* females die at gastrulation, irrespective of their own genotype. However, the embryos can be rescued by injecting oocytes from *oo* females with an extract of nuclei from either o^+o^+ or o^+o eggs. Injection of cytoplasm is not as effective. Suggest an explanation for these results.

14.21 The nuclei of brain cells in the adult frog normally do not synthesize DNA or undergo mitosis. However, when transplanted into developing oocytes, the brain cell nuclei behave as follows: (a) In rapidly growing pre-

meiosis oocytes, they synthesize RNA. (b) In more mature oocytes, they do not synthesize DNA or RNA, but their chromosomes condense and they begin meiosis. How would you explain these results?

14.22 The mRNA from the *bicoid* gene includes a sequence at the 3' end that results in its being localized at the anterior end of the egg. What phenotype would you search for in a mutant screen for genes necessary for this localization?

14.23 What is the expected floral structure arising from each whorl in an *Arabidopsis* mutant in which *ap1* (*apetala-1*) is restricted to whorl 2? (*Hint:* The default state is "leaf.")

14.24 What is the expected floral structure arising from each whorl in an *Arabidopsis* mutant in which *ag* (*agamous*) is restricted to whorl 3? (*Hint:* The default state is "leaf.")

14.25 Explain why, in accounting for the phenotypes of floral mutants in *Arabidopsis*, it was necessary to postulate that:

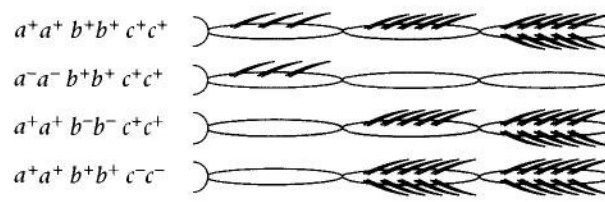
(a) In *agamous* mutants, the domain of expression of *apetala-1* expands to whorls 3 and 4.

(b) In *apetala-1* mutants, the domain of expression of *agamous* expands to whorls 1 and 2.

14.26 If *agamous* expression did not expand into whorls 1 and 2 in *apetala-1* mutants, could the phenotype of whorls 1 and 2 be predicted? Why or why not?

14.27 The autosomal gene *rosy* (*ry*) in *Drosophila* is the structural gene for the enzyme xanthine dehydrogenase (XDH), which is necessary for wildtype eye pigmentation. Flies of genotype *ry*/*ry* lack XDH activity and have rosy eyes. The X-linked gene *maroonlike* (*mal*) is also necessary for XDH activity, and *mal*/*mal*; ry^+/ry^+ females and *mal* Y; ry^+/ry^+ males also lack XDH activity; they have maroon-like eyes. The cross mal^+/*mal*; *ry*/*ry* × *mal* Y; ry^+/ry^+ produces *mal*/*mal*; ry^+/*ry* females and *mal* Y; ry^+ *ry* males that have wildtype eye color even though they lack active XDH enzyme. Suggest an explanation.

14.28 Genetic analysis is being carried out on a specialized appendage of a marine invertebrate that is used in immobilizing prey. The wildtype appendage, shown at the top in the accompanying diagram, has three segments. The proximal segment, attached directly to the body, has a single row of sparse bristles; the medial segment has a



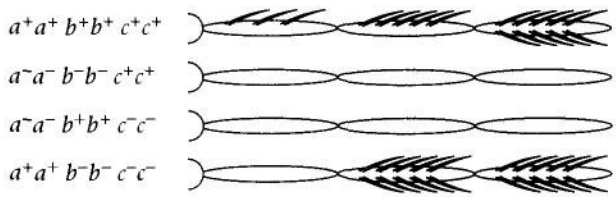
single row of dense bristles; and the distal segment has two symmetrical rows of dense bristles. Three mutations that affect the appendage are isolated; they cause the phenotypes shown above. The presence of a segment lacking bristles may be regarded as the default state of development of this appendage. Suggest a hypothesis to explain how the genes *a*, *b*, and *c* are involved in the developmental determination of this appendage.

14.29 To gain further insight into the genetic control of development of the appendage discussed in the previous problem, all possible double mutants are constructed. They have the phenotypes shown here.

(a) Are these phenotypes consistent with the inferences drawn from those of the single mutants?

(b) What additional information is provided by the double mutants?

(c) Propose a detailed model of which gene products are necessary for the development of each segment.



14.30 From the model for the genetic control of appendage development that emerges from data in the previous two problems, sketch the phenotype associated with each of the following types of mutations.

(a) A mutation in which the domain of expression of c^+ expands into the distal segment.

(b) A b^- mutation in which the domain of expression of a^+ expands into the proximal segment.

(c) An a^- mutation in which the domain of expression of b^+ expands into the medial segment.

Challenge Problems

14.31 You wish to demonstrate that during segmentation of the *Drosophila* embryo, normal pair-rule patterns of expression require normal expression of the gap-genes, whereas gap gene expression does not require pair-rule expression. You have the following four mutations available:

(a) A mutation in the zygotic-effect gap gene *knirps* (*kni*).

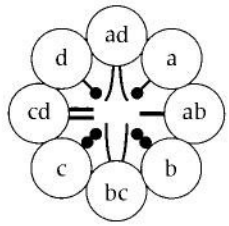
(b) A mutation in the zygotic-effect pair-rule gene *fushi tarazu* (*ftz*).

(c) A transgene consisting of a reporter gene (*lacZ*) fused to the enhancer elements of *kni*.

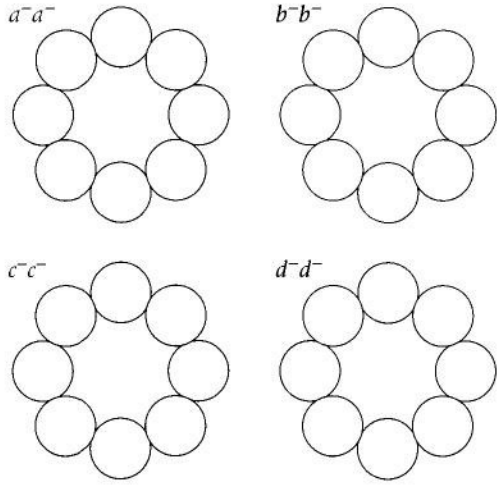
(d) A transgene consisting of a reporter gene (*lacZ*) fused to the enhancer elements of *ftz*.

Describe the strains you would need and how you would use them to show that *kni* is epistatic to *ftz*. You do not need to give details of the crosses.

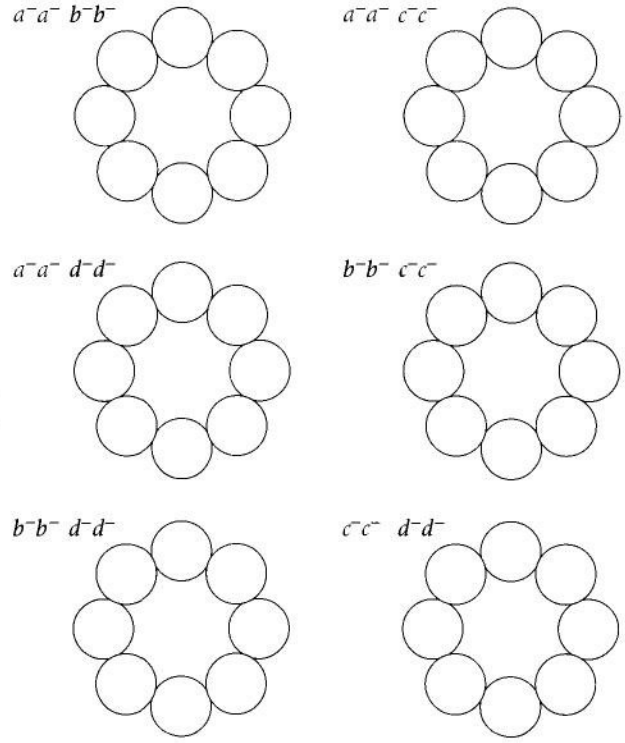
14.32 A type of jellyfish has a structure composed of eight circularly arranged cells from which projections emerge, forming a sort of pore. Whereas some of the cells have identical projections, others are quite different, as shown in the accompanying diagram. The letters a through d indicate the products of four developmental-control genes, *a* through *d*, that are found in each cell around the circle.



In the following diagrams, sketch the expected phenotype of each of the single mutants, a^- through d^-, assuming that the default state of a cell, with none of the gene products present, is the absence of projections.



14.33 For the developmental-control system described in the previous problem, sketch the expected phenotypes of each of the six possible types of double mutants:



Further Reading

Capecchi, M. R., ed. 1989. *The Molecular Genetics of Early Drosophila and Mouse Development*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Davis, G. K., and N. H. Patel. 1999. The origin and evolution of segmentation. *Trends in Biochemical Sciences* 24: M68.

De Robertis, E. M., G. Oliver, and C. V. E. Wright. 1990. Homeobox genes and the vertebrate body plan. *Scientific American*, July.

Duke, R. C., D. M. Ojcius, and J. D. E. Young. 1996. Cell suicide in health and disease. *Scientific American*, December.

Gaul, U., and H. Jäckle. 1990. Role of gap genes in early *Drosophila* development. *Advances in Genetics* 27: 239.

Grunert, S., and D. St. Johnston. 1996. RNA localization and the development of asymmetry during *Drosophila* oogenesis. *Current Opinion in Genetics & Development* 6: 395.

Kaufman, T. C., M. A. Seeger, and G. Olsen. 1990. Molecular and genetic organization of the *Antennapedia* gene complex of *Drosophila melanogaster*. *Advances in Genetics* 27: 309.

Kennison, J. A. 1995. The *Polycomb* and *Trithorax* group proteins of *Drosophila*: Trans-regulators of homeotic gene function. *Annual Review of Genetics* 29: 289.

Kornfeld, K. 1997. Vulval development in *Caenorhabditis elegans*. *Trends in Genetics* 13: 55.

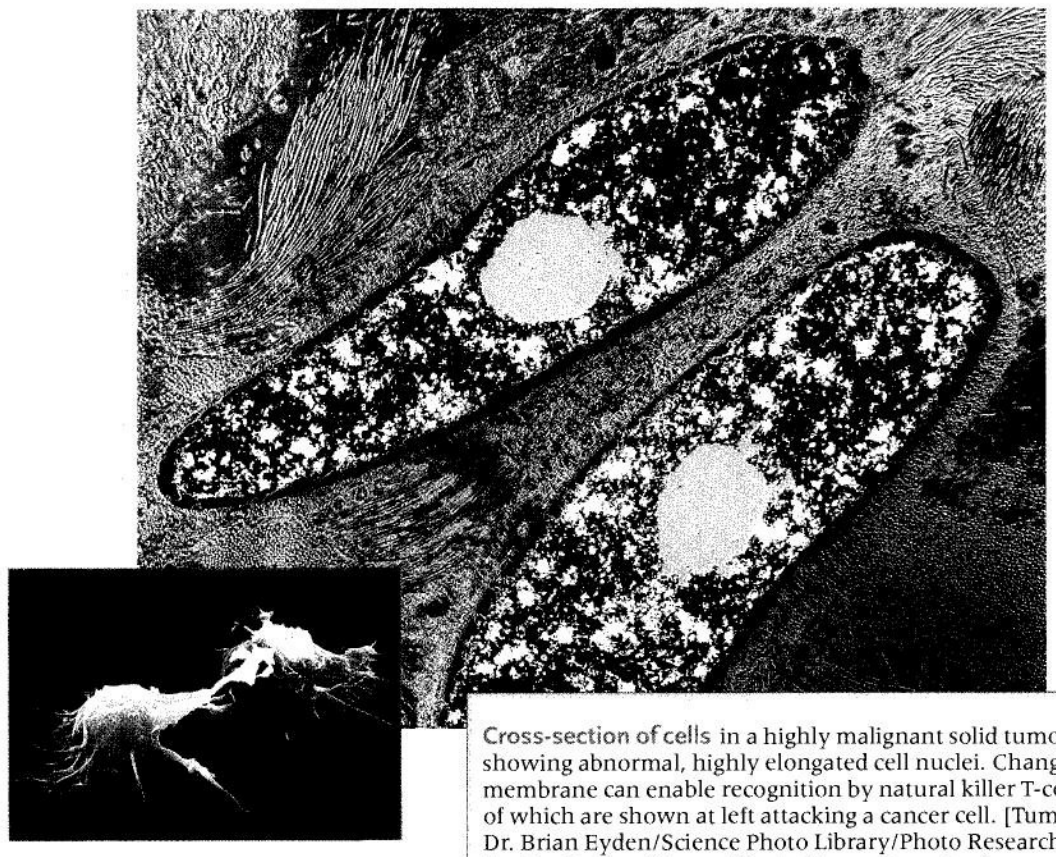
Lawrence, P. A. 1992. *The Making of a Fly: The Genetics of Animal Design*. Oxford, England: Blackwell.

McCall, K., and H. Steller. 1997. Facing death in the fly: Genetic analysis of apoptosis in *Drosophila*. *Trends in Genetics* 13: 222.

Meyerowitz, E. M. 1996. Plant development: Local control, global patterning. *Current Opinion in Genetics & Development* 6: 475.
Meyerowitz, E. M. 1999. Plants, animals and the logic of development. *Trends in Biochemical Sciences* 24: M65.
Nüsslein-Volhard, C. 1996. Gradients that organize embryo development. *Scientific American*, August.
Riverapomar, R., and H. Jäckle. 1996. From gradients to stripes in *Drosophila* embryogenesis: Filling in the gaps. *Trends in Genetics* 12: 478.
Sternberg, D. W. 1990. Genetic control of cell type and pattern formation in *Caenorhabditis elegans*. *Advances in Genetics* 27: 63.
Weigel, D. 1995. The genetics of flower development: From floral induction to ovule morphogenesis. *Annual Review of Genetics* 29: 19.
Wieschaus, E. 1996. Embryonic transcription and the control of developmental pathways. *Genetics* 142: 5.
Wolpert, L. 1996. One hundred years of positional information. *Trends in Genetics* 12: 359.
Wood, W. B., ed. 1988. *The Nematode* Caenorhabditis elegans. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
Wright, T. R. F., ed. 1990. *Genetic Regulatory Hierarchies in Development*. New York: Academic Press.

CHAPTER 15

Molecular Genetics of the Cell Cycle and Cancer



Cross-section of cells in a highly malignant solid tumor (above) showing abnormal, highly elongated cell nuclei. Changes in the membrane can enable recognition by natural killer T-cells, two of which are shown at left attacking a cancer cell. [Tumor cells: Dr. Brian Eyden/Science Photo Library/Photo Researchers, Inc.; T-cells: Meckes/Ottawa/Photo Researchers, Inc.]

CHAPTER OUTLINE

15.1 The Cell Cycle
- Key Events in the Cell Cycle
- Transcriptional Program of the Cell Cycle

15.2 Genetic Analysis of the Cell Cycle
- Mutations Affecting Progression Through the Cell Cycle

15.3 Progression Through the Cell Cycle
- Cyclins and Cyclin-Dependent Protein Kinases
- Targets of the Cyclin–CDK Complexes
- Triggers for the G_1/S and G_2/M Transitions
- Protein Degradation Helps Regulate the Cell Cycle

15.4 Checkpoints in the Cell Cycle
- The DNA Damage Checkpoint
- The Centrosome Duplication Checkpoint
- The Spindle Checkpoint

15.5 Cancer Cells
- Oncogenes and Proto-Oncogenes
- Tumor-Suppressor Genes

15.6 Hereditary Cancer Syndromes
- Defects in Cell-Cycle Regulation and Checkpoints
- Defects in DNA Repair

15.7 Genetics of the Acute Leukemias

PRINCIPLES

- About 10–15 percent of the genes in a yeast cell function in cell division and in the transition from one phase of the cell cycle to the next; their transcription is initiated once per cell cycle.
- Progression through the cell cycle is controlled by cyclin-dependent protein kinases and by protein degradation.
- Checkpoints monitor a dividing cell for DNA damage, cellular defects, other abnormalities in the cell cycle, and cell size. Detection of abnormalities elicits a response that arrests the cell cycle, allowing time for repair of defects (or cell death—apoptosis) and ensuring that the phases remain in correct order.
- Cancer cells show uncontrolled growth and proliferation and loss of contact inhibition.
- Progression from a normal to a cancerous state requires several genetic changes. Most cancers are sporadic (not inherited).
- A small proportion of cancers are associated with mutations transmitted through the germ line, which predispose the somatic cells of people carrying them to undergo cancer progression.
- The genetic changes that take place in cancer progression often involve defects or overexpression of genes that function in cell cycle regulation or checkpoint control.

CONNECTIONS

Cycle-Ops
Leland H. Hartwell, Joseph Culotti, John R. Pringle, and Brian J. Reid 1974
Genetic Control of the Cell Division Cycle in Yeast

Sick of Telomeres
William C. Hahn, Christopher M. Counter, Ante S. Lundberg, Roderick L. Beijersbergen, Mary W. Brooks, and Robert A. Weinberg 1999
Creation of Human Tumor Cells with Defined Genetic Elements

Two Hits, Two Errors
Alfred G. Knudson 1971
Mutation and Cancer: Statistical Study of Retinoblastoma

Cancer is a disease characterized by the uncontrolled proliferation of cells. The normal mechanisms that regulate cellular growth and division break down. Cancer is a genetic disease. It results from mutations that overcome the normal limits to the number of cell divisions that can take place before a cell dies. These mutations usually occur in somatic cells, and full-blown malignant cancer usually requires a number of sequential mutations to get started. But occasionally a mutation affecting cell-cycle regulation is inherited through the germ line, and persons who inherit the mutation have a greatly increased risk (sometimes approaching 100 percent) of developing malignancies due to additional somatic mutations. To understand cancer, therefore, one must understand the normal mechanisms that govern cellular proliferation. This is where we begin.

15.1 The Cell Cycle

The essential functions of the mitotic cell cycle are

1. To ensure that each chromosomal DNA molecule is replicated once and only once per cycle.
2. To ensure that the identical replicas of each chromosome (the sister chromatids) are distributed equally to the two daughter cells.

As we saw in Chapter 4, the cell cycle is divided into four phases: the *M (mitosis) phase,* in which the sister chromatids are physically separated into the two daughter nuclei, and a three-part *interphase* composed of G_1 *(gap 1), S (DNA synthesis),* and G_2 *(gap 2),* which occur in that order. The cell cycle may thus be illustrated as shown in Figure 15.1 for a mammalian cell. In this representation, *cytokinesis* (the division of the cytoplasm into two approximately equal parts containing the daughter nuclei) is included in the M period.

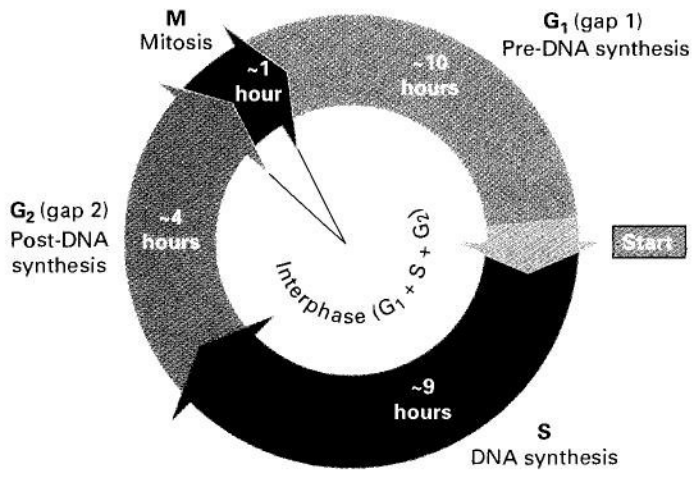


Figure 15.1 The cell cycle of a typical mammalian cell growing in tissue culture with a generation time of 24 hours.

Key Events in the Cell Cycle

Figure 15.2 shows a number of key events that must occur to ensure the proper duplication and distribution of the chromosomes. The spindle pole is organized around a small region of clear cytoplasm near the interphase nucleus called the **centrosome;** in many organisms this role is played by a pair of **centrioles,** which is more particulate in appearance. Both are microtubule organizing centers that must be duplicated and positioned. The duplication begins late in G_1 and is completed during S phase (in some cells, duplication may occur somewhat later). The duplicated poles then slowly begin to migrate to positions on opposite sides of the nucleus. Meantime, within the nucleus during the S phase, DNA replication takes place. Completion of DNA replication marks the beginning of G_2. Soon after the M phase commences, the centrosomes reach their final destinations and the chromosomes begin to condense. Each centrosome organizes one pole of the bipolar spindle by nucleating the formation of spindle and astral microtubules, and the condensed chromosomes become attached to spindle microtubules on both sides of the centromere (kinetochore) as the nuclear envelope breaks down. Each chromosome is thus physically attached to both spindle poles and is maneuvered to a position approximately halfway between them. As anaphase begins the spindle elongates, the centromeres separate, and the sister chromatids migrate toward opposite poles. Once the daughter chromosomes reach the poles, the spindle disassembles, the chromosomes decondense, and the nuclear membrane is formed again. These events return the cell to the G_1 phase.

The mechanisms by which cell growth is regulated and the cell cycle controlled have

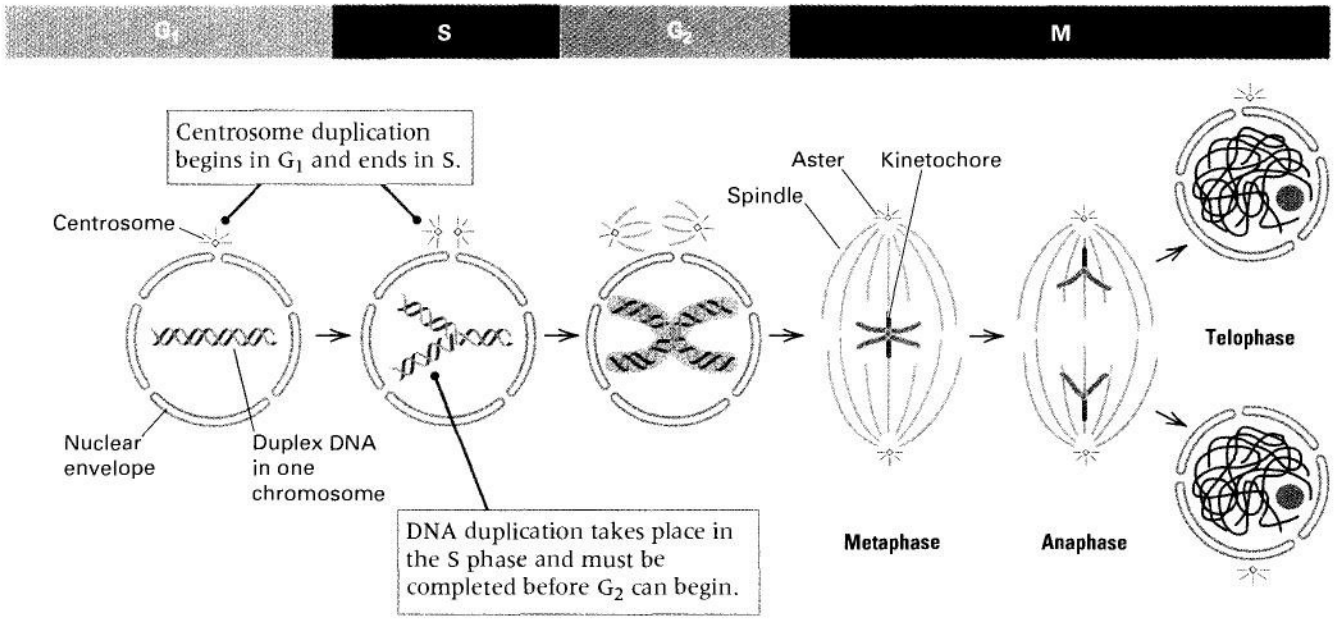


Figure 15.2 Major events in the cell cycle. In yeast, the spindle pole body serves the same function as the centrosome in many other organisms: Both are microtubule organizing centers from which the spindle emerges. In most other organisms, nuclear division takes place and the pinching off of cells (cytokinesis) follows. In yeast, the "shell" of the daughter cell forms and enlarges before nuclear division takes place. The nucleus (its membrane never breaks down) moves into the bridge between the mother and daughter cell, and nuclear division occurs there. After the daughter nuclei move into the two cell bodies, a septum is laid down between the two cells.[Adapted from L. H. Hartwell and M. Kastan. 1994. *Science* 266: 1821.]

been approached via the methods of biochemistry, cell biology, and genetics. The most extensive genetic studies have focused on two yeasts, the budding yeast (*Saccharomyces cerevisiae*) and the fission yeast (*Schizosaccharomyces pombe*). As its common name implies, *S. cerevisiae* multiplies by budding. The position of a cell within the cell cycle can be monitored visually by the size of the bud. This convenient feature of the cell cycle is depicted diagrammatically in **Figure 15.3**. The bud first emerges shortly after entry into S phase and grows throughout the cell cycle. Mitosis takes place *within* the nucleus without breakdown of the nuclear envelope. In this case the spindle poles, which are known as *spindle pole bodies*, are embedded within the nuclear envelope. To position the spindle, one spindle pole body moves while still within the nuclear envelope to a position exactly opposite the second spindle pole body across the way. By this time the nucleus already contains the duplicated chromosomes, and it moves to the neck of cytoplasm connecting the mother and daughter cell at G_2. Mitosis

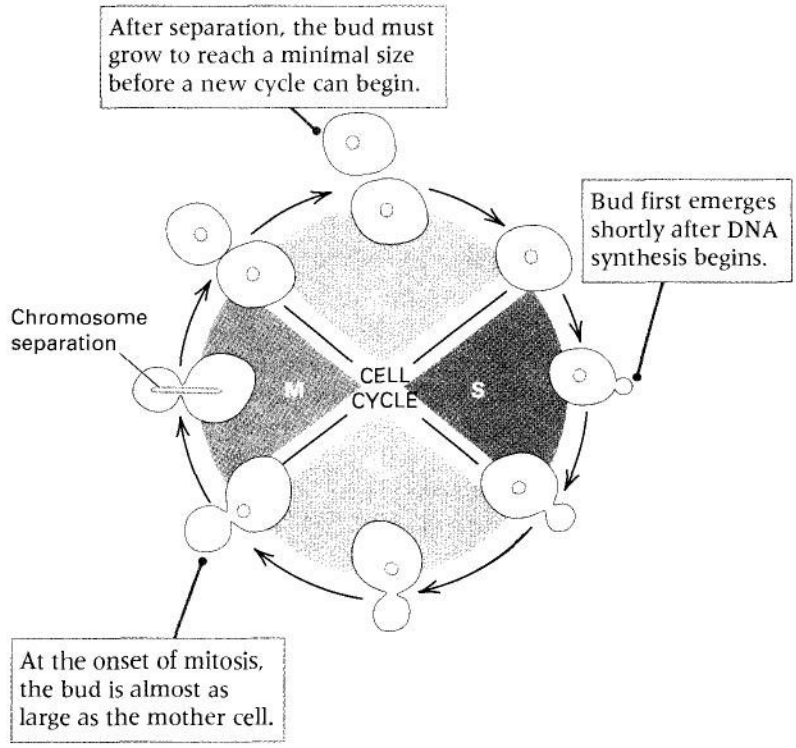


Figure 15.3 The cell cycle of budding yeast, *Saccharomyces cerevisiae*.

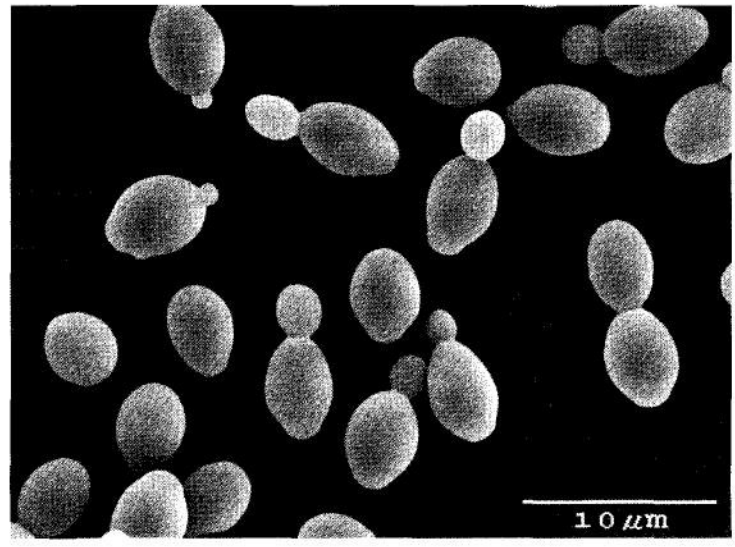


Figure 15.4 Scanning electron micrograph of cells of diploid budding yeast in various stages of the cell cycle. Bud size is correlated with the position of the cell within the cell cycle. Here the buds range in size from quite small to nearly as large as the mother cell.

occurs when the bud has grown to a size nearly as large as the mother cell. Shortly after chromosome separation, a barrier or *septum* is laid down between the mother and daughter cells. At cell separation, the daughter cell is only slightly smaller than the mother cell, but it typically must grow this extra bit to achieve the optimal size before it can start its own cell cycle. The various stages of bud growth from beginning to end can be identified in the electron micrograph in **Figure 15.4**.

Transcriptional Program of the Cell Cycle

In budding yeast, because the DNA sequence of the entire genome is known, it is possible to analyze the transcription patterns of all of the 6400 genes in the cell in a single experiment using the type of high-density gene arrays (DNA chips) discussed in Section 13.8 and illustrated in Figure 13.30 on page 579. Such experiments have shown that the transcript levels of about 800 genes vary in a periodic or cyclic pattern through the cell cycle. This result is dramatically illustrated in **Figure 15.5**. Each horizontal stripe represents the expression pattern of a single gene through two or more cell cycles in synchronized cells. (In a culture of *synchronized cells*, all cells are at the same stage in the cell cycle.) The colored bands across the top indicate the stage of the cell cycle (using the same colors for the stages as labeled on the right), and the designations Alpha, *cdc15*, and *cdc28* refer to different ways in which the cells were synchronized. For each gene, red indicates overexpression relative to the level of expression of the same gene in a nondividing cell, and green indicates underexpression. The genes have been grouped according to the similarity in their pattern of transcription through the cell cycle. Transcription of each of these 800 genes is initiated once per cell cycle, an event indicated by the red portion of the stripe as the cell cycle progresses. Some genes are transcribed in G_1, some in G_2, a few in S, and some in M. The genes are arranged in the order in which they are transcribed, and their principal stage of expression is indicated at the far right. Typically, genes encoding proteins that are needed in one part of the cycle are transcribed in the immediately preceding period. For example, enzymes needed for synthesis of the trinucleotide precursors of DNA and for DNA replication are made in G_1 immediately prior to their use in S phase. Similarly, the histone proteins are synthesized during S phase immediately prior to their incorporation into chromatin and their use in chromosome condensation.

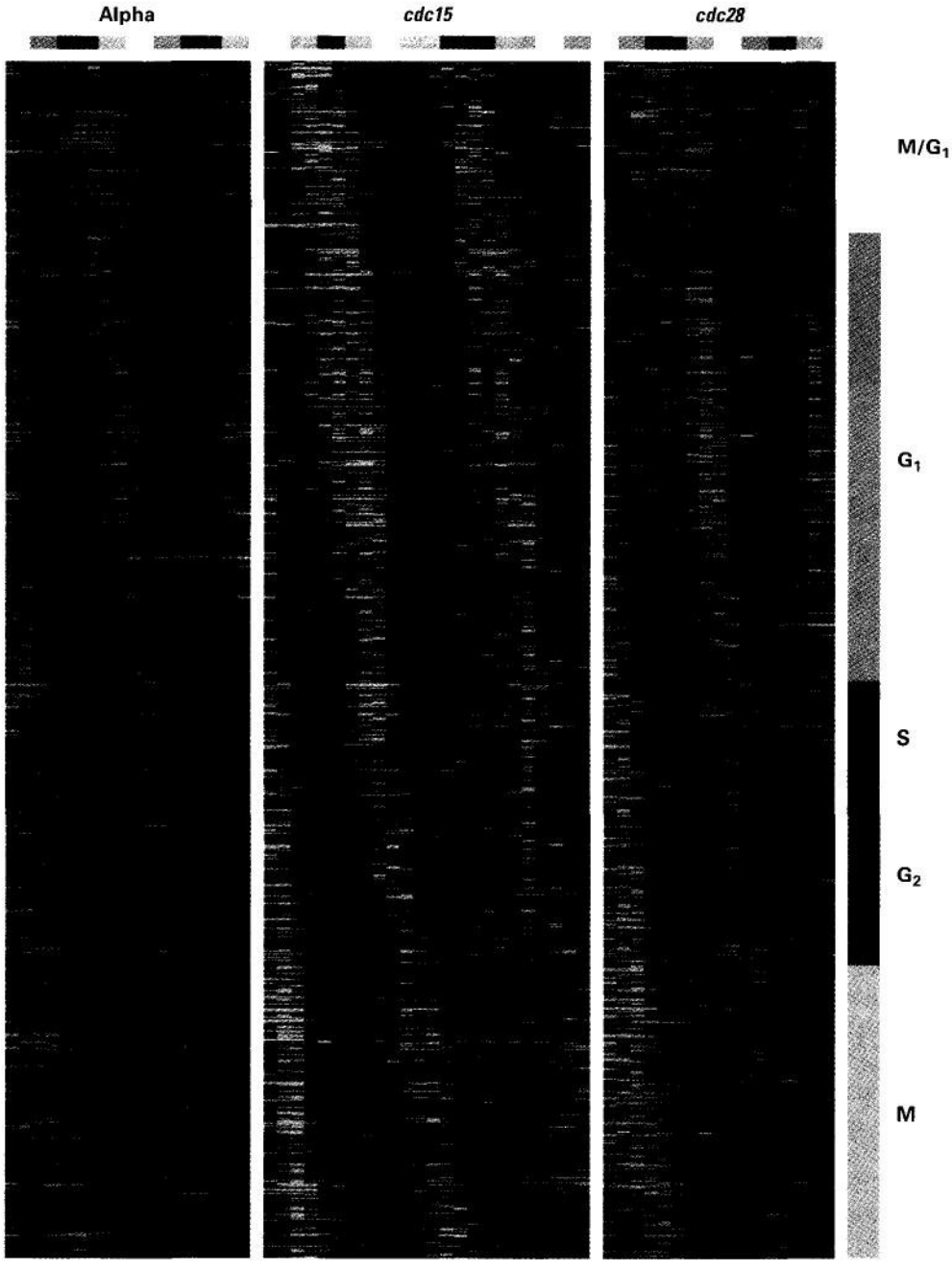
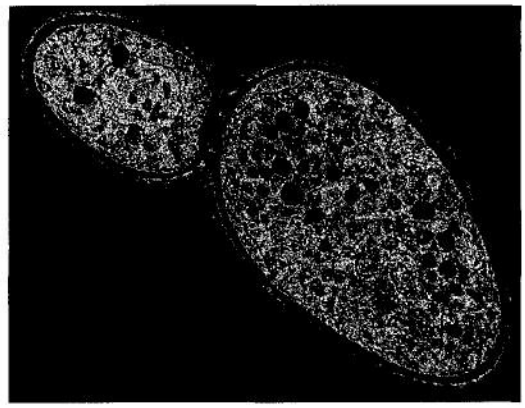


Figure 15.5 Expression patterns of about 800 yeast genes whose level of transcription varies systematically through the cell cycle. Each horizontal stripe indicates the transcription pattern of a different gene through the cell cycle. Green indicates underexpression relative to a nondividing cell, red overexpression. The color bars on the right indicate the phase in which each gene is maximally expressed: M/G_1 transition, G_1, S, G_2, or M. These same colors indicate the cell-cycle phase along the top. Alpha, *cdc15*, and *cdc28* indicate different ways in which the cell cycle has been synchronized. [Courtesy of Patrick O. Brown, from P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. 1998. *Molecular Biology of the Cell* 9: 3273.]



Cross-section of a budding yeast cell. In nuclear division, the spindle stretches through the neck connecting the mother cell and the daughter cell. [Biophoto/Photo Researchers, Inc.]

15.2 Genetic Analysis of the Cell Cycle

The ready assignment of a yeast cell to a position in the cell cycle on the basis of the relative sizes of the mother cell and its bud has made possible extensive genetic analyses of the cell cycle through the isolation and study of temperature-sensitive mutants. Temperature-sensitive **cell division cycle (*cdc*) mutants** are typically wildtype at 23°C (the *permissive* temperature) but unable to complete the cell cycle at 36°C (the *restrictive* temperature). At the higher temperature, mutant cells accumulate at a characteristic stage in the cell cycle. This is the stage at which their progression in the cell cycle is halted. The stage-specific stop is exceptional among mutations affecting cellular functions. For example, temperature-sensitive mutants with defects in protein synthesis do not cease growing abruptly when the temperature is raised. Each cell continues along until it runs out of functional proteins, and new proteins need to be synthesized for continued progression in the cell cycle. The stopping point differs from cell to cell, so the mutant cells stop growing at different stages in the cell cycle. In the microscope one sees unbudded cells, cells with small buds, and cells with large buds—essentially the same distribution one sees in an asynchronous cell population. But for each *cdc* mutant the stop is at a specific stage, which differs from one mutant to the next and is related to the function of the mutant gene product. Once it has been established where in the cell cycle a *cdc* mutant is blocked, further analysis can be used to determine whether specific processes can occur in the mutants, such as DNA synthesis, spindle formation, or formation of the division septum. In the next section we consider the analysis of two *cdc* mutants that stop, or **arrest,** at the G_2/M boundary.

Mutations Affecting Progression Through the Cell Cycle

The micrograph in **Figure 15.6**, part A, shows cells of a mutant designated *cdc13* grown in an unsynchronized culture at the permissive temperature of 23°C. The morphology and variation in size are normal; some unbudded cells are present, as well as dividing cells with buds ranging in size from very small to nearly as large as the mother cell. When this culture was placed at the restrictive temperature of 36°C for 6 hours, the cells took on the appearance shown in Figure 15.6B. Two types of configurations are seen. One consists of a pair of large cells. These are cells that were not dividing or had small buds at 23°C; after the temperature shift they continued in the cell cycle until they arrested at the large-bud stage. The other configuration consists of four large cells (*quartets*). These derive from cells that, prior to the temperature shift, were nearing the end of division and had large buds; after the temperature shift they completed their current cycle, and both daughter cells, still side-by-side, initiated one more cell cycle and arrested at the large-bud stage, yielding the quartet.

Figure 15.7 illustrates how these results may be interpreted. Single mutant cells that are transferred to 36°C prior to the time that the affected protein is needed arrest as large budded cells, because in the absence of functional protein they cannot proceed beyond this point. Cells that are transferred to 36°C when the protein is no longer needed in the current cycle (some time late in the S phase in this example) complete the current cycle and initiate a new cycle that is terminated when both daughter cells arrest with large buds because the affected protein is now unavailable. The *cdc13* mutant showing the pattern in Figure 15.6 has a defect in telomere processing, which occurs late in the S phase; when *cdc13* cells are grown at the restrictive temperature, abnormally extended single-stranded DNA is found at the telomeres.

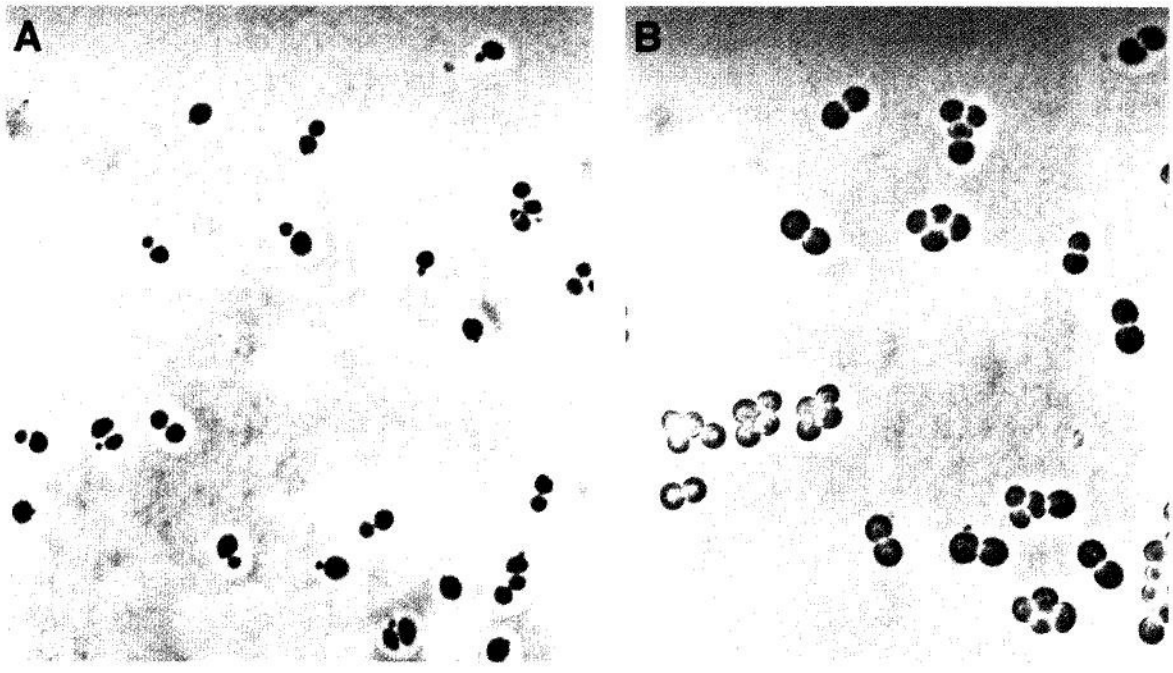


Figure 15.6 Cell-cycle arrest in temperature-sensitive *cdc13* mutants. (A) An asynchronous culture of a *cdc13* mutant grown at 23°C (the permissive temperature). The cells either are unbudded or have buds ranging in size from very small to nearly the size of the mother cell. (B) The same cells after incubation for 6 hours at 36°C (the restrictive temperature). Some cells are arrested with very large buds. Others have completed their current division and started another. These are the cells arrested in quartets. [From J. Culotti and L. H. Hartwell. 1971. *Experimental Cell Research* 67:389.]

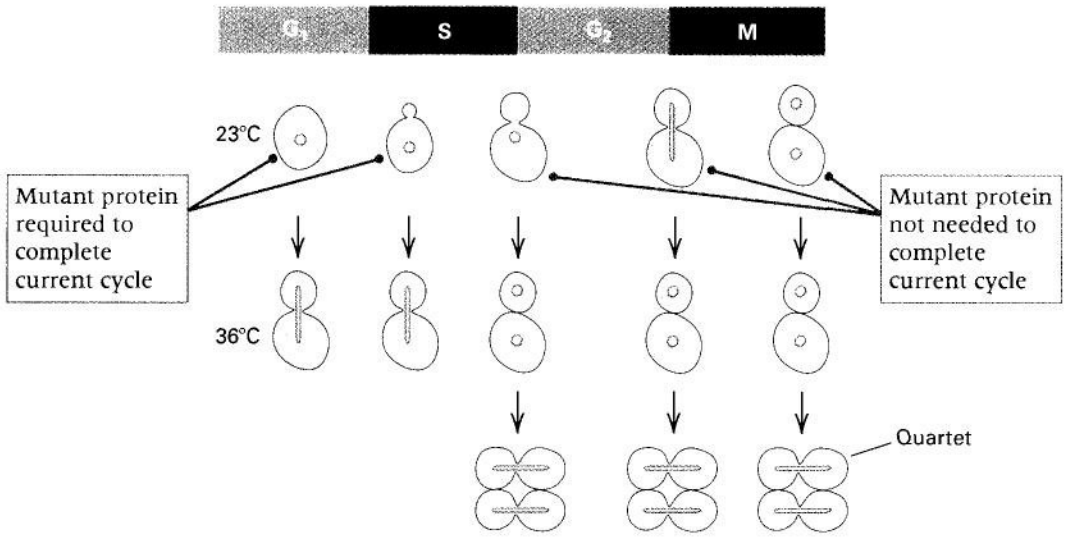


Figure 15.7 Determining the time of function of a cell-cycle protein by shift of a temperature-sensitive mutant to the restrictive temperature. Cells in a stage prior to when the defective protein is needed undergo arrest in the current division cycle at the stage when the defective protein is needed. These yield the doublet configurations. Cells that have passed the time when the protein is needed in the current cycle complete the current cycle, begin another, and arrest in this second cycle at the stage when the defective protein is needed. These yield the quartet configurations. [Adapted from L. H. Hartwell. 1991. *Accomplishments in Cancer Research*, p. 105 ff.]

Figure 15.8 shows the results of an experiment with another cell-division-cycle mutant; this one is designated *mob1-77*, which also arrests at the large-bud stage like the cells in Figure 15.5B. These data indicate that *mob1-77* cells, although they cannot complete the division process, can replicate their DNA at the restrictive temperature. At the permissive temperature (the upper graphs in Figure 15.8), asynchronous cultures of both wildtype and mutant cells include some cells with *1C* (haploid) and some cells with *2C* (diploid) DNA content, depending on whether they have completed DNA replication. At 36°C, the *mob1-77* mutant cells all become arrested at the large-bud stage with the *2C* DNA content characteristic of cells at the G_2/M interface. These and other experiments allow the mechanism of a mutation-induced block in the cell cycle to be progressively narrowed down. Ultimately, the protein affected and the nature of the mutation are identified via cloning and sequencing of the mutant allele. On the basis of this information, further experiments are designed to zero in on the protein's exact function.

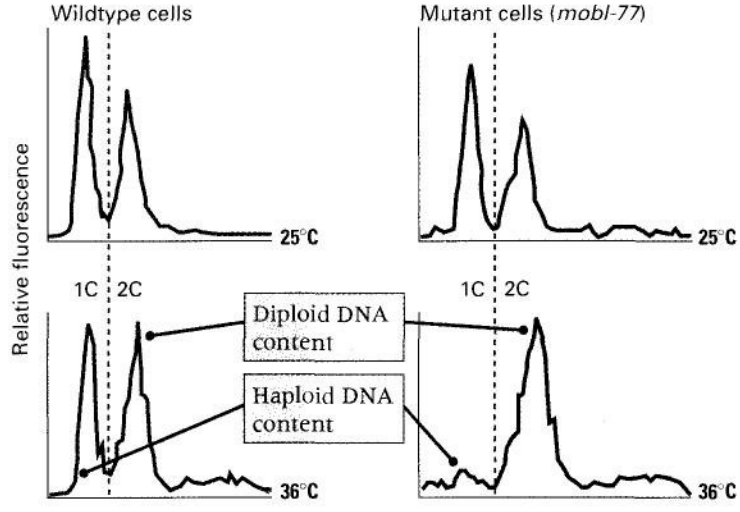


Figure 15.8 Replication of DNA in *cdc* mutant cells, starting with a population of haploid (*1C*) cells. At the permissive temperature (top panels), both mutant and wildtype cells have about 67 percent of cells with a haploid DNA content and 33 percent with a diploid DNA content; the latter have completed S phase but have not yet completed M. At the restrictive temperature (bottom panels), the mutant cells (right) accumulate with a diploid DNA content (*2C*), indicating that they are unable to complete the cell cycle. [After F. C. Luca and M. Winey. 1998. *Molecular Biology of the Cell* 9: 29.]

15.3 Progression Through the Cell Cycle

Much of the molecular machinery that regulates the cell cycle has been conserved in evolution. For this reason, our current understanding of the way the cell cycle works includes inferences made from studies of cells of organisms as diverse as yeasts, frogs, and human beings.

Cyclins and Cyclin-Dependent Protein Kinases

In the early stages of the cell cycle, progression from one phase to the next is controlled by characteristic protein complexes that are called **cyclin–CDK complexes** because they are composed of **cyclin** subunits combined with **cyclin-dependent protein kinase (CDK)** subunits (**Figure 15.9**). All eukaryotes regulate progression through the cell cycle by means of cyclin–CDK complexes, although the details of their structure and their mechanisms of action may differ slightly from one organism to another. Typically, a eukaryotic cell has a small *family* of genes whose individual members encode cyclins that function in specific portions of the cell cycle, or subprograms of differentiation, as shown in **Table 15.1**. In unicellular eukaryotes, a single CDK can interact with any of several different cyclins at different times in the cell cycle; this allows a single CDK to play multiple roles, because interaction with the cyclin determines the substrate specificity of the CDK. Such multiple roles are played by Cdc28 in budding yeast and by Cdc2 in fission yeast (Table 15.1). On the other hand, even unicellular eukaryotes can have more than one cyclin for a specific stage in the cell cycle; for example, cyclins Cln1, Cln2, and Cln3 all function in the G_1 stage in budding yeast (Table 15.1). Mammalian cells express different CDKs through the cell cycle, and some CDKs interact with different cyclins in different stages. For example, Cdk2 interacts with cyclin E in G_1 but with cyclin A in S (Table 15.1). Despite these differences, the function of some CDKs has been remarkably conserved in evolution. A striking illustration of this principle is the finding that coding sequences for the human proteins Cdk2 and Cdc2 are able

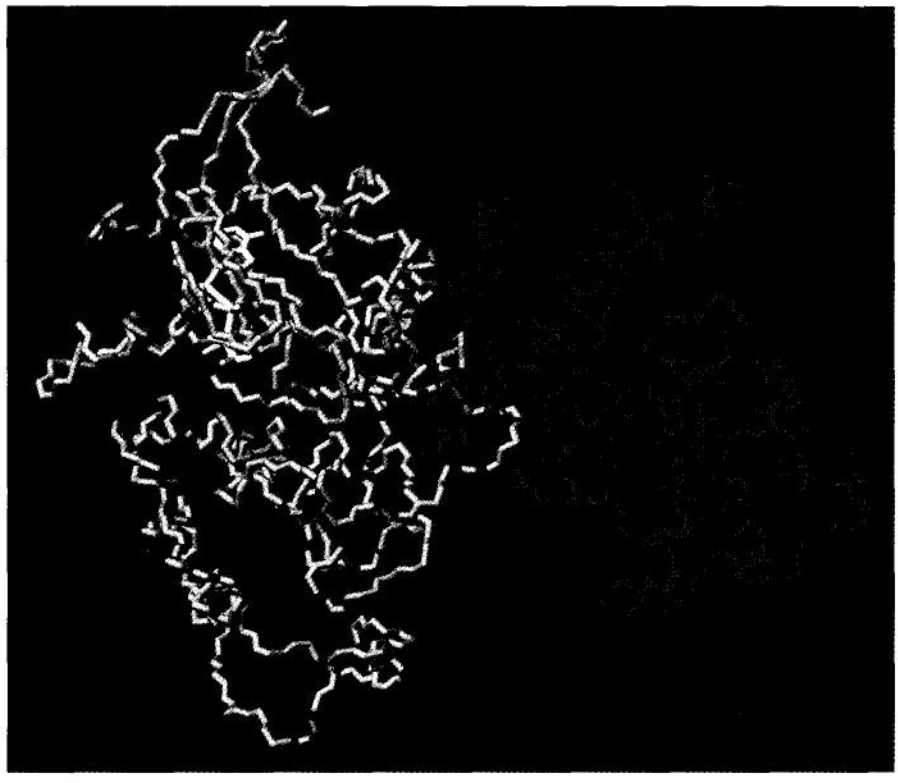


Figure 15.9 Structure of cyclin A (right) complexed with Cdk2 (left). Cyclin A binds to one side of the catalytic cleft in Cdk2, inducing a large conformational change that opens the active site and activates the kinase domain. The ATP (white) bound to Cdk2 is the phosphate donor in the kinase reaction. [Courtesy of Carlos Bustamante, from coordinate data of P. D. Jeffrey et al. 1995. *Nature* 376: 294.]

to substitute for the *CDC28* gene in budding yeast, and the human proteins allow the yeast cell to grow and mate essentially normally. Among all the enzyme complexes listed in Table 15.1, the mammalian cyclin B–Cdc2 complex that forms in G_2/M is especially noteworthy, because it controls the passage from G_2 into M. Even though the cyclin B–Cdc2 complex is assembled prior to mitosis, its catalytic activity is restricted to mitosis, because until then, it is inhibited by phosphorylation.

Table 15.1 Cyclin–CDK interactions in eukaryotic cells

	Stage of cell cycle					
	G_1 phase		S phase		G_2/M phase	
Organism	cyclin	CDK	cyclin	CDK	cyclin	CDK
Budding yeast	*Cln1, Cln2, Cln3*	*Cdc28*	*Clb5, Clb6*	*Cdc28*	*Clb2*	*Cdc28*
Fission yeast	*Cig1*	*Cdc2*	*Cig2*	*Cdc2*	*Cdc13*	*Cdc2*
Higher eukaryotes	*Cyclins D* *Cyclins E*	*Cdk4, Cdk6* *Cdk2*	*Cyclin A*	*Cdk2*	*Cyclin B*	*Cdc2 (Cdk1)*
	These regulate growth and morphogenesis		These regulate DNA replication		These regulate mitosis	

CoNnection
Cycle-Ops

Leland H. Hartwell, Joseph Culotti, John R. Pringle, and Brian J. Reid 1974
University of Washington,
Seattle, Washington
Genetic Control of the Cell Division Cycle in Yeast

This classic paper is one of the high points of genetic analysis. It shows how the phenotypes of mutants can be used to dissect a complex temporal process, in this case the cell cycle. The results demonstrate that the yeast cell cycle includes two parallel pathways emerging from a "start" that defines the point at which a cell becomes irreversibly committed to mitosis rather than to mating.

Two features which distinguish the cell cycle of *Saccharomyces cerevisiae* from most other eukaryotes are particularly useful for an analysis of the gene functions that control the cell division cycle. First, the fact that both haploid and diploid cells undergo mitosis permits the isolation of recessive mutations in haploids and their analysis by complementation in diploids. Second, the daughter cell is recognizable at an early stage of the cell cycle as a bud on the surface of the parent cell. Since the ratio of bud size to parent cell size increases progressively during the cycle, this ratio provides a visual marker of the position of the cell cycle. We have taken advantage of these features in the isolation and characterization of 150 temperature-sensitive mutants of the cell division cycle (*cdc* mutants). These mutants are temperature-sensitive in the sense that they are unable to reproduce at 36°C (the restrictive temperature) but do grow normally at 23°C (the permissive temperature). . . . These mutations define 32 genes, each of whose products plays an essential role in the completion of one event in the mitotic cycle. . . . The phenotypes of the *cdc* mutants suggest that the following events are ordered in a single dependent pathway: "start," initiation of DNA synthesis, DNA synthesis, medial nuclear division, late nuclear division, cytokinesis, and cell separation. Hence, the temporal sequence of these events is easily accounted for by the fact that no event in this pathway can occur without the prior occurrence of all preceding events. A second dependent pathway is comprised of the events "start," bud emergence, nuclear migration, cytokinesis, and cell separation. Thus, the temporal sequence of these five events is also assured. Furthermore, the integration of the two pathways is accomplished by the facts that both diverge from a common event, "start," and that both converge on a common event, cytokinesis.

These mutations define 32 genes, each of whose products plays an essential role in the completion of one event in the mitotic cycle.

Source: Science 183: 46–51

The term *cyclin* is apt because the abundance of these proteins changes cyclically in the cell cycle, and some are present only at specific times. Most cyclins appear abruptly and disappear a short time later (**Figure 15.10**). Each cyclin appears at its characteristic time in the cell cycle because its transcription is linked to the cell cycle via a previously expressed cyclin. Typically, the presence of an active cyclin–CDK complex results in the activation of a transcription factor that leads to transcription of the next cyclin needed in the cell cycle. The cyclin disappears when its gene is no longer transcribed and the previously made mRNA and protein are degraded. Active cyclin–CDK complexes also cause the transcriptional activation of genes other than cyclins.

Targets of the Cyclin–CDK Complexes

How do the cyclin–CDK complexes control progression through the cell cycle? Through their protein kinase activity, they add phosphate groups to the hydroxyl groups present in the amino acids serine, threonine, and tyrosine found in proteins. The cyclin component of the complex binds to the protein substrate and tethers it, allowing the CDK component to phosphorylate the tethered substrate. Once the targeted protein is phosphorylated, it dissociates from the cyclin–CDK complex. The activities of the phosphorylated form of a protein often differ dramatically from those of the unphosphorylated form. Phosphorylation may acti-

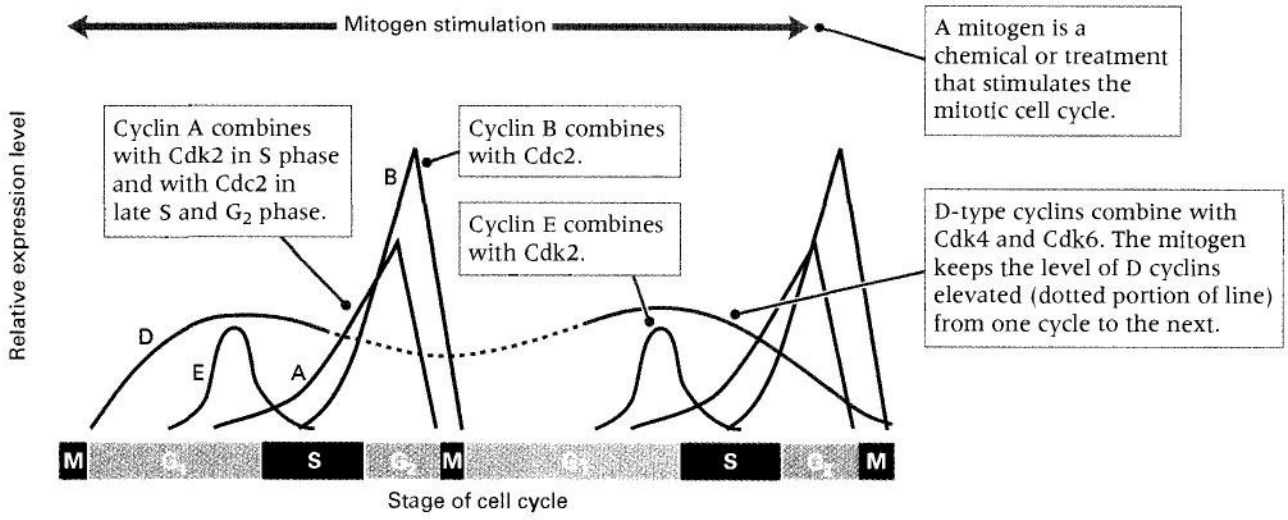


Figure 15.10 Fluctuations of cyclin levels during the cell cycle. Expression of cyclins E, A, and B (mitotic cyclins) are periodic, whereas cyclin D is expressed throughout the cell cycle in response to mitosis-stimulating drugs (mitogens). [Adapted from C. Sherr. 1996. *Science* 274: 1672.]

vate one enzyme but inactivate another. Even in a single enzyme molecule, phosphorylation at different sites may have opposite effects. Another route of cell-cycle regulation is mediated through *phosphatase* enzymes that dephosphorylate proteins that the cyclin-dependent kinases have phosphorylated. Reversing the effects of CDKs, the phosphatases activate enzymes that are inactive or deactivate ones that are active.

In mammalian cells, specific cyclin–CDK complexes can be detected at different times in the cell cycle, as illustrated in **Figure 15.11**. The cyclin D–Cdk4 and cyclin D–Cdk6 complexes appear in the early or middle part of G_1, whereas cyclin E–Cdk2 and cyclin A–Cdk2 appear later in G_1. The cyclin A–Cdk2 complex is present throughout the S phase and into M (mitosis), and the cyclin B–Cdc2 complex carries the cell through the G_2/M transition. The functional activity of the cyclin–CDK complexes is controlled in part by their state of phosphorylation, but the activity of cyclin D–Cdk4 and cyclin D–Cdk6 are also controlled by a protein designated the p16 inhibitor.

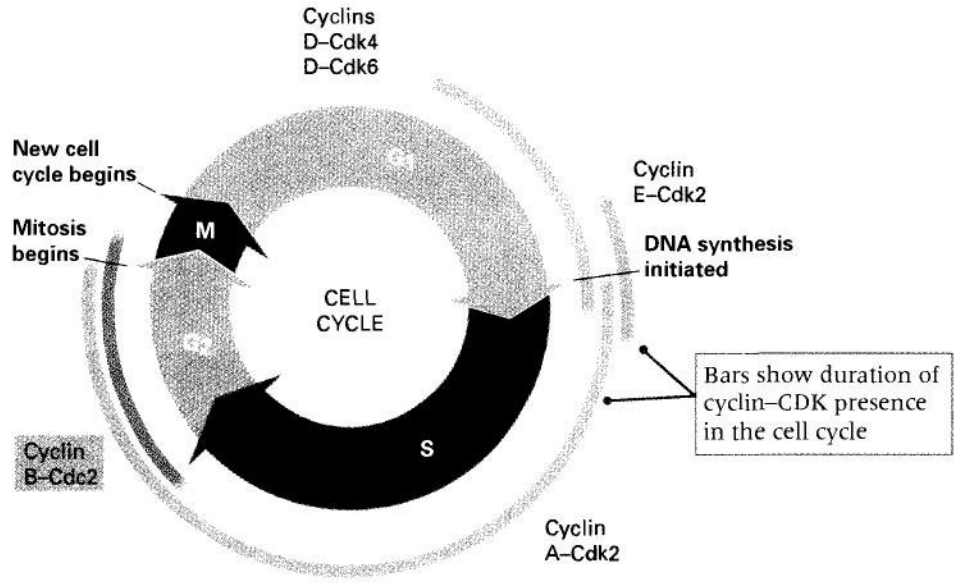


Figure 15.11 The temporal expression pattern of activities of the cyclin–CDKs in mammalian cells.

Triggers for the G_1/S and G_2/M Transitions

In the cell cycle, the cells monitor their internal and external conditions. Until the conditions are suitable to initiate a division cycle, the cells accumulate at a point in G_1 called the **G_1 restriction point** or *start*. In animal cells, a protein called the **retinoblastoma (RB) protein** holds cells at the restriction point by binding to and sequestering the transcription factor E2F, which is needed for further progression (**Figure 15.12**, part A). The RB protein was first identified in human pedigrees because mutant forms of the protein are associated with the formation of malignant tumors in the retina, which can require surgical removal of the eyes.

Growth is triggered by a growth factor combining with its receptor. This event activates a signal-transduction pathway that culminates in the production of cyclin D. In the middle of G_1 phase, the RB protein begins to be phosphorylated by both the cyclin D–Cdk4 kinase and the cyclin D–Cdk6 kinase; late in G_1 the RB phosphorylation is completed by the cyclin E–Cdk2 kinase as cells approach the **G_1/S transition** when they become committed to DNA replication. Phosphorylation of RB inactivates the protein and frees the bound E2F transcription factor (Figure 15.12A). Release of the E2F results in transcription of the genes and translation of the enzymes responsible for DNA replication, including DNA polymerase. The cells therefore begin to accumulate the precursors and enzymes re-

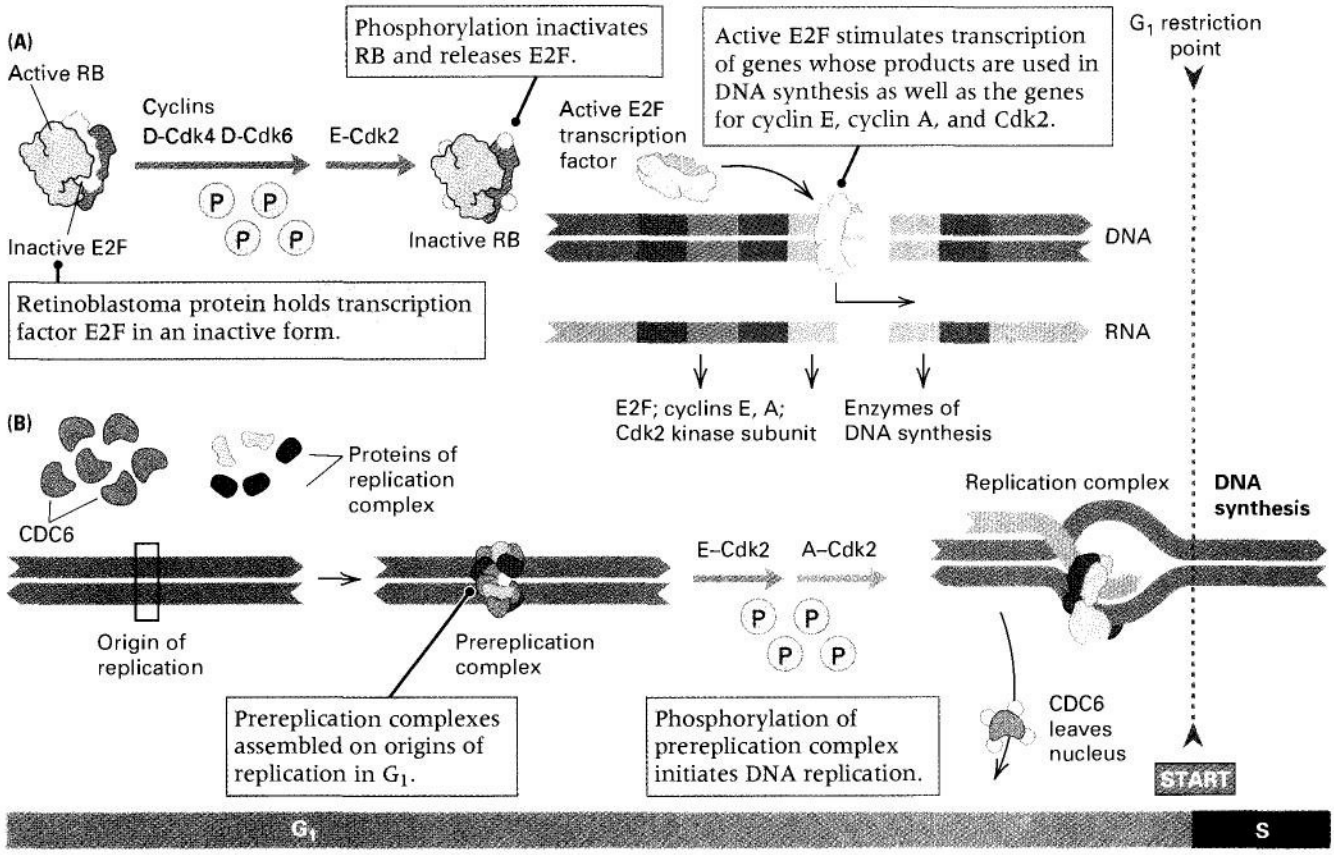


Figure 15.12 (A) Role of the retinoblastoma protein RB in controlling the transition from G_1 phase to S phase. The cyclin D-dependent kinases Cdk4 and Cdk6 initiate phosphorylation of RB in mid-G_1 phase, a process completed by cyclin E–Cdk2; this frees the transcription factor E2F, which activates transcription of enzymes for DNA synthesis. (B) The free E2F also activates transcription of the genes for cyclin A, cyclin E, and Cdc2, which help convert prereplication complexes to replication complexes for transition to S phase.

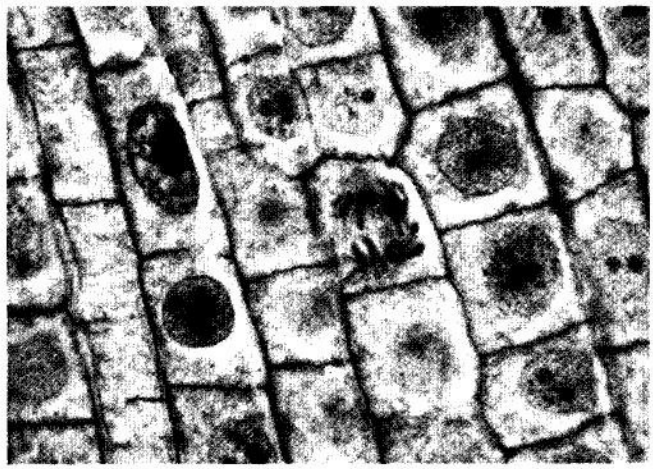
quired for DNA synthesis. E2F also activates transcription of the gene for E2F itself (an example of positive autoregulation), as well as those for the cyclins E and A and the Cdk2 kinase subunit.

Cells in G_1 also prepare for DNA synthesis by assembling **pre-replication complexes** at origins of replication. Assembly of each of the complexes is initiated by association of six proteins constituting the **origin recognition complex (ORC)** with an origin of replication in the DNA. CDC6 protein (named after a protein in yeast) attaches to each ORC and recruits other components, which completes the assembly of a pre-replication complex poised to initiate DNA synthesis. Cyclin E–Cdk2 kinase and cyclin A–Cdk2 kinase (Figure 15.12B) phosphorylate components of the complexes, including the CDC6 protein, that in turn trigger the initiation of DNA synthesis and thus the G_1/S transition. Once phosphorylated, CDC6 leaves the initiation complex (this event correlates with the actual initiation of DNA synthesis) and moves from the nucleus to the cytoplasm. The exit of CDC6 appears to ensure that DNA synthesis will not reinitiate at the same origin of replication. Once cells have entered S phase, the cyclin A–Cdk2 phosphorylates E2F and inhibits its binding to DNA, thus inactivating its function as a transcription factor. The cyclin A–Cdk2 activity is required throughout S phase, apparently to keep the RB protein heavily phosphorylated.

The progression from G_2 to M (the **G_2/M transition**) is controlled by a cyclin B–Cdc2 complex also known as *maturation-promoting factor.* After synthesis and assembly in the cytoplasm, the cyclin B–Cdc2 complex remains in the cytoplasm in an inactive form, because its rate of import into the nucleus is smaller than its rate of export from the nucleus. The balance is tipped in favor of import when cyclin B is phosphorylated just prior to the G_2/M transition, which masks the nuclear export signal. Reducing nuclear export ensures that cyclin B–Cdc2 accumulates in the nucleus. The cyclin B–Cdc2 complex is activated by dephosphorylation at different sites, and it then carries out phosphorylation of its substrates to bring about events that complete the G_2/M transition. These include (1) phosphorylation of a kinesin-related motor protein associated with duplication of the spindle pole and formation of the bipolar spindle and (2) phosphorylation of nuclear lamins, causing their disassembly and ultimately the breakdown of the nuclear envelope. At this time the chromosomes condense and assemble onto the spindle, and the chromosome segregation machinery becomes active.

Protein Degradation Helps Regulate the Cell Cycle

A fundamental feature of the cell cycle is that it is a true cycle: It is not reversible. The cycle is propelled forward by a process of protein degradation that complements the periodic activation of cyclin–CDK complexes. Protein degradation (*proteolysis*) eliminates proteins that were used in the preceding phase as well as proteins that would inhibit progression into the next phase. In the early stages of the cell cycle, progression requires the sequential activation of cyclin–CDKs. Entry into each new phase also requires destruction of the cyclins used in the preceding phase. In the later stages of the cell cycle, progression is propelled by proteolysis alone. This process is best understood in yeast, and we describe it here using the yeast terminology.



Cells in the root tip of an onion plant showing various stages in mitosis. Protein degradation is an important mechanism for propelling a cell from one stage of nuclear division to the next. [J. L. Carson/Custom Medical Stock Photo.]

In the completion of mitosis and the return to G_1 phase, two key regulatory events must occur:

1. The sister chromatids must separate (marking the onset of anaphase).
2. The cells must exit from mitosis, which entails chromosome decondensation, spindle disassembly, inactivation of the chromosome segregation machinery, and cytokinesis.

Both of these key events are triggered by protein degradation, as indicated in **Figure 15.13** for yeast. Exit from mitosis requires the destruction of cyclin B (yeast Clb). Cyclin B is most abundant when cells enter mitosis, but it disappears after chromosome disjunction has occurred at the transition from metaphase to anaphase. Cyclin B is marked for destruction by the **anaphase-promoting complex (APC/C)**, which is a ubiquitin–protein ligase responsible for ubiquitinating its target proteins. In Chapter 1 we explained how ubiquitin, a 76-residue protein, is transferred to certain abnormal proteins to mark them for degradation into amino acids through digestion in a *proteasome*—a large, multi-functional, multi-subunit complex responsible for most of the cytoplasmic proteolysis in the cell beyond that which takes place in lysosomes. The same process is used to mark certain proteins for destruction when their function is no longer needed. Among the substrates for anaphase-promoting complex is cyclin B, which when ubiquitinated is marked for degradation via the proteasome. Another substrate is the protein Pds1p, which stabilizes the protein Scc1p that forms the "glue" holding sister chromatids together. (Scc1p is a member of the cohesion complex that condenses chromosomes.) Destruction of Pds1p makes possible the breakdown of Scc1p, which allows the sister chromatids to come apart during anaphase. Thus the metaphase/anaphase and mitosis/G_1 transitions do not require cyclin-dependent protein kinases, but rather the activity of the anaphase-promoting complex to target specific proteins for degradation.

15.4 Checkpoints in the Cell Cycle

Cells monitor their external environment as well as their internal physiological state and functions. In the absence of needed nutrients or growth factors, animal cells may exit from the cell cycle and become quiescent in a state called G_0. Upon growth

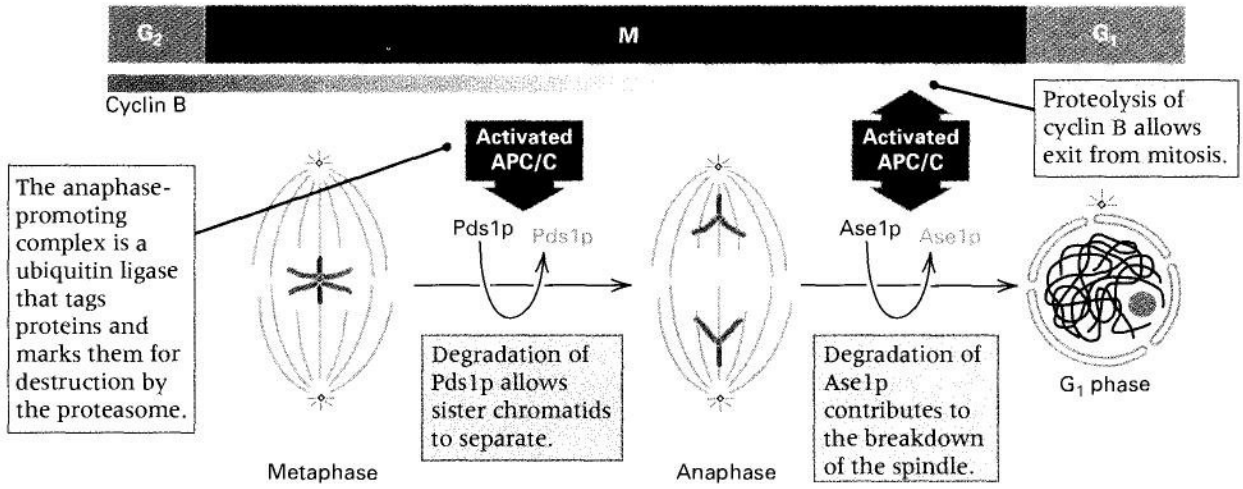


Figure 15.13 Role of activated anaphase-promoting complex (APC/C) in controlling the proteolysis necessary for the transition to anaphase and the exit from mitosis. APC/C is a ubiquitin ligase that marks proteins for proteasome destruction. Its substrates include Pds1p, which unless destroyed inhibits breakdown of the protein "glue" (Scc1p) that holds the sister chromatids together. Destruction of cyclin B somewhat later allows the exit from mitosis.

stimulation, they reenter the cell cycle in a process that requires cyclin D–CDK activity. Cells also have mechanisms that respond to symptoms of stress, including DNA damage, oxygen depletion, inadequate pools of nucleoside triphosphates, and (in the case of animal cells) loss of intercellular adhesion. Inside the cell, several key events in the cell cycle are monitored. When defects are identified, progression through the cell cycle is halted at a **checkpoint.** By allowing time for correction and repair, the checkpoints serve to maintain the correct order of steps with respect to each another in the cell cycle. Three principal checkpoints are

1. A DNA damage checkpoint
2. A centrosome duplication checkpoint
3. A spindle checkpoint

These checkpoints are summarized in **Figure 15.14**. All three types of checkpoints are important in maintaining the stability of the chromosome complement. When there is an error in any of the three processes monitored by the checkpoints, failure to stop at the checkpoint may lead to aneuploidy, polyploidy, or an increase in mutation rate. In the following sections, we shall examine how some of these checkpoints work.

The DNA Damage Checkpoint

A **DNA damage checkpoint** arrests the cell cycle when DNA is damaged or replication is not completed. DNA damage includes either modification of nitrogenous bases or breakage of the phosphodiester backbone. When modified nucleotides are repaired by an excision repair pathway (Chapter 7), the repair entails removal of the affected nucleotides, followed by resynthesis with a repair polymerase and finally ligation. DNA molecules broken in the backbone are repaired by homology-based recombination, nonhomologous end joining, or addition of new telomeres. In animal cells, a DNA damage checkpoint acts at three stages in the cell cycle: at the G_1/S

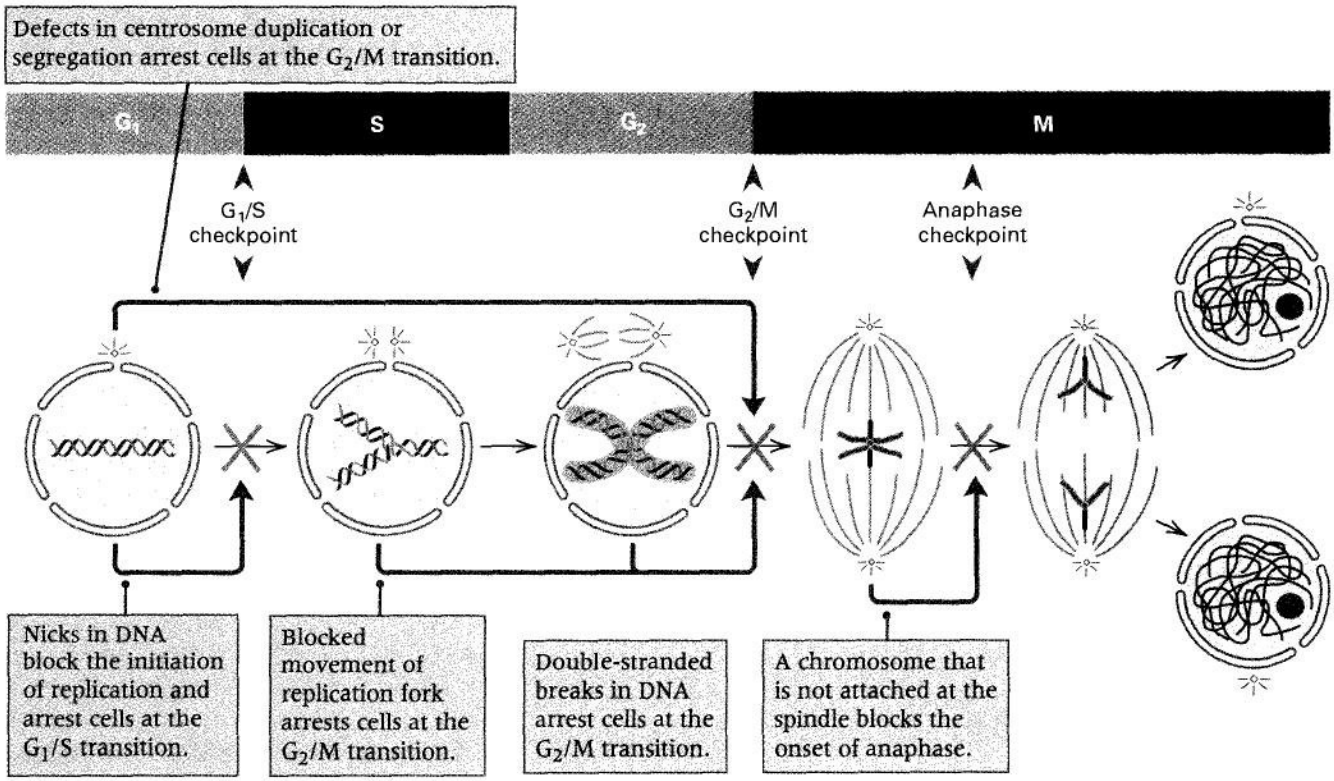


Figure 15.14 Cell-cycle checkpoints. The text in the boxes explains the events monitored and the steps affected. [Adapted from L. H. Hartwell and M. Kastan. 1994. *Science* 266: 1821.]

transition, in S phase, and at the G_2/M boundary. The S-phase checkpoint continuously monitors the progression of DNA synthesis throughout S. Although there are three DNA damage checkpoints, each monitors DNA damage or incomplete replication. If either type of problem is detected, the DNA checkpoint acts to block the cell cycle at multiple points.

A key protein in the mammalian cell's response to stress in general, and to DNA damage in particular, is a protein called the **p53 transcription factor.** In normal cells, the level of activated p53 protein is very low, even though substantial amounts of p53 mRNA and protein are present. The activity of p53 is kept low by another protein called Mdm2. To function as a transcription factor, p53 must be activated first by phosphorylation and then by acetylation. Mdm2 binds to p53 and prevents phosphorylation and subsequent steps in the activation of p53 as a transcription factor. In addition, Mdm2 continuously shuttles between the nucleus and the cytoplasm, and in this process it continuously exports p53 from the nucleus for degradation by the proteasome in the cytoplasm. The p53–Mdm2 export cycle is illustrated in **Figure 15.15**.

Table 15.2 Products of genes transcriptionally activated by p53

Gene product	Function
p21	Inhibits several cyclin-dependent kinases; arrests cells at G_1/S boundary.
14-3-3σ	Predicted to bind to and sequester phosphorylated Cdc25C phosphatase in the cytoplasm, which prevents Cdc25C from activating the cyclin B–Cdc2 kinase; arrests cells at the G_2/M boundary.
GADD45	Binds to proliferating cell nuclear antigen (PCNA), blocking its role as a processivity factor for DNA polymerase and hence blocking DNA replication; functions directly in DNA repair.
Bax	Acts as a positive regulator of apoptosis (programmed cell death).

When cells are treated with agents that damage DNA, some cells arrest in G_1 and others arrest in G_2. The DNA damage signal is sensed and transmitted, and this causes p53 to become activated by protein kinases and acetylases, overriding the inhibitory effect of Mdm2. Activation of p53 causes its release from Mdm2, which stabilizes the active p53 transcription factor and results in increased levels of the p53 protein. The activated p53 then triggers the transcription of a number of genes and the repression of others. Some of the key proteins whose genes are transcriptionally activated by p53 are listed in **Table 15.2**, and a flow chart of how these proteins affect processes in the cell cycle is given in **Figure 15.16**.

Increased transcription of the genes for p21, GADD45, and 14-3-3σ and decreased transcription of the gene for cyclin B all serve to block the cell cycle at particular points, as illustrated in **Figure 15.17**. The G_1/S transition checkpoint is mediated by the increased levels of p21, which result in inhibition of the G_1 cyclin–CDKs and in this way block the G_1/S transition. Hence if DNA damage is detected in G_1, the cell is blocked in the G_1/S transition. The S-phase response to DNA damage is mediated by the p21 protein and GADD45, which form a complex with proliferating cell nuclear antigen (PCNA) and reduce the processiv-

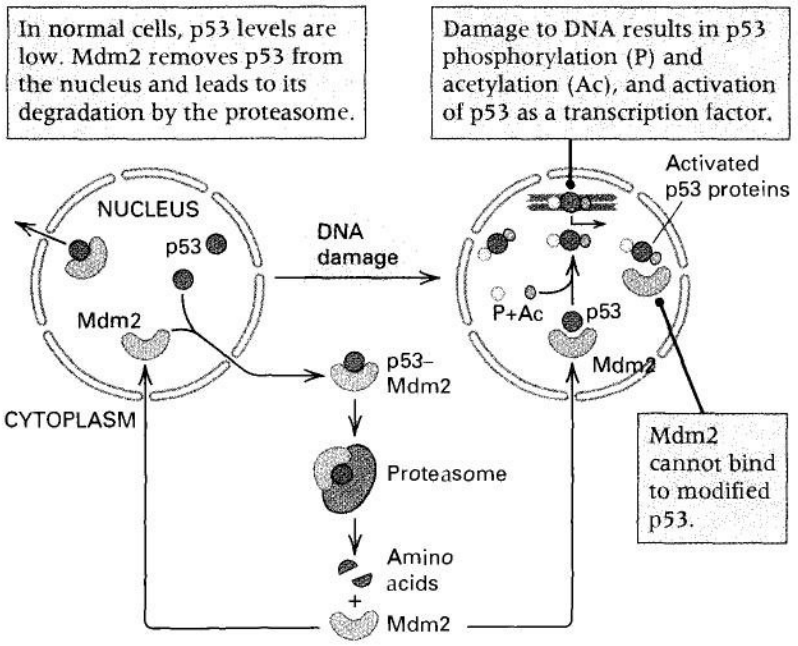
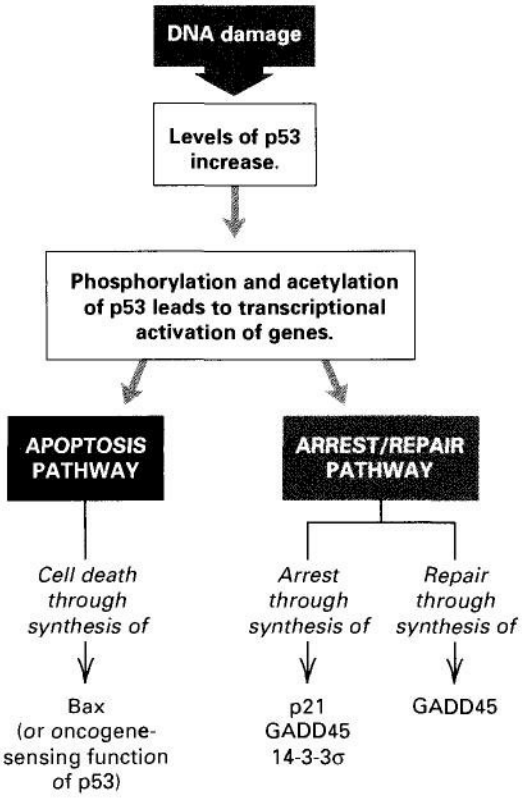


Figure 15.15 The effect of DNA damage on transcription factor p53. In normal cells the level of p53 is low, in part because Mdm2 shuttles it to the cytoplasm where it is destroyed by the proteasome. DNA damage results in phosphorylation and acetylation of p53, rendering it unable to bind Mdm2. Hence p53 levels rise.



ity of DNA polymerase. The **processivity** of a DNA polymerase is the number of consecutive nucleotides in the template strand that are replicated before the polymerase detaches from the template. Decreasing the processivity of DNA polymerase therefore slows down DNA synthesis, in effect buying time for the cell to repair DNA damage. The G_2/M checkpoint is mediated by the 14-3-3σ protein, which hinders activation of cyclin B–Cdc2 by preventing dephosphorylation of Cdc2 by the phosphatase Cdc25C, thus blocking the G_2/M transition. At the same time, the decrease in the level of cyclin B reduces the level of the active cyclin B–Cdc2 complex, which also ensures

Figure 15.16 Downstream events triggered by p53 include transcriptional activation of the genes for p21, GADD45, 14-3-3σ, and Bax. Activation of the apoptosis pathway by p53 follows transcriptional activation of Bax and, possibly, direct sensing of activated oncogenes (cancer-causing genes) of either cellular or viral origin.

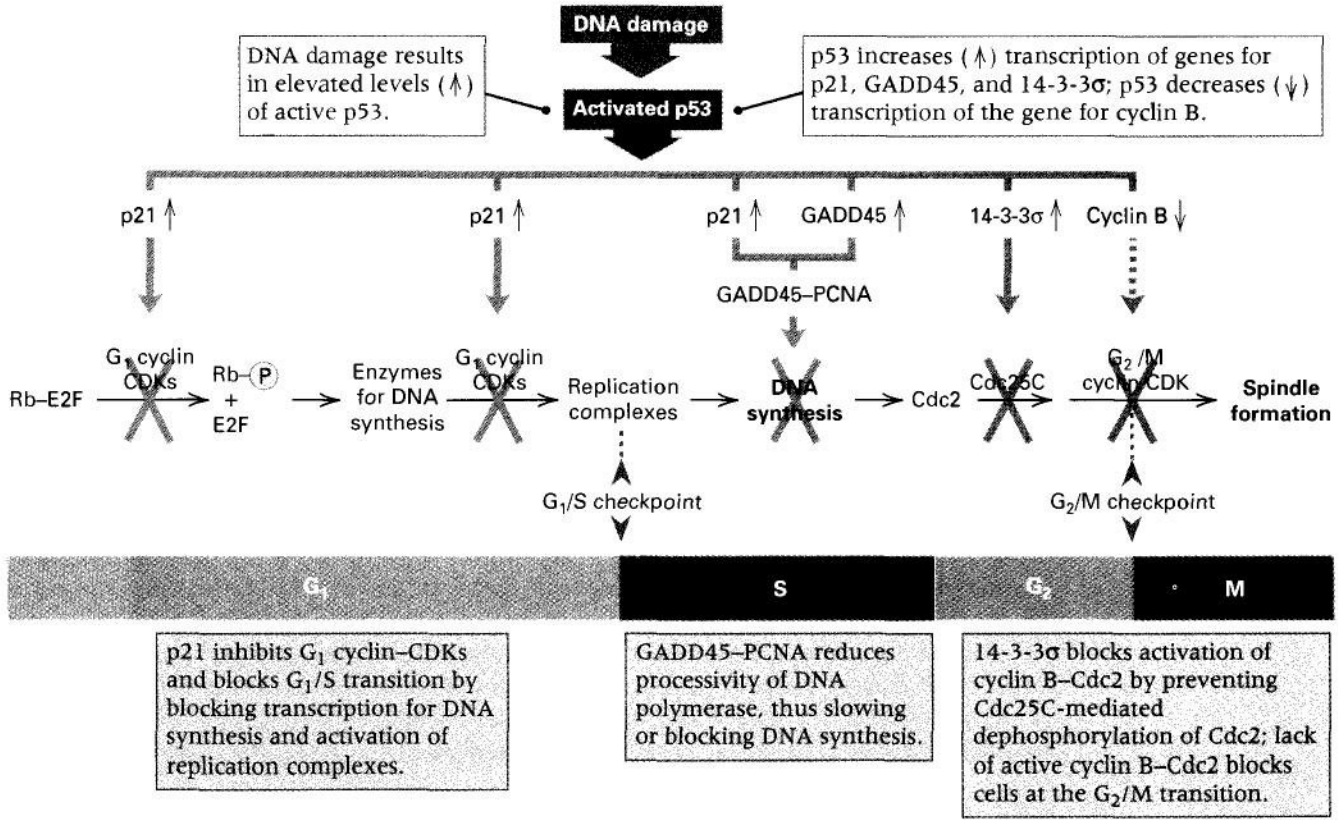


Figure 15.17 Role of p53 in the DNA damage checkpoints. DNA damage results in an increase in the level of active p53, which increases transcription of the genes for p21, GADD45, and 14-3-3σ and decreases transcription of the genes for cyclin B.

CoNnection

Sick of Telomeres

William C. Hahn,[1,2] Christopher M. Counter,[3] Ante S. Lundberg,[1,2] Roderick L. Beijersbergen,[1] Mary W. Brooks,[1] and Robert A. Weinberg[1] 1999

[1] Massachusetts Institute of Technology, Cambridge, Massachusetts

[2] Harvard Medical School, Boston, Massachusetts

[3] Duke University Medical Center, Durham, North Carolina

Creation of Human Tumor Cells with Defined Genetic Elements

One day, about 10 years ago, at a well-known midwestern medical school, the head of the Department of Medicine confronted the head of the Department of Genetics and said, "Your Professor Blank up the hall studies telomeres. Who cares about telomeres? Nobody ever gets sick because of their telomeres!" We recount this story to emphasize that the directions of basic research that prove most important in the long run cannot usually be predicted, even by experts. As this important paper shows, lots *of people get sick—very sick—because of their telomeres.*

During malignant transformation, cancer cells acquire genetic mutations that override the normal mechanisms controlling cellular proliferation. Primary rodent cells are efficiently converted into tumorigenic cells by the coexpression of cooperating oncogenes [genes that cause cancer when mutated]. However, similar experiments with human cells have consistently failed to yield tumorigenic transformants, indicating a fundamental difference in the biology of human and rodent cells. . . . One ostensibly important difference between rodent and human cells derives from their telomere biology. Murine somatic cells express telomerase activity and have much longer telomeres than their human counterparts, which lack telomerase activity. Because normal human cells progressively lose telomeric DNA with passage in culture, telomeric erosion is thought to limit cellular life span. Ectopic expression of the *hTERT* gene, which encodes the catalytic subunit of the telomerase, enables some . . . pre-senescent primary human cells to multiply indefinitely. . . .

It is now highly likely that telomere maintenance contributes directly to oncogenesis

To determine whether *hTERT* human cells were tumorigenic, . . . we also introduced large-T antigen [which inhibits both the retinoblastoma and p53 tumor-suppressor proteins] and an oncogenic *ras* [signal transduction] mutation. . . . We observed efficient colony formation in soft agar [an assay for the tumorigenic state] only with cells expressing the combination of large-T, *ras* and *hTERT*. . . . When these cells were introduced into [immunologically deficient] mice, rapidly growing tumors were repeatedly observed with high efficiency. . . . We conclude that ectopic expression of a defined set of genes . . . suffices to convert normal human cells into tumorigenic cells. . . . It is now highly likely that telomere maintenance contributes directly to oncogenesis by allowing precancerous cells to proliferate beyond the number of replicative doublings allotted to their normal precursors.

Source: Nature 400: 464–468

that the cell remains in G_2. Hence if DNA damage is detected in the S or the G_2 phase, the cell cycle arrests at the G_2/M checkpoint.

DNA damage also triggers activation of another pathway, a pathway for **apoptosis,** or **programmed cell death.** When the apoptotic pathway is activated, a cascade of proteolysis is initiated that culminates in cell suicide. The proteases involved are called *caspases*. Their activation ultimately results in destruction of the cellular DNA, internal organelles, and the actin cytoskeleton, and it is accompanied by nuclear condensation and usually followed by engulfment of the cellular remnants by phagocytes. The p53 transcription factor activates the apoptotic pathway by activating transcription of *Bax*. As illustrated in **Figure 15.18**, Bax normally exists in a heterodimer with an inhibitor of apoptosis called Bcl2. When p53 activates transcription of *Bax*, the balance is tilted in favor of Bax homodimers and against the Bcl2 heterodimer, which promotes apoptosis and self-destruction of the cell. On the other hand, activation of **oncogenes,** which are genes associated with cancers, can increase the level of activated (phosphorylated) Bcl2, which prevents apoptosis and allows the affected cells to grow and divide indefinitely. Cellular immortality does not always follow

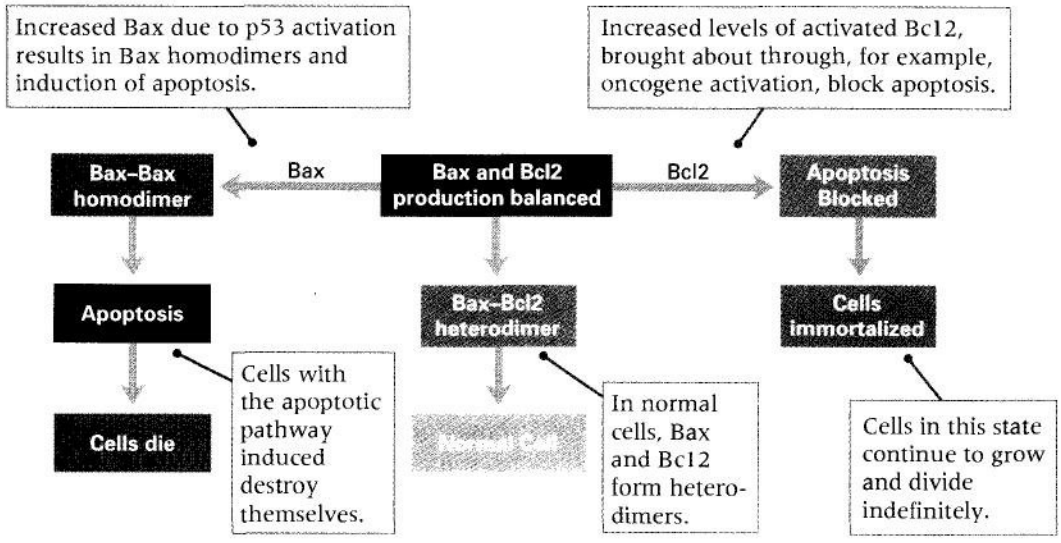
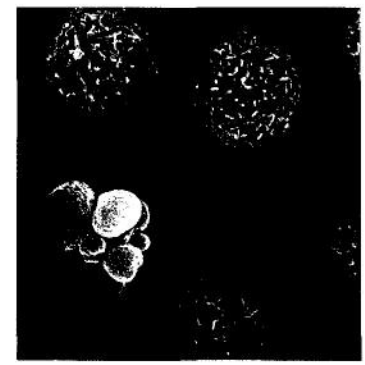


Figure 15.18 How Bax and Bcl2 interact to regulate apoptosis.

oncogene activation, because in some cases it activates the apoptotic pathway, possibly through an oncogene-sensing function of p53 illustrated in Figure 15.16. In all of its functions, p53 acts to protect the integrity of the genome with respect to nucleotide sequence and strand integrity and with respect to euploidy. Instability in the genome, unbalanced genomes, and damaged DNA pose a hazard to the organism. Apoptosis is a mechanism for killing such damaged—and thus dangerous—cells.

To gain an appreciation of the importance of the DNA checkpoint, look again at Figure 15.17 and consider what would happen if part or all of this fail-safe mechanism should malfunction. Suppose, for example, that cells lacked p21 protein but retained the rest of the mechanism. If DNA damage were properly sensed, and p53 functioned as usual to increase levels of 14-3-3σ and GADD45, cells would accumulate in G_2 and be unable to undergo mitosis. However, DNA would continue to be synthesized. In addition, lacking p21 protein, the cells would have a defective G_1/S checkpoint and so could embark on additional rounds of DNA synthesis. The expected result of p21 malfunction would therefore be polyploid cells, and this is the result observed.

To take another example, consider the consequence of loss of p53 function. Even if DNA damage were detected, the cell would be unable to mount a response—it would be unable to buy time to repair the DNA damage. The cells could initiate new rounds of synthesis with damaged chromosomes. Such synthesis would not only result in an increased frequency of mutation but would also permit gene amplification. Amplification of genes encoding cyclin D or Cdk4 would allow cells to escape the normal controls on DNA synthesis and proliferation. Cells already in S or G_2 would enter mitosis with damaged chromosomes, because not enough time for repair of lesions would have elapsed. In addition, the organism would have lost its ultimate protection against such damaged cells: apoptosis. The absence of p53 means that transcription of *Bax* would not be increased and that the apoptotic pathway would not be turned on; the normal balance between Bax and Bcl2 would be maintained, ensuring the survival of these damaged cells and thus putting the organism at risk. If the description of these runaway cells reminds you of the unchecked proliferation of cancer cells, this is because cancer cells *become* cancer cells by subverting the checkpoint mechanisms.



The forming of large blebs ("blisters") from the cell membrane is a characteristic of apoptosis. [K. G. Murti/Visuals Unlimited.]

The Centrosome Duplication Checkpoint

A **centrosome duplication checkpoint** monitors formation of a bipolar spindle. It seems to be coordinated with entry into

mitosis, because activation of cyclin B–Cdc2 kinase is correlated with centrosome duplication and formation of the spindle. This duplication checkpoint may also be coordinated, in some organisms, with the spindle checkpoint and the exit from mitosis.

The Spindle Checkpoint

The **spindle checkpoint** monitors assembly of the spindle and attachment of the kinetochores to the spindle. (The *kinetochore* is the spindle fiber attachment site on the chromosome.) Improper or incomplete assembly triggers a block to anaphase by preventing activation of the anaphase-promoting complex (Figure 15.13) and thus separation of the sister chromatids.

Cells can detect a single unattached or misattached chromosome and delay anaphase. Studies on insect and mammalian chromosomes suggest that the absence of tension at the kinetochore is the initiating signal for cell-cycle arrest. Tension on the kinetochore is related to the level of phosphorylation at kinetochores; unattached kinetochores have relatively more of a phosphorylated entity than do attached kinetochores. If all chromosomes form stable, bipolar attachments to spindle microtubules at their kinetochores, the Cdc20 protein (initially identified and named in yeast) is able to activate the anaphase-promoting complex (Figure 15.13). The anaphase-promoting complex not only degrades cyclin B, allowing exit from mitosis, but also degrades proteins that inhibit the separation of sister chromatids. Thus the anaphase-promoting complex promotes anaphase by allowing the separation of chromosomes. As illustrated in **Figure 15.19**, when an unattached kinetochore is detected, the Bub, Mad, and Mps1 proteins act to block the onset of anaphase, apparently by inhibiting the protein Cdc20. Because Cdc20 is not activated, the anaphase-promoting complex is not activated, and the cells remain at metaphase until all chromosomes form stable, bipolar attachments to spindle microtubules.

If any one of the Mad, Bub, or Mps1 proteins is defective, Cdc20 is still activated even if a microtubule has failed to attach to a kinetochore. Hence the spindle checkpoint is not activated. The Cdc20 is activated normally, and in turn the anaphase-promoting complex is activated, and anaphase takes place. The result is two aneuploid daughter cells, one lacking a chromosome and one with an extra copy.

Failure of any checkpoint results in genetic instability. The particular kinds of genomic instability associated with defects in the three checkpoints discussed are summarized in **Figure 15.20**. Malfunctioning of the spindle itself can lead to aneuploidy, whereas a failure to duplicate a spindle pole can lead to polyploidy (Chapter 9). Defects in the DNA damage checkpoints can result in chromosomal aberrations of various kinds, including translocations, deletions,

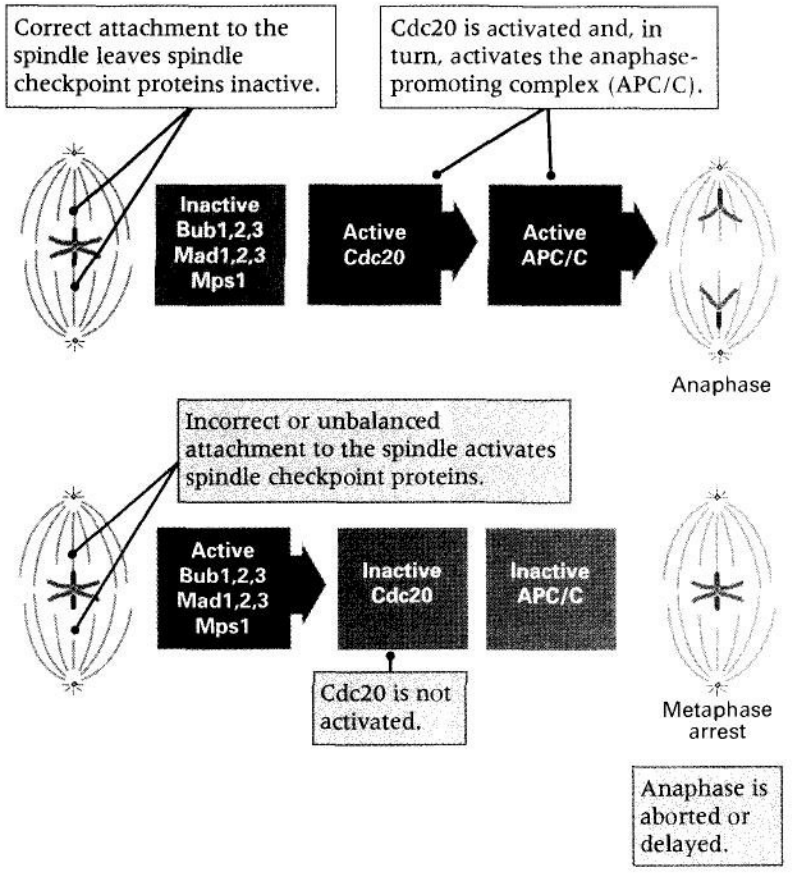


Figure 15.19 The spindle checkpoint. All chromosomes must form stable, bipolar attachments to spindle microtubules to activate the anaphase-promoting complex, which promotes the onset of anaphase and separation of the sister chromatids.

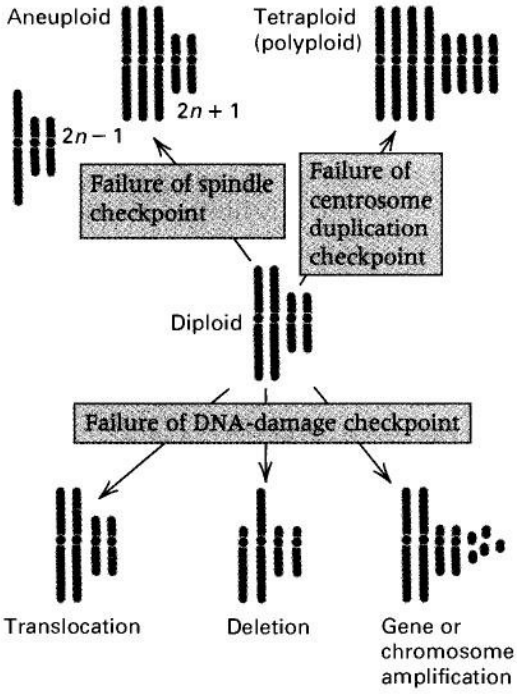


Figure 15.20 Contributions of checkpoint failures to genetic instability. [Adapted from L. H. Hartwell and M. Kastan. 1994. *Science* 266: 1821.]

and amplification of genes or subregions of chromosomes. The amplified genes may be found as tandem repeats within a chromosome or as extrachromosomal circles lacking a centromere and teleomeres.

15.5 Cancer Cells

Cancer is not one disease but rather many diseases that share similar attributes: The cells show uncontrolled growth as a result of mutations that affect a limited number of genes. It is a disease of somatic cells. About 1 percent of cancer cases are **familial,** which means there is clear evidence for segregation of a gene in the pedigree that predisposes cells in affected individuals to progress to the cancerous state. The other 99 percent of cases are called **sporadic,** which in this context means *not familial,* and are the result of genetic changes in somatic cells. Cancer cells share several properties not found in normal cells. One of these is the effect of cell-to-cell contact. In normal cells, cell-to-cell contact inhibits further growth and division, a process called **contact inhibition.** Cancer cells have lost contact inhibition: They continue to grow and divide, and they even pile on top of one another. This phenomenon is shown in **Figure 15.21**. The normal cells in A are flat and form a sheet; they have ceased to grow where they touch. In contrast, the tumorigenic cells in panel B fail to show contact inhibition. The cells look very different from the normal cells, particularly in the protruding bubbles that indicate membrane activity. If badly damaged or starved for nutrients, normal cells undergo apoptosis and self-destruct; cancer cells do

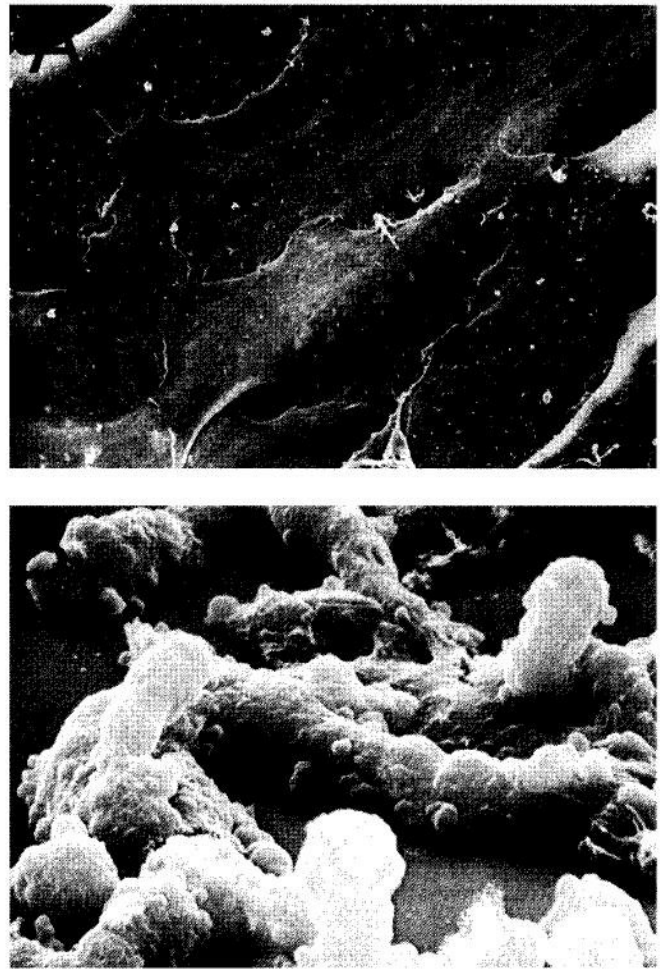


Figure 15.21 Scanning electron micrographs of (A) normal and (B) adenovirus-transformed (tumorigenic) hamster cells. [From R. Goldman et al. 1974. *Cold Spring Harbor Symposium on Quantitative Biology* 39, 601.]

not undergo apoptosis. Even in the absence of damage, normal cells cease to divide in culture after about 50 doublings (a process called **cell senescence**), whereas cancer cells in culture divide indefinitely—they are immortal. The senescent behavior of normal cells is associated with a loss of *telomerase* activity: The telomeres are no longer elongated, which contributes to the onset of senescence and cell death. Cancer cells have high levels of telomerase, which help to protect them from senescence, making them immortal.

Within an organism, tumor cells are clonal. They descend from a single ancestral cell that became cancerous. This conclusion is based in part on the observation that tumor cells in a female express the genes in only one of the X chromosomes, as a result of the normal inactivation of the other X chromosome (Chapter 9). If the tumor cells were not clonal, gene expression from both X chromosomes would be expected, reflecting the random inactivation of one X chromosome in each cell lineage. The finding of expression from a single X chromosome must therefore reflect clonality.

The conversion of normal somatic cells to cancer cells is a process that requires multiple steps. An important contributor to cancer conversion is genetic instability in the precursor cell population. The genetic instability may occur at the level of nucleotide sequences in the DNA or at the level of chromosome structure or number. The genetic instability is manifested as mutations, gene amplification, chromosomal rearrangements, or aneuploidy. The mutations result in a cell population that is genetically heterogeneous. In such a mixed population, any cell that has a proliferative advantage will contribute a greater fraction of descendants to the future cell population than will its neighbors, and because of this advantage, its clone expands at the expense of others. Subsequent mutations in the descendants of this cell, and further clonal expansions of the derivative cells, can give rise to a clone of cells with the proliferative capacity typical of cancer cells.

Many cancers are the result of alterations in cell-cycle control, particularly in control of the G_1-to-S transition, and in the G_1/S checkpoint associated with this transition. These alterations also affect apoptosis through their interactions with p53. **Figure 15.22** summarizes the key elements of the regulatory circuitry that governs the G_1/S transition, including the proteins that promote cell-cycle progression, that activate the checkpoint, and that govern the apoptotic pathway. Tumor cells commonly show altered expression or inactivation of function of one or more of these genes. Some genes for which alterations have been detected in cancer cells are listed in **Table 15.3**. The major mutational targets for the multistep cancer progression are of two types: *proto-oncogenes* and *tumor-suppressor genes*. These are discussed in the following sections.

Oncogenes and Proto-Oncogenes

Oncogenes are gain-of-function mutations associated with cancer progression. They are derived from normal cellular genes called **proto-oncogenes.** Oncogenes are gain-of-function mutations because they improperly enhance the expression of genes that promote cell proliferation or inhibit apoptosis. In Figure 15.22, the proteins enclosed in rectangles are the products of proto-oncogenes, which when mutated can give rise to oncogenes. In this section, some of the oncogenes identified in Figure 15.22 and Table 15.3 are discussed in more detail. These include cyclins and CDKs, growth-factor receptors, Ras, and Mdm2.

Cyclin D and Cdk4 Overexpression of cyclin D promotes unscheduled entry into S phase. Overexpression is often the result of gene amplification. Amplification of the cyclin D gene is found in about 35 percent of esophageal carcinomas, 15 percent of bladder cancers, and about 15 percent of breast cancers. On the other hand, although only 15 percent of breast cancers show amplification of the cyclin D gene, more than 50 percent of breast cancers show overexpression of cyclin D; this discrepancy implies that other mechanisms of overexpression also occur. Paralleling the increase in cyclin D expression, the gene for cyclin D's CDK partner, Cdk4, is amplified in some tumors, including 12 percent of gliomas (brain tumors) and 11 percent of sarcomas (muscle or connective tissue cancer).

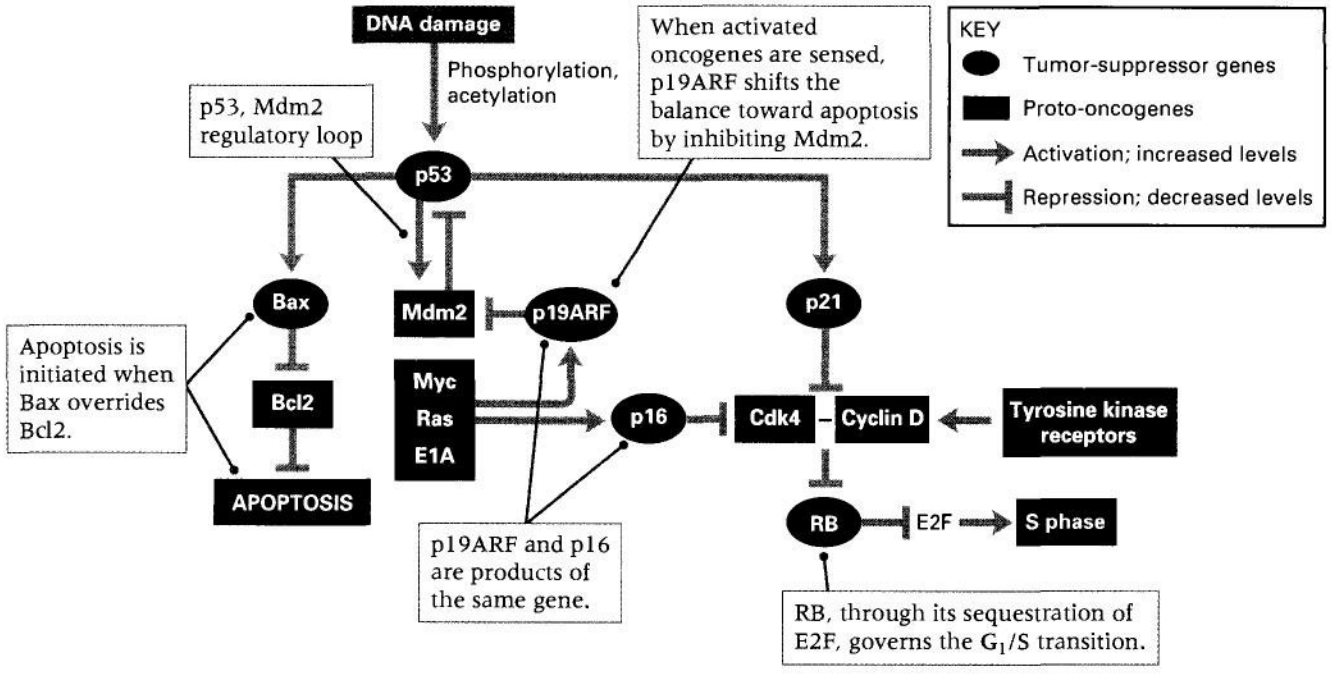


Figure 15.22 Interactions between the p53 pathway and the RB (retinoblastoma protein) pathway in controlling apoptosis or DNA synthesis.

Table 15.3 Cell cycle regulatory genes affected in tumors

Protein oncogenes	Alteration	Consequence
Cyclin D	Amplification or overexpression	Promotes entry into S phase.
Cdk4	Amplification	Promotes entry into S phase.
Cdk4	Mutation	Cdk4 resistant to inhibition by p16; promotes entry into S phase.
EGFR (epidermal growth-factor receptor, a tyrosine kinase receptor)	Amplification	Promotes proliferation by constitutive activation of pathway from growth-factor receptor.
FGFR (fibroblast growth-factor receptor)	Amplification	Promotes proliferation by constitutive activation of pathway from growth factor receptor.
Ras	Amplification	Promotes proliferation by constitutively transmitting growth signal.
Ras	Mutation	Inactivates GTPase activity; constitutive activation of pathway from growth-factor receptors.
Bcl2	Overexpression by translocation next to strong enhancer	Blocks apoptosis.
Mdm2	Amplification	Mimics loss of p53 with loss of G_1/S, S, and G_2/M checkpoint functions.
Tumor-suppressor genes	**Alteration**	**Consequence**
p53	Mutation	Loss of G_1/S,S and G_2/M checkpoint functions.
p21	Mutation	Loss of G_1/S and S checkpoint functions.
RB	Mutation	Promotes proliferation; E2F uninhibited.
Bax	Mutation	Failure to promote apoptosis of damaged cells.
Bub1p	Mutation	Loss of spindle checkpoint function.

Growth-factor receptors and Ras Growth factors stimulate growth by binding to a growth-factor receptor at the membrane. The binding activates a signal transduction pathway that acts through Ras (discussed below), cyclin D, and its partner CDKs. The gene that encodes the receptor for epidermal growth factor (EGFR) is amplified in 45 percent of malignant astrocytomas (a kind of brain tumor), 35–50 percent of glioblastomas (another kind of brain tumor), 20 percent of breast cancers, 15–30 percent of ovarian cancers, 10 percent of head and neck cancers, and melanomas. The genes for fibroblast growth factor receptor (FGFR) are amplified in about 20 percent of breast cancers.

Activated tyrosine kinase receptors such as EGFR and FGFR activate a signal transduction pathway that activates a small signaling protein called a **G protein,** as illustrated in **Figure 15.23** for the G protein Ras. Ras is usually bound to GDP, and in this state it is inactive. Receptor activation results in activation of a protein that stimulates the exchange of GTP to replace GDP in the Ras protein, and the Ras protein is active when it is bound to GTP. The Ras–GTP, in turn, activates a downstream protein, propagating a growth signal. The activation of Ras as Ras–GTP is transient: Yet another protein (GAP) activates the intrinsic GTPase activity of Ras, so Ras hydrolyzes the GTP and returns itself to the inactive state of Ras–GDP. Mutations in the *RAS* gene occur that inactivate its GTPase function; its GTPase activity cannot be activated by GAP, and it remains in its active form of Ras–GTP, whether or not the cell is receiving signals from growth-factor receptors. A signal for growth is therefore transmitted constitutively, and unrestrained growth and division take place.

EGFR and FGFR Amplification or overexpression of EGFR or FGFR results in autoactivation of the receptor and transmission of a constitutive growth signal acting through the Ras pathway. Overexpression of Ras also leads to enhanced signal transduction and renewed cycles of proliferation.

Mdm2 Amplification of the gene that encodes Mdm2 acts to promote cell proliferation, but it does so indirectly. Overexpression of Mdm2 effectively overwhelms any increase in production of p53, because the Mdm2 shuttles the extra p53 from the nucleus to the cytoplasm, where it is destroyed by the proteasome (Figure 15.15). This process effectively prevents functioning of the G_1/S checkpoint. Hence, amplification of the *Mdm2* gene and the resultant overexpression of Mdm2 protein is equivalent in its effects to inactivation of the gene for p53. Amplification of the *Mdm2* gene has been found in 19 of 28 tumor types examined, most commonly in tumors of adipose tissue (42 percent), soft-tissue sarcomas (20 percent), osteosarcoma (16 percent) and esophageal carcinoma (13 percent).

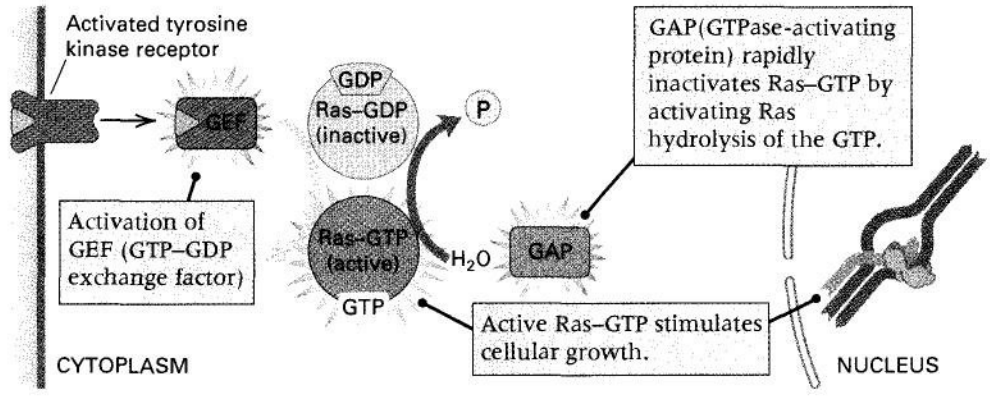


Figure 15.23 Function of the Ras protein, which acts as a switch in stimulating cellular growth in the presence of growth factors. Ras is the product of a proto-oncogene. Certain mutant Ras proteins, such as a G19V (valine substitution for glycine at position 19), lack GTPase activity and remain in the form of Ras–GTP; hence a growth-promoting signal is present even in the absence of growth factors.

Tumor-Suppressor Genes

Tumor-suppressor genes are genes that normally negatively control cell proliferation or that activate the apoptotic pathway, in which loss-of-function mutations contribute to cancer progression. Some of the key functions of tumor-suppressor genes are illustrated in Figure 15.22 and listed in Table 15.3. They are examined in more detail in this section.

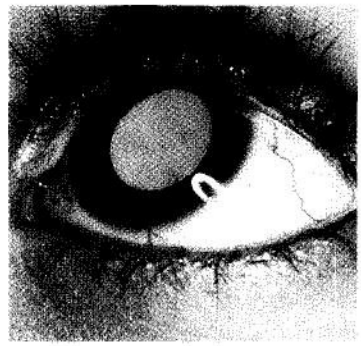
p53 Loss of function of p53 eliminates the DNA checkpoint that monitors DNA damage in G_1 and S, as we saw earlier in Figures 15.14 and 15.17. In the absence of functional p53, the proteins responsible for arresting cells in G_1 or G_2 (p21, 14-3-3σ, and GADD45) are not synthesized in response to DNA damage. There is therefore no block to a cell's proceeding into S phase or into M phase even if it has damaged chromosomes, altered ploidy, or amplified genes. In addition, loss of p53 function costs the organism its ultimate defense against aberrant cells, because DNA damage is no longer able to trigger enhanced expression of Bax and hence the self-destruction of aberrant cells: The damaged cells survive and proliferate. Furthermore, their genetic instability increases the probability of additional genetic changes and thus progression toward the cancerous state. Given this scenario, and the key role of p53 in protecting the cell against the consequences of DNA damage, it is not surprising that p53 proves to be nonfunctional in 55 percent of all cancers. In addition, as mentioned previously, Mdm2 overexpression—equivalent to loss of p53 function—occurs in other tumors. Because loss of p53 and overexpression of Mdm2 rarely occur in the same tumor, the actual loss of p53 function by one mechanism or another is probably substantially greater than 55 percent. Mutant p53 proteins are found frequently in melanomas, several kinds of lung cancers, colorectal tumors, bladder and prostate cancers, and astrocytomas.

Testicular embryonal carcinoma is one of the most curable of cancers, with a cure rate of 90–95 percent. A striking observation is that these carcinomas never include mutations in the p53 gene. The curability and the lack of p53 involvement may be causally related. The p53 gene is not expressed in embryonal testicular cells, possibly because the process of recombination with breakage and rejoining of DNA strands would be sensed as DNA damage, and imposition of the checkpoints would block meiosis. Because the p53 gene is not expressed, there can be no selective advantage conferred on cells by mutations in the p53 gene in embryonal testicular cells. The current interpretation of the high success rate with treatment using certain drugs (for example, *cisplatin*) is that the drug induces the expression of p53, which activates the apoptotic pathway and leads to suicide of the cancer cells.

p21 Loss of p21 function results in renewed rounds of DNA synthesis without accompanying mitosis, and the level of ploidy of the cell increases. Mutations in the gene that encodes p21 occur in some prostate cancers.

p16/p19ARF The p16 and p19ARF proteins are products of the same gene; they derive from two distinct transcripts from different promoters, each with a different 5′ exon. The p16 product can inhibit the cyclin D–Cdk4 complex and help control entry into S phase; the gene encoding p16 is a very frequent target for inactivation during tumor progression. The gene is deleted in 55 percent of gliomas (a form of brain tumor), 55 percent of mesotheliomas, more than 50 percent of melanomas, 40 percent of nasopharyngeal carcinomas, 50 percent of biliary-tract carcinomas, and 30 percent of esophageal carcinomas. Figure 15.22 shows the functional interactions of p19ARF as a tumor suppressor. If cellular or viral oncogenes are being overexpressed, the synthesis of p19ARF is elevated, and at this higher concentration p19ARF binds to Mdm2 and prevents it from binding with p53 and shuttling p53 from the nucleus. Given this mode of action, one might expect that some tumors might lack p19ARF. In fact, no mutation has yet been reported that inactivates p19ARF without also inactivating p16, although the reverse kind of mutation has been observed.

RB The retinoblastoma protein controls the transition from G_1 to S phase by controlling the activity of the transcription factor E2F, as illustrated in Figure 15.12. Loss



A retinoblastoma tumor first appears as a white mass in the retina but may grow to fill the entire eye, as in this case, giving the pupil a white or yellow cast. [Custom Medical Stock Photo.]

of RB function frees E2F to initiate transcription of the enzymes for DNA synthesis at all times in the cell cycle; hence excessive rounds of DNA synthesis are continuously being initiated. Loss of RB function is found in melanomas, in small-cell lung carcinoma, in osteosarcoma, and in liposarcomas. Just as loss of p53 and overexpression of Mdm2 appear to be equivalent in their effects, so do loss of either of the tumor-suppressor proteins RB or p16 and overexpression of either cyclin D or Cdk4. One basis for this conclusion is the observation that tumor cells are nearly always altered for RB, p16, cyclin D, or Cdk4, but rarely for more than one. This finding suggests that loss of function of either p16 or RB is equivalent to overexpression of either cyclin D or Cdk4. Any one of these defects leads to the same result: The cells lose control of the G_1/S transition and embark upon unscheduled rounds of DNA synthesis.

Bax The Bax tumor-suppressor protein promotes apoptosis (Figure 15.18). Loss of Bax function is found particularly in gastric adenocarcinomas and in colorectal carcinomas, particularly those associated with microsatellite instability because of defective mismatch repair. Cells that are defective in mismatch repair are especially prone to undergo replication slippage leading to deletions or additions of nucleotides in runs of short tandem repeats (discussed in Chapter 7). One of the sequences prone to replication slippage is a run of eight consecutive G nucleotides within the *Bax* gene, which makes this gene especially susceptible to frameshift mutation by the addition or deletion of a base pair in this region.

Bub1 Bub1 is a protein that is primarily involved in the spindle checkpoint (Figure 15.19). A subset of colon cancers show chromosomal instability: They do not maintain a constant karyotype but, rather, are aneuploid. Some of these unstable lines are defective in Bub1.

15.6 Hereditary Cancer Syndromes

As we have noted, approximately 99 percent of cancers are sporadic: All of the genetic changes occur in somatic cells. Furthermore, the change from a normal cell into a tumor cell is progressive. It goes step by step as each new somatic mutation along the way compromises a different mechanism of cell-cycle control.

Defects in Cell-Cycle Regulation and Checkpoints

Approximately 1 percent of cancers are familial. In these cases, one of the cancer-progression mutations that may occur in sporadic tumors is inherited through the germ line. The presence of this mutation predisposes the individual to cancer, because it makes fewer somatic mutations necessary for a precancerous cell to progress to malignancy. **Figure 15.24** illustrates some of the genetic mutations and changes in cell morphology that take place in the progression of a type of familial colon cancer called adenomatous polyposis. In this case the inherited mutation is in the gene *APC* in chromosome 5, which is a tumor-suppressor gene whose normal function is to transduce the signal of contact inhibition into the cell to inhibit further growth and division. Subsequent mutations in the progression to malignancy include an oncogenic mutation in a *Ras* gene and loss or mutation of the gene encoding p53. This is only one possible route of progression. Different tumors may progress by different pathways, depending on what mutations occur and in what order.

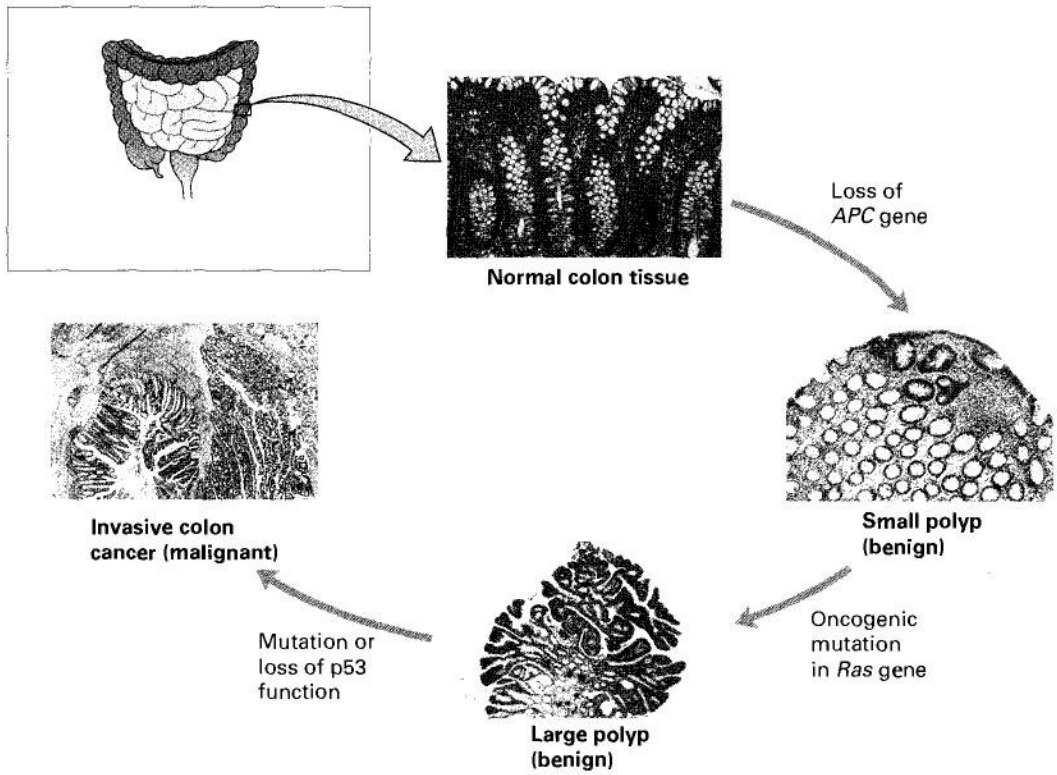


Figure 15.24 Somatic mutations and cell morphology in the progression of invasive colon cancer. [Photographs courtesy of Kathleen R. Cho.]

Familial adenomatous polyposis of the colon and some other important familial cancer syndromes are identified in **Table 15.4**. The Li-Fraumeni syndrome, familial retinoblastoma, and familial melanoma are all associated with germ-line mutations in genes that are often found to be mutant in sporadic tumors. The genes affected in these syndromes are the *p53* gene, the *RB1* gene, and the *p16* gene, respectively. A pedigree of a cancer-prone family segregating for a mutation in the *p53* gene (Li-Fraumeni

Table 15.4 Inherited cancer syndromes

Syndrome	Primary tumors	Associated tumors	Chromosome	Gene	Proposed function
Li–Fraumeni syndrome	Sarcomas, breast cancer	Brain tumors, leukemia	*17p13*	*p53*	Transcription factor
Familial retinoblastoma	Retinoblastoma	Osteosarcoma	*13q14*	*RB1*	Cell-cycle regulator
Familial melanoma	Melanoma	Pancreatic cancer	*9p21*	*p16*	Inhibitor of Cdk4 and Cdk6
Hereditary nonpolyposis colorectal cancer (HNPCC)	Colorectal cancer	Ovarian, glioblastoma	*2p22* *3p21* *2q32* *7p22*	*MSH2* *MLH1* *PMS1* *PMS2*	DNA-mismatch repair
Familial breast cancer	Breast cancer	Ovarian cancer	*17q21*	*BRCA1*	Repair of DNA double-strand breaks
Familial adenomatous polyposis of the colon	Colorectal cancer	Other gastrointestinal tumors	*5q21*	*APC*	Regulation of β-catenin
Xeroderma pigmentosum	Skin cancer		Several complementation groups	*XPB* *XPD* *XPA*	DNA-repair helicases, nucleotide excision repair

syndrome) is given in **Figure 15.25**. This syndrome shows clear autosomal dominant inheritance. However, the affected individuals show a range of different tumors and often have more than one, including osteosarcoma, leukemia, breast cancer, lung cancer, soft tissue sarcoma, and brain tumors. A large fraction of Li-Fraumeni families show segregation for a mutation in the *p53* gene. For this family, affected members are heterozygous for a single nucleotide pair in the *p53* gene, which changes codon 245 from GGC (glycine) to GAC (aspartic acid). This change is evident in the sequencing gel in **Figure 15.26**, which shows a DNA-sequencing gel with template DNA from the affected father and two sons in Figure 15.25, along with DNA from an unaffected individual from the same pedigree. In the affected individuals, the black arrows show the presence of both a G and an A at the same nucleotide site, which indicates that the individuals are heterozygous at this site. The heterozygous site is the second nucleotide in codon 245 of the *p53* gene; the flanking white arrows indicate that the first and third nucleotides (G and C, respectively) are not heterozygous. Hence the germ-line mutation in this pedigree results in a GAC codon, which is heterozygous in individuals who are at risk, whereas nonmutant alleles have the sequence GGC for codon 245. The observation that the affected individuals are heterozygous is consistent with the dominant autosomal inheritance.

A situation analogous to the human Li-Fraumeni syndrome has been created in mice by experimental knockout (loss of function) of the *p53* gene via the germ-line transformation methods discussed in Chapter 13. Animals heterozygous or homozygous for the p53 knockouts were compared to wildtype mice. Normal mice do not develop tumors; 50 percent of them live to the age of 120 weeks, and about 5 percent live for at least 160 weeks. In contrast, half of the heterozygous p53 knockout mice are dead by 55 weeks, and almost none of them live as long as 120 weeks. Furthermore, 50 percent of these mice develop tumors (mostly osteosarcomas) by 18 months of age. The homozygous knockout mice fare even worse: 50 percent are dead by 20 weeks, and all are dead by 40 weeks; among the homozygous animals, 75 percent develop tumors by 6 months, and by 10 months all have developed tumors or are dead. In the homozygous genotype, the tumors are lymphomas rather than osteosarcomas, possibly because normal development of the immune system requires massive apoptosis within the thymus gland, and in the p53 homozygous knockouts, apoptosis is abolished because Bax is not up-regulated as in normal mice and p53 knockout heterozygotes.

The fact that animals heterozygous for the p53 knockout are severely affected for both longevity and incidence of tumors does not necessarily imply that the *p53* mutation is dominant at the level of the individual cell. In fact, in the case of *p53*, the effects of the mutation are manifested only in somatic cells that have become homozygous for the knockout mutation and in those that have lost (or undergone somatic mutation at) the wildtype allele in the homologous chromosome. In a heterozygous animal, only one copy of the wildtype *p53* gene is present to protect the cell; inactivation of the lone wildtype allele disables the checkpoints that depend on the p53 protein. Because there are so many somatic cells in which such a rare aberration can occur, it is nearly certain that inactivation or loss of *p53* will take place somewhere in the heterozygous organism, initiating the sequence of mutations that results in cancer.

The idea that loss of the wildtype allele of a tumor-suppressor gene might be the triggering event at the cellular level for tumors in heterozygous genotypes was first suggested from studies on retinoblastoma by Alfred Knudson in 1971, long before the function of the *RB* gene was identified. As the excerpts in the Connection on page 665 entitled *Two Hits, Two Errors* indicate, Knudson noted that sporadic cases of retinoblastoma usually had a single tumor, whereas familial retinoblastoma cases usually had bilateral tumors and more than one tumor. On the basis of a statistical analysis of the time of tumor diagnosis, he suggested that genesis of a tumor in familial cases required a "single hit" in a somatic cell, whereas genesis of a tumor in sporadic cases required "two hits," which would happen only rarely. The two-hit model would also explain why sporadic cases rarely have a second tumor: Each individual "hit" is a rare event.

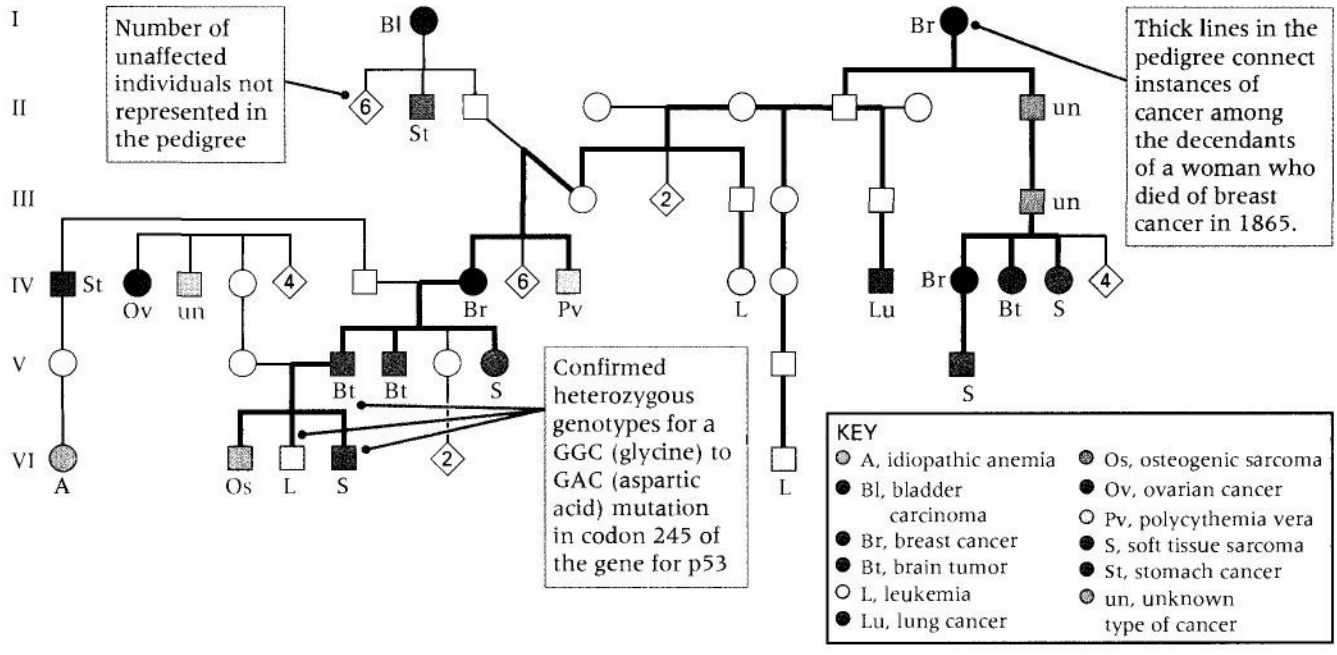


Figure 15.25 Pedigree of Li-Fraumeni syndrome associated with mutation in the gene for p53. Individuals are prone to develop any of a variety of cancers. [Data from W. A. Blattner, D. B. McGuire, J. J. Mulvihill, B. C. Lampkin, J. Hananian, and J. F. Fraumeni. 1979. *J. American Medical Assoc.* 241: 259.]

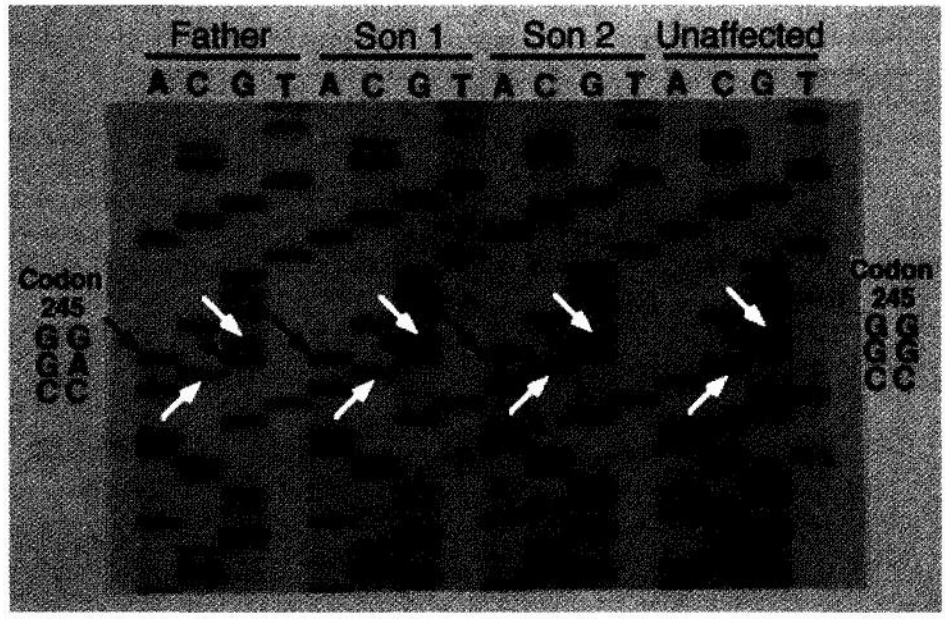


Figure 15.26 Part of the DNA sequence of the gene encoding p53 in the father and two sons indicated in the pedigree in Figure 15.25. In this case the mutation changes codon 245 from GGC to GAC. Black arrows indicate the heterozygous nucleotide in the father and his two sons; white arrows indicate the flanking nucleotides. The four lanes at the far right show the sequence in an individual homozygous for the wildtype allele. [Courtesy of Esther H. Chang.]

Retinoblastoma, like the p53 deficiency, is inherited in pedigrees as a simple Mendelian dominant. But Knudson's hypothesis implied that even in familial cases, there must be another mutational event that triggers tumor development. Once the gene itself, called *RB1*, was located, analysis of genetic markers around the gene in tumor cells revealed that the triggering event is the loss of the wildtype *RB1* allele. Any of several mechanisms can uncover the mutant allele, including chromosome loss, mitotic recombination, deletion, and inactivating nucleotide substitutions. These can be distinguished from one another on the basis of the genetic markers that are lost or retained in the tumor cells; several of the mechanisms known to occur are diagrammed in **Figure 15.27.** At the organismic level, expression of the mutant gene is dominant. However, at the cellular level, expression of the mutant gene is recessive. Uncovering of the recessive allele by various mechanisms is called **loss of heterozygosity.**

The inherited cancer syndromes listed in Table 15.4 result from mutations in tumor-suppressor genes. All show autosomal dominant inheritance except for xeroderma pigmentosum, which is inherited as a recessive. In all cancer syndromes that show dominant inheritance, loss of heterozygosity is required for manifestation of the tumor phenotype. Expression of the mutant allele *predisposes* the cell to become cancerous but is not in itself sufficient for generation of a cancer cell. Tumor progression still occurs only when additional somatic mutations and clonal expansions take place. The germ-line mutations do not themselves cause cancer; they merely make it much more likely that the progression will occur, because the tumor, in effect, has been given a "head start."

Defects in DNA Repair

Genetic instability clearly contributes to the genesis of tumor cells. We know from studies on yeast and bacteria that cells have extensive mechanisms for repairing DNA lesions (see Chapter 7). Defects in these processes result in greatly elevated mutation rates and genetic instability. It was therefore not surprising when hereditary cancer syndromes that result from inherited defects in DNA repair were discovered.

Several of the inherited cancer syndromes listed in Table 15.4 are the consequence of defects in DNA repair. Defects in any of four genes that encode proteins involved in DNA mismatch repair cause hereditary nonpolyposis colorectal cancer, which shows autosomal dominant inheritance. Mutant cells have higher mutation rates, which promote progression toward the cancerous state. Inherited breast cancer syndromes prove to be associated with mutations in either of two genes, *BRCA1* or *BRCA2*. Less is known about the function of these genes. Current evidence suggests that *BRCA1* is somehow involved in repair of double-strand breaks. Inherited skin cancer

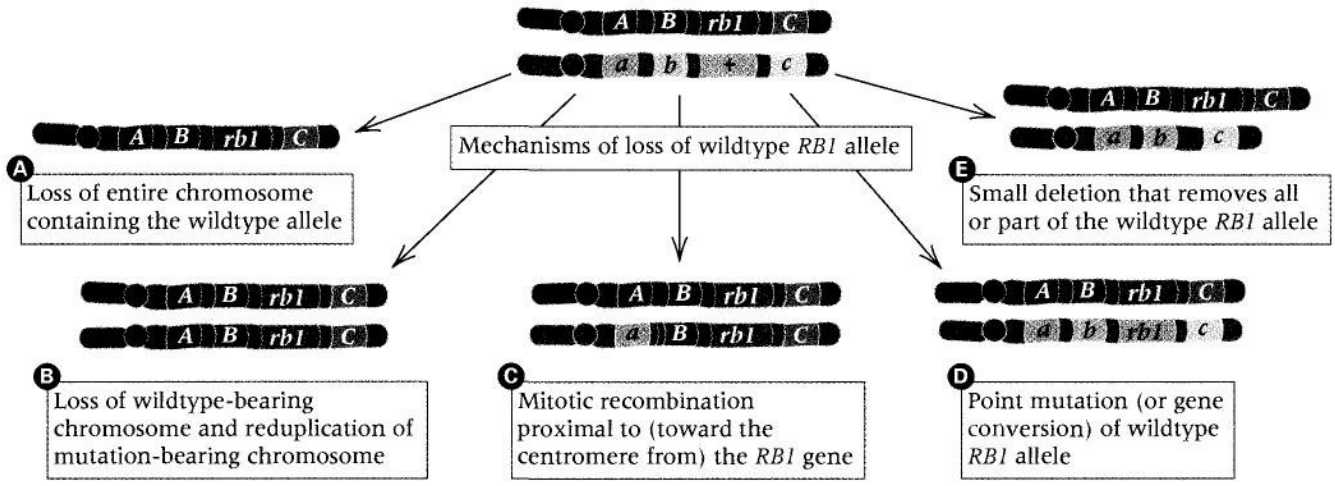


Figure 15.27 Genetic mechanisms for loss of heterozygosity of the wildtype *RB1* allele in patients with familial retinoblastoma.

CoNnection

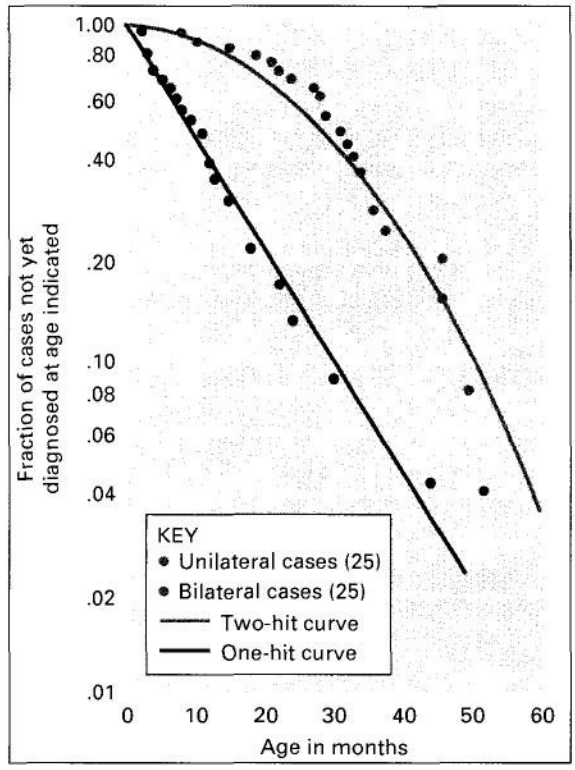
Two Hits, Two Errors

Alfred G. Knudson 1971
M. D. Anderson Hospital
The University of Texas,
Houston, Texas
Mutation and Cancer:
Statistical Study of Retinoblastoma

This landmark paper was a turning point in thinking about genetics and cancer. Pedigree studies had already shown that there were hereditary predispositions to certain cancers, such as retinoblastoma. But "sporadic" cases also occurred, in which only one member of a kindred was affected. What is the relationship between these two forms? By an ingenious statistical analysis, Knudson showed that the sporadic cases exhibit "two-hit" kinetics as a function of age, indicating that two independent mutations in the same retinal cell are involved, whereas familial cases exhibit "one-hit" kinetics. The simplest interpretation is that in sporadic cases, the first mutation knocks out a key gene and the second knocks out its allele; in familial cases, the initial mutant allele is inherited, and consequently only one mutation (in the remaining functional allele) is needed to cause the disease. At the level of the individual cell, therefore, a mutation in the retinoblastoma gene (RB1) is recessive, whereas at the familial level, the disease is inherited as a dominant, because almost every person who inherits one mutant allele will undergo the second mutation in at least one retinal cell.

The hypothesis is developed that retinoblastoma is a cancer caused by two mutational events. In the dominantly inherited form, one mutation is inherited via the germinal cells and the second occurs in somatic cells. In the nonhereditary form, both mutations occur in somatic cells. . . . Several authors have concluded that retinoblastoma may be caused by either a germinal or a somatic mutation. . . . All bilateral cases [25–30 percent] should be counted as hereditary because the proportion of affected offspring closely approximates the 50 percent expected with dominant inheritance. . . . If a second, single event is involved [in the hereditary cases], the distribution of bilateral cases with time should be an exponential function, i.e., the fraction of the total cases that develop should be constant, as expressed in the relationship $dS/dt = -kS$, and $\ln(S) = -kt$, where S is the fraction of survivors not yet diagnosed at time t, and dS is the change in this fraction in the interval dt. As shown in the figure, this is indeed the case. By contrast, the fractional decrease in unilateral cases per unit time does not show this relationship. . . . The exponential decline in new hereditary cases with time reflects the occurrence of a second event at a constant rate . . . of the order of 2×10^{-7} per year. . . . The two-mutation hypothesis is consistent with current thought . . . that the common cancers are produced by about 3–7 mutations. Interestingly, one of the lowest estimates [is] for brain tumors, which are, like retinoblastoma, derived from neural elements.

Source: Proceedings of the National Academy of Sciences USA 68: 820–823



syndromes are called xeroderma pigmentosum. Xeroderma pigmentosum cells are defective in nucleotide excision repair; they are unable to repair defects such as thymine dimers that are induced by ultraviolet light. Individuals with this syndrome are very sensitive to the ultraviolet light that is present in sunlight and emitted by fluorescent lights.

15.7 Genetics of the Acute Leukemias

The human acute leukemias arise in progenitors of the white blood cells. The initial genetic events that result in the acute leukemias are quite different from those indicated in the cancers examined previously. They do not arise as a consequence of alterations in cell cycle regulation or checkpoints, nor are they familial.

Up to 65 percent of the acute leukemias arise as a consequence of chromosomal translocations involving genes that encode transcription factors that play a role in blood-cell development (hematopoiesis). The translocations are of the two types illustrated in Figure 15.28. In a **promoter fusion** (part A), the coding region for a gene that encodes a transcription factor is translocated near an enhancer for an immunoglobulin heavy-chain gene or a T-cell receptor gene. The result is overexpression of the transcription factor, derangement of normal hematopoiesis, and overproduction of lymphocytes. One important aspect of normal hematopoiesis is the apoptotic destruction of progenitor cells that fail to rearrange their antigen-receptor genes productively (Chapter 12). Estimates are that 75 percent of B-cell and 95 percent of T-cell precursors self-destruct during normal development. *Bcl2* (*Bcl* stands for B-cell lymphoma) is an example of an oncogene in Table 15.4 that was originally discovered in a promoter fusion that placed *Bcl2* next to an immunoglobulin heavy-chain gene

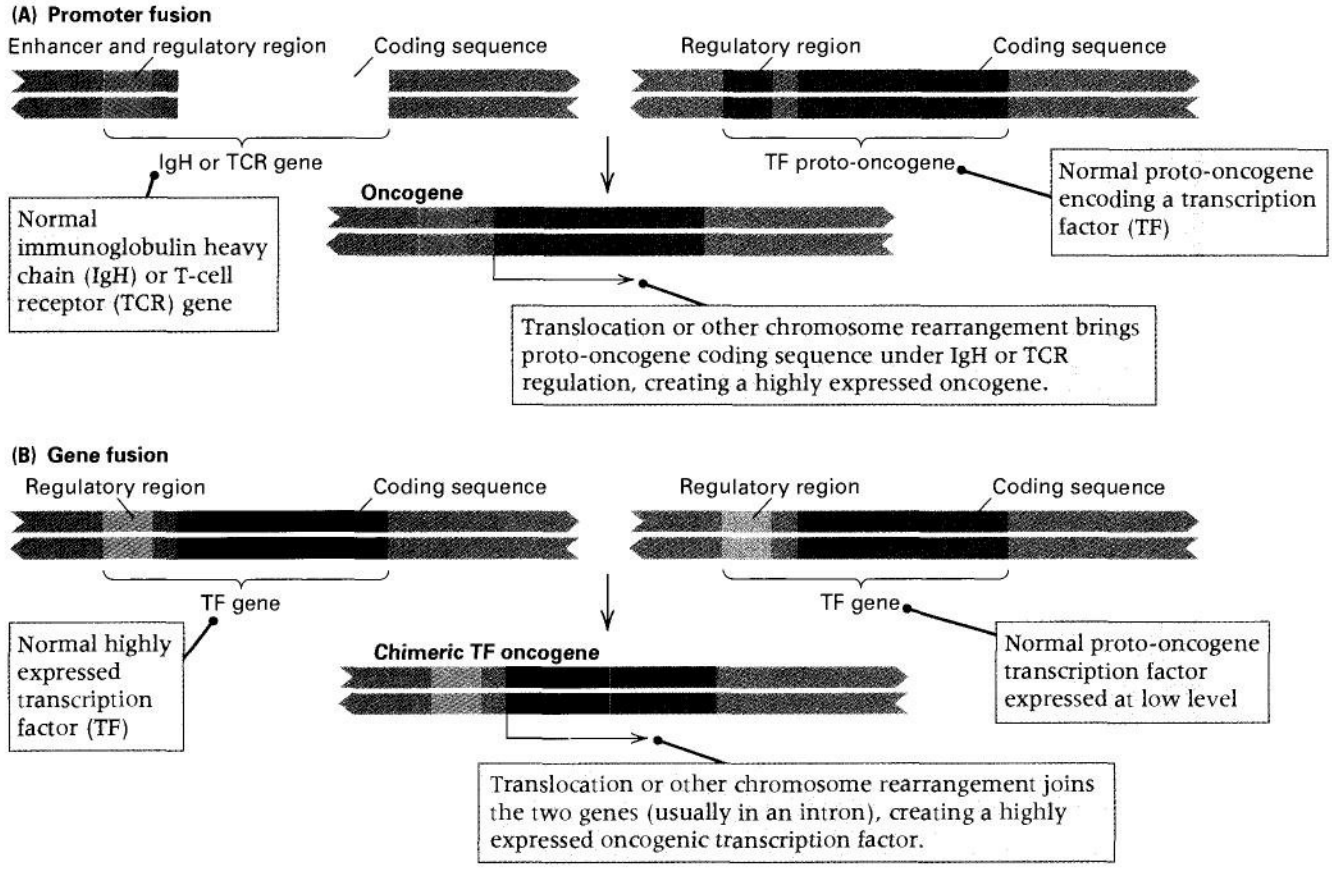


Figure 15.28 Translocations that aberrantly activate transcription factors in acute leukemias. (A) Promoter fusion. (B) Gene fusion.

enhancer. Overexpression of Bcl2 blocks apoptosis by preventing formation of Bax homodimers (Figure 15.18). These "undeservedly alive" lymphocytes proliferate and come to dominate the population of white blood cells, but they are useless in fighting infection because they have not rearranged their immunoglobulin genes properly. In such cases the bone marrow becomes almost totally occupied by the cancerous leukocytes, leading to severe anemia and bleeding. Chemotherapy can sometimes be quite effective in treating acute leukemias in young children by activating the *p53* gene, provided that the aberrant leukocytes still carry the wildtype alleles encoding p53.

The second type of translocation, a **gene fusion** (part B of Figure 15.28), is more common than a promoter fusion. Most commonly, breakpoints occur in introns of two different transcription factor-encoding genes on different chromosomes. The result is a fusion gene that encodes a chimeric transcription factor with altered function, which may have the capacity to interfere with normal hematopoiesis. The uniqueness of these chimeric proteins, and the fact that they are present only in cancer cells and not in normal cells, makes them inviting targets for drugs or chemotherapy. If one could successfully attack cells expressing the chimeric protein, one could selectively kill the cancer cells. For example, in one of the chronic myelogenous leukemias a chimeric protein called BCR-ABL is made. It has been associated with a very high risk of treatment failure. Recent reports suggest that a drug called ST1571, which is specifically directed against BCR-ABL, shows great promise as a therapeutic agent.

Chapter Summary

The essential function of the cell cycle is to ensure that each chromosomal DNA molecule is replicated once in the cell cycle and that the genetically identical daughter chromosomes are distributed equally to daughter cells. The cell cycle comprises four phases: G_1 (gap/growth 1), S (DNA synthesis), G_2 (gap/growth 2), and M (mitosis). Key events in the cycle that ensure proper replication and segregation of chromosomes include centrosome duplication and positioning, replication of the DNA, formation of the bipolar spindle, attachment of each centromere to spindle fibers from each pole, proper functioning of the spindle, and disjunction of the chromosomes.

The cell cycle has been studied genetically in both budding and fission yeast. Conditional mutants defective in the cell division cycle (*cdc* mutants) arrest at particular places in the cell cycle under restrictive conditions. Arrest of the cell cycle in *cdc* mutants takes place at a limited number of points, which correspond to the position of checkpoints within the cell cycle.

Progression from one stage of the cell cycle to the next is controlled by protein complexes called cyclin-dependent kinase (CDK) complexes. These are made up of a cyclin component, which is stage-specific, and a cyclin-dependent protein kinase. The CDK complexes phosphorylate stage-specific protein substrates to change their properties and thus make possible particular reaction sequences. Cyclins are synthesized immediately prior to their use and degraded immediately thereafter.

Prior to the G_1/S transition in animal cells, the cyclin–CDKs phosphorylate the retinoblastoma protein, RB, which changes its conformation and thereby frees the E2F transcription factor, which until this time is rendered inactive by binding with RB. The free E2F activates transcription of enzymes needed for synthesis of DNA. The event responsible for the onset of S phase is phosphorylation of one or more proteins in the protein complex assembled at the origins of replication, and DNA replication ensues.

The G_2/M transition involves phosphorylation of cyclin B, resulting in its retention in the nucleus. Cyclin B–CDK phosphorylates proteins required for assembly of the spindle and dissolution of the nuclear membrane. The onset of anaphase and the exit from mitosis are controlled by proteolysis via activity of the anaphase-promoting complex (APC). Proteolysis ensures that sister chromatids separate and that cyclin B is degraded, propelling the cell into the next G_1 phase.

Cells monitor external and internal conditions. If defects are detected, checkpoint mechanisms halt cell-cycle progression, allowing time for repair or correction. Major checkpoints monitor the DNA in G_1 and S, duplication of the centrosome, and attachment of chromosomes to the spindle.

In animal cells, DNA damage activates a signal transduction pathway that results in phosphorylation and acetylation of p53 protein, preventing its being bound by Mdm2 and shuttled from the nucleus and destroyed by the proteasome. This results in both stabilization of p53 and an increase in the level of p53.

The p53 protein activates transcription of several genes whose products block progression of the cell cycle at the G_1/S and G_2/M boundaries, in effect through inhibition of the cyclin–CDKs at these steps. The p53 protein also activates transcription of the gene for Bax, which propels the cell toward apoptosis. Detection of a defect in attachment of a chromosome to the spindle blocks activation of the anaphase-promoting complex and thereby prevents sister chromatid separation and exit from mitosis.

Cancer cells exhibit uncontrolled proliferation. Most cancers are sporadic; they involve a sequence of genetic changes in somatic cells. Tumors are clonal, deriving from a single cell. The genetic changes in cancer cells include the dominant oncogenes, which derive from normal cellular genes called proto-oncogenes by mutation or overexpression. Other genetic changes include the recessive tumor-suppressor genes. Many of the genetic changes involved in progression from the normal to the cancerous state are changes in cell-cycle regulatory genes or checkpoint genes. These changes often result in genetic instability, which enhances the probability of further changes.

About 1 percent of cancer syndromes are heritable. These often can be traced to germ-line mutations in the same cell-cycle regulatory or checkpoint genes that are mutated in sporadic cancers. Such germ-line changes often show a dominant inheritance pattern at the organismic or pedigree level. At the cellular level, expression of the mutant phenotype is recessive. For expression, the cell must undergo a loss of heterozygosity through any of several mechanisms. Other inherited cancer syndromes can be traced to defects in DNA-repair genes. Defects in mismatch repair show dominant inheritance, whereas defects in nucleotide excision repair show recessive inheritance.

A large fraction of the acute leukemias involve chromosomal translocations of either of two types. In the promoter fusion type of translocation, a proto-oncogene is translocated near the strong enhancer of an immunoglobulin heavy-chain gene (IgH) or a T-cell receptor gene, resulting in high-level expression of the oncogene. In the gene fusion type of translocation, genes for two different transcription factors become fused, giving rise to a chimeric transcription factor.

Key Terms

anaphase-promoting complex (APC/C)
apoptosis
cancer
cell cycle arrest
cell division cycle *(cdc)*
mutant
cell senescence
centriole
centrosome
centrosome duplication checkpoint
checkpoint
contact inhibition
cyclin
cyclin–CDK complex
cyclin-dependent protein kinase (CDK)
DNA damage checkpoint
familial
G protein
G_1 restriction point (start)
G_1/S transition
G_2/M transition
gene fusion
loss of heterozygosity
oncogene
origin recognition complex (ORC)
p53 transcription factor
pre-replication complex
processivity
programmed cell death
promoter fusion
proto-oncogene
retinoblastoma (RB) protein
spindle checkpoint
sporadic
tumor-suppressor gene

Review the Basics

- What role does the centriole (in some organisms, the centrosome) play in cell division?
- What molecular process that takes place in the nucleus defines the S period?
- In yeast, what phenotype defines a *cdc* mutant? For a temperature-sensitive *cdc* mutant grown at the restrictive temperature, why do cells arrest at a particular stage in the cell cycle?
- What does the abbreviation CDK stand for? What is a cyclin–CDK complex? What determines the substrate specificity of a cyclin–CDK complex?
- What is the role of proteolysis in regulating the cell division cycle?
- What is a cell cycle checkpoint? Which three checkpoints are emphasized in this chapter, and what does each "check" for?
- What is apoptosis and what role does it play in preserving the integrity of the genome of a multicellular organism?
- Mutations in genes whose products are involved in DNA repair are often associated with an increased risk of cancer. What does this observation imply about the

role of spontaneous mutations in the development of cancer?

- How can one reconcile the following statements: "Cancer is a genetic disease" and "Most cancers are sporadic (not inherited)."
- Distinguish between proto-oncogenes and tumor-suppressor genes, and give one example of each. In which class of genes does a loss-of-function mutation predispose to cancer? a gain-of-function mutation? Explain your answer.
- What is "loss of heterozygosity," and how is this phenomenon related to the progression of some types of cancer?

Guide to Problem Solving

Problem 1 Distinguish between an oncogene and a tumor-suppressor gene. Why are oncogene mutations associated with cancer gain-of-function mutations, whereas tumor-suppressor mutations associated with cancer are loss-of-function mutations?

Answer Oncogenes are mutant genes whose normal counterparts, called proto-oncogenes, encode products that promote cell proliferation or inhibit apoptosis. The products of tumor-suppressor genes inhibit cell proliferation or activate the apoptotic pathway. Oncogenes are gain-of-function mutations because their enhanced expression makes uncontrolled cell division possible or prevents the apoptosis pathway. Tumor-suppressor mutations associated with cancer are loss-of-function mutations because absence of a functional gene product fails in inhibiting cell division or in activating the apoptosis pathway.

Problem 2 The pedigree shown below includes individuals affected with retinoblastoma, and the diagram of the gel shows a restriction fragment from a number of alleles of the *RB1* gene, mutant forms of which are associated with this cancer. Three sizes of restriction fragments (a, b, and c) are observed. Individuals in generations I and II are old enough that they will have developed the cancer if they carry a mutant *RB1* allele, but the individuals in generation III are all too young to have developed the disease. Identify the high-risk individuals in generation III and those who are not at risk.

Answer First compare the pedigree with the gel patterns to identify which restriction fragment is associated with the mutant *RB1* allele in this pedigree. Because the father in generation I (individual 10) is affected, the mutant allele must be associated with either band a or band b. All of the affected individuals in generation II have band b, so this band must mark the mutant *RB1* allele. Therefore, the individuals at high risk in generation III are those with band b, whereas the others are not at risk. The high-risk individuals are numbers 1, 2, 8, 11, 13, and 18; those not at risk are individuals 3, 6, 7, 12, 16, and 17.

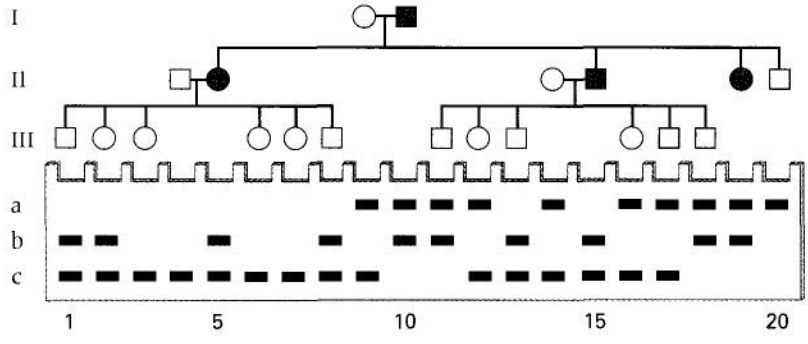
Problem 3 The gel diagram on the next page shows the pattern of bands observed in a Southern blot for a 12-kb *Eco*RI restriction fragment from a nonmutant $RB1^+$ allele and a mutant allele $RB1^-$ in which there is an internal 10-kb deletion in the *Eco*RI fragment. A scale showing the electrophoretic mobility of DNA fragments of various sizes appears at the right. Show what pattern of bands would be expected in:

(a) Normal retinal cells from a heterozygous RB^+/RB^- individual.

(b) Cells of retinoblastoma tumors caused by loss of the wildtype RB^+-bearing chromosome.

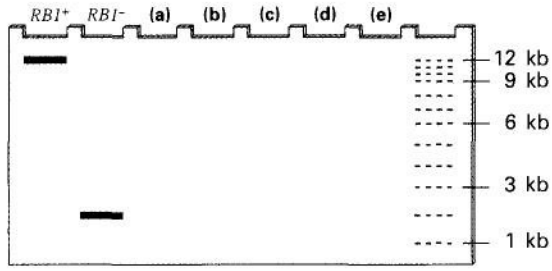
(c) Cells of retinoblastoma tumors caused by mitotic recombination.

(d) Cells of retinoblastoma tumors caused by a new missense substitution in $RB1^+$.



Problem 2

(e) Cells of retinoblastoma tumors caused by a 6-kb deletion in the *Eco*RI fragment (not including the region that hybridizes with the probe).



Answer To solve this type of problem, first think through the consequences of each type of event with respect to the *RB1* alleles present in the cells with loss of heterozygosity.

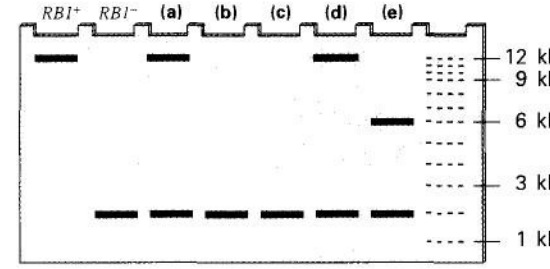
(a) Normal heterozygous cells will contain both DNA fragments.

(b) Loss of the wildtype chromosome eliminates the $RB1^+$ allele.

(c) Mitotic recombination makes the $RB1^-$ allele homozygous in the tumor cells.

(d) A missense mutation in *RB1* does not change the size of the *Eco*RI fragment (except in the extremely unlikely case that the mutation hits one of the *Eco*RI sites).

(e) In the new *RB1* deletion mutation, the size of the *Eco*RI fragment will be 6 kb insted of 12 kb. Therefore, the expected gel patterns are as shown in the illustration.



Analysis and Applications

15.1 What phenotype would you expect of a yeast strain carrying a temperature-sensitive mutation in a gene required for formation of the septum between mother cell and daughter cell:

(a) At the permissive temperature.

(b) At the restrictive temperature.

15.2 A yeast cell carries a temperature-sensitive mutation in a gene necessary to repair double-stranded breaks in DNA. If cells were irradiated at the restrictive temperature, at what stage of the cell division cycle would you expect the mutant cells to accumulate?

15.3 How do cyclin–CDK complexes exert their effects on cells? What counterpart class of enzymes reverses the effects of cyclin–CDK complexes?

15.4 How does the normal retinoblastoma protein function to hold mammalian cells at the G_1 restriction point ("start")?

15.5 Name two proteins that are targets of the anaphase-promoting complex (APC/C), and explain how this complex acts on these proteins.

15.6 What three cell cycle checkpoints were highlighted in this chapter? What event or events cause each checkpoint to be activated to stop the cell division cycle?

15.7 What protein is the major player in activating a DNA damage checkpoint? How is this protein normally kept from triggering the checkpoint? What happens to the protein when there is DNA damage?

15.8 The protein GADD45, for which p53 is a transcription factor, reduces the processivity of the DNA polymerase. What does *processivity* mean? How does a reduction in processivity help in the repair of damaged DNA?

15.9 What are the roles of Bax and Bcl2 proteins in programmed cell death? How is the balance in the amounts of these proteins affected by activated p53? by certain cellular oncogenes?

15.10 If cancer is a "genetic disease," how can it be true that most cases are sporadic (that is, not familial)?

15.11 Li–Fraumeni syndrome is a familial cancer syndrome in which affected relatives show early onset of a diverse set of tumors, including brain, breast, and lung cancers and sarcomas. Nonsense and missense mutations and mutations altering splice donor sites of the *p53* gene have been found segregating in diverse Li–Fraumeni families. In about half of the families showing the syndrome, no mutation in the *p53* coding regions or in splice donor or acceptor sites has been detected. The question is

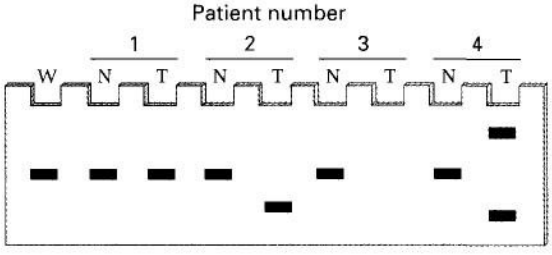
whether, in these families, the syndrome is due to changes in *p53* or to some other cause.

(a) How would you determine whether the mutations in such families are actually associated with the *p53* gene in any way?

(b) If the disease gene is *p53*-associated, what kinds of changes would you look for?

(c) Suppose that the evidence indicates that the segregating mutation is not linked to the *p53* gene. What other possible causes of the syndrome might you suggest? (Remember that inheritance of the syndrome is dominant.)

15.12 You use a cloned cDNA encoding p53 to try to determine what kinds of *p53* mutations are present in several cancer patients by using Southern blot analysis. None of the patients has relatives with cancer, and the tumors are classified as sporadic. DNA is isolated from normal and tumor cells from the four patients, cleaved with *Not*I (an eight-base cutter), subjected to agarose gel electrophoresis, transferred to a nitrocellulose filter, and probed with radioactive *p53* cDNA. The autoradiogram is depicted here. The lane labeled W is the pattern obtained from the wildtype *p53* gene; the other lanes contain DNA from nontumor (N) and tumor (T) tissue from four patients.



Assuming that all four patients do carry *p53* mutations in their tumor cells, what kinds of mutations have occurred in the *p53* genes of these patients?

15.13 Many types of cancer cells have defects in the G_1/S checkpoint. These also tend to have abnormalities in chromosome number or structure. Why would chromosome abnormalities be expected in such cases?

15.14 The p53 protein is defective in more than half of all cancers. Why might this be expected?

15.15 What role does Ras–GTP play in intracellular signalling that makes it a proto-oncogene?

15.16 Draw a diagram showing how recombination between homologous chromosomes during mitosis can result in a cell lineage with loss of heterozygosity for a *p53* mutation. What other genes also lose heterozygosity in this process?

15.17 What does it mean to say that mutations in the retinoblastoma gene *RB1* are "dominant at the organismic level but recessive at the cellular level"?

15.18 In familial retinoblastoma, there is an average of 3 retinal tumors per heterozygous carrier of the mutation. Assuming that the number of retinal cells at risk is 2×10^6 in each eye, and that each tumor results from an independent loss of heterozygosity, what is the estimated rate of loss of heterozygosity per cell? Should this be regarded as a "mutation rate"? Why or why not?

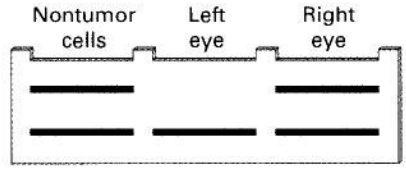
15.19 Considering the data on familial retinoblastoma in the previous problem, what is the probability that an individual who inherits an *RB1* mutation will not develop any retinoblastimas? What is the "penetrance" of familial retinoblastoma?

15.20 DNA from cells of a patient with retinoblastoma was analyzed using a Southern blot with a probe for a particular restriction fragment in the *RB1* gene. The result from nontumor cells and the results from cells taken from a tumor in each eye are shown in the accompanying diagram.

(a) Which band should be associated with the mutant allele and which with the nonmutant allele?

(b) How is it possible for the bands from the tumor in the left eye to be different from those from the tumor in the right eye?

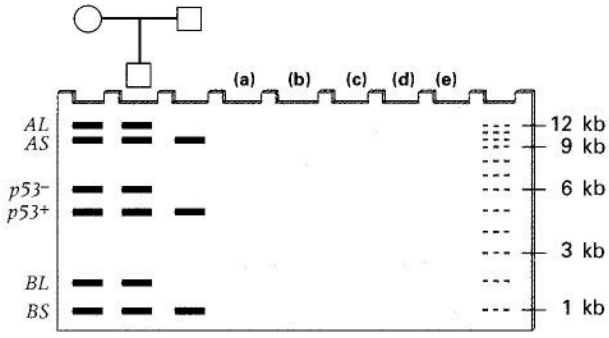
(c) Explain how cells in the tumor in the right eye can have a "loss of heterozygosity" even though the bands are indistinguishable from those observed from nontumor cells.



15.21 In patients with bilateral retinoblastoma, would the mechanism of loss of heterozygosity in tumors in different eyes be expected to be the same or different? Explain your answer.

15.22 The gel diagram on the next page shows the pattern of bands observed in a Southern blot for three linked RFLPs in and flanking the *p53* gene in a family in which Li–Fraumeni syndrome is found. Shown are an affected mother, an unaffected father, and one affected son. The bands AL and AS are for one RFLP (with alleles *AL* and *AS*), BL and BS for another RFLP (with alleles *BL* and *BS*), and $p53^+$ and $p53^-$ the RFLP for *p53*, where the wildtype $p53^+$ allele yields the band $p53^+$, and a mutant *p53* allele, which has a 1-kb insertion in the restriction fragment, yields the band $p53^-$. (The blot was probed with a fragment from each of the three RFLPs, so all three can be visualized in the same blot.) The order of the genes along the chromosome arm is *centromere–A–p53–B–telomere*. Considering the mother and the father, which RFLP alleles are linked with $p53^+$ in the son? If DNA from tumor cells in the son are assayed for these RFLPs, what pattern of bands would be expected from each of the following cells?

(a) Cells that had lost the $p53^+$-bearing chromosome and had a reduplication of the $p53^-$-bearing homolog.
(b) Cells that had undergone mitotic recombination in the *centromere–A* interval.
(c) Cells that had undergone mitotic recombination in the *A–p53* interval.
(d) Cells that had a new nonsense mutation in the $p53^+$ allele.
(e) Cells that had undergone gene conversion of $p53^+$ to $p53^-$.



15.23 The pedigree shown below includes individuals affected with adenomatous polyposis, and the diagram of the gel shows a restriction fragment from a number of alleles of *APC*, mutant forms of which are associated with this cancer. Four sizes of restriction fragments (a through d) are observed. Individuals in generations I and II are old enough that they will have developed the cancer if they carry a mutant *APC* allele, but the individuals in generation III are all too young to have developed the disease. Identify the high-risk individuals in generation III and those who are not at risk. (Note that a mutant allele and a nonmutant allele can yield the same size restriction fragment.)

15.24 Mutagenesis of a *RAS* gene of budding yeast yields a temperature-sensitive conditional mutation.
(a) Would you expect a cell that is carrying a mutation that prevents Ras from exchanging GTP for GDP at 36°C to continue to divide at the restrictive temperature?
(b) Would this mutation be dominant or recessive?
(c) Would you expect a cell that is carrying a mutation that inactivates the GTPase activity of Ras at 36°C to continue to divide at the restrictive temperature?
(d) Would this mutation be dominant or recessive?

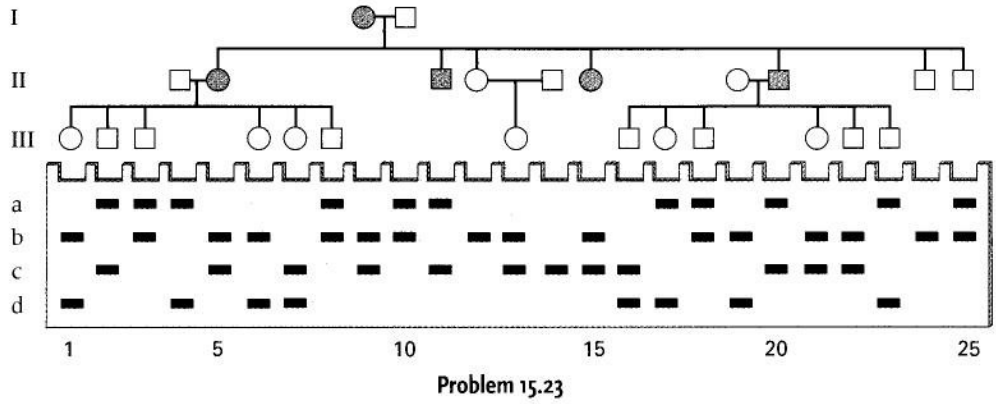
15.25 You are studying a particular leukemia in humans that is associated with a translocation between chromosomes 8 and 14. How would you determine experimentally whether the translocation is likely to be of the promoter-fusion or the gene-fusion type?

15.26 Human papilloma virus (HPV) is present in greater than 90% of cervical cancers. HPV encodes two proteins, E6 and E7, that are potent contributors to its tumorigenicity. E7 is known to disable RB; E6 binds to p53 and targets it for degradation. Discuss how these activities might contribute to the development of the cancerous state in infected cells.

15.27 A number of *cdc* mutants of budding yeast that arrest at the G_2/M boundary were recovered. They were crossed in all pairwise combinations, and the diploids were tested for the ability to grow at 36°C. The results are given below.

	1	2	3	4	5	6	7	8	9	10
1	−	+	+	+	+	+	+	+	+	+
	2	−	+	+	+	−	+	+	+	+
		3	−	+	+	+	+	+	+	+
			4	−	−	+	+	−	+	+
				5	−	+	+	−	+	+
					6	−	+	+	+	+
						7	−	+	−	+
							8	−	+	+
								9	−	+
									10	−

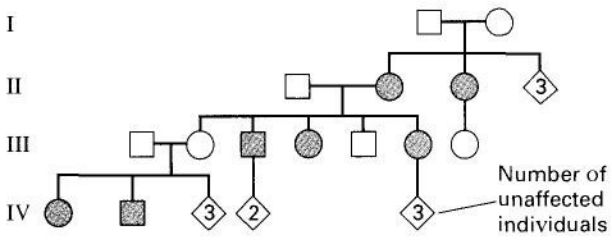
(a) How many genes are represented among these 10 mutants?
(b) You cross the 10 mutants to known *cdc13* and *mob1* mutants that arrest at the G_2/M boundary, test these diploids for growth at 36°C, and obtain the results shown on the next page. Interpret these results in light of those of part **(a)**.



Problem 15.23

	1	2	3	4	5	6	7	8	9	10
cdc13	+	+	+	−	−	+	+	−	+	+
mob1	+	+	+	+	+	+	−	+	−	+

15.28 Among 24 kindreds that manifested a syndrome of sarcoma, breast cancer, and other cancers in young patients, and that showed autosomal dominant inheritance, one family had the following pedigree.



If inheritance is due to an autosomal dominant gene, which individuals must be heterozygous for the disease allele but nevertheless be cancer-free?

15.29 *cdc9* mutants of budding yeast are defective for DNA ligase activity at 36°C but have normal activity at 23°C. If *cdc9* cells grown at 23°C are transferred to 36°C, they arrest at the G_2/M boundary. Upon return to 23°C, they resume the cell cycle after a delay during which they ligate all the nicks in the newly synthesized DNA. The cells are viable. *rad9* mutants are defective in the G_2/M DNA checkpoint. The *rad9 cdc9* double mutants grow normally at 23°C. If transferred to 36°C, the cells continue through mitosis. Upon plating at 23°C, however, most of the cells prove to be inviable. If, prior to transfer to 36°C and throughout incubation at 36°C, the double mutant is incubated with benzimidazole, a microtubule poison akin to colchicine, then upon return to 23°C accompanied by washout of the benzimidazole, the cells behave like the *cdc9* mutant cells. Propose an explanation for these observations.

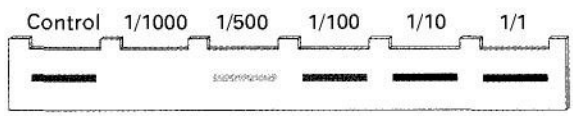
15.30 The properties of the *rad9 cdc9* double mutants led Weinert and Hartwell to investigate whether other *cdc rad9* double mutants continued to cycle and lose viability upon transfer to 36°C, in order to determine the specificity of the *RAD9* checkpoint. The results of these analyses are summarized in the table below:

Double mutant	Function of the *CDC* gene	Dies rapidly at 36°C?
cdc28 rad9	Cyclin-dependent kinase	No
cdc7 rad9	Spindle pole body duplication (centrosome)	No
cdc8 rad9	Thymidylate kinase (TTP synthesis)	No
cdc6 rad9	ORC component	No
cdc2 rad9	DNA polymerase subunit	Yes
cdc13 rad9	Telomere-binding protein	Yes
cdc17 rad9	DNA polymerase subunit	Yes
cdc16 rad9	Anaphase-promoting complex (APC/C) component	No
cdc20 rad9	Required for exit from mitosis	No
cdc23 rad9	APC/C component	No
cdc15 rad9	Required for degradation of B cyclins	No

What would you deduce about *RAD9*'s role in the several checkpoints in the yeast cell cycle?

Challenge Problems

15.31 The human oncogene *MYC* in chromosome 8 (near 8q24) was first discovered in association with Burkitt lymphoma. The diagram illustrates the result of a Southern blot obtained when DNA from noncancer cells (the control) was compared with various dilutions (up to 1000-fold) of DNA isolated from an equivalent number of cancer cells from the same patient. The blot was hybridized with a labeled probe for *MYC*. The dilutions clearly differ in the intensity (darkness) of the hybridization signal. Propose a hypothesis to account for these results.



15.32 The cDNAs for mammalian Cdk2 and Cdc2 can substitute for the *CDC28* gene in budding yeast. Design an experiment to test whether the cDNAs for human cyclin B and cyclin D can substitute for Clb2 in budding yeast.

15.33 You are given a diploid strain of budding yeast and are told that it is homozygous for a mutation that inactivates a checkpoint. For chromosome V, the strain has the genotype



The *can1* mutation results in recessive resistance to canavanine, *ura3* results in a recessive requirement for uracil and resistance to 5-fluoro-orotic acid, and *his1* results in a recessive requirement for histidine.

(a) What property would you expect the strain to exhibit (for these genetic markers) if it were homozygous for a mutation that inactivates the spindle checkpoint?

(b) What property would you expect the strain to show if it were homozygous for a mutation affecting a DNA checkpoint?

GeNETics on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes,* choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

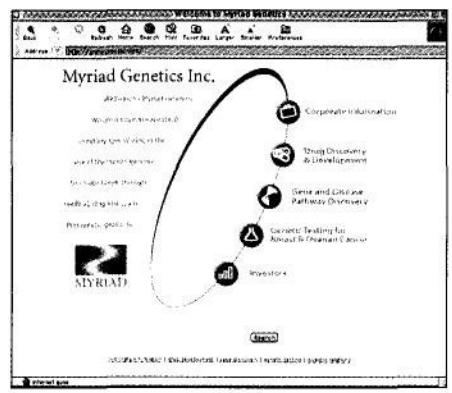
• Among 400 people, about 200 will develop at least one tumor of the colon (large bowel), 20 of these cases will progress to malignancy, and 1 will be associated with an inherited genetic abnormality. Among the inherited forms of **colon cancer**, two important types are familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC). They demonstrate the difference between tumor initiation and tumor progression. Patients with FAP develop thousands of colon polyps, only a few of which progress to cancer, whereas patients with HNPCC have only a few polyps, but each has a high probability of progressing to cancer. Consult this keyword site to learn more about these and other types of colon cancer.

• Mutations in the human genes ***BRCA1*** and ***BRCA2*** are major genetic risk factors for hereditary breast cancer in women and prostate cancer in men. Among carriers of either mutation, the risk of breast cancer by age 70 is 56 percent, that of ovarian cancer 16 percent, and that of prostate cancer 16 percent. The frequency of carriers differs among population groups. Among Ashkenazi Jews, about 1 person in 50 carries a mutation in either *BRCA1* or *BRCA2*. These keyword sites contain more information on the genetics and molecular biology of these genes.

• In the United States the frequency of **retinoblastoma** is about 1 in 23,000 live births. It is a disease of young children, because the great majority of cases are diagnosed prior to the age of 5. In many cases there are treatment options other than surgical removal of the eye. These include cryotherapy to destroy the cancer cells with extreme cold, laser light to target the tumors and the surrounding blood vessels that feed them, and radiation and chemotherapy. This keyword site aims to provide information and links for anyone interested in any aspect of the disease.

• The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 15 and you will be linked automatically.

• The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 15.



Further Reading

Cahill, D. P., K. W. Kinzler, B. Vogelstein, and C. Lengauer. 1999. Genetic instability and Darwinian selection in tumours. *Trends in Biochemical Sciences* 24: M57.

Cavenee, W. K., and R. L. White. 1995. The genetic basis of cancer. *Scientific American*, March.

Chin, L., and R. A. DePinho. 2000. Flipping the oncogene switch: Illumination of tumor maintenance and regression. *Trends in Genetics* 16: 147.

Clarke, D. J. and J. F. Gimenez-Abian. 2000. Checkpoints controlling mitosis. *Bioessays* 22: 351.

Cox, E. C. 1997. *mutS*, proofreading, and cancer. *Genetics* 146: 443.

Fearon, E. R. 1997. Human cancer syndromes: Clues to the origin and nature of cancer. *Science* 278:1043.

Fearon, E. R., and C. V. Dang. 1999. Cancer genetics: Tumor suppressor meets oncogene. *Current Biology* 9: R62.

Golub, T. R., D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, A. Caligiuri, C. D. Bloomfield, and E. S. Lander. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286: 531.

Hartwell, L. H., and M. B. Kastan. 1994. Cell cycle control and cancer. *Science* 266: 1821.

Hartwell, L. H., and T. A. Weinert. 1989. Checkpoints: Controls that ensure the order of cell cycle events. *Science* 246: 629.

Jacks, T. 1996. Tumor suppressor gene mutations in mice. *Annual Review of Genetics* 30: 603.

Kaelin, W. G. 1999. Functions of the retinoblastoma protein. *Bioessays* 21: 950.

Koepp, D. M., J. W. Harper, and S. J. Elledge. 1999. How the cyclin became a cyclin: Regulated proteolysis in the cell cycle. *Cell* 97: 431.

Levine, A. J. 1997. p53, the cellular gatekeeper for growth and division. *Cell* 88: 323.

Lew, D. J., T. A. Weinert, and J. R. Pringle. 1997. Cell cycle control in *Saccharomyces cerevisiae*. In *The Molecular and Cellular Biology of the Yeast Saccharomyces*. Vol. 3: *Cell Cycle and Cell Biology*, ed. J. R. Pringle, J. R. Broach, and E. W. Jones. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, p.67.

Look, A. T. 1997. Oncogenic transcription factors in the human acute leukemias. *Science* 278: 1059.

MacNeill, S. A., and P. Nurse. 1997. Cell cycle control in fission yeast. In *The Molecular and Cellular Biology of the Yeast Saccharomyces*. Vol. 3: *Cell Cycle and Cell Biology*, ed. J. R. Pringle, J. R. Broach, and E. W. Jones. Cold Spring Harbor NY: Cold Spring Harbor Laboratory Press, p. 697.

McCormick, F. 1999. Signalling networks that cause cancer. *Trends in Biochemical Sciences* 24: M53.

Murray, A. W., and T. Hunt. 1993. *The Cell Cycle: An Introduction*. New York: Oxford University Press.

Rahman, N., and M. R. Stratton. 1998. The genetics of breast cancer susceptibility. *Annual Review of Genetics* 32: 95.

Song, Z. W., and H. Steller. 1999. Death by design: Mechanism and control of apoptosis. *Trends in Biochemical Sciences* 24: M49.

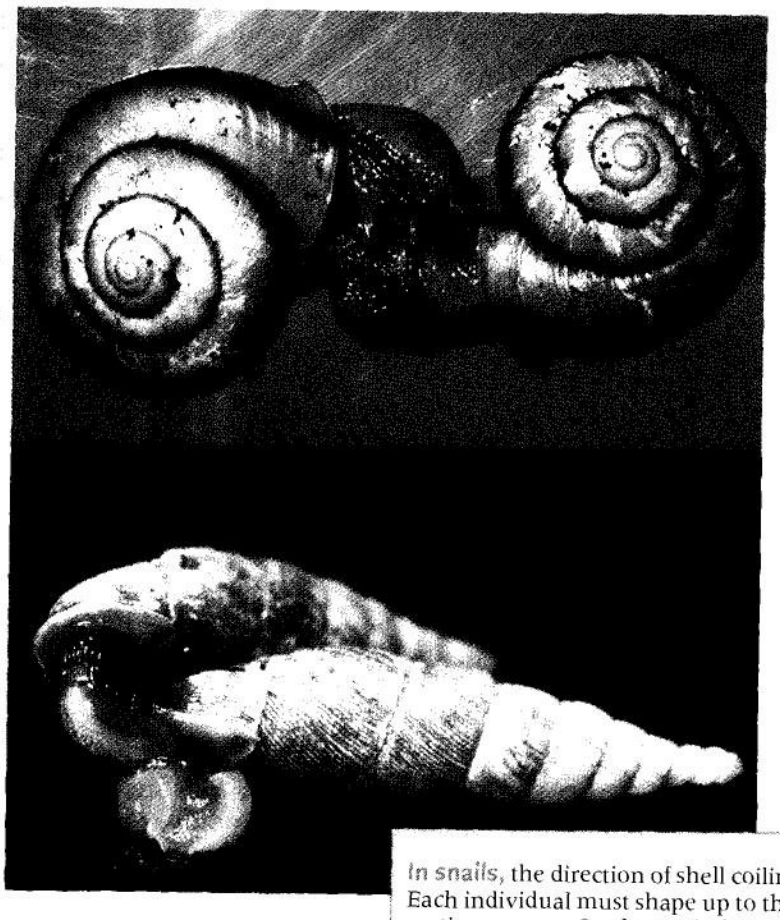
Vogelstein, B., and K. W. Kinzler., eds. 1998. *The Genetic Basis of Human Cancer*. New York: McGraw-Hill.

Welcsh, P. L., K. N. Owens, and M. C. King. 2000. Insights into the functions of BRCA1 and BRCA2. *Trends in Genetics* 16: 69.

Weinberg, R. A. 1996. How cancer arises. *Scientific American*, September.

CHAPTER 16

Extranuclear Inheritance



In snails, the direction of shell coiling is genetically determined. Each individual must shape up to the requirements of its mating partner. On the top is the species *Euhadra congenita*, which has right-handed shell coiling; the genital organs open on the right side of the body, so a "face-to-face" orientation allows mating to take place. On the bottom is the species *Stereophaedusa gouldi*, which has left-handed shell coiling; the genital organs open on the left side of the body, so in this case a mounted posture is necessary for mating to take place. [Courtesy of Takahiro Asami.]

CHAPTER OUTLINE

16.1 Patterns of Extranuclear Inheritance
Mitochondrial Genetic Diseases
Heteroplasmy
Maternal Inheritance and Maternal Effects

16.2 Organelle Heredity
RNA Editing
The Genetic Codes of Organelles
Leaf Variegation in Four-O'clock Plants
Drug Resistance in *Chlamydomonas*
Respiration-Defective Mitochondrial Mutants
Cytoplasmic Male Sterility in Plants

16.3 The Evolutionary Origin of Organelles

16.4 Cytoplasmic Transmission of Symbionts

16.5 Maternal Effect in Snail Shell Coiling

16.6 In Search of Mitochondrial "Eve"

PRINCIPLES

- Some inherited traits are determined by DNA molecules present in organelles, principally mitochondria and chloroplasts. A hallmark of such traits is uniparental inheritance, either maternal or paternal.
- Progeny resembling the mother is characteristic of both maternal inheritance and maternal effects. Although maternal inheritance results from the transmission of non-nuclear genetic factors, maternal effects result from nuclear genes expressed in the mother that determine the phenotype of the progeny.
- Mitochondria, photosynthetic plastids, and certain other cellular organelles are widely believed to have originated through independent episodes of symbiosis in which a prokaryotic cell, the progenitor of the present organelle, became an intracellular symbiont of the proto-eukaryotic cell.
- Evidence from DNA sequence variation in human mitochondrial DNA suggests that the mitochondrial DNA present in all modern human beings originated from a single female who lived approximately 150,000–300,000 years ago, probably in Africa.

CONNECTIONS

***Chlamydomonas* Moment**
Ruth Sager and Zenta Ramanis 1965
Recombination of Nonchromosomal Genes in Chlamydomonas

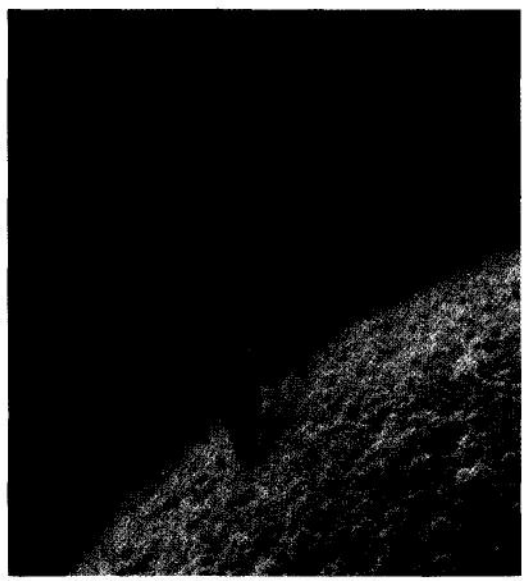
A Coming Together
Lynn Margulis (formerly Lynn Sagan) 1967
The Origin of Mitosing Cells

Although most hereditary traits in eukaryotes are determined by genes in the nucleus, some are not. The exceptions include traits determined by the DNA in mitochondria and in chloroplasts, the self-replicating cytoplasmic **organelles** specialized for respiration (mitochondria) and for photosynthesis (chloroplasts). Each of these organelles contains its own DNA that codes for numerous proteins and RNA molecules. The DNA is replicated within the organelles and transmitted to daughter organelles as they form. The organelle genetic system is separate from that of the nucleus, and traits determined by organelle genes show patterns of inheritance quite distinct from those of nuclear genes.

Many organisms also contain intracellular parasites or symbionts, including cytoplasmic bacteria, viruses, and other elements. In some cases, such cytoplasmic entities confer inherited phenotypic traits on the infected cell or organism. Organelle heredity and other examples of the diverse phenomena that make up **extranuclear inheritance** (also called **cytoplasmic inheritance**) are presented in this chapter.

16.1 Patterns of Extranuclear Inheritance

No criterion other than a non-Mendelian pattern of inheritance is universally applicable to distinguish extranuclear from nuclear inheritance. In higher organisms, extranuclear inheritance is usually indicated by **uniparental inheritance,** which means transmission through only one parent. Genetic transmission of cytoplasmic factors, such as mitochondria or chloroplasts, is determined by maternal and paternal contributions at the time of fertilization, by mechanisms of elimination from the zygote, and by the irregular sorting of cytoplasmic elements during cell division. Uniparental transmission through the mother constitutes **maternal inheritance;** uniparental transmission through the father is **paternal inheritance.**



A human sperm and egg. The volume of the egg cell is about 5000 times the volume of the sperm head and contributes virtually all of the cytoplasm to the zygote, including the mitochondria. [D. W. Fawcett/Photo Researchers, Inc.]

Mitochondrial Genetic Diseases

In higher animals, mitochondria are typically inherited through the mother because the egg is the major contributor of cytoplasm to the zygote. Therefore, mitochondria usually show maternal inheritance: The progeny of a mutant mother and a normal father are mutant, whereas the progeny of a normal mother and a mutant father are normal. A typical pattern of maternal inheritance of mitochondria is shown in the human pedigree in **Figure 16.1**. When human mitochondrial DNA is cleaved with the restriction enzyme *Hae*II, the cleavage products include either one fragment of 8.6 kb or two smaller fragments of 4.5 kb and 4.1 kb (Figure 16.1A). The pattern with two smaller fragments is typical, and it results from the presence of a *Hae*II cleavage site within the 8.6-kb fragment. Maternal inheritance of the 8.6-kb mitochondrial DNA fragment is indicated in Figure 16.1B. The key feature is that females (I-3, I-5, and II-8) transmit the pattern to *all* of their progeny, whereas males (I-2, II-7, and II-10) transmit the pattern to *none* of theirs. Although the mutation in the *Hae*II site yielding the 8.6-kb fragment is not associated with any disease, a number of other mutations in mitochondrial DNA do cause diseases and exhibit similar patterns of mitochondrial inheritance. Most of these conditions decrease the ATP-generating capacity of the mitochondria and affect the function of muscle and nerve cells, particularly in the central nervous system, leading to blindness, deafness, or stroke. **Table 16.1** lists some of the human diseases caused by

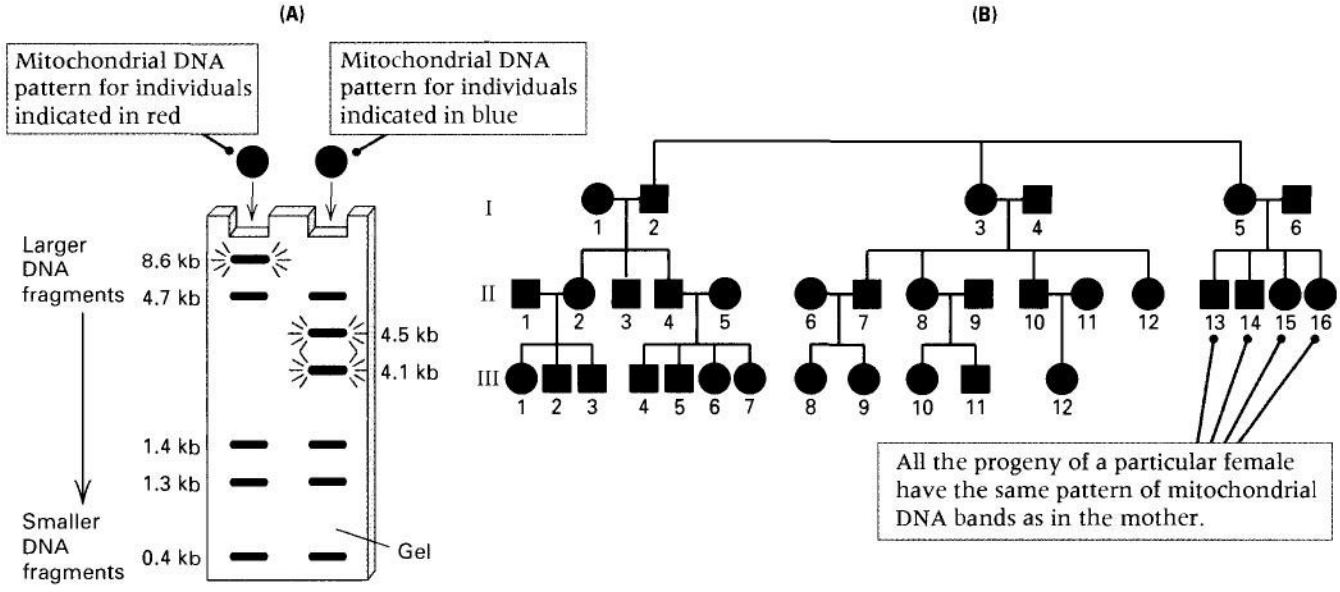


Figure 16.1 Maternal inheritance of human mitochondrial DNA. (A) Pattern of DNA fragments obtained when mitochondrial DNA is digested with the restriction enzyme *Hae*II. The DNA type at the left includes a fragment of 8.6 kb (red). The DNA type at the right contains a cleavage site for *Hae*II within the 8.6-kb fragment, which results in smaller fragments of 4.5 kb and 4.1 kb (blue). (B) Pedigree showing maternal inheritance of the DNA pattern with the 8.6-kb fragment (red symbols). The mitochondrial DNA type is transmitted only through the mother. [After D. C. Wallace. 1989. *Trends in Genetics* 5: 9.]

Table 16.1 Phenotypes associated with some mitochondrial mutations

Nucleotide changed	Mitochondrial component affected	Phenotype[a]
3460	ND1 of Complex I[b]	LHON
11778	ND4 of Complex I	LHON
14484	ND6 of Complex I	LHON
8993	ATP6 of Complex V[b]	NARP
3243	$tRNA^{Leu(UUR)}$[c]	MELAS, PEO
3271	$tRNA^{Leu(UUR)}$	MELAS
3291	$tRNA^{Leu(UUR)}$	MELAS
3251	$tRNA^{Leu(UUR)}$	PEO
3256	$tRNA^{Leu(UUR)}$	PEO
5692	$tRNA^{Asn}$	PEO
5703	$tRNA^{Asn}$	PEO, myopathy
5814	$tRNA^{Cys}$	Encephalopathy
8344	$tRNA^{Lys}$	MERRF
8356	$tRNA^{Lys}$	MERRF
9997	$tRNA^{Gly}$	Cardiomyopathy
10006	$tRNA^{Gly}$	PEO
12246	$tRNA^{Ser(AGY)}$[c]	PEO
14709	$tRNA^{Glu}$	Myopathy
15923	$tRNA^{Thr}$	Fatal infantile multisystem disorder
15990	$tRNA^{Pro}$	Myopathy

[a]LHON Leber's hereditary optic neuropathy; NARP Neurogenic muscle weakness, ataxia, retinitis pigmentosa; MERRF Myoclonic epilepsy and ragged-red fiber syndrome; MELAS Mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes; PEO Progressive external ophthalmoplegia

[b]Complex I is NADH dehydrogenase. Complex V is ATP synthase.

[c]In $tRNA^{Leu(UUR)}$, the R stands for either A or G; in $tRNA^{Ser(AGY)}$, the Y stands for either T or C.

mitochondrial mutations. Many of these conditions are lethal unless some of the mitochondria are normal, and there is variable expressivity because the proportions of normal and mutant mitochondria differ among affected persons.

An example of a human pedigree for inheritance of a mitochondrially inherited disease is shown in **Figure 16.2**. This disease, which is known as MERRF (myoclonic epilepsy associated with ragged-red muscle fibers), affects the central nervous system and skeletal muscle. The cause is a mutation in a $tRNA^{Lys}$ gene encoded in mitochondrial DNA (mtDNA). **Figure 16.3** shows the organization of genes in human mtDNA. The $tRNA^{Lys}$ mutation is highly pleiotropic because it affects the synthesis of all proteins encoded in the mitochondrial DNA in all cells. Like many mitochondrial mutations, MERRF is lethal in the absence of normal mitochondria. All of the individuals indicated by red symbols in the pedigree in Figure 16.2B manifest some symptoms of the disease, which include hearing loss, seizures, fatigue, dementia, tremors, and jerkiness of movement. The affected individuals in the pedigree vary considerably in the symptoms they express. The variable expressivity is due to differences in the proportions of normal and mutant mitochondria in the cells of affected persons. Those most severely affected have the highest proportion of mutant mitochondria; those least affected have the lowest, as indicated in Figure 16.2B. An electron micrograph of an abnormal mitochondrion from one of the affected patients is shown in Figure 16.2A. It contains numerous paracrystalline inclusion bodies formed from abnormal protein aggregates.

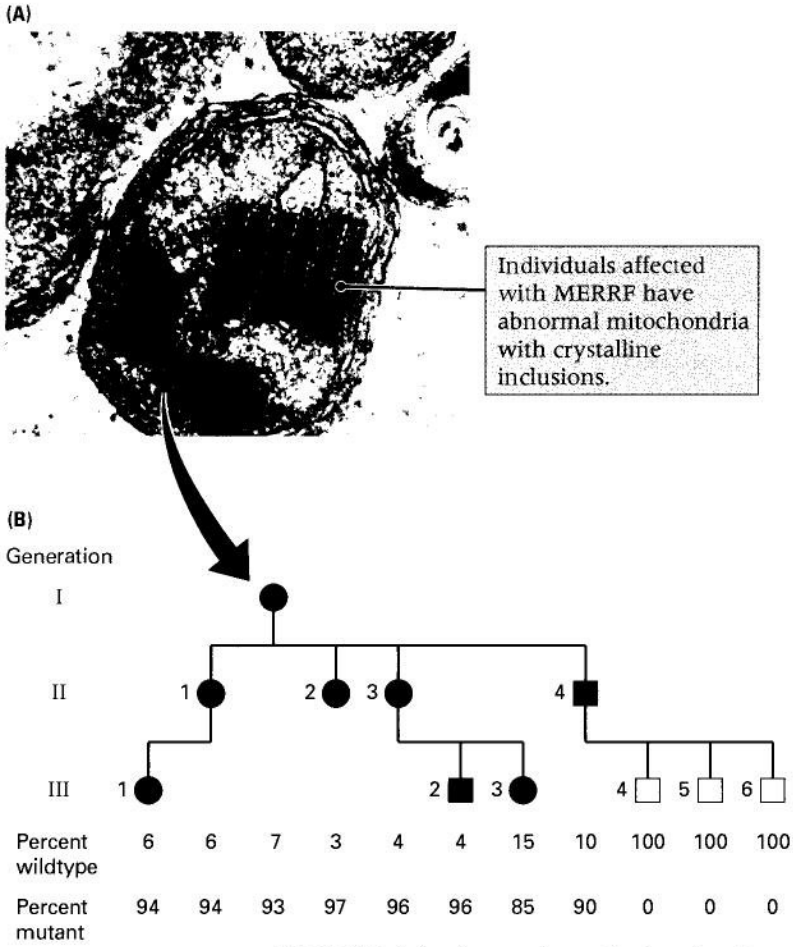


Figure 16.2 Inheritance of myoclonic epilepsy with ragged-red fiber disease (MERRF) in humans. (A) Electron micrograph of an abnormal MERRF mitochondrion containing paracrystalline inclusions. (B)The pedigree shows inheritance of MERRF in one family and the percentage of the mitochondria in each person found to be wildtype or mutant. [Micrograph courtesy of D. C. Wallace, from J. M. Shoffner, M. T. Lott, A. M. S. Lezza, P. Seibel, S. W. Ballinger, and D. C. Wallace. 1990. *Cell* 61: 931.]

Consideration of the phenotypes caused by the mutations listed in Table 16.1 and the variable expressivity seen in pedigrees of mitochondrial diseases indicates that mutations in different mitochondrial genes can result in similar phenotypes and that mutations in the same mitochondrial gene can result in different phenotypes. Some of the variation appears to arise from differences among affected persons in the ratio of mutant to wildtype mitochondria, as seen in Figure 16.2B. Another source of variation results from differences in the structural and metabolic consequences of each individual mutation.

Heteroplasmy

The condition in which two or more genetically different types of mitochondria (or other organelle) are present in the same cell is known as **heteroplasmy.** Heteroplasmy of mitochondria is unusual among animals. For example, in a typical human cell containing 1000–10,000 mitochondria, only one mtDNA sequence is usually represented.

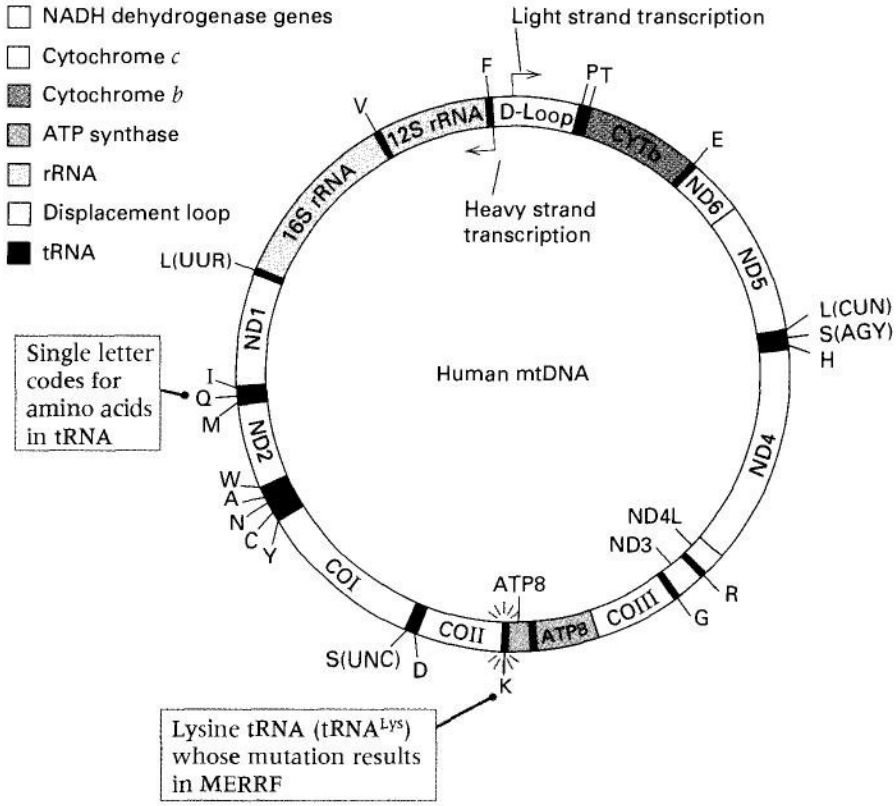


Figure 16.3 Genes in human mitochondrial DNA. The tRNA genes are indicated by the one-letter amino acid symbols; hence $tRNA^{Lys}$ is denoted K. The positions of these and other genes in the mitochondrial DNA are indicated by color according to the key at the upper left. The arrows indicate the promoters for transcription of the heavy and light strands. [Courtesy of N-G. Larsson and D. A. Clayton. With permission, from the *Annual Review of Genetics* 29: 151. Copyright 1995 by Annual Reviews, www.AnnualReviews.org.]

The rarity of heteroplasmy made it possible to identify the remains of the last Russian tsar, Nicholas II, who, together with his family and servants, was secretly executed by a Bolshevik firing squad on the night of July 16, 1918. Their bodies were hidden in a shallow grave. DNA analysis of bone samples exhumed in 1992 and thought to be from members of the murdered family resulted in identification of nine bodies, five of whom were related. The mother, Tsarina Alexandra, and three of her daughters were identified because their mitochondrial DNA was identical to that of the present Duke of Edinburgh, the maternal grandnephew of the tsarina. The father, Tsar Nicholas, proved to be heteroplasmic for mitochondrial DNA. To establish this identification, the body of the tsar's brother, Georgij, who had died of tuberculosis in 1899, was exhumed. The mitochondrial DNA sequence of Georgij Romanov not only matched that of the tsar but was also heteroplasmic at the identical base pair. The two brothers differed in the ratios of the two mitochondrial types: Georgij Romanov had about 62 percent T and 38 percent C at the heteroplasmic base pair, whereas Nicholas had 28 percent T and 72 percent C. The heteroplasmy is not present in two living members of this matrilineage, who are separated by two or three additional generations from the brothers. This suggests that random segregation of mitochondria within the lineage has finally resulted in homoplasmy.

Maternal Inheritance and Maternal Effects

A maternal pattern of inheritance usually indicates extranuclear inheritance, but not always. The difficulty is in distinguishing

maternal inheritance from *maternal effects*. A **maternal effect** is an influence of the mother's nuclear genotype on the phenotype of the progeny. Such effects may be mediated either through substances present in the egg that affect early development or through nurturing. Examples of developmental and nurturing effects are, respectively, the intrauterine environment of female mammals and their ability to produce milk. The distinction between maternal inheritance and maternal effects is as follows:

1. In *maternal inheritance*, the hereditary determinants of a trait are extranuclear, and genetic transmission is only through the maternal cytoplasm; Mendelian segregation is not observed.
2. In a *maternal effect*, the nuclear genotype of the mother determines the phenotype of the progeny. The hereditary determinants are nuclear genes transmitted by both sexes and, in suitable crosses, undergo Mendelian segregation.

In higher plants, organelle inheritance is often uniparental, but the particular type of uniparental inheritance depends on the organelle and the organism. For example, among angiosperms, mitochondria typically show maternal inheritance, but chloroplasts may be inherited maternally (the predominant mode), paternally, or from both parents, depending on the species. In conifers, chloroplast DNA is usually inherited paternally, but a few offspring result from maternal transmission. The redwood *Sequoia sempervirens* shows paternal transmission of both mitochondrial and chloroplast DNA. Whatever the pattern of cytoplasmic transmission, most species have a few exceptional progeny, which indicates that the predominant mode of transmission is not absolute.

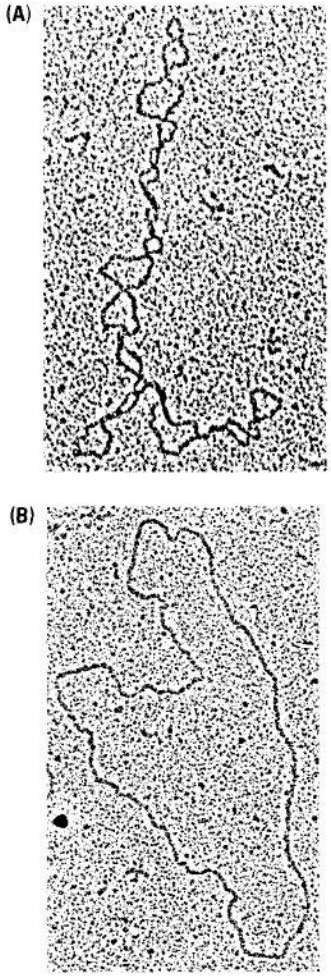


Figure 16.4 Mitochondrial DNA from rat liver. (A) Native supercoiled configuration. (B) Relaxed configuration produced by a nick in the DNA strand in the process of isolation. The size of each molecule is 16,298 nucleotide pairs. [Courtesy of David Wolstenholme.]

16.2 Organelle Heredity

As noted earlier, mitochondria are respiratory organelles, and chloroplasts are photosynthetic organelles. The DNA of these organelles is usually in the form of supercoiled circles of double-stranded molecules (**Figure 16.4**). The chloroplast DNA in most plants ranges in size from 120 to160 kb. Mitochondrial genomes are usually smaller and have a greater size range. For example, the mitochondrial genome in mammals is about 16.5 kb, that in *D. melanogaster* about 18.5 kb, and that in some higher plants more than 100 kb. The mitochondrial genomes of higher plants are exceptional in consisting of two or more circular DNA molecules of different sizes. Several copies of the genome are present in each organelle. There are typically from 2 to 10 copies in mitochondria and from 20 to 100 copies in chloroplasts. In addition, most cells contain multiple copies of each organelle.

Organelle genes code for DNA polymerases that replicate the organelle DNA. They also code for other components essential to their function or to replication, but many organelle components are determined by nuclear genes. These components are synthesized in the cytoplasm and imported into the organelle. Mitochondrial DNA contains a relatively small number of genes. For example, both the human mitochondrial genome and that of yeast contain approximately 40 genes. Chloroplast genomes are generally larger than those of mitochondria and contain more genes. Many genes in mitochondria and chloroplasts code for the ribosomal RNA and transfer RNA components used in protein synthesis. **Figure 16.5** shows the genetic functions of the genes in the 120-kb chloroplast genome of the liverwort *Marchantia polymorpha*.

RNA Editing

The mRNAs of mitochondria and chloroplasts often are modified posttranscriptionally in processes called **RNA editing.** The mechanisms of modification as well as the predominant patterns of modification differ between plants and animals. In most land plants, except algae and mosses, mRNAs may be extensively modified by conversion of cytosine residues to uracils. As a result of RNA editing, a CGG codon for arginine present in an organelle gene is changed into a UGG codon for tryptophan in the mRNA. The modification appears to be accomplished by deamination of selected cytosine residues in the mRNA, leaving uracil residues in their places. In the mitochondria of the protozoans that cause sleeping sickness (*Trypanosoma brucei*), the insertion or deletion of uridine residues is directed by a guide RNA in a process that resembles mRNA splicing (Chapter 11). Whatever the mechanism, the outcome is an mRNA whose sequence may differ substantially from that of the DNA from which it was transcribed.

The Genetic Codes of Organelles

Translation in organelles follows the standard decoding process in which the organelle tRNA molecules translate codon by codon along the mRNA. However, the genetic code in animal mitochondria often differs somewhat from that in nuclear genes, prokaryotes, and archaeans. Human mitochondria depart from the standard code in three principal ways:

- The UGA codon is not a stop codon but encodes tryptophan instead.
- The AGA and AGG codons are stop codons rather than arginine codons.
- AUA encodes methionine rather than isoleucine.

The genetic code in yeast mitochondria is identical to the standard code except that UGA is a tryptophan codon and all four codons of the form CUN code for threonine rather than leucine. In animals, the mito-

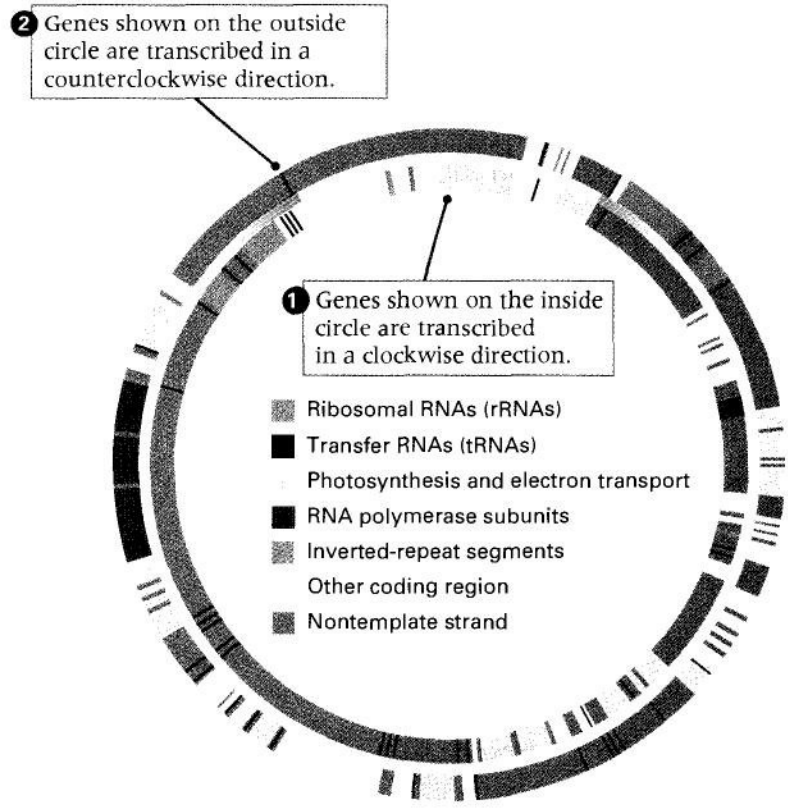


Figure 16.5 Organization of genes in the chloroplast genome of the liverwort *Marchantia polymorpha*. The large inverted-repeat segments (green) that include the genes coding for ribosomal RNA (light brown) are a feature of many other chloroplast genomes. The red lines show the locations of the genes for transfer RNA. Yellow indicates the genes for components of the photosynthetic and electron transport systems; purple indicates the genes for subunits of RNA polymerase. Genes shown inside the circle are transcribed in a clockwise direction; those outside are transcribed counterclockwise. The length of the entire molecule is 121,024 nucleotide pairs. [Data from K. Ohyama, H. Fukuzawa, T. Kohchi, H. Shirai, T. Sano, S. Sano, K. Umesono, Y. Shiki, M. Takeuchi, Z. Chang, S. Aota, H. Inokuchi, and H. Ozeki. 1986. *Nature* 322: 752.]

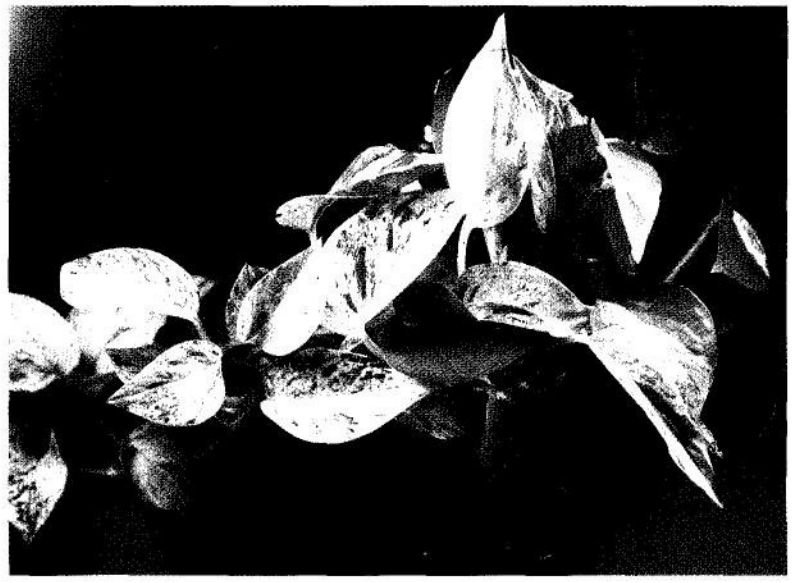


Figure 16.6 Leaf variegation in *Mirabilis jalapa*. The sectors result from cytoplasmic segregation of normal (green) and defective (white) chloroplasts.

chondrial genetic code can differ slightly from one species to the next, and some such codes are identical to the standard code. In plants, the standard genetic code is used in both mitochondria and chloroplasts.

Because the structure and function of mitochondria and chloroplasts depend on both nuclear and organelle genes, it is sometimes difficult to distinguish these contributions. Another complication is that numerous mitochondria or chloroplasts are present in a zygote, but only one of them may contain a mutation. The result is that traits determined by organelle genes exhibit a pattern of determination and transmission very different from simple Mendelian traits determined entirely by nuclear genes. Some examples of traits determined by chloroplast and mitochondrial genes are examined in the following sections.

Leaf Variegation in Four-O'clock Plants

Chloroplast transmission accounts for the unusual pattern of inheritance observed with leaf variegation in a strain of the four-o'clock plant, *Mirabilis jalapa*. Variegation refers to the appearance of white regions in the leaves and stems that result from the lack of green chlorophyll (**Figure 16.6**). On the variegated plants, some branches are completely green, some completely white, and others variegated. Branches of all three types produce flowers, and these flowers can be used to perform nine possible crosses (**Table 16.2**). From each cross, the seeds are collected and planted, and the phenotypes of the progeny are examined. The results of the crosses are summarized in Table 16.2. Two significant observations are that

1. Reciprocal crosses yield different results. For example, the cross

   green ♀ × white ♂

   yields green plants, whereas the cross

   white ♀ × green ♂

   yields white plants. (These usually die shortly after germination because they lack chlorophyll and hence photosynthetic activity.)
2. The phenotype of the branch bearing the female parent in each case determines the phenotype of the progeny (compare columns 1 and 3 in Table 16.2). The male makes no contribution to the phenotype of the progeny.

Furthermore, when flowers present on either the green or the variegated progeny plants are used in subsequent crosses, the patterns of transmission are identical with those of the original crosses. These observations suggest direct transmission of the trait through the mother, or maternal inheritance.

The genetic explanation for the variegation is as follows:

- Green color depends on the presence of chloroplasts, and pollen contains no chloroplasts.
- Segregation of chloroplasts into daughter cells is determined by cytoplasmic division and is somewhat irregular.
- The cells of white tissue contain mutant chloroplasts that lack chlorophyll.

These points are depicted in the model in **Figure 16.7**. Variegated plants germinate from embryos that are heteroplasmic for green and white (mutant) chloroplasts. Random segregation of the chloroplasts in cell division results in some branches derived from cells that contain only green chloroplasts and some branches derived from cells that contain only mutant chloroplasts. The flowers on green branches produce ovules with normal chloroplasts, so all progeny are green. The flowers on white branches produce ovules with only mutant chloroplasts, so only white progeny are formed. The ovules formed by flowers on variegated branches are of three types: those with only chlorophyll-containing chloroplasts, those with only mutant chloroplasts, and those that are heteroplasmic. The heteroplasmic class again yields variegated plants.

Table 16.2 Crosses and progeny phenotypes in variegated four-o'clock plants

Phenotype of branch bearing egg parent	Phenotype of branch bearing pollen parent	Phenotype of progeny
white	white	white
white	green	white
white	variegated	white
green	white	green
green	green	green
green	variegated	green
variegated	white	variegated, green, or white
variegated	green	variegated, green, or white
variegated	variegated	variegated, green, or white

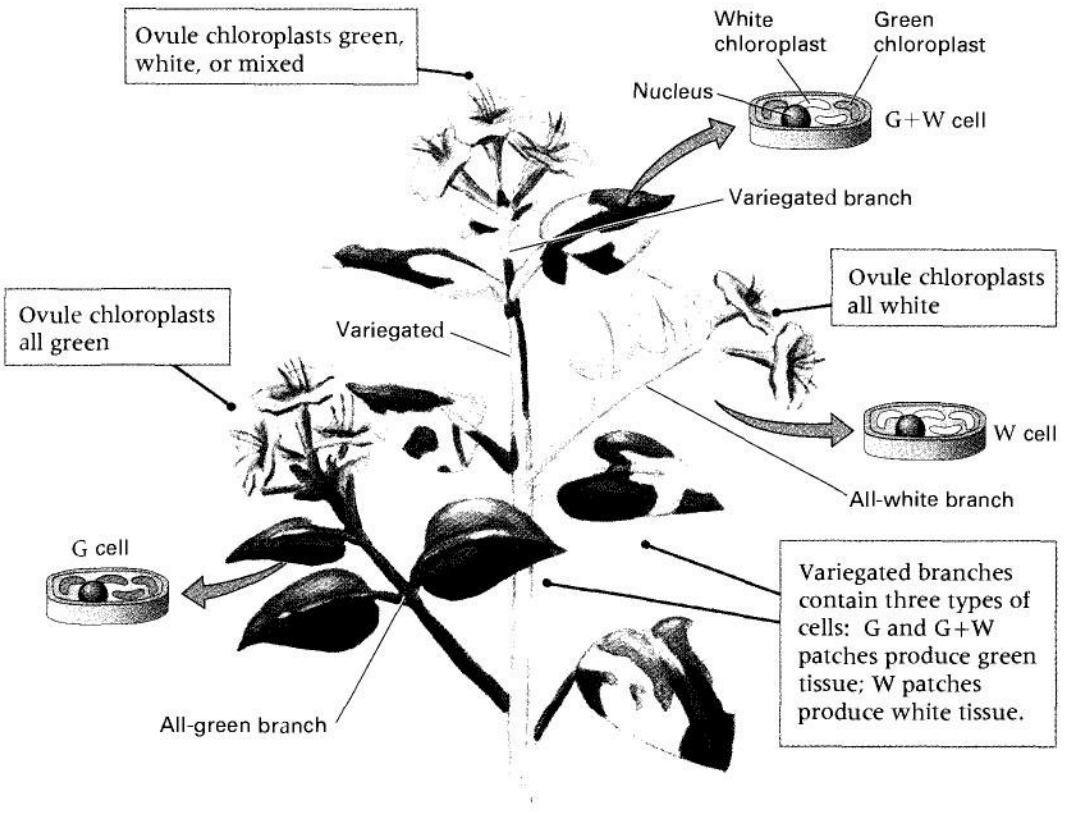


Figure 16.7 Genetic model for leaf variegation. Branches that are all green, all white, and variegated form on the same plant. Flowers form on all three kinds of branches. The insets show the chloroplast composition of cells in each type of branch. Cells in all-green or all-white branches contain only green or white chloroplasts, respectively, whereas cells in variegated branches are heteroplasmic.

CoNnection

Chlamydomonas Moment

Ruth Sager and Zenta Ramanis 1965
Columbia University,
New York, New York
Recombination of Nonchromosomal Genes in Chlamydomonas

Before the widespread use of polymorphic DNA markers in mitochondria and chloroplasts, the study of extranuclear inheritance was hindered in most organisms not only by uniparental inheritance of cellular organelles but also by the lack of mutant phenotypes showing extranuclear transmission. The alga Chlamydomonas *proved to be suitable for studies of extranuclear inheritance. Mutants showing nonchromosomal inheritance could be obtained, and strains were discovered in which the inheritance of nonchromosomal genes was biparental, making it possible to examine segregation of nonchromosomal alleles. The compelling evidence for the existence of genetic systems apart from genes in the nucleus was the discovery of intragenic recombination in nonchromosomal genes. In the matings described here, the parental genotypes for the nonchromosomal genes were* ac_1 sd *(slow growth on acetate, streptomycin requiring) and* ac_2 sr *(acetate-requiring, streptomycin-resistant). The intragenic recombinant products were either wildtype* ac^+ *(acetate-independent) versus the double mutant* ac_1 ac_2 *or they were wildtype* ss *(streptomycin-sensitive) versus the double mutant* sd sr. *The authors leave no doubt about their opinion (which turned out to be correct) that the nonchromosomal genetic systems are due to DNA in the cellular organelles.*

The analysis of nonchromosomal heredity has made slow progress, often under severe attack, [in part because of] the difficulty in obtaining mutations of nonchromosomal genes. . . . Within the past few years, a concerted attack has been made on this problem with the alga *Chlamydomonas*. . . . Two findings have been of key importance in our investigation. First, streptomycin was developed as [an agent for selecting] nonchromosomal (NC) genes. . . . Second, the existence of an exceptional class of zygotes was discovered that transmits NC genes from both parents to the progeny, in contrast to the standard pattern of maternal inheritance. With this material it was established that the NC genes are particulate in nature, genetically autonomous, and stable. . . . In this paper we describe what is to our knowledge the first evidence of recombination of nonchromosomal genes. . . . The occurrence of linked recombination demonstrates that NC genes, like the chromosomal ones, are capable of close pairing and precise reciprocal exchange. . . . The salient features of our results are the following: (1) Segregation of NC genes did not occur during meiosis. . . . (2) NC gene segregation began at the first or second mitotic division after meiosis, with pure clones arising so long as any heterozygous cells remained. (3) The average segregation ratio is 1 : 1 although the progeny of individual zygotes gave ratios significantly different from 1 : 1. (4) The acetate alleles segregated independently of the streptomycin alleles. (5) Novel classes of progeny were recovered that were phenotypically indistinguishable from wildtype ac^+ and from wildtype ss. New mutant phenotypes were also observed with new levels of acetate requirement and new levels of streptomycin resistance and dependence. [These were the reciprocal, double-mutant types.] . . . From existing knowledge, we infer that intragenic recombination, in which the wildtype gene is reconstituted, requires a close intermolecular pairing and exchange of the sort only known to occur between nucleic acids. Consequently, we view our results as providing evidence of the nucleic acid nature of a nonchromosomal genetic system. These genetic findings have been paralleled by the discovery of DNA of high molecular weight in chloroplasts and mitochondria. It is a reasonable surmise that these organelle DNA's carry primary genetic information.

We view our results as providing evidence of the nucleic acid nature of a nonchromosomal genetic system.

Source: Proceedings of the National Academy of Sciences of the USA 53: 1053–1061

Drug Resistance in *Chlamydomonas*

Cytoplasmic inheritance has been studied extensively in the unicellular green alga *Chlamydomonas* (Figure 16.8). Cells of this organism have a single large chloroplast containing 75–80 copies of a duplex DNA molecule approximately 195 kb in size. Cells of the two mating types, designated mt^+ and mt^-, are of equal size and appear to contribute the same amount of cytoplasm to the zygote. After cells of opposite mating type fuse, the diploid zygote undergoes meiosis to form a tetrad of four haploid

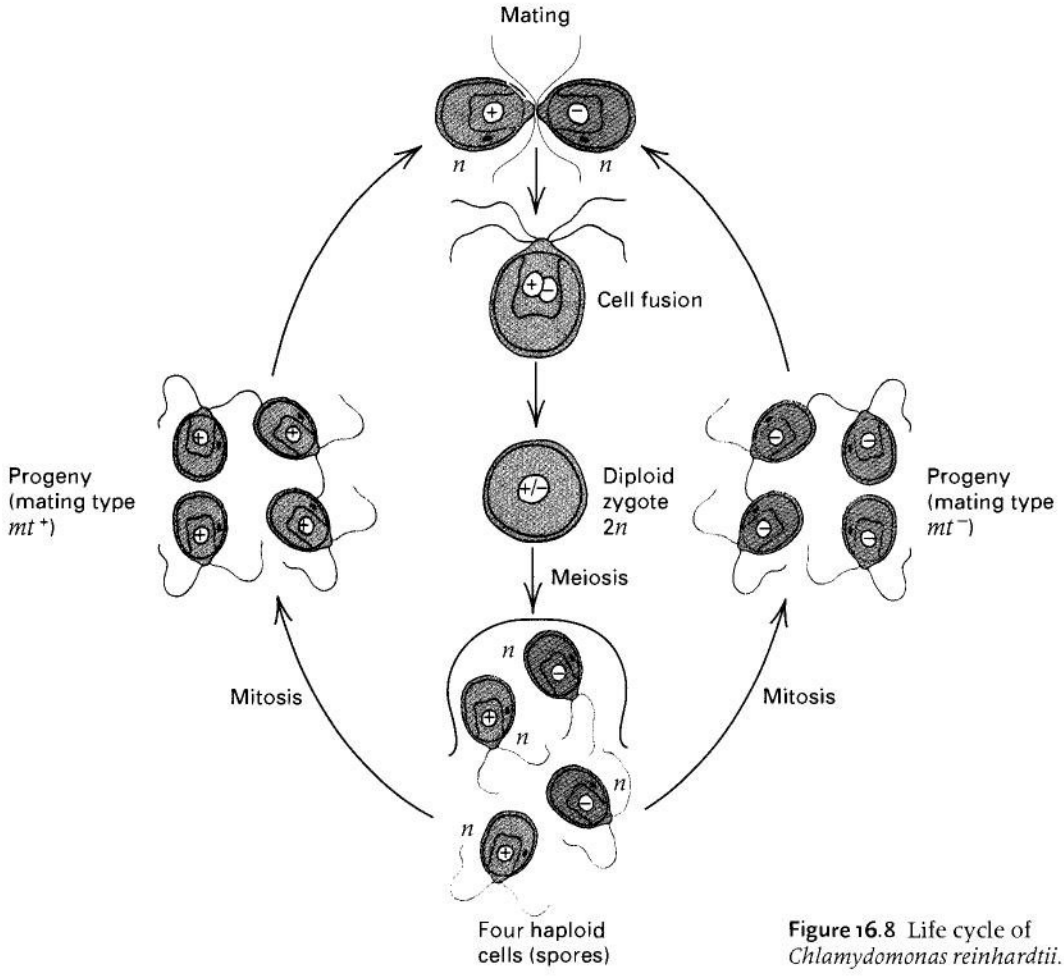


Figure 16.8 Life cycle of *Chlamydomonas reinhardtii*.

cells. The cells can be grown on solid growth medium to form visible clusters of cells (colonies). Cytoplasmic inheritance was first suggested by differences in the result of reciprocal crosses between mutant, antibiotic-resistant strains and wildtype, antibiotic-sensitive strains. For example, for streptomycin resistance (*str-r*) versus sensitivity (*str-s*),

$$str\text{-}r\ mt^{+} \times str\text{-}s\ mt^{-}$$
yields only *str-r* progeny

$$str\text{-}s\ mt^{+} \times str\text{-}r\ mt^{-}$$
yields only *str-s* progeny

This is a clear example of uniparental inheritance, because the resistance or sensitivity comes from the mt^{+} parent. In contrast, the *mt* alleles behave in strictly Mendelian fashion, yielding the 1 : 1 progeny ratio expected of nuclear genes.

A large number of antibiotic-resistance markers with uniparental inheritance have been examined in *Chlamydomonas*. In each case, the antibiotic resistance has been traced to the chloroplast. Haploid *Chlamydomonas* contains only one chloroplast. If the chloroplast could be derived from the mt^{+} or mt^{-} cell with equal probability, then uniparental inheritance would not be observed because the chloroplast could come from either parent. But direct tracking of chloroplast DNA via molecular markers indicates that the chloroplast of the mt^{-} parent is preferentially lost after mating. This accounts for the fact that the antibiotic-resistance phenotype of the mt^{+} parent is transmitted to the progeny.

About 5 percent of the progeny from *Chlamydomonas* crosses are heteroplasmic. Both parental chloroplasts are retained, although they ultimately segregate in later cell divisions. The heteroplasmic cells have been valuable for genetic analysis, because recombination can take place between the DNA in the two chloroplasts. Analysis of

the recombination frequencies for the various phenotypes has made possible the construction of fairly detailed genetic maps of the chloroplast genome.

The inheritance of mitochondria in *Chlamydomonas* is also uniparental, but it is through the mt^- parent. This pattern is observed for *MUD2* mutations, which occur in the mitochondrial cytochrome *b* gene and confer resistance to the antibiotic myxothiazol. When crosses are made between mutant, antibiotic-resistant (*MUD2*) and wildtype, antibiotic-sensitive (*mud2*) strains, the results are as follows:

$$MUD2\ mt^+ \times mud2\ mt^-$$
yields only *mud2* (sensitive) progeny

$$mud2\ mt^+ \times MUD2\ mt^-$$
yields only *MUD2* (resistant) progeny

In each cross, the tetrads yield 2 mt^+ : 2 mt^- spores. However, all spore progeny have the resistance phenotype of the mt^- parent. Inheritance of resistance to myxothiazol is therefore uniparental, because the phenotype of the progeny is the same as that of the mt^- parent. Because resistance to myxothiazol results from a change in the mitochondrial DNA, we can also infer that the progeny must receive their mitochondrial DNA from the mt^- parent. In contrast, as indicated by the streptomycin example above, they receive their chloroplast DNA from the mt^+ parent. It is not presently known whether having mitochondria come from one parent and chloroplasts from the other has any biological significance, or whether the patterns are merely a matter of evolutionary chance.

Respiration-Defective Mitochondrial Mutants

In budding yeast, *Saccharomyces cerevisiae*, very small colonies are occasionally observed when the cells are grown on solid medium containing glucose. These are called **petite mutants.** Microscopic examination indicates that the cells are of normal size even though the colonies are small. Physiological studies show that the colonies grow normally during the early stages when they are fermenting glucose and making ethanol, but they soon stop growing because of a defect in the oxygen-requiring respiration that is needed for further metabolism of the ethanol. The colonies are small because the cells obtain ATP only via fermentation and are unable to obtain the much larger yields of ATP via metabolism of the ethanol.

Among several types of petite mutants, two major types can be distinguished by their behavior in crosses with wildtype nonpetites. The results are illustrated in Figure 16.9. One type, called **segregational petites,** exhibits the typical segregation of a simple Mendelian recessive. In a cross between a segregational petite and wildtype, the diploid zygotes are normal; and when the diploids undergo meiosis, half of the ascospores in an ascus produce petite colonies and half form wildtype colonies. The 1 : 1 ratio (actually 2 : 2) indicates that these petites are the result of a nuclear mutation and that the allele for petite is recessive.

For the second type, **neutral petites,** the result is completely different. In a cross with wildtype, all diploid zygotes produce wildtype ascospores (4 : 0 segregation). The same pattern is found when the progeny from such a cross are backcrossed with neutral petites. In both cases, the phenotype that is inherited is from the normal parent, so the inheritance is uniparental. The explanation for these results is that the majority of neutral petites have deletions of most of the mitochondrial DNA genes involved in oxidative respiration. When a neutral-petite cell mates with a wildtype cell, the cytoplasm of the wildtype cell contributes normal mitochondrial DNA, which is transmitted to the progeny spores.

Petites arise at a frequency of roughly 10^{-5} per generation. The neutral petite phenotype is the result of large deletions in mitochondrial DNA. For unknown reasons, yeast mitochondria often fuse and fragment, which may cause occasional deletions of DNA. The production of neutral petites is apparently a result of the segregation of aberrant mitochondrial DNA from normal DNA, and the further sorting out of mutant mitochondria, in the course of cell division.

Cytoplasmic Male Sterility in Plants

An example of extranuclear inheritance that is important in agriculture is **cytoplasmic male sterility,** a condition in

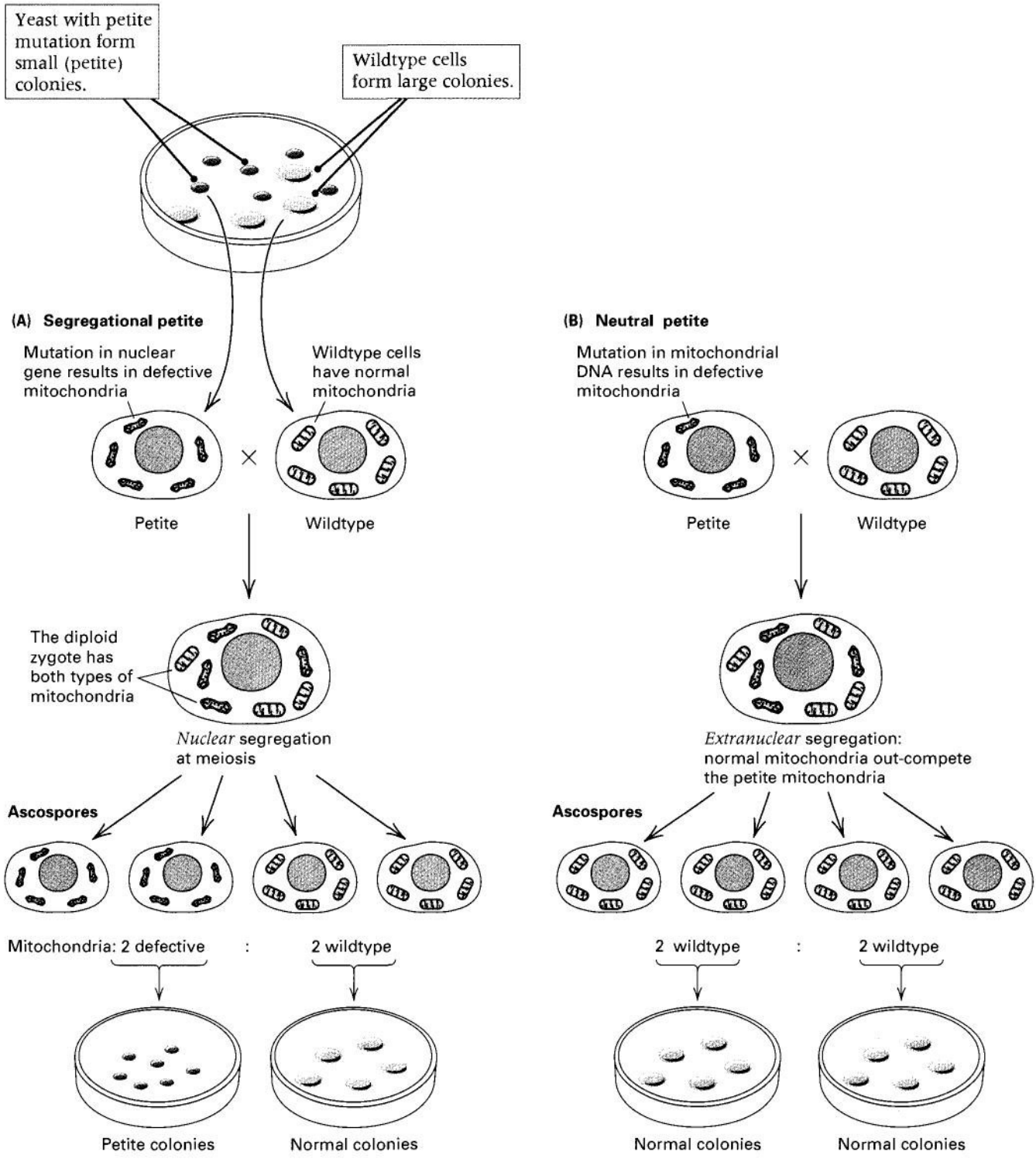


Figure 16.9 Behavior of petite mutations in genetic crosses. The brown circles represent petite cells. The brown nucleus in part A contains an allele resulting in petite colonies. The blue cells have normal mitochondria.

which a plant does not produce functional pollen, but the female reproductive organs and fertility are normal. This type of sterility is used extensively in the production of hybrid corn seed, because it circumvents the need for manual "detasseling" of the plants to be used as females to produce the hybrid. The tassel is the pollen-bearing organ, and in male-fertile varieties, detasseling is necessary to prevent self-fertiliziation of female plants.

The genetically transmitted male sterility is not controlled by nuclear genes but is transmitted through the egg cytoplasm from generation to generation. The pattern of inheritance of cytoplasmic male sterility in corn is summarized in **Figure 16.10**. The key observation is that repeated backcrossing of a male-sterile variety, using pollen from a male-fertile variety, does not restore male fertility. The resulting plants, in which nearly all nuclear genes from the male-sterile variety have been replaced with those from the male-fertile variety, remain male-sterile. Because a small amount of functional pollen is produced by the male-sterile variety, it is also possible to carry out the reciprocal cross:

male-fertile ♀ × male-sterile ♂

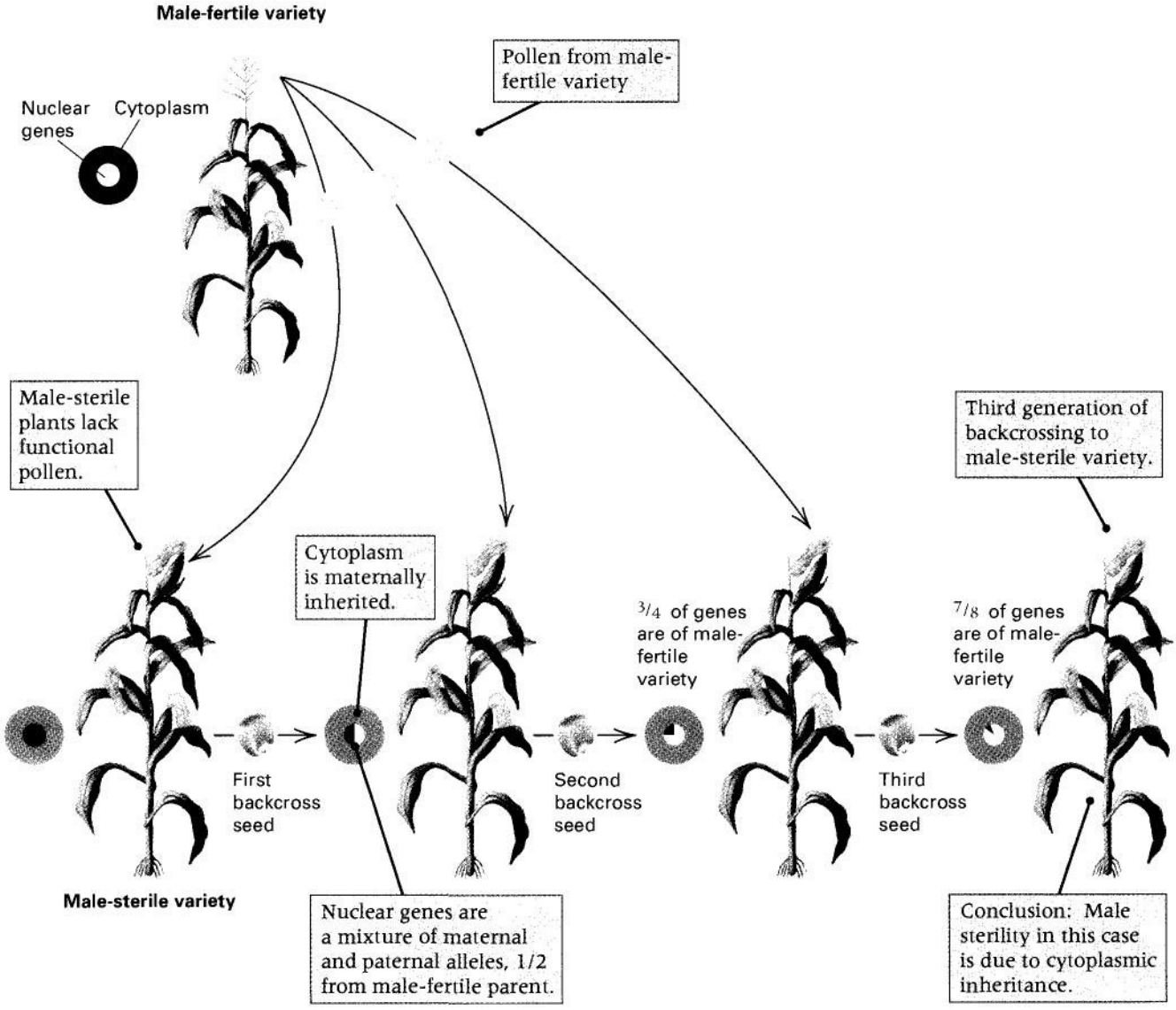


Figure 16.10 Backcross experiment demonstrating maternal inheritance of male sterility (orange cytoplasm) in maize.

In this case, the progeny are fully male-fertile. The difference between the reciprocal crosses means that male sterility in these varieties is maternally inherited.

Cytoplasmic male sterility in maize results from rearrangements in mitochondrial DNA. For example, one DNA rearrangement fuses two mitochondrial genes and creates a novel protein that causes male sterility. Although cytoplasmic male sterility is maternally inherited, certain nuclear genes called **fertility restorers** can suppress the male-sterilizing effect of the cytoplasm. Different types of cytoplasmic male sterility can be classified on the basis of their response to restorer genes. For example, in one case, restoration of fertility requires the presence of dominant alleles of two different nuclear genes, and plants with male-sterile cytoplasm and both restorer alleles produce normal pollen. In another case, suppression requires the dominant allele of a different restorer gene, but this gene acts only in the gametophyte; hence plants with male-sterile cytoplasm that are heterozygous for the restorer allele produce a 1 : 1 ratio of normal to aborted pollen grains.

16.3 The Evolutionary Origin of Organelles

A widely accepted theory holds that eukaryotic organelles originally evolved from prokaryotes that lived inside primitive cells in **symbiosis,** a mutually beneficial interaction. This is the **endosymbiont theory.** Mitochondria are thought to have originated from bacteria and chloroplasts from cyanobacteria (formerly called blue-green algae). The archaeans, rather than eubacteria or eukaryotes, are thought to be good candidates for the primitive hosts of the endosymbionts. The archaeans share some features with eubacteria, including their prokaryotic cytological organization, but their mechanisms of macromolecular syntheses—replication, transcription, and translation—resemble those of eukaryotes. Ultimately this symbiotic association evolved into eukaryotes, including animals, plants, and fungi.

One line of evidence for the endosymbiont theory is that mitochondrial and chloroplast organelles of eukaryotes share with prokaryotes numerous features that distinguish them from eukaryotic cells. Among these common features:

- The genomes are composed of circular DNA that is not extensively complexed with histone-like proteins.
- The genomes are organized with functionally related genes close together and often expressed as a single unit.
- The ribosome particles on which protein synthesis takes place have major subunits whose size is similar in organelles and in prokaryotes but differs from that in cytoplasmic ribosomes in eukaryotic cells.
- The nucleotide sequences of the key RNA constituents of ribosomes are similar in chloroplasts, cyanobacteria such as *Anacystis nidulans,* and even bacteria such as *E. coli.*

The genomes of today's organelles are small compared with those of bacteria and cyanobacteria. The chloroplast genome is only 3–5 percent of that of cyanobacteria. Organelle evolution was probably accompanied by major genome rearrangements, some genes being transferred into the nuclear genome of the host and others being eliminated. The transfer of organelle genes into the nucleus differed from one lineage to the next. For example, the gene for one subunit of the mitochondrial ATP synthase is located in the mitochondrial genome in yeast but in the nuclear genome in *Neurospora.*

16.4 Cytoplasmic Transmission of Symbionts

In eukaryotes, a variety of cytoplasmically transmitted traits result from the presence of bacteria and viruses living in the cytoplasm of certain cells. One of the best known is the "killer" phenomenon found in certain strains of the protozoan *Paramecium aurelia.* Killer strains of *Paramecium* release to the surrounding medium a substance that is lethal to many other strains of the protozoan. The killer phenotype requires the presence of cytoplasmic bacteria referred to as **kappa particles,** whose maintenance is dependent on a dominant nuclear gene, *K.* The kappa particles are cells of *Caedobacter taeniospiralis,* which produce the killer substance. Why killer strains themselves are immune to the substance has not yet been determined.

Paramecium is a diploid protozoan that undergoes sexual exchange through a mating process called **conjugation** (part A of Figure 16.11). Initially, each cell has two diploid micronuclei. Many protozoans contain micronuclei (small nuclei) and a macronucleus (a large nucleus with specialized functions), but only the micronuclei are relevant to the genetic processes described here. When two cells come into contact for conjugation, the two micronuclei in each cell undergo meiosis, forming eight micronuclei in each cell. Seven of the micronuclei and the macronucleus disinte-

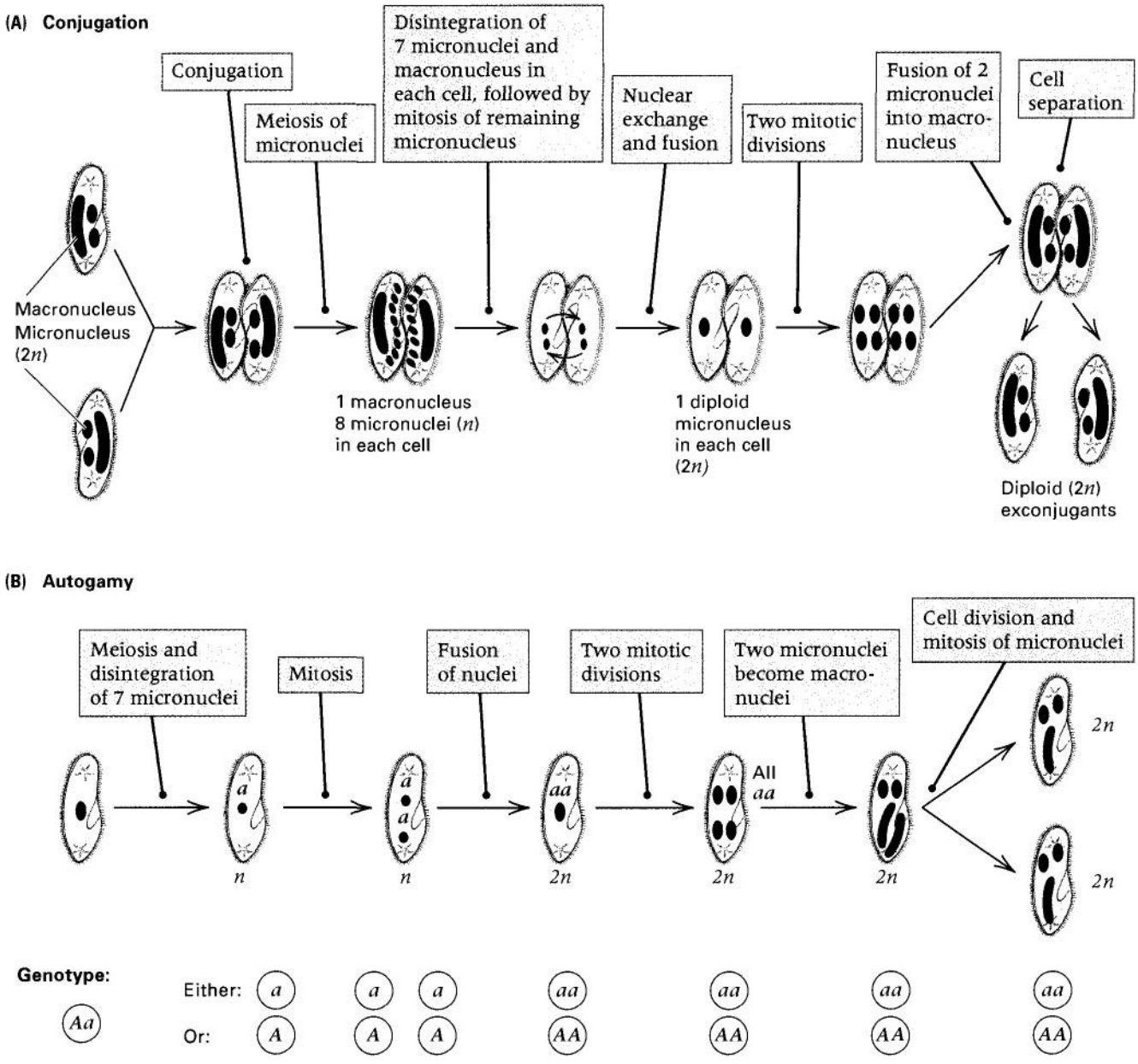


Figure 16.11 (A) Conjugation in *Paramecium* results in reciprocal fertilization and the formation of two genetically identical exconjugant cells. (B) Autogamy in *Paramecium*. The alleles in the heterozygous *Aa* cell segregate in meiosis, with the result that the progeny cells become homozygous for either *A* or *a*. The production of *aa* progeny cells is shown in the diagram. The initial steps resemble those in conjugation in part A: The micronuclei undergo meiosis, after which seven of the products and the macronucleus degenerate.

grate, leaving each cell with one micronucleus, which then undergoes a mitotic division. The cell membrane between the two cells breaks down slightly, and the cells exchange a single micronucleus; then the nuclei fuse to form a diploid nucleus. After nuclear exchange, the two new cells (the exconjugants) are genotypically identical. In this sequence of events, only the micronuclei are exchanged, without any mixing of cytoplasm.

Single cells of *Paramecium* also occasionally undergo an unusual nuclear phenomenon called **autogamy** (Figure 16.11B). Meiosis takes place, and again seven of the micronuclei and the macronucleus disintegrate. The surviving nucleus then undergoes mitosis and nuclear fusion to recreate the diploid state, and a new macronucleus is formed. The genetic importance of autogamy is that, even if the initial cell is heterozygous, the newly formed diploid nucleus becomes homozygous because it is derived from a single haploid meiotic product. In a population of cells undergoing autogamy, the surviving micronucleus is selected randomly, so for any pair of alleles, half of the new cells contain one allele and half contain the other.

In conjugation, the exchange of micronuclei is occasionally accompanied by small amounts of cytoplasmic mixing. A comparison of the phenotypes following conjugation of killer (*KK*) cells and sensitive (*kk*) cells, with and without cytoplasmic exchange, yields evidence for cytoplasmic inheritance of the killer phenotype (**Figure 16.12**). Without cytoplasmic mixing (Figure 16.12A), the expected 1 : 1 ratio of killer to sensitive exconjugant cells is observed; with cytoplasmic mixing (Figure 16.12B), both exconjugants are killer cells because each cell has kappa particles derived from the cytoplasm of the killer parent. Conjugation is eventually followed by autogamy of each of the exconjugants

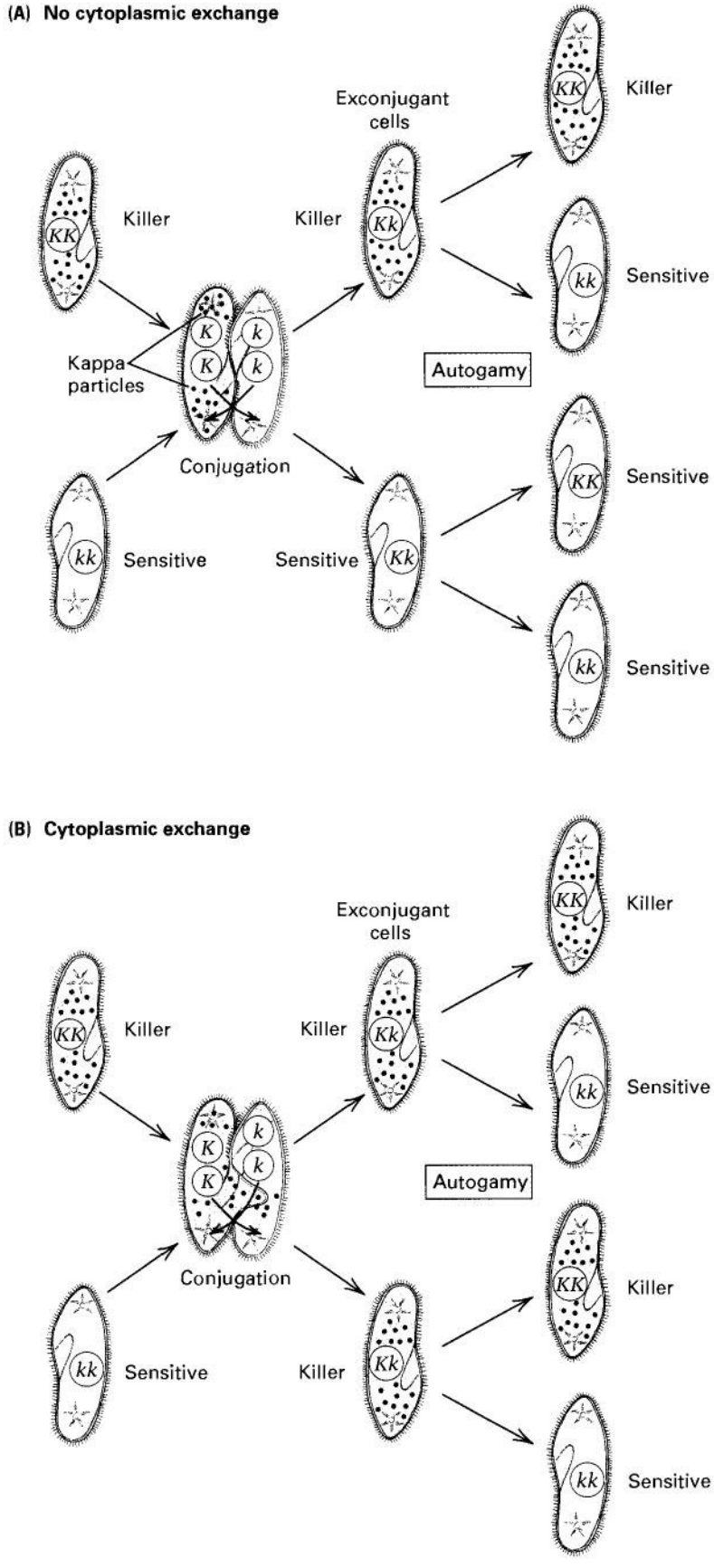


Figure 16.12 Crosses between killer (*KK*) and sensitive (*kk*) *Paramecium*. (A) The result when no cytoplasmic exchange takes place in conjugation. (B) The result when cytoplasmic exchange occurs. Cells containing kappa are shown in yellow. The genotype and phenotype of each cell are indicated.

CoNnection

A Coming Together

Lynn Margulis (formerly Lynn Sagan) 1967
Boston University, Boston, Massachusetts
The Origin of Mitosing Cells

By the 1960s certain features of the genetic apparatus of cellular organelles were already known to resemble those in prokaryotes. A number of scientists had suggested that the similarities may reflect an ancient ancestral relationship. Lynn Margulis took these ideas a giant step further by proposing a comprehensive hypothesis of the origin of the eukaryotic cell as a series of evolved symbiotic relationships. The main players are the mitochondria, the kinetosomes (basal bodies) of the flagella and cilia, and the photosynthetic plastids. All are hypothesized to have originated as prokaryotic symbionts. When the paper was submitted for publication, the anonymous reviewers were so strongly negative that it was almost rejected. Through the years, specific predictions of the hypothesis have been experimentally verified. A few revisions have been necessary; for example, current evidence suggests that the earliest symbiont was kinetosome-bearing, not a protomitochondrion. However, no major inconsistencies have been uncovered. Today the serial endosymbiont theory is widely accepted.

This paper presents a theory of the origin of the discontinuity between eukaryotic and prokaryotic cells. Specifically, the mitochondria, the basal bodies [kinetosomes] of the flagella, and the photosynthetic plastids can be considered to have derived from free-living cells, and the eukaryotic cell is the result of the evolution of ancient symbioses. Although these ideas are not new, in this paper they have been synthesized in such a way as to be consistent with recent data on the biochemistry and cytology of subcellular organelles. . . . Many aspects of this theory are verifiable by modern techniques of molecular biology. . . . Prokaryotic cells containing DNA, synthesizing protein on ribosomes, and using messenger RNA as intermediate between DNA and protein are ancestral to all extant cellular life. Such cells arose under reducing conditions of the primitive terrestrial atmosphere. . . . Eventually, a population of cells arose using photoproduced ATP with water as the source of hydrogen atoms in the reduction of CO_2 for the production of cell material. This led to the formation of gaseous oxygen as a by-product of photosynthesis. . . . The continued production of free oxygen resulted in a crisis. . . . It is suggested that the first step in the origin of eukaryotes from prokaryotes was related to survival in the new oxygen-containing atmosphere: an aerobic prokaryotic microbe (the protomitochondrion) was ingested into the cytoplasm of a heterotrophic anaerobe. This symbiosis became obligate and resulted in the first aerobic amoeboid [without, at this stage, mitosis]. . . . Some of the amoeboids ingested certain motile prokaryotes. The genes of the parasite coded for its characteristic morphology, (9 + 2) fibrils in cross section. This parasite also became symbiotic, forming primitive amoeboflagellates . . . that could actively pursue their own food. . . . The replicating nucleic acid of the endosymbiont genes (which determines its characteristic 9 + 2 structure of sets of microtubules) was eventually utilized to form the chromosomal centromeres and centrioles of eukaryotic mitosis and to distribute newly synthesized host nuclear chromatin to host daughters. . . . Eukaryotic plant cells acquired photosynthesis by symbiosis with photosynthetic prokaryotes (protoplastids), which themselves evolved from organisms related to cyanobacteria. . . . In order to document this theory, the following, at the very least, are required: (1) The theory must be consistent with geological and fossil records. (2) Each of the three symbiotic organelles (the mitochondria, the (9 + 2) homologues, and the plastids) must demonstrate general features characteristic of cells originating in hosts as symbionts. None may have features that conflict with such an origin. (3) Predictions based on this account of the origin of eukaryotes must be verified. The rest of this paper [24 pages] discusses the evidence for this theory in terms of assumptions based on molecular biology which can be made concerning evolutionary mechanisms.

The eukaryotic cell is the result of the evolution of ancient symbioses.

Source: Journal of Theoretical Biology 14: 225–274

(Figure 16.12). Autogamy of *Kk* killer cells results in an equal number of *KK* and *kk* cells, but the *kk* cells cannot maintain the kappa particle and so become sensitive.

Another example of a symbiont-related cytoplasmic effect is a condition in *Drosophila* called **maternal sex ratio,** which is characterized by the production of almost no male progeny. The daughters of sex-ratio females pass on the trait, whereas the sons that are occasionally produced do not. Cytoplasm taken from the eggs of sex-ratio

females transmits the condition when injected into the female embryos from unaffected cultures of the same or other *Drosophila* species. A species of bacteria has been isolated from the cytoplasm of sex-ratio females; when such bacteria are allowed to infect other *Drosophila* females, these acquire the sex-ratio trait. The causative agent of the sex-ratio condition is not the bacteria themselves but rather a virus that multiplies in them. When released by the bacterial cells, this virus kills most male *Drosophila* embryos. Why female embryos are not killed by the virus is still a mystery.

16.5 Maternal Effect in Snail Shell Coiling

Maternal transmission is not always mediated by cytoplasmic organelles or symbionts; in some cases, it is mediated through the cytoplasmic transmission of the products of nuclear genes. An example is determination of the direction of coiling of the shell in the snail *Limnaea peregra*. The direction of coiling, as viewed by looking into the opening of the shell, may be either to the right (dextral coiling) or to the left (sinistral coiling). Reciprocal crosses between homozygous strains give the following results:

dextral ♀ × sinistral ♂ → all F_1 dextral
sinistral ♀ × dextral ♂ → all F_1 sinistral

In these crosses, the F_1 snails have the same genotype, but the reciprocal crosses give different results: The direction of progeny coiling is the same as that of the mother. This result is typical of traits with maternal inheritance. However, in this case, all F_2 progeny from both crosses exhibit dextral coiling, a result that is inconsistent with cytoplasmic inheritance. The F_3 generation provides the explanation.

The F_3 generation, obtained by self-fertilization of the F_2 snails (the snail is hermaphroditic), indicates that the inheritance of coiling direction depends on nuclear genes rather than on extranuclear factors. Three-fourths of the F_3 progeny have dextrally coiled shells and one-fourth have sinistrally coiled shells (**Figure 16.13**). This is a typical 3 : 1 ratio with dextral coiling (+/+ and *s*/+) dominant over sinistral coiling (*s*/*s*). The ratio indicates Mendelian segregation, but it is observed in the F_3 generation, not in the F_2 generation. The Mendelian ratio is delayed for a generation because:

> The coiling phenotype of an individual snail is determined by the genotype of its mother.

This means that the direction of shell coiling is not a case of maternal inheritance but rather a case of maternal effect (Section 16.1). Cytological analysis of developing

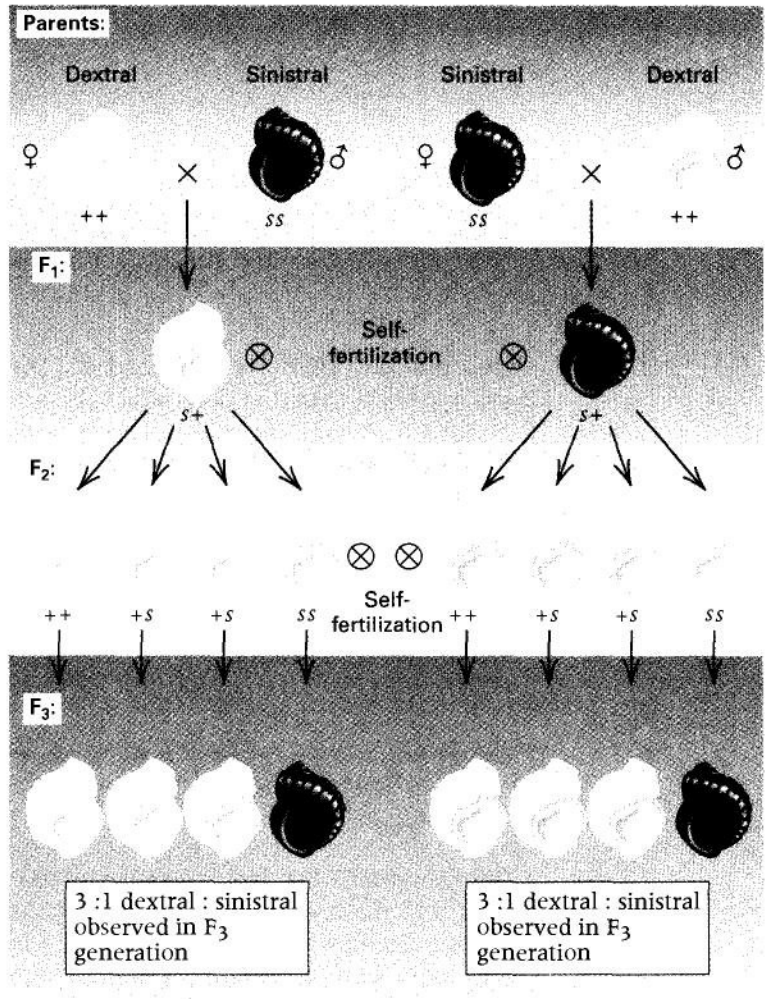


Figure 16.13 Inheritance of the direction of shell coiling in the snail *Limnaea*. Sinistral coiling is determined by the recessive allele *s*, dextral coiling by the wildtype + allele. However, the direction of coiling is determined by the nuclear genotype of the mother, not by the genotype of the snail itself or by extranuclear inheritance. The F_2 and F_3 generations can be obtained by self-fertilization because the snail is hermaphroditic and can undergo either self-fertilization or cross-fertilization.

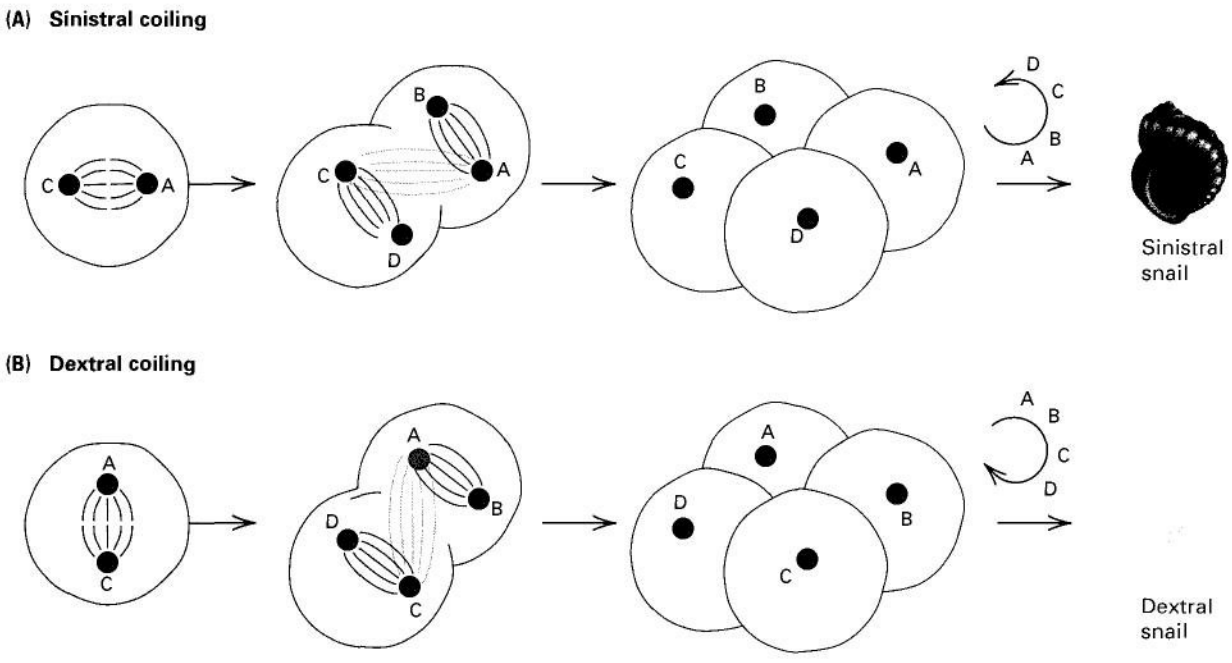


Figure 16.14 The direction of shell coiling in *Limnaea* is determined by the orientation of the mitotic spindles during the second cleavage division in the zygote, which is predetermined in the egg cytoplasm as a result of the mother's genotype. The spiral patterns established by the cleavage result in (A) sinistral (leftward) or (B) dextral (rightward) coiling.

eggs has provided the explanation: The genotype of the mother determines the orientation of the spindle in the initial mitotic division after fertilization, and this in turn controls the direction of shell coiling of the offspring (**Figure 16.14**). This classic example of maternal effect indicates that more than a single generation of crosses is needed to provide conclusive evidence of extranuclear inheritance of a trait.

16.6 In Search of Mitochondrial "Eve"

Mitochondrial DNA has a number of features that make it useful for studying the genetic relationships among organisms in natural populations, including human populations. First, it is maternally inherited and (possibly with rare exceptions) does not undergo genetic recombination. Second, mammalian mitochondrial DNA evolves considerably faster than that of nuclear genes; for example, differences in mitochondrial DNA sequences accumulate among human lineages at a rate of approximately 1 change per mitochondrial lineage every 1500–3000 years.

To put this rate of evolution into perspective, consider the people of Australia and New Guinea, who have been relatively isolated genetically from each other since an aboriginal people colonized these areas approximately 30,000 years ago (Australia) and 40,000 years ago (New Guinea). The total time of separation between the populations is therefore 70,000 years (30,000 years in Australia and 40,000 years in New Guinea). If one nucleotide change accumulates every 1500–3000 years, the total number of differences between the mitochondrial DNA of Australian aborigines and that of New Guineans is expected to range from 23 nucleotides (calculated as 70,000\3000) to 47 nucleotides (calculated as 70,000\1500). This example shows how the rate of mitochondrial DNA evolution can be used to predict the number of differences between populations. In practice, the calculation is done the other way around, and the observed number of differences between pairs of populations is used to estimate the rate of mitochondrial evolution.

Nucleotide differences in mitochondrial DNA have been used to reconstruct the probable historical relationships among human populations. **Figure 16.15** is an example that includes people from five native populations: African, Asian, Australian, Caucasian (European), and New Guinean. The diagram is *not* a pedigree, although it superficially resembles one. Rather, the diagram is a **phylogenetic tree,** which depicts the lines of descent that connect the mitochondrial DNA sequences. The "true" phylogenetic tree connecting the mitochondrial DNA sequences is unknown, so the tree illustrated should be regarded as an approximation (or an "estimate") of the true tree. It is approximate because it is inferred from analysis of the mitochondrial DNA, and this analysis requires a number of assumptions, including, for example, uniformity of rates of evolution in different branches. Nevertheless, the tree in Figure 16.15 has two interesting features: (1) Each population contains multiple mitochondrial types connected to the tree at widely dispersed positions, and (2) one of the primary branches issuing from the inferred common ancestor is composed entirely of Africans. These features can be explained most readily by postulating that the mitochondrial DNA type of the common ancestor was African. Furthermore, because the inferred common ancestor links mitochondrial DNA types that differ at an average of 95 nucleotides, the time since the existence of the common ancestor can be estimated as between $95 \times 1500 = 142{,}500$ years and $95 \times 3000 = 285{,}000$ years. Therefore, the common ancestor of all human mitochondrial DNA was a woman who lived in Africa 142,500–285,000 years ago. This woman has been given the name Eve by the popular press, although she was not the only female alive at the time. Nevertheless, Eve's mitochondrial DNA became the common ancestor of all human mitochondrial DNA that exists today, because during the intervening years, all other mitochondrial lineages except hers became extinct (probably by chance).

The hypothesis of an African mitochondrial Eve has been very controversial. Later analysis of the same data suggested that other trees were almost as good as the one in Figure 16.15 and that some trees were

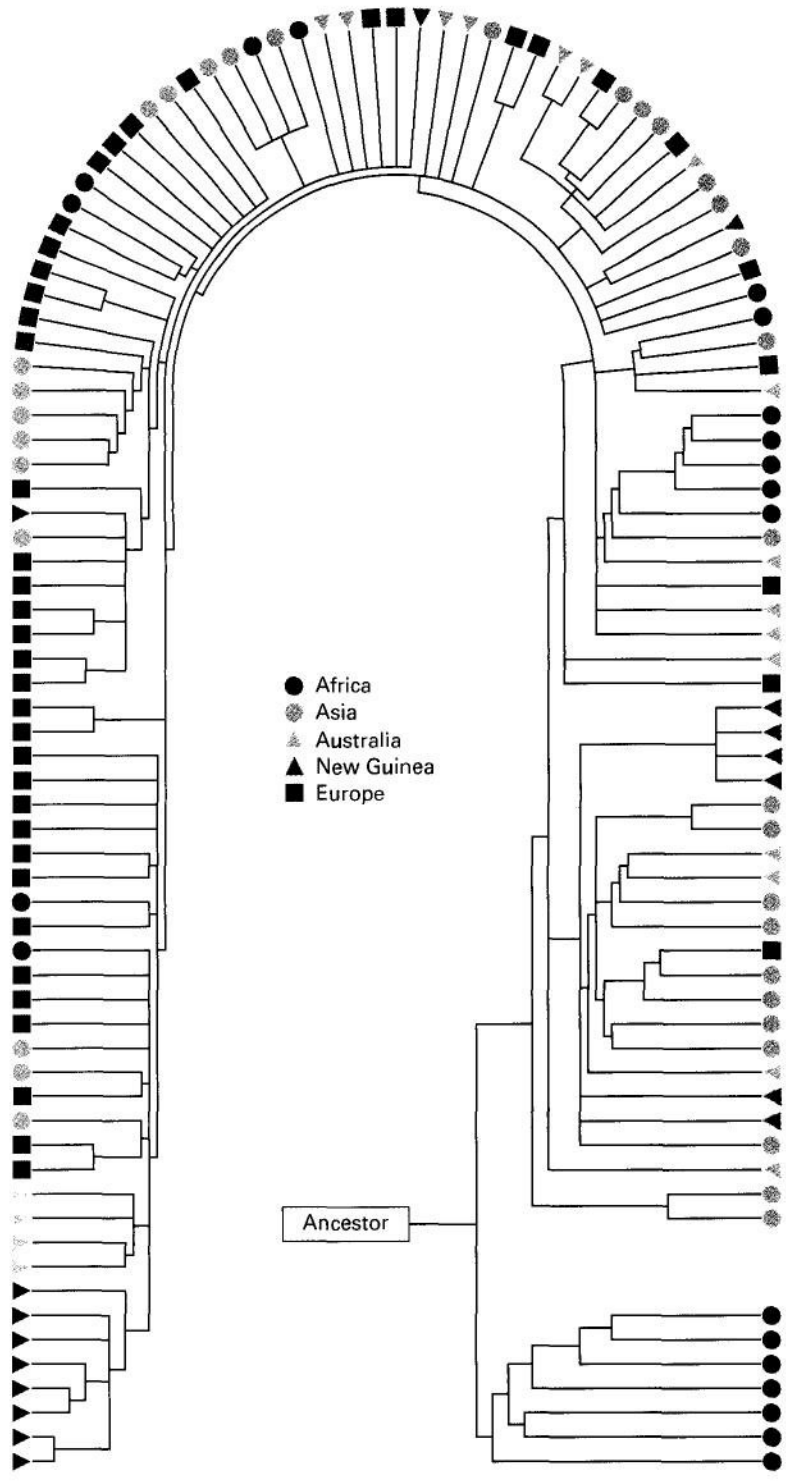


Figure 16.15 Possible phylogenetic tree of human mitochondria based on analysis of presence or absence of restriction sites for 12 different restriction enzymes in mitochondrial DNA from 134 persons. [Reprinted by permission from *Nature* 325: 31, R. L. Cann, M. Stoneking, and A. C. Wilson. Copyright 1987 Macmillan Magazines, Ltd.]

even better. The main criterion for comparing hypothetical trees based on DNA sequences is the total number of mutational steps in all the branches together: Trees with fewer mutational steps are generally regarded as better. Therefore, the finding of better trees than the one in Figure 16.15 was a serious blow to the hypothesis of an African Eve, because none of the better trees had an earliest branch composed exclusively of Africans. However, much of the uncertainty is intrinsic in the data themselves. Another hypothetical phylogenetic tree, with fewer mutational steps than that in Figure 16.15, is shown in **Figure 16.16**. This tree is drawn with the lengths of the branches proportional to the number of mutational steps, and it resembles a starburst with the viewer looking back in time. The difficulty of identifying the earliest common ancestor is evident from this style of representation. Although there is an early branch (at a position near one o'clock in the starburst) composed primarily of Africans, this branch also includes two Asians, one Caucasian, and one Australian aborigine. Nevertheless, considerable evidence of other kinds supports Africa as the cradle of human evolution and colonization.

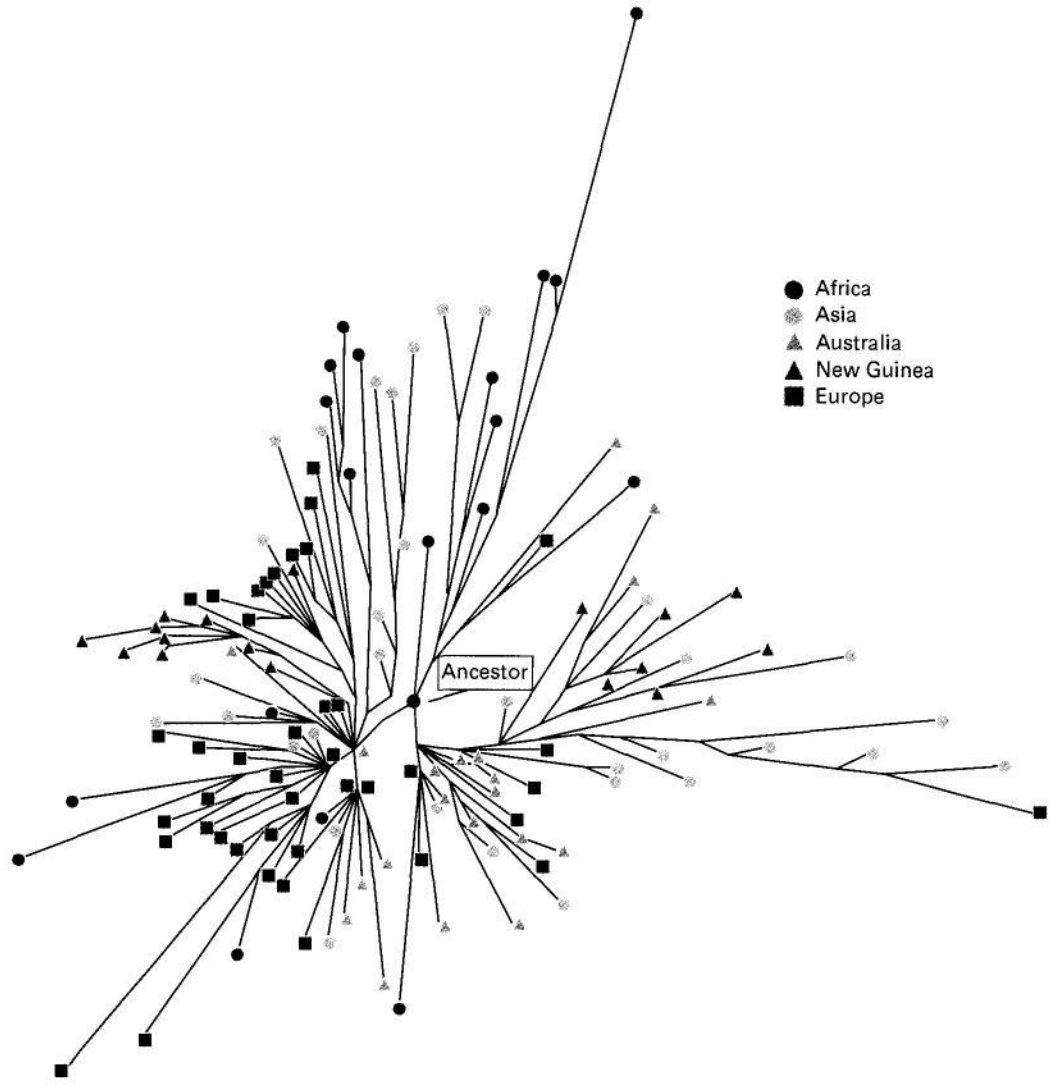


Figure 16.16 An improved phylogenetic tree of human mitochondrial DNA presented as a starburst with the reader looking back in time. The symbols are the same as those in Figure 16.15. [Reprinted by permission from *Nature* 356: 389, C Wills. Copyright 1992 Macmillan Magazines, Ltd.]

Chapter Summary

The cytoplasm of eukaryotic cells contains complex organelles, the most prevalent of which are mitochondria, found in both plant and animal cells, and chloroplasts, found in plant cells. Mitochondria are responsible for respiratory metabolism. Chloroplasts are responsible for photosynthesis. Both mitochondria and chloroplasts contain DNA molecules that encode a variety of proteins. The genes contained in mitochondrial and chloroplast DNA are not present in chromosomal DNA. Hence, traits determined by these genes are not inherited according to Mendelian principles but show extranuclear (cytoplasmic) inheritance instead. Extranuclear inheritance is also observed for traits that are determined by bacteria and viruses transmitted in the cytoplasm from one generation to the next. In eukaryotic microorganisms, cytoplasmic mixing usually takes place in zygote formation, and both parents contribute cytoplasmic particles to the zygote. In multicellular animals, however, the egg contains a large amount of cytoplasm and the sperm very little, so cytoplasmic transmission results in maternal inheritance.

Progeny resembling the mother is characteristic of both maternal inheritance and maternal effects. Maternal inheritance results from the transmission of extranuclear factors by the mother. Maternal effects occur when the nuclear genotype of the mother determines the phenotype of the offspring—for example, through the molecular organization of the egg or by nurturing ability. The direction of shell coiling in snails results from a maternal effect on the orientation of the spindle in the first mitotic division in the zygote.

Variegation in the four-o'clock plant is caused by mutations in chloroplast DNA that eliminate chlorophyll. Variegated plants contain patches of green and white derived from wildtype cells and from cells with mutant chloroplasts, respectively. Inheritance of the variegated phenotype depends on the chloroplast content of the egg because pollen does not contain chloroplasts. The inheritance of resistance to certain antibiotics in *Chlamydomonas* is also cytoplasmic and depends on mutations in the chloroplast.

Mitochondrial defects account for the cytoplasmic inheritance of some petite mutations in yeast. Petites produce very small colonies because the defective mitochondria are unable to carry out aerobic metabolism. Petite cells can use only anaerobic metabolism, which is so inefficient that small colonies are produced. There are two major types of petites. Segregational petites result from a mutation in a nuclear gene and show strictly Mendelian inheritance. Neutral petites show non-Mendelian inheritance (4 : 0 segregation in ascospores) and result from large deletions in mitochondrial DNA. Male sterility in maize also results from defective mitochondria that contain certain types of deletions or rearrangements in the DNA. Male sterility is very useful in plant breeding, because it makes it possible to cross-pollinate plants without removing immature anthers from the recipient plants.

The killer phenotype in *Paramecium* exemplifies cytoplasmic inheritance through a symbiotic bacterium that produces a product that is toxic to many other paramecia. Cells containing the bacterium are resistant to the killer substance. *Paramecium* mates by two cells coming in contact and exchanging haploid nuclei. After the nuclei exchange, the cells separate. Normally, the exchange takes place without any cytoplasmic mixing, in which case all segregating traits are inherited according to Mendelian principles. Sometimes cytoplasmic mixing occurs, and in this case one sees cytoplasmic inheritance. Maternal sex ratio in *Drosophila*, expressed as a deficiency of males among progeny, is another phenomenon determined by the presence of a symbiotic bacterium. In this case, the bacterium contains a virus that inhibits the development of male embryos.

Human mitochondrial DNA types have been organized into a hypothetical phylogenetic tree in which one branch issuing from the inferred common ancestor is composed entirely of mitochondrial DNA from Africans. The woman whose mitochondrial DNA is thought to be the common ancestor of all human mitochondria has been named "Eve." The all-African branch in the phylogenetic tree and the number of nucleotide differences observed in mitochondrial DNA have been interpreted to mean that the mitochondrial "Eve" was a woman who lived in Africa 142,500–285,000 years ago. On the other hand, additional analyses of the same data have yielded hypothetical phylogenetic trees that have fewer mutational steps overall and that do not contain exclusively African branches.

Key Terms

autogamy
conjugation
cytoplasmic inheritance
cytoplasmic male sterility
endosymbiont theory
extranuclear inheritance
fertility restorer
heteroplasmy
kappa particle
maternal effect
maternal inheritance
maternal sex ratio
neutral petite
organelle
paternal inheritance
petite mutant
phylogenetic tree
RNA editing
segregational petite
symbiosis
uniparental inheritance

Review the Basics

- What are the normal functions of mitochondria and chloroplasts?
- Can eukaryotic cells survive without mitochondria? Can plant cells survive without chloroplasts?
- Give an example of a human disease caused by a mutation in mitochondrial DNA.
- Distinguish between maternal inheritance and maternal effect.
- What is the endosymbiont theory of the origin of certain cellular organelles?
- What is variegation?
- How does leaf variegation in four-o'clock plants demonstrate heteroplasmy of chloroplasts in certain cell lineages and homoplasmy in others?
- Of what practical use is cytoplasmic male sterility in maize breeding?
- What does the occurrence of segregational petites in yeast imply about the relationship between the nucleus and mitochondria?
- What is meant by the term *Eve* with reference to human mitochondrial DNA? Do you think *Eve* is a suitable term for this concept? Why or why not?

Guide to Problem Solving

Problem 1 Two strains of maize exhibit male sterility. One is cytoplasmically inherited; the other is Mendelian. What crosses could you make to identify which is which?

Answer Cytoplasmic male sterility is transmitted through the female, so outcrosses (crosses to unrelated strains) using the cytoplasmic male-sterile strain as the female parent will yield male-sterile progeny. Repeated outcrosses using the strain with Mendelian male sterility will result in either no male sterility (if the sterility gene is recessive) or 50 percent sterility (if it is dominant).

Problem 2 A snail of the species *Limnaea peregra* in which the shell coils to the left undergoes self-fertilization, and all of the progeny coil to the right. What is the genotype of the parent? If the progeny are individually self-fertilized, what coiling phenotypes are expected?

Answer In this species, shell coiling is determined by the genotype of the mother. Because the coiling of the mother is leftward, its own mother had the genotype *ss*, so the individual must carry at least one *s* allele. Because the progeny coil to the right, the mother's genotype must be s^+. The progeny genotypes are 1/4 s^+s^+, 1/2 s^+s, and 1/4 *ss*. When self-fertilized, the s^+s^+ and s^+s individuals yield rightward-coiled progeny, and the *ss* individuals yield leftward-coiled progeny.

Problem 3 Consider an inherited trait in *Drosophila* with the following characteristics: (1) It is transmitted from father to son to grandson. (2) It is not expressed in females. (3) It is not transmitted through females.

(a) Is the trait likely to result from extranuclear inheritance?
(b) Can male-limited expression explain the data?
(c) What hypothesis of Mendelian inheritance can explain the data?
(d) What observations could be made to confirm the Mendelian hypothesis? (*Hint*: Think about the consequences of nondisjunction.)

Answer

(a) The trait is very unlikely to result from extranuclear inheritance. First, it would be very unusual for extranuclear inheritance in animals to be transmitted through the male. Second, if it were transmitted through sperm, both sexes should be affected.
(b) The trait is also unlikely to have male-limited expression. Although the observation that only males are affected is consistent with this hypothesis, a trait with male-limited expression could be transmitted through females.
(c) The father → son → grandson pattern of inheritance parallels that of the Y chromosome. Therefore, the data fit a Y-linked model of inheritance.
(d) Nondisjunction in an XY male results in XXY zygotes (which in *Drosophila* are female) and XO zygotes (which in *Drosophila* are male). Therefore, if the trait is Y-linked, then XXY females resulting from nondisjunction should be affected, and XO males resulting from nondisjunction should be unaffected. Furthermore, the affected XXY females should give rise to affected sons (XY) and some affected daughters (XXY).

Analysis and Applications

16.1 Why are so many inherited mitochondrial diseases characterized by muscle insufficiency or weakness?

16.2 Why are diseases that result from mitochondrial mutations highly variable in severity from one patient to the next?

16.3 Is the genetic code used in human mitochondrial DNA completely redundant in the sense that, at any third-codon position, either purine or either pyrimidine is equivalent? How does this situation differ from the standard genetic code?

16.4 In the course of evolution, many genes originally present in mitochondrial DNA have been transferred to the nucleus. If a present-day human mitochondrial coding sequence were to be transferred to the nucleus and translated successfully, what constraints on codon usage would be necessary?

16.5 The following mRNA is part of a coding sequence isolated from a human cell. Is it likely to be from a nuclear gene or a mitochondrial gene? Explain your answer.

5'-AUGAGACAUAAUGAUAGAGGGAGGCCCAAUCGC-3'

16.6 In the human mitochondrial genetic code, what fraction of transition mutations in the third-codon position are synonymous? What fraction of transversion mutations in the third position are synonymous?

16.7 A mutant plant is found with yellow instead of green leaves. Microscopic and biochemical analyses show that the cells contain very few chloroplasts and that the plant manages to grow by making use of other pigments. The plant does not grow very well, but it does breed true. The inheritance of yellowness is studied in the following crosses:

green ♂ × yellow ♀ → all yellow
yellow ♂ × green ♀ → all green

What do these results suggest about the type of inheritance?

16.8 For the phenotype in the preceding problem, suppose that variegated progeny are found in the crosses at a frequency of about 1 per 1000 progeny. How might this be explained?

16.9 Some plant species have distinct sexes determined by a mechanism analogous to the XX–XY chromosomal system in animals. In these species, how could paternal transmission of the Y chromosome be distinguished from paternal transmission of an organelle?

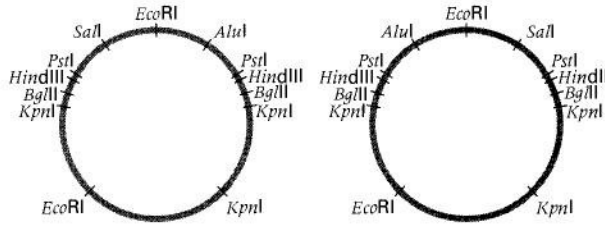
16.10 What kinds of progeny would be expected to result from the following crosses with the four o'clock plant *Mirabilis*?

(a) green female × white male
(b) white female × green male
(c) variegated female × green male
(d) green female × variegated male

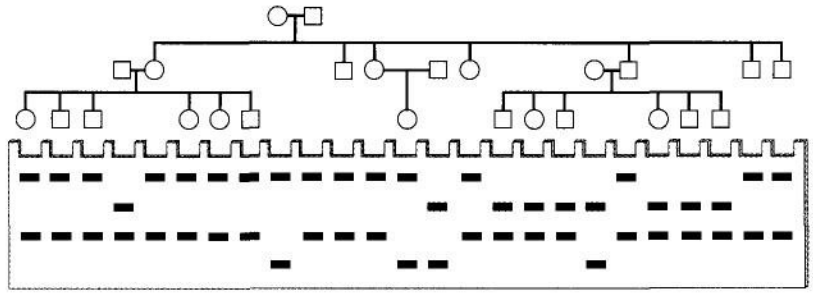
16.11 In the maternally inherited leaf variegation in *Mirabilis*, how does the F_2 of the cross green ♀ × white ♂ differ from that which would be expected if white leaves were due to a conventional X-linked recessive gene in animals?

16.12 Assuming that chloroplasts duplicate and segregate randomly in cell division, what is the probability that both replicas of a single mutant chloroplast will segregate to the same daughter cell during mitosis? After two divisions, what is the probability that at least one daughter cell will lack the mutant chloroplast?

16.13 Analysis of the mitochondrial DNA of a marine snail consistently yields two different restriction maps in the mtDNA present in a single animal. These are shown in the accompanying diagram. Suggest a hypothesis for the difference in the restriction maps, and explain why both types of mtDNA are found in the same individual.



16.14 The pedigree illustrated below shows the pattern of bands observed in several generations of crosses of a certain plant species. Are the patterns consistent with Mendelian inheritance? Explain why or why not. If not, suggest an explanation for the patterns.



Problem 16.14

GeNETics on the web

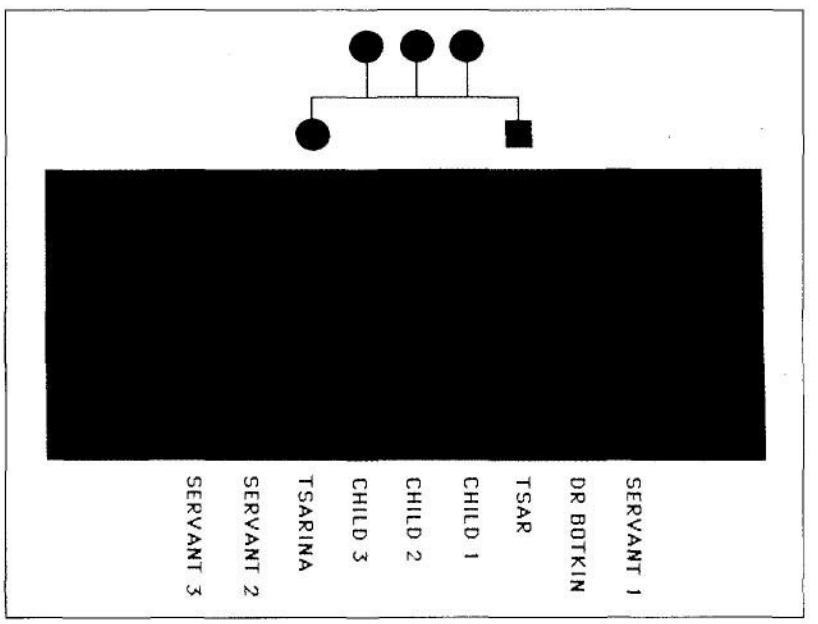
GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by-chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- Each **mitochondrion** contains about 3000 different proteins and about 5 copies of its own circular DNA. The mitochondrial DNA itself encodes only 13 proteins. The others are encoded in nuclear genes and imported from the cytoplasm. Perhaps 20–30 percent of all proteins encoded in a typical cell are destined for the mitochondria. Little wonder that more than 50 inherited diseases affect mitochondrial function. These maladies affect approximately 4000 children born in the United States each year. For more on the diversity of types of mitochondrial diseases, their symptoms, and their treatment, consult this keyword site.

- The skeletons of six adults and three children were found in a shallow, unmarked grave near Ekaterinburg, Russia, in 1979, but remained hidden until July 1991 when they were revealed under the openness policy (*glasnost*) of Mikhail Gorbachev. The remains have been positively identified as those of **Tsar Nicholas II**, Tsarina Alexandra, and their daughters Olga, Tatiana, and Anastasia, who had been killed by a firing squad in the basement of a merchant's house and their bodies dumped in a pit north of the city. The other skeletons were those of Anna Demidova, the tsarina's lady-in-waiting; Eugene Bodkin, the family's physician; Ivan Kharitonov, their cook; and Alouzy Trupp, the Tsar's valet. The remains of Marie and Alexei, who was afflicted with hemophilia, were never found and are believed to have been burned. How were the re-



[Courtesy of Peter D. Gill.]

Tsar Nicolas II of Russia, his wife the Tsarina Alexandra, their five children, and several servants were secretly executed in 1918 by order of Vladmir Lenin, leader of the Russian Revolution, and their bodies buried in a common, shallow, unmarked grave. Remains discovered in 1991 provided sufficient material for DNA typing. Shown here in blue are polymorphic nuclear DNA markers. The mother in the pedigree is Alexandra, who was homozygous for this marker, and the father is Nicholas II, who was heterozygous. The remains of three immature females each exhibit a band identical with that in the tsarina and a second band identical with one of those in the tsar, consistent with their being the daughters of the Royal family. The remains of the other two children have not been found. Sex was identified separately by PCR amplification of genes in the X and Y chromosomes. The identities of the tsar, tsarina, and Dr. Botkin were inferred from mitochondrial DNA typing of living relatives. The red bands across the gel are size markers that are the same in all lanes. Learn more of the story by visiting GeNETics on the Web for Chapter 14; see also page 681 of the text.

mains identified? By mitochondrial DNA typing, which you can learn about at this keyword site.

- Amniocentesis for the prenatal diagnosis of chromosome disorders is usually performed between 15 and 18 weeks after a woman's last menstrual period. A somewhat different procedure called chorionic villus sampling (**CVS**) can be carried out earlier in pregnancy, usually between 10 and 12 weeks after a woman's last menstrual period. At this keyword site you can find a nontechnical description of CVS and a discussion of its uses and potential dangers.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 16 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured at right, select the **PIC Site** for Chapter 16.



16.15 If *a* and *b* are chloroplast markers in *Chlamydomonas*, what genotypes of progeny are expected from the cross $a^+ b^- mt^+ \times a^- b^+ mt^-$?

16.16 Consider three types of antibiotic resistance (*r*) versus sensitivity (*s*) in *Chlamydomonas:*

(a) *cpl-r* results from a mutation in chloroplast DNA.

(b) *mit-r* results from a mutation in mitochondrial DNA.

(c) *nuc-r* results from a mutation in nuclear chromosomal DNA.

If a cross of resistant $mt^+ \times$ sensitive mt^- is carried out for each of these three types of resistances individually, what types of progeny are expected, and in what proportions? Explain your answers.

16.17 For the antibiotic-resistance markers described in the previous problem, what phenotypes would be expected, and in what ratio, from the following crosses?

(a) *cpl-r mit-s nuc-r* $mt^+ \times$ *cpl-s mit-r nuc-s* mt^-

(b) *cpl-s mit-r nuc-s* $mt^+ \times$ *cpl-r mit-s nuc-r* mt^-

16.18 A new antibiotic-resistance, *klv-r*, is found in *Chlamydomonas*. Random spores chosen from the cross *klv-r* $mt^+ \times$ *klv-s* mt^- yield the following progeny:

92 *klv-r* mt^+
88 *klv-s* mt^-
11 *klv-r* mt^-
9 *klv-s* mt^+

What type of inheritance does the antibiotic resistance show? How can the association with mating type be explained?

16.19 Why are segregational petites in yeast called segregational?

16.20 What is the phenotype of a diploid yeast produced by crossing a segregational petite with a neutral petite?

16.21 An antibiotic-resistant haploid strain of yeast is isolated. Mating with a wildtype (antibiotic-sensitive) strain produces a diploid that, when grown for some generations and induced to undergo sporulation, produces tetrads containing either four antibiotic-resistant spores or four antibiotic-sensitive spores. What can you conclude about the inheritance of the antibiotic resistance?

16.22 What are the possible genotypes of the progeny of an *Aa Paramecium* that undergoes autogamy?

16.23 Equal numbers of *AA* and *aa Paramecium* undergo conjugation in random pairs. What are the expected matings? What genotypes and frequencies are expected after the conjugating pairs separate? What genotypes and frequencies are expected after the cells undergo autogamy?

16.24 In conjugation between a *KK* killer strain of *Paramecium* and a *kk* nonkiller strain, why is cytoplasmic exchange necessary to obtain a 1 : 1 ratio of killer : nonkiller after autogamy of the exconjugants has taken place? What is the ratio of killer : nonkiller in the absence of cytoplasmic exchange?

16.25 A cross is carried out in maize using, as the female parent, a variety with cytoplasmically inherited male sterility and, as the male parent, a variety homozygous for a dominant allele, *Rsf*, that overrides the cytoplasmic sterility and restores fertility. The genotype of the female parent is homozygous *rsf*. From this cross, F_1 plants are self-fertilized. What are the expected genotypes and fertilities of the female and male parts of the plants in the next generation?

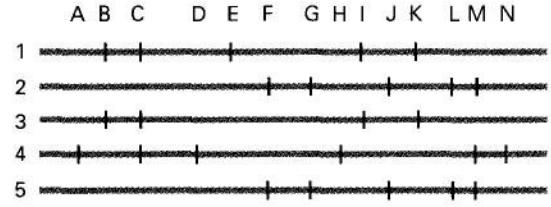
16.26 Corn plants of genotype *AA bb* with cytoplasmic male sterility and normal *aa BB* plants are planted in alternate rows, and pollination is allowed to occur at random. What progeny genotypes would be expected from the two types of plants?

16.27 A strain of *Drosophila* recovered from nature has the following properties. In every generation, about 90 percent of the progeny are females. When the females are crossed to males from a normal strain in successive generations, in each generation they produce about 90 percent female progeny. The males from the original strain, as well as those from the backcross generations, produce normal sex ratios when mated with normal females. Suggest an explanation for these results.

16.28 Females of a certain strain of *Drosophila* are mated with wildtype males. All progeny are female. There are three possible explanations: (1) The females are homozygous for an autosomal recessive allele that produces lethal male embryos (that is, a maternal effect). (2) The females are homozygous for an X-linked allele that is lethal in males. (3) The females carry a cytoplasmically inherited factor that kills male embryos. A cross is carried out between F_1 females and wildtype males, and only female progeny are produced. Does this result help us decide among these hypotheses?

16.29 A rightward-coiled snail, when self-fertilized, has all leftward-coiled progeny. What is the genotype of the rightward-coiled parent? What was the genotype of the parent's mother?

16.30 The illustration below shows the locations of restriction sites A through N in the mitochondrial DNA molecules of five species of crabs. The presence of a restriction site is represented by a short vertical line. Assuming that the proportion of shared restriction sites between two molecules is an indication of their closeness of relationship, draw a diagram showing the inferred relationships among the species.



Challenge Problems

16.31 Genetic suppression of nonsense (chain-terminating) mutations is often due to mutant tRNA molecules whose anticodon binds with the chain-terminating codon, allowing a fraction of the mRNA molecules to translate through the termination codon and continue polypeptide synthesis. Would you expect nonsense suppression through tRNA mutations to be possible in human mitochondrial genes? Explain your answer.

16.32 There are an average of 78 nucleotide differences in mitochondrial DNA between two Africans, an average of 58 differences between two Asians, and an average of 40 differences between two Caucasians. Assuming a rate of evolution of mitochondrial DNA of 1 nucleotide difference per 1500 to 3000 years, what is the estimated average age of the common ancestor of the mitochondrial DNA among Africans? Among Asians? Among Caucasians?

16.33 In yeast, an *ade6*$^-$ mutation results in a requirement for adenine. An *ade6*$^-$ strain with normal colony size was crossed to each of three *ade6*$^+$ petite strains. Among 100 colonies arising from random spores examined from each cross, the nutritional status and sizes were as follows:

(a) 26 Ade$^-$ petite : 23 Ade$^-$ normal : 24 Ade$^+$ petite : 27 Ade$^+$ normal.

(b) 46 Ade$^-$ normal : 54 Ade$^+$ normal.

(c) 7 Ade$^-$ petite : 45 Ade$^-$ normal : 40 Ade$^+$ petite : 8 Ade$^+$ normal.

State what type of petite is involved in each cross, and indicate whether anything further can be said about the genetic basis of the petite phenotype.

Further Reading

André, C., A. Levy, and V. Walbot. 1992. Small repeated sequences and the structure of plant mitochondrial genomes. *Trends in Genetics* 8: 128.

Cann, R. L., M. Stoneking, and A. C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325: 31.

Capaldi, R. A. 2000. The changing face of mitochondrial research. *Trends in Biochemical Sciences* 25: 212.

Gillham, N. 1978. *Organelle Heredity.* New York: Raven.

Grivell, L. A. 1983. Mitochondrial DNA. *Scientific American*, March.

Howe, C. J., A. C. Barbrook, and P. J. Lockhart. 2000. Organelle genes: Do they jump or are they pushed? *Trends in Genetics* 16: 65.

Jacobs, H. T., and D. M. Lonsdale. 1987. The selfish organelle. *Trends in Genetics* 3: 337.

Lang, B. F., M. W. Gray, and G. Burger. 1999. Mitochondrial genome evolution and the origin of eukaryotes. *Annual Review of Genetics* 33: 351.

Larsson, N. G., and D. A. Clayton. 1995. Molecular genetic aspects of human mitochondrial disorders. *Annual Review of Genetics* 29: 151.

Laughnan, J. R., and S. Gabay-Laughnan. 1983. Cytoplasmic male sterility in maize. *Annual Review of Genetics* 17: 27.

Palmer, J. D. 1997. The mitochondrion that time forgot. *Nature* 387: 454.

Preer, J. R., Jr. 1971. Extrachromosomal inheritance: Hereditary symbionts, mitochondria, chloroplasts. *Annual Review of Genetics* 5: 361.

Preer, J. R., Jr. 1997. Whatever happened to *Paramecium* genetics? *Genetics* 145: 217.

Race, H. L., R. G. Herrmann, and W. Martin. 1999. Why have organelles retained genomes? *Trends in Genetics* 15: 364.

Rochaix, J. D. 1995. *Chlamydomonas reinhardtii* as the photosynthetic yeast. *Annual Review of Genetics* 29: 209.

Stoneking, M., and H. Soodyall. 1996. Human evolution and the mitochondrial genome. *Current Opinion in Genetics & Development* 6: 731.

Sturtevant, A. H. 1923. Inheritance of the direction of coiling in *Limnaea*. *Science* 58: 269.

Tattersall, I. 1997. Out of Africa again . . ., and again. *Scientific American*, April.

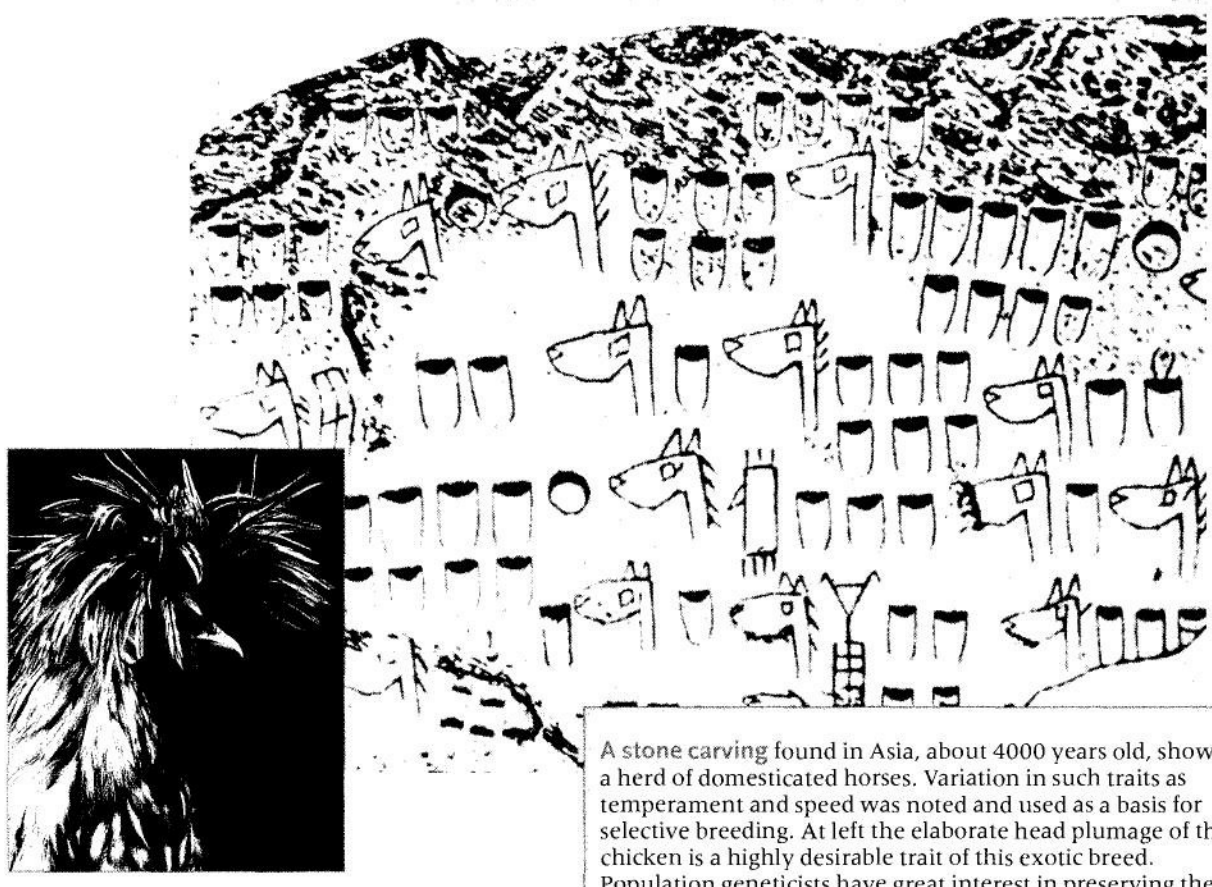
Thorne, A. G., and M. H. Wolpoff. 1992. The multiregional evolution of humans. *Scientific American*, April.

Wallace, D. C. 1993. Mitochondrial diseases: Genotype versus phenotype. *Trends in Genetics* 9: 128.

Wilson, A. C., and R. L. Cann. 1992. The recent African genesis of humans. *Scientific American*, April.

CHAPTER 17

Population Genetics and Evolution



A stone carving found in Asia, about 4000 years old, showing a herd of domesticated horses. Variation in such traits as temperament and speed was noted and used as a basis for selective breeding. At left the elaborate head plumage of this chicken is a highly desirable trait of this exotic breed. Population geneticists have great interest in preserving the genetic variation within and among such specialized breeds, as well as among endangered species. [Reproduction of carving courtesy of Dorsey Stuart; chicken: Eric Sanders.]

CHAPTER OUTLINE

17.1 Allele Frequencies and Genotype Frequencies

17.2 The Matter of Mating
- Random Mating and the Hardy–Weinberg Principle
- Implications of the Hardy–Weinberg Principle
- A Test for Random Mating
- Frequency of Heterozygous Genotypes
- Multiple Alleles
- DNA Typing
- DNA Exclusions
- Differences Among Populations
- X-Linked Genes

17.3 Inbreeding
- The Inbreeding Coefficient
- Allelic Identity by Descent
- Calculation of the Inbreeding Coefficient from Pedigrees
- Effects of Inbreeding

17.4 Genetics and Evolution

17.5 Mutation and Migration
- Irreversible Mutation
- Reversible Mutation

17.6 Natural Selection
- Selection in a Laboratory Experiment
- Selection in Diploid Organisms
- Components of Fitness
- Selection–Mutation Balance
- Heterozygote Superiority

17.7 Random Genetic Drift

PRINCIPLES

- Many genes in natural populations are polymorphic—they have two or more common alleles. Genetic polymorphisms can be used as genetic markers in pedigree studies and for individual identification (DNA typing).
- The Hardy–Weinberg principle asserts that, with random mating, the frequencies of the genotypes *AA*, *Aa*, and *aa* in a population are expected to be p^2, $2pq$, and q^2, respectively, where p and q are the allele frequencies of *A* and *a*. One implication of this principle is that the great majority of rare, harmful recessive alleles are present in heterozygous carriers.
- Inbreeding means mating between relatives. Through inbreeding, replicas of an allele present in a common ancestor may be passed down both sides of a pedigree and come together in an inbred organism. Such alleles are said to show identity by descent. Because inbreeding can result in alleles that are identical by descent, inbreeding results in an excess of homozygous genotypes relative to the frequencies of genotypes expected with random mating.
- Mutation and migration both introduce new alleles into populations. Because mutation rates are typically small, mutation is a weak force for changing allele frequencies.
- Natural selection can change allele frequency in a systematic direction because of the differential reproductive success of certain genotypes. Selection can result in a stable genetic polymorphism if the heterozygous genotype has the highest fitness. Selection against recurrent harmful mutations leads to a selection–mutation balance in which the allele frequency of the harmful recessive remains constant generation after generation.
- Random genetic drift consists of random changes in allele frequency that occur in finite populations as a result of chance excess reproduction (or lack of reproduction) among certain genotypes, because all breeding adults cannot be expected to contribute equally to the next generation.

CONNECTIONS

Twin Bill
Wilhelm Weinberg 1908
On the Demonstration of Heredity in Man

Resistance in the Blood
Anthony C. Allison 1954
Protection Afforded by Sickle-Cell Trait Against Subtertian Malarial Infection

In most organisms, matings are not under the control of the investigator, and the familial relationships among individuals are unknown. This situation is typical of studies of organisms in their natural habitat. Most organisms live in *populations* of individuals, and the genealogical relationship between one individual and the next is obscure. At first, it might seem that classical genetics with its simple Mendelian ratios could have little to say about such complex situations, but this is not the case. The principles of Mendelian genetics can be used both to interpret data collected in natural populations and to make predictions about the genetic composition of populations. Application of genetic principles to entire populations of organisms constitutes the subject of **population genetics.**

17.1 Allele Frequencies and Genotype Frequencies

The term *population* refers to a group of organisms of the same species living within a prescribed geographical area. Sometimes the area is large, as when we refer to the population of sparrows in North America; it may even include the entire earth, as when we speak of the "human population." More commonly, the area is considered to be of a size within which individuals in the population are likely to find mates. Many geographically widespread species are subdivided into more or less distinct breeding groups, called **subpopulations,** that live within limited geographical areas. Each subpopulation is a *local population.* The complete set of genetic information contained within the individuals in a population is called the **gene pool.** The gene pool includes all alleles present in the population.

The basic concepts of population genetics can be illustrated with respect to an AIDS-resistance phenotype determined by a relatively common mutant allele. The gene in question is a chemokine receptor gene, *CCR5,* found in the human population. (*Chemokines* are molecules that white blood cells of the immune system use to attract one another.) The *CCR5* receptor enables the HIV virus to combine with the plasma membrane and infect the CD4(+) class of T cells of the immune system, which is necessary for an HIV infection to progress. Most human subpopulations contain a *CCR5* allele known as Δ*32* because it has a 32-bp deletion within the coding region. The molecular consequences of the deletion are shown in **Figure 17.1**. The relevant part of the normal *CCR5* DNA sequence is shown, grouped into codons, along with the amino acid sequence of the polypeptide, given in the single-letter abbreviations. The nucleotides missing in the Δ*32* deletion are highlighted in color. Above the DNA sequence is the normal CCR5 polypeptide; below is the Δ*32* mutant polypeptide. The deletion creates a frameshift in translation following codon 184, which results in the insertion of 31 incorrect amino acids until a termination codon is encountered after amino acid 215. The truncated protein is nonfunctional and does not support HIV entry into CD4(+) cells. The Δ*32* mutation was originally discovered among persons at high risk of HIV-1 infection who had remained uninfected. The usual frequency of heterozygous genotypes, as we shall see, is about 20 percent, but among AIDS nonprogressors the frequency is approximately 40 percent. The homozygous Δ*32* genotypes, which are much less frequent, seem to have even greater protection.

Why is the Δ*32* homozygous genotype much less frequent than the Δ*32* heterozygous genotype? And what is the relationship between the homozygous and heterozygous frequencies? To begin to answer these questions, we should look at some data. For convenience, we will symbolize the normal *CCR5* allele as *A* and the Δ*32* deletion allele as *a*. In one study of 1000 French people whose DNA was genotyped for *CCR5*, the numbers of homozygous normal (*AA*), heterozygous Δ*32* (*Aa*), and homozygous Δ*32* (*aa*) individuals were as follows:

795 *AA* 190 *Aa* 15 *aa*

These numbers contain a great deal of information about the population, such as whether it is a single homogeneous population or a mixture of genetically somewhat different subpopulations. To interpret this information, first note that the sample contains two types of data: (1) the number of each of the three genotypes, and (2) the

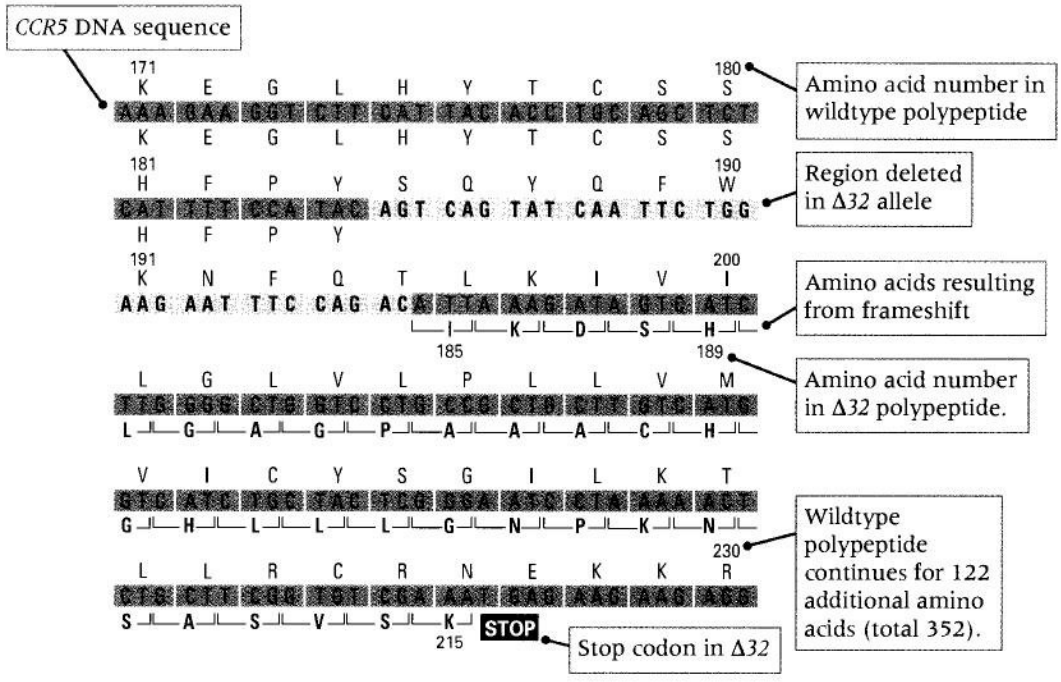


Figure 17.1 DNA sequence of a portion of the normal *CCR5* chemokine receptor gene and of the Δ*32* deletion allele, along with their products of translation.

number of individual *CCR5* normal (*A*) and Δ*32* (*a*) alleles. Furthermore, the 1000 genotypes represent 2000 alleles of the *CCR5* gene, because each human genome is diploid. These alleles break down as shown in **Figure 17.2**. Each homozygous *AA* genotype represents two *A* alleles, each homozygous *aa* genotype two *a* alleles, and each heterozygous *Aa* genotype one allele of each type. By the kind of allele counting in Figure 17.2, the sample of 1000 people therefore represents 1780 *CCR5* normal (*A*) alleles and 220 Δ*32* (*a*) alleles.

Usually, it is more convenient to analyze the data in terms of relative frequency than in terms of the observed numbers. For genotypes, the **genotype frequency** in a population is the proportion of organisms that have the particular genotype. For each allele, the **allele frequency** is the proportion of all alleles that are of the specified type. In the *CCR5* example, the genotype frequencies are obtained by dividing the observed numbers by the total sample size—in this case, 1000. Therefore, the genotype frequencies are

$$0.795\ AA \quad 0.190\ Aa \quad 0.015\ aa$$

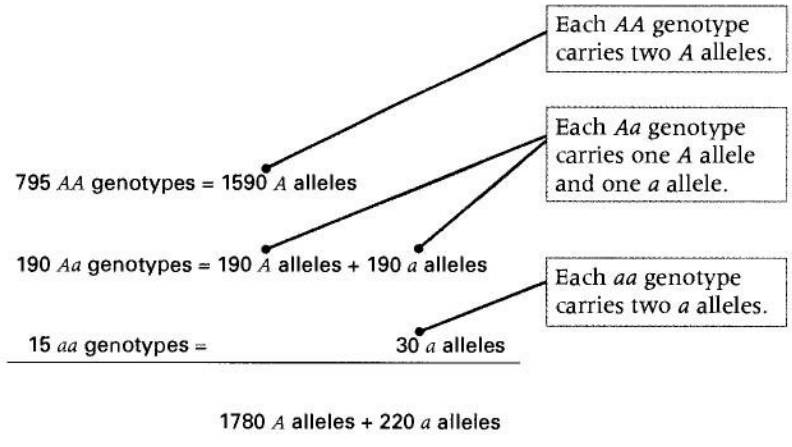


Figure 17.2 Analysis of the alleles present among a sample of 1000 people genotyped for the *CCR5* receptor gene. The symbol *a* refers to the Δ*32* allele.

Similarly, the allele frequencies are obtained by dividing the observed number of each allele by the total number of alleles (in this case, 2000):

$$\text{Allele frequency of } A = \frac{1780}{2000} = 0.89$$

$$\text{Allele frequency of } a = \frac{220}{2000} = 0.11$$

Note that the genotype frequencies add up to 1.0, as do the allele frequencies. This is a consequence of their definition in terms of proportions: They must add up to 1.0 when all of the possibilities are taken into account. A population with an allele frequency of 1.0 for some allele is said to be **fixed** for the allele. If an allele frequency becomes 0, the allele is said to be **lost.**

One question we can ask at this stage is whether the genotypes in the sample have the frequencies we might expect if the alleles were joined in pairs at random. Two *A* alleles (*AA*) would then be paired with a frequency of $0.89 \times 0.89 = 0.7921$, an *A* allele would be paired with an *a* allele (*Aa*) with a frequency of $2 \times 0.89 \times 0.11 = 0.1958$, and two *a* alleles (*aa*) would be paired with a frequency of $0.11 \times 0.11 = 0.0121$. These values are, in fact, quite close to the observed genotype frequencies. It is also noteworthy that the allele frequencies among adults, as we have calculated them in Figure 17.2, are the same as those among the gametes produced by the adults. This is true because Mendelian segregation ensures that each heterozygous genotype will produce equal numbers of each type of gamete (**Figure 17.3**). Thus when we calculate the gametic frequencies, the heterozygous genotypes contribute equally to both classes of gametes, whereas the homozygous *AA* and homozygous *aa* genotypes contribute only *A*-bearing or only *a*-bearing gametes, respectively. Consequently, when the genotype frequencies among the parents are taken into account, as well as Mendelian segregation in the heterozygous *Aa* genotypes, the gametes produced in the population have the composition deduced in **Figure 17.4**. The equality between the allele frequencies among the adults and those among the gametes produced by the adults, namely 0.89 *A* and 0.11 *a*, must be true whenever each adult in the population produces the same number of functional gametes. The apparent random pairing of alleles in the adult genotypes is also important, because it is a clue as to how the gametes are united in fertilization. This issue is examined in the next section.

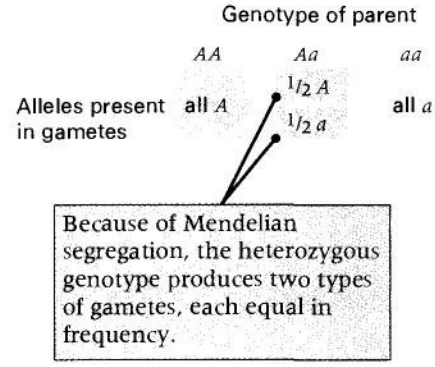


Figure 17.3 Mendelian considerations in population genetics. Each homozygous genotype produces gametes containing a single allele, but a heterozygous genotype, because of Mendelian segregation, produces two types of gametes equal in frequency.

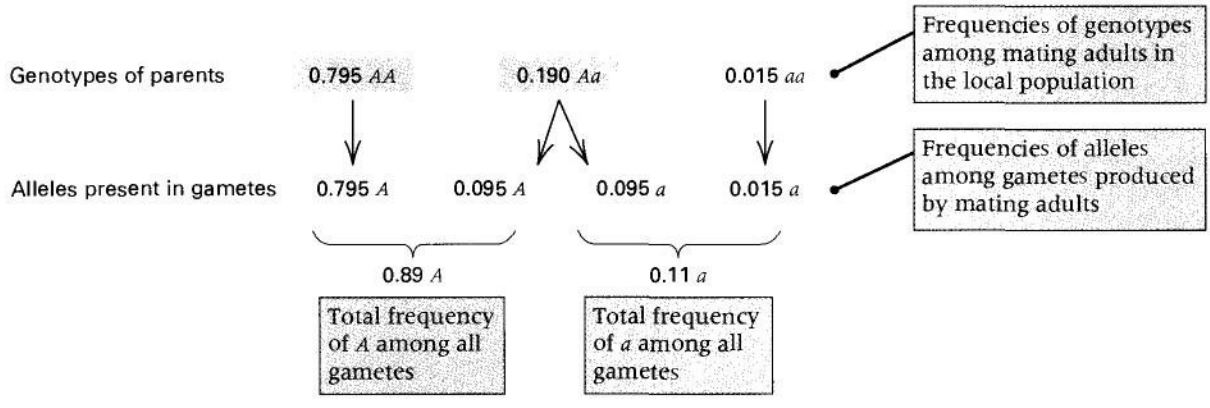


Figure 17.4 Calculation of the allele frequencies in gametes. In any population, the frequency of gametes containing any particular allele equals the frequency of the genotypes that are homozygous for the allele, plus one-half the frequency of all genotypes that are heterozygous for the allele.

17.2 The Matter of Mating

The transmission of genetic material from one generation to the next must be analyzed in terms of alleles rather than genotypes, because genotypes are disrupted in each generation by the processes of segregation and recombination. When gametes unite in fertilization, the genotypes of the next generation are formed. The genotype frequencies in the zygotes of the progeny generation are determined by the frequencies with which the parental gametes come together, and these frequencies are, in turn, determined by the **mating system** in which the genotypes pair as mates. Three types of mating systems predominate in natural populations: random mating, assortative mating, and inbreeding.

In **random mating,** the mating pairs are formed without regard to genotype: Two genotypes mate in proportion to their relative frequencies in the population. A population can undergo random mating with respect to some traits but nonrandom mating with respect to other traits at the same time. **Assortative mating** is a type of nonrandom mating in which mates are formed on the basis of their degree of similarity in phenotype. We speak of *positive assortative mating* if the mating pairs are more similar in phenotype than would be expected by chance, and of *negative assortative mating* if they are more dissimilar. Positive assortative mating takes place for many traits in human populations; for example, married couples tend to resemble each other more in skin color and height than would be expected by chance. Negative assortative mating is not so common as positive assortative mating, but a famous example is found in some species of primroses (*Primula*) and their relatives. In most populations of these species, two types of flowers are found in approximately equal proportions (**Figure 17.5**). One type, known as *pin*, has a long style and short stamens. The anthers are located atop the stamens and produce pollen; the stigma is located atop the style and receives pollen. Pin flowers produce pollen low in the flowers and receive pollen high in the flowers. In the other type of plant, known as *thrum*, the relative positions of stigma and anthers are reversed. Consequently, insect pollinators that work low in the flowers pick up mainly pin pollen and deposit it mainly on thrum stigmas, whereas pollinators that work high in the flowers pick mainly thrum pollen and deposit it mainly on pin stigmas. This behavior promotes negative assortative mating, which is further enhanced because pin pollen grows better in thrum stigmas, and vice versa. The genetic basis of the flower type resides in a single chromosomal region: Pin is homozygous *ss*, and thrum is heterozygous *Ss*.

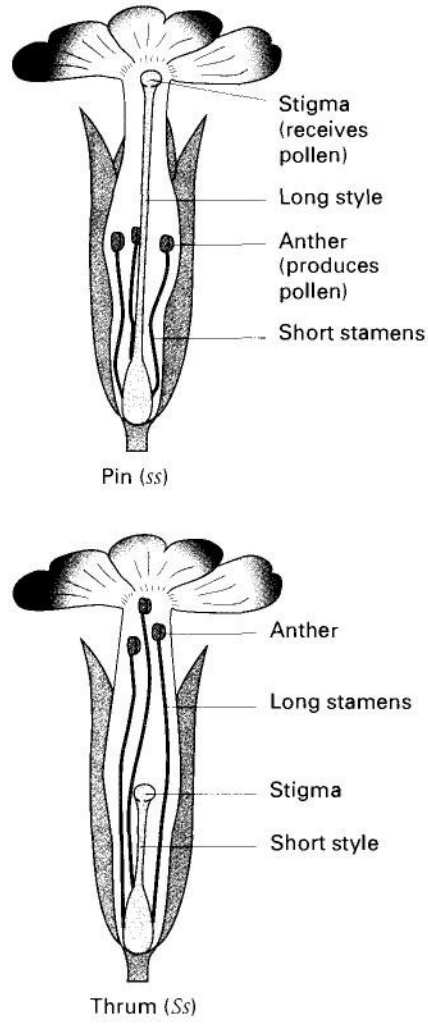


Figure 17.5 Pin and thrum forms of flowers of *Primula officalis*. Negative assortative mating is promoted by the anatomy of these flowers. Pollinators that work low in the flowers tend to pick up pin pollen and deposit it on thrum stigmas, and pollinators that work high in the flowers tend to do the reverse.

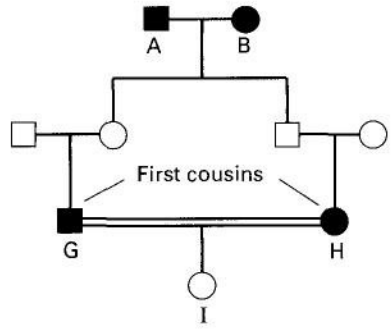


Figure 17.6 An inbreeding pedigree in which individual I is the result of a mating between first cousins.

A third important type of mating system is **inbreeding,** mating between relatives. In human pedigrees, a mating between relatives is often called **consanguineous mating.** Figure 17.6 represents an example of inbreeding. In this case, the female I is the offspring of a first-cousin mating (G with H). The closed loop in the pedigree is diagnostic of inbreeding, and the individuals designated A and B are called *common ancestors* of I, because they are ancestors of both of I's parents. Because A and B are common ancestors, a particular allele present in A (or in B) could, by chance, be transmitted in inheritance down both sides of the pedigree, to meet again in the formation of I. This possibility, which is the most important characteristic and consequence of inbreeding, will be discussed later in this chapter.

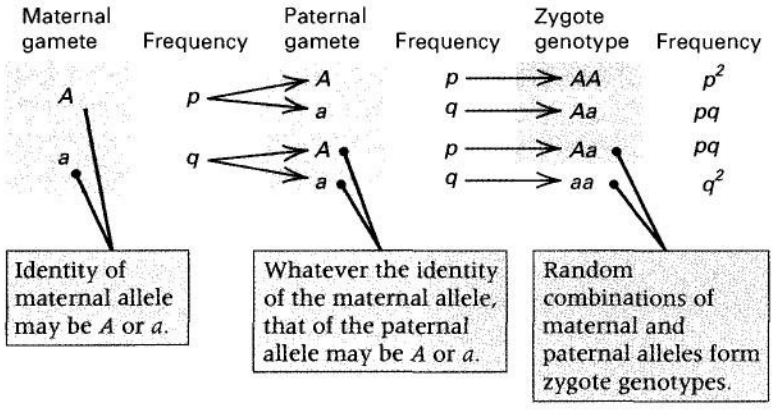


Figure 17.7 When gametes containing either of two alleles unite at random to form the next generation, the genotype frequencies among the zygotes are given by the ratio $p^2 : 2pq : q^2$. This constitutes the Hardy–Weinberg principle.

Random Mating and the Hardy–Weinberg Principle

Random mating is by far the most prevalent mating system for most species of animals and plants, except for plants that regularly reproduce through self-fertilization. In random mating, each type of mating pair is formed as often as would be expected through chance encounters between the genotypes.

With random mating, the relationship between allele frequency in one generation and genotype frequency in the next generation is particularly simple, because random mating of individuals is equivalent to random union of gametes. Conceptually, we may imagine all the gametes of a population as present in a large container. To form zygote genotypes, pairs of gametes are withdrawn from the container at random. To be specific, consider two alleles A and a with allele frequencies p and q, respectively, where $p + q = 1$. In the sample of *CCR5* receptor genotypes considered earlier in this chapter, we calculated

$$p = 0.89 \text{ for the } A \text{ allele}$$
$$q = 0.11 \text{ for the } a \text{ allele}$$

The genotype frequencies expected with random mating can be deduced from the tree diagram in Figure 17.7. The gametes at the left represent the eggs and those in the middle the sperm. The genotypes that can be formed with two alleles are shown at the right, and with random mating, the frequency of each genotype is calculated by multiplying the allele frequencies of the corresponding gametes. However, the genotype Aa can be formed in two ways: The A allele could have come from the mother (top part of diagram) or from the father (bottom part of diagram). In each case, the frequency of the Aa genotype is pq; considering both possibilities, we find that the frequency of Aa is $pq + pq = 2pq$. Consequently, the overall genotype frequencies expected with random mating are

$$AA: p^2 \quad Aa: 2pq \quad aa: q^2 \qquad (1)$$

The frequencies p^2, $2pq$, and q^2 result from random mating for a gene with two alleles; they constitute the **Hardy–Weinberg principle,** after Godfrey Hardy and Wilhelm Weinberg, who derived it

independently of each other in 1908. Sometimes the Hardy–Weinberg principle is demonstrated by the type of Punnett square illustrated in **Figure 17.8**. Such a square is completely equivalent to the tree diagram used on the previous page. Although the Hardy–Weinberg principle is exceedingly simple, it has a number of important implications that are not obvious. These are described in the following sections.

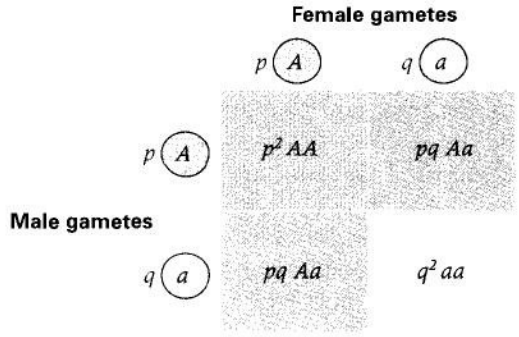


Figure 17.8 A Punnett square showing, in a cross-multiplication format, the ratio $p^2 : 2pq : q^2$ characteristic of random mating.

Implications of the Hardy–Weinberg Principle

One important implication of the Hardy–Weinberg principle is that *the allele frequencies remain constant from generation to generation.* To understand this point, note that with random mating and gametic frequencies p and q, the frequencies of zygotes with the genotypes *AA*, *Aa*, and *aa* are p^2, $2pq$, and q^2, respectively. Assuming equal *viability* (ability to survive) among the zygotes, these frequencies will equal those among adults. Furthermore, if all of the adult genotypes are equally fertile, then the frequency, p', of the allele *A* among gametes of the next generation will be

$$p' = p^2 + 2pq/2 = p(p + q) = p$$

because $p + q = 1$. This argument shows that the frequency of allele *A* remains constant at the value of p through the passage of one (or any number of) complete generation(s). This principle depends on certain assumptions, of which the most important are the following:

- Mating is random; there are no subpopulations that differ in allele frequency.
- Allele frequencies are the same in males and females.
- All the genotypes are equal in viability and fertility (*selection* does not operate).
- Mutation does not occur.
- Migration into the population is absent.
- The population is sufficiently large that the frequencies of alleles do not change from generation to generation merely because of chance.

A Test for Random Mating

As a real application of the Hardy–Weinberg principle, consider again the *CCR5* receptor discussed in Section 17.1. The frequencies of the *CCR5* normal allele (*A*) and *Δ32* allele (*a*) among 1000 adults were 0.89 and 0.11, respectively; assuming random mating, the expected genotype frequencies can be calculated from Equation (1) as

$$AA: \quad (0.89)^2 = 0.7921$$
$$Aa: \quad 2(0.89)(0.11) = 0.1958$$
$$aa: \quad (0.4575)^2 = 0.0121$$

Note that these are the same frequencies calculated earlier on the basis of the supposition that the alleles were joined in pairs at random. Now we know what this means: It means genotype frequencies according to the Hardy–Weinberg principle. Because the total number of people in the sample is 1000, the expected numbers of the genotypes are 792.1 *AA*, 195.8 *Aa*, and 12.1 *aa*. The *observed* numbers were 795 *AA*, 190 *Aa*, and 15 *aa*. Goodness of fit between the observed and expected numbers can be determined by means of the χ^2 test described in Chapter 4. In this case,

$$\chi^2 = \frac{(795 - 792.1)^2}{792.1} + \frac{(190 - 195.8)^2}{195.8} + \frac{(15 - 12.1)^2}{12.1}$$
$$= 0.877$$

The fit is obviously satisfactory, but to evaluate it quantitatively we need to use the chart in Figure 4.22 on page 163. There is also an adjustment needed in the degrees of freedom. Normally, with 3 classes of data, we would have 2 degrees of freedom. But when any quantity, such as an allele

frequency, is estimated from the data, 1 degree of freedom must be deducted for each quantity estimated. In this case we estimated the gene frequency p ($q = 1 - p$ follows automatically), so we lose 1 of the 2 degrees of freedom we would otherwise have, and thus the appropriate number of degrees of freedom for the test is 1. For this we obtain a P value of 0.35, which means that the hypothesis of random mating can account for the data. On the other hand, the χ^2 test detects only deviations that are rather large, so a good fit to Hardy–Weinberg frequencies should not be over interpreted, because:

> It is entirely possible for one or more assumptions of the Hardy–Weinberg principle to be violated, including the assumption of random mating, and still not produce deviations from the expected genotype frequencies that are large enough to be detected by the χ^2 test.

The use of a χ^2 test to determine departures from random mating may seem somewhat circular, inasmuch as the data themselves are used in calculating the expected genotype frequencies, but the use of the data is precisely compensated for by the simple expedient of deducting 1 degree of freedom for every quantity estimated from the data. To verify that the χ^2 test as employed is not circular, consider the following observed numbers of genotypes in the flower *Phlox cuspidata* for a gene coding for phosphoglucomutase: $Pgm^a\ Pgm^a$, 15; $Pgm^a\ Pgm^b$, 6; and $Pgm^b\ Pgm^b$, 14. You should verify for yourself that the allele frequencies of Pgm^a and Pgm^b are 0.51 and 0.49, respectively, and that the expected numbers of the genotypes with random mating are 9.1, 17.5, and 8.4, respectively. The χ^2 value is 17.1, which has 1 degree of freedom because there are 3 classes of data (the observed number of each of the three genotypes), and 1 parameter (the allele frequency of either Pgm^a or Pgm^b) had to be estimated from the data. The corresponding probability is about 0.0002. The probability associated with the χ^2 value is less than 0.05 (in this case, very much less), so these data must be judged significantly different from the Hardy–Weinberg expectations. The reason, as we will soon see, is inbreeding: In this plant, some of the ovules are fertilized by self-fertilization.

Frequency of Heterozygous Genotypes

Another important implication of the Hardy–Weinberg principle is that *for a rare allele, the frequency of heterozygotes far exceeds the frequency of the rare homozygote.* For example, when the frequency of the rarer allele is $q = 0.1$, the ratio of heterozygous genotypes to homozygous genotypes is

$$2pq/q^2 = 2p/q = 2(0.9)/0.1$$

or approximately 20; when $q = 0.01$, this ratio is about 200; and when $q = 0.001$, it is about 2000. In other words:

> When an allele is rare, there are many more heterozygotes than there are homozygotes for the rare allele.

The reason for this perhaps unexpected relationship is shown in **Figure 17.9**, which plots the frequencies of homozygous and heterozygous genotypes with random mating. Note that at allele frequencies near 0 or 1, the frequency of the heterozygous genotype goes to 0 as a linear function, whereas that of the rarer, homozygous genotype goes to 0 more rapidly.

One practical implication of this principle is seen in the example of cystic fibrosis, an inherited secretory disorder of the pancreas and lungs. Cystic fibrosis is one of the most common recessive severe genetic disorders among Caucasians; it affects about 1 in 1700 newborns. This is the frequency of homozygous recessives, but what are the allele frequencies? The calculation is straightforward, because with random mating, the frequency of homozygous recessives must correspond to q^2. In the case of cystic fibrosis,

$$q^2 = 1/1700 = 0.00059 \quad \text{or}$$
$$q = (0.00059)^{1/2} = 0.024$$

and consequently,

$$p = 1 - q = 1 - 0.024 = 0.976$$

The frequency of heterozygous carriers of the allele for cystic fibrosis is therefore

$$2pq = 2(0.976)(0.024) = 0.047 = 1/21$$

This calculation implies that, for cystic fibrosis, although only 1 person in 1700 is affected with the disease (homozygous), about 1 person in 21 is a carrier (hetero-

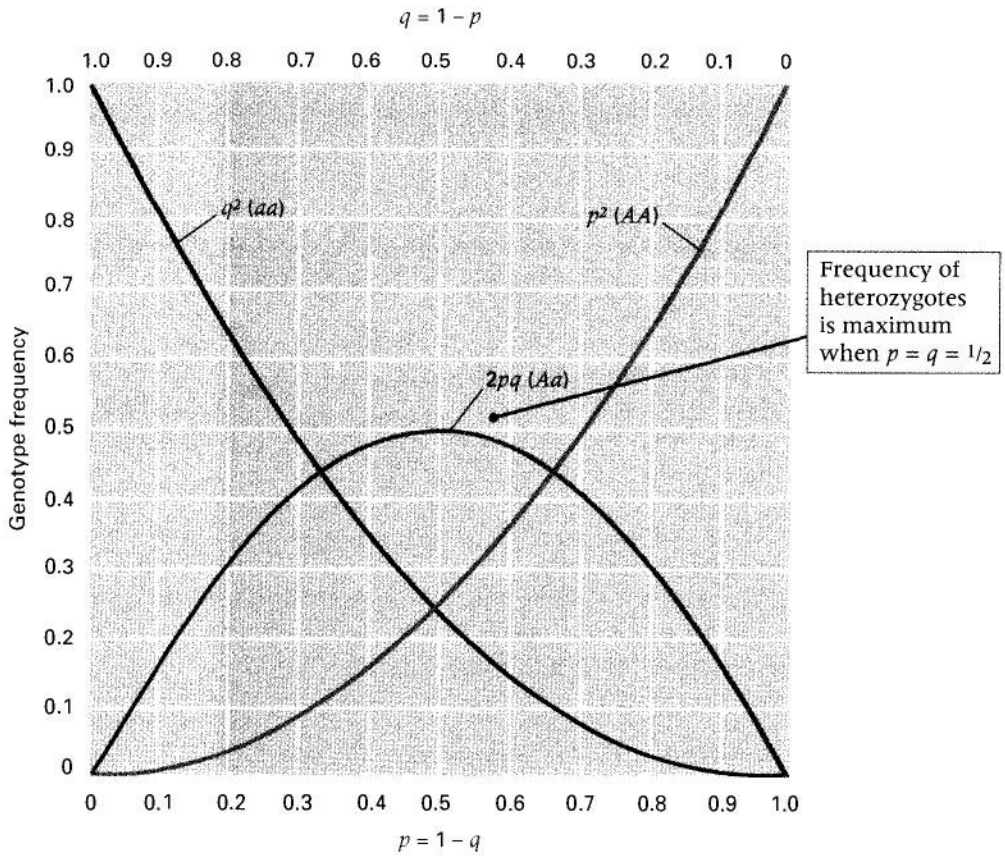


Figure 17.9 Graphs of p^2, $2pq$, and q^2. If the allele frequencies are between 1/3 and 2/3, then the heterozygous genotype has the greatest frequency.

zygous). The calculation should be regarded as approximate because it is based on the assumption of Hardy–Weinberg genotype frequencies. Nevertheless, considerations like these are important in predicting the outcome of population screening for the detection of carriers of harmful recessive alleles, which is essential in evaluating the potential benefits of such screening programs.

Multiple Alleles

Extension of the Hardy–Weinberg principle to multiple alleles of a single autosomal gene can be illustrated by a three-allele case. **Figure 17.10** shows the situation with three alleles, designated A_1, A_2, and A_3, with allele frequencies p_1, p_2, and p_3. As always, the allele frequencies must sum to 1, so $p_1 + p_2 + p_3 = 1.0$. As in Figure 17.8, the entry in each square is obtained by multiplying the frequencies of the alleles at

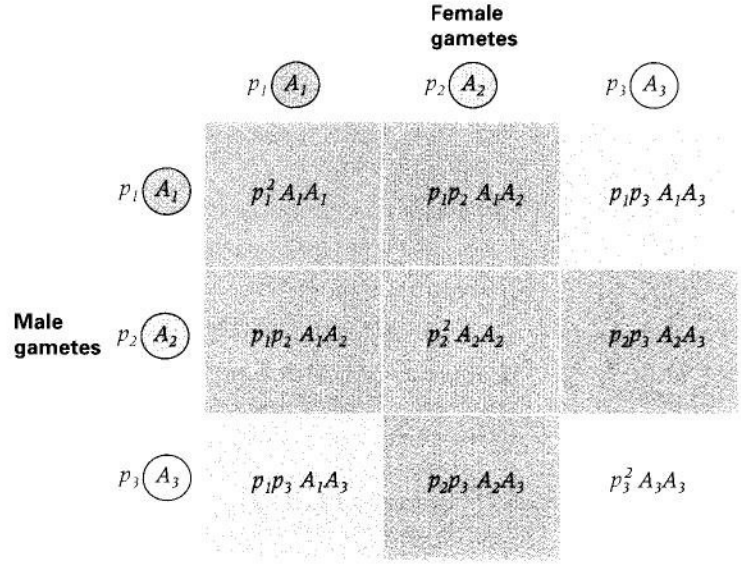


Figure 17.10 Punnett square showing the results of random mating with three alleles.

CoNnection
Twin Bill

Wilhelm Weinberg 1908
Stuttgart, Germany
On the Demonstration of Heredity in Man

This paper is regarded as a classic even though its main conclusion is completely wrong! Weinberg was a practicing physician who took a great interest in the genetic basis of twin births in human beings. His studies, summarized in this excerpt, led him to conclude that twinning was due to a simple Mendelian recessive gene. We now know that this is totally incorrect. Twinning is actually a multifactorial trait, which means that it is affected by several genes, and probably also by environmental factors. However, while on the way to this erroneous conclusion, Weinberg correctly deduced the numerical frequencies with which various genotypes are to be expected under the condition of random mating (then called panmixis*). Weinberg used the symbols* m *and* n *for allele frequencies, where we now use* p *and* q*, and he used* AA, AB, *and* BB *for genotypes, where we now use* AA, Aa, *and* aa*. We have substituted the modern symbols. The right arrows in the excerpt replace equals signs in the original. To the left of the arrows are the frequencies of all possible matings pairs, assuming random mating; to the right of the arrows are the frequencies of all possible progeny, taking Mendelian segregation into account. The population frequencies of* AA, Aa, *and* aa *are obtained by summing the progeny frequency of each of these genotypes across all possible mating pairs—for example, for* AA, $p^4 + 2p^3q + p^2q^2$.

How does the numerical influence of Mendelian inheritance behave under the influence of panmixis?

If we construct the genealogical table of an individual, that is, a schematic synopsis of his ancestors, we have a synopsis of those individuals who may have influenced, through the germ plasm, certain qualities of the individual under consideration. However, among all theoretical possibilities only a few are actually to be taken into consideration. . . . According to the Mendelian rule it appears that each quality in the final analysis is determined [by the alleles present in] only two ancestors. . . . [In my considerations of twinning] I was confronted

the corresponding margins; any homozygote (such as A_1A_1) has a random-mating frequency equal to the square of the corresponding allele frequency (in this case, p_1^2). Any heterozygote (such as A_1A_2) has a random-mating frequency equal to twice the product of the corresponding allele frequencies (in this case, $2p_1p_2$). The extension to any number of alleles is straightforward:

Frequency of any homozygote =
square of allele frequency
Frequency of any heterozygote =
$2 \times$ product of allele frequencies (2)

Multiple alleles determine the human ABO blood groups (Chapter 3), and there are three principal alleles, designated I^A, I^B, and I^O. In one study of 3977 Swiss people, the allele frequencies were found to be 0.27 I^A, 0.06 I^B, and 0.67 I^O. Applying the rules for multiple alleles, we can expect the genotype frequencies resulting from random mating to be as shown in **Figure 17.11**. Because both I^A and I^B are dominant to I^O, the expected frequency of blood-group *phenotypes* is that shown at the right in the illustration. Note that the majority of A and B phenotypes are heterozygous for the I^O allele; this is because the I^O allele has such a high frequency in the population.

DNA Typing

Genetic variation in the form of DNA polymorphisms is common in the human genome as well as in genomes of many other organisms. This much is clear from the application of the variety of experimental methods by which DNA polymorphisms can be detected, which we examined in Chapter 2. In the human genome there is so much genetic variation that, except for identical twins and other multiple births that arise from a single zygote, no two human beings are (or ever have been) genetically identical. Each human genotype is unique. One practical application of the principle of genetic uniqueness has come through **DNA typing,** a procedure in which biological samples of unknown hu-

with the question: How does the numerical influence of Mendelian inheritance behave under the influence of panmixis? . . . I start with the general hypothesis that there are initially a proportion p of males and females who are pure representatives of type AA, and likewise a proportion q of individuals who are pure representatives of type aa. If these are crossed at random the composition of the daughter generation is obtained by using the symbolism of the binomial theorem:

$$(p\,AA + q\,aa)^2 = p^2\,AA + 2pq\,Aa + q^2\,aa$$

If the male and female members of the first generation are crossed at random the following frequenc[ies] of the different combinations of crosses are obtained:

$$p^2 \times p^2\,(AA \times AA) \rightarrow p^4\,AA$$
$$2 \times p^2 \times 2pq\,(AA \times Aa) \rightarrow 2p^3q\,AA + 2p^3q\,Aa$$
$$2 \times p^2 \times q^2\,(AA \times aa) \rightarrow 2p^2q^2\,Aa$$
$$2pq \times 2pq\,(Aa \times Aa) \rightarrow p^2q^2\,AA + 2p^2q^2\,Aa + p^2q^2\,aa$$
$$2 \times 2pq \times q^2(Aa \times aa) \rightarrow 2pq^3\,Aa + 2pq^3\,aa$$
$$q^2 \times q^2\,(aa \times aa) \rightarrow q^4\,aa$$

or the relative frequencies

$$AA: p^2$$
$$Aa: 2pq$$
$$aa: q^2$$

We thus obtain the same distribution of pure types and hybrids for each generation under the influence of panmixis and therewith the possibility of calculating, for each generation, how the representation of these types is arranged . . . among the parents, sibs, and children of the various types and hybrids. . . . [These formulas] have enabled me to utilize not only the absolutely few cases in which children of mothers of twins married each other but also my entire earlier material. . . . The situation found in the inheritance of twinning best finds its explanation in the assumption that the trait for twinning is inherited according to the Mendelian rule and is recessive.

Source: Naturkunde in Württemberg, Stuttgart. 64: 368-382. Original in German. Excerpt from translation in S. H. Boyer, IV. 1963. *Papers on Human Genetics* (Englewood Cliffs, NJ: Prentice-Hall), pp. 4–15.

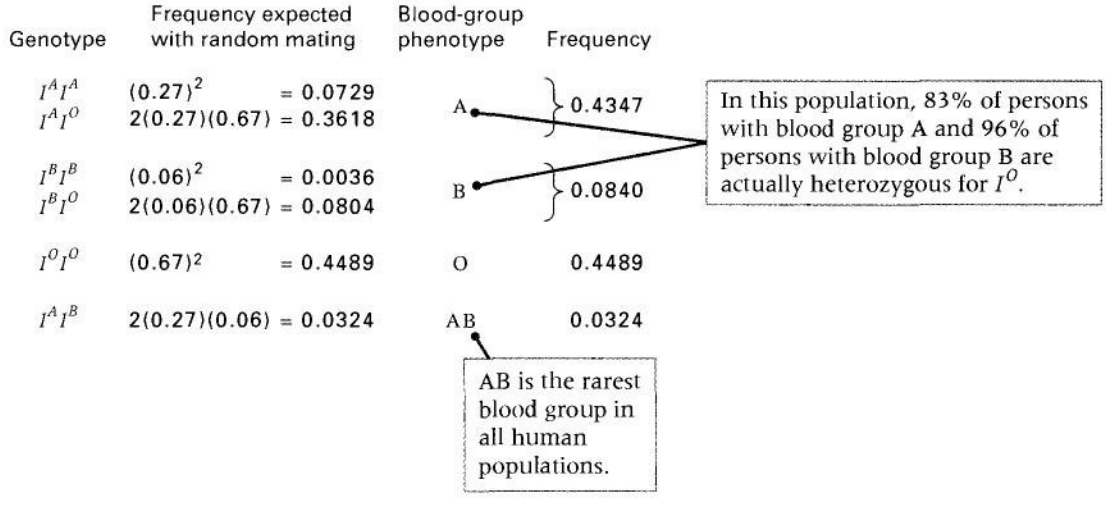


Figure 17.11 Random-mating frequencies for the three alleles governing the ABO blood groups.

man origin are matched with their source through the use of polymorphic DNA markers. Even small samples of human material often contain enough DNA that the genotype can be determined for a number of markers and matched against those present among a group of suspects. Typical examples of crime-scene evidence include blood, semen, hair roots, and skin cells. Even a small number of cells is sufficient,

because predetermined regions of DNA can be amplified by the polymerase chain reaction (Chapter 2).

If a suspect's DNA contains sequences that are clearly not present in a sample collected at the crime scene, then the sample must have originated from a different person. On the other hand, if a suspect's DNA *does* match that from the crime scene, then the suspect could be the source. The strength of the DNA evidence depends on the number of polymorphisms examined and on the number of alleles present in the population. The greater the number of polymorphisms that match, especially if they are highly polymorphic, the stronger the evidence linking the suspect to the sample taken from the scene of the crime. With a sufficient number and quality of matching DNA markers, DNA typing can be as reliable a method of individual identification as fingerprints.

Figure 17.12 shows one type of polymorphism used in DNA typing. The restriction fragments that correspond to each allele differ in length because they contain different numbers of units repeated in tandem; the polymorphism is said to be a *variable number of tandem repeats (VNTR) polymorphism* (see Figure 2.26 on page 70 for an illustration). Such markers are of value in DNA typing because many alleles are possible, owing to the variable number of repeating units. In Figure 17.12, the lanes in the gel labeled M contain multiple DNA fragments of known size to serve as molecular-weight markers. Each of lanes 1–9 contains DNA from a single person. Two typical features of VNTRs are to be noted:

1. Most people are heterozygous for VNTR alleles that produce restriction fragments of different sizes. Heterozygosity is indicated by the presence of two distinct bands. In Figure 17.12, only the person numbered 1 appears to be homozygous for a particular allele.
2. The restriction fragments from different people cover a wide range of sizes. The variability in size indicates that the population as a whole contains many VNTR alleles. Although many alleles can be present in the population as a whole, each person can have no more than two alleles.

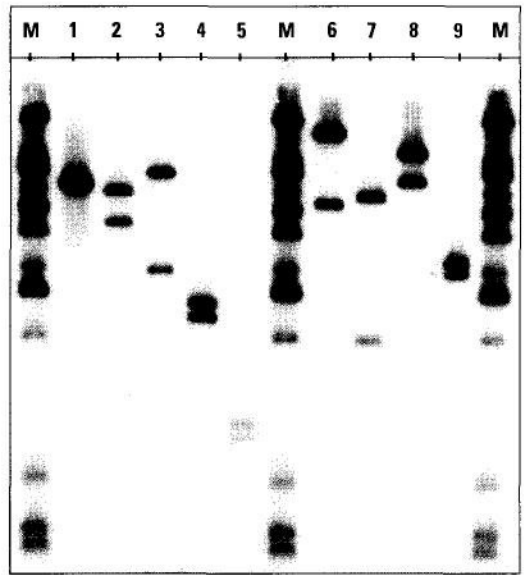


Figure 17.12 Genetic variation in a VNTR used in DNA typing. Each numbered lane contains DNA from a single person; the DNA has been cleaved with a restriction enzyme, separated by electrophoresis, and hybridized with radioactive probe DNA. The lanes labeled M contain molecular-weight markers. [Courtesy of R. W. Allen.]

The reason why VNTRs are useful in DNA typing is also evident in Figure 17.12: Each of the nine people tested has a different pattern of bands and thus could be identified uniquely by means of this VNTR. On the other hand, the uniqueness of each person in Figure 17.12 is due in part to the high degree of polymorphism of the VNTR and in part to the small sample size. If more people were examined, then pairs that matched in their VNTR types by chance would certainly be found. This is why matches at multiple, independent loci are needed for reliable DNA typing. The greater the number of loci, the less likely it is that a match of DNA types is due to chance. With enough matching markers, it becomes virtually certain that the two DNA samples came from the same person. **Figure 17.13** shows an example of evidence in a serial rape case in Minnesota in which semen samples from seven victims (U1 through U7) were shown to match perfectly with DNA from one suspect (S2), who was also implicated on the basis of other evidence from some of the crime scenes. The suspect was convicted of 81 criminal charges and sentenced to 139 years in prison. He will become eligible for parole in the year 2087.

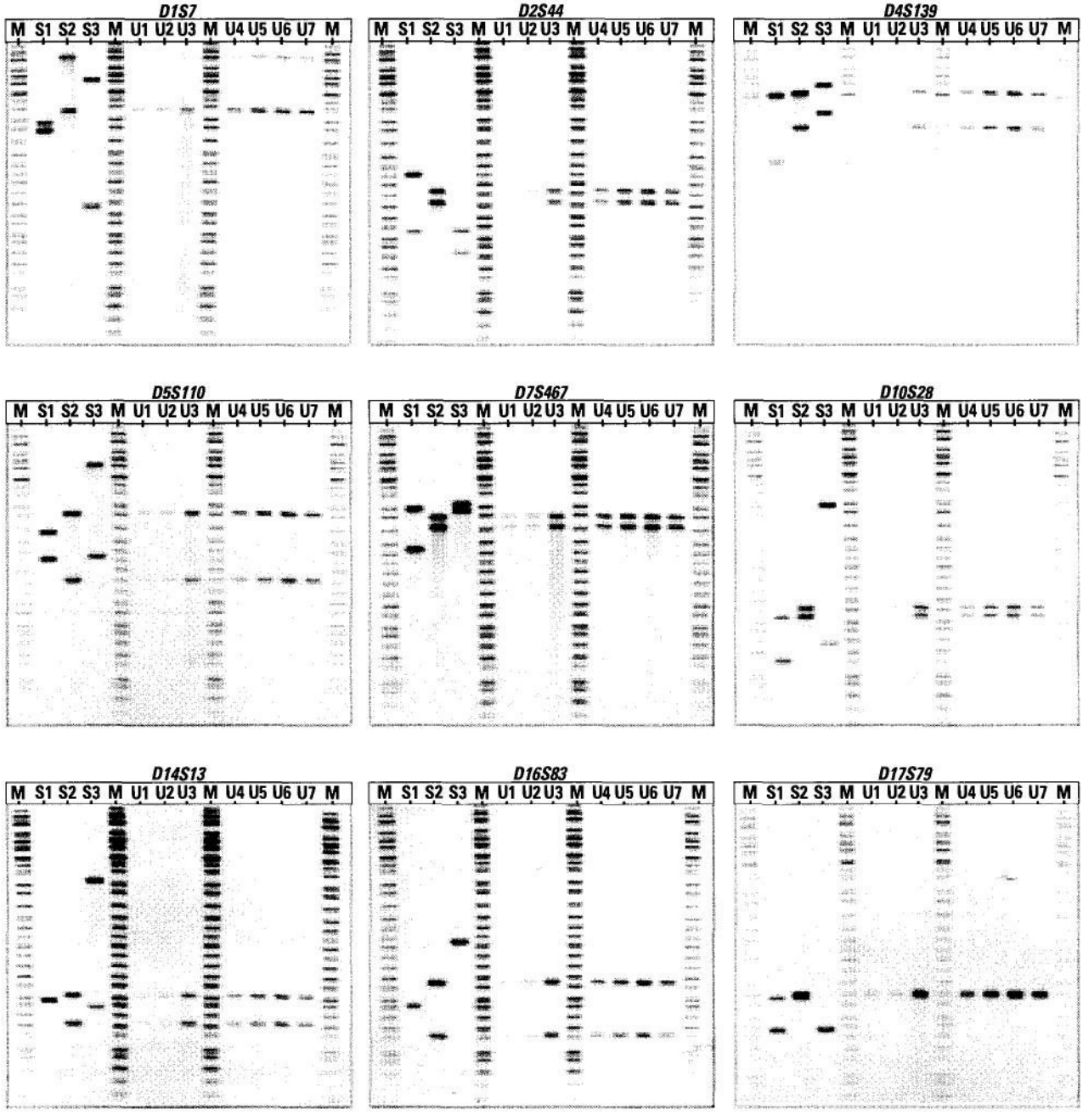


Figure 17.13 An example of DNA typing in a criminal case. Each panel is the result of DNA typing for a different VNTR. The lanes marked S1, S2, and S3 contain DNA from blood samples of three male suspects; those in columns U1 through U7 contain DNA from semen samples collected from seven female victims of rape. The lanes marked M contain molecular-weight markers. In each case, the DNA from suspect S2 matches the samples obtained from the victims. [Courtesy of Steven J. Redding, Office of the Hennepin County District Attorney, Minneapolis, and Lowell C. Van Berkom and Carla J. Finis, Minnesota Bureau of Criminal Apprehension.]

DNA Exclusions

DNA types that do *not* match can be as important in exonerating innocent suspects as those that *do* match are in implicating the guilty. An example that illustrates the use of DNA typing in paternity testing is shown in **Figure 17.14**. The numbers 1 and 2 designate different cases, in each of which a man was accused of fathering a child. In each case, DNA was obtained from the mother (M), the child (C), and the accused man (A). The pattern of bands obtained for one VNTR is shown. The lanes labeled A + C contain a mixture of DNA from the accused man and the child. In case 1, the lower band in the child was inherited from the mother and the upper band from the father; because the upper band is the same size as one of those in the accused, the accused man could have contributed this allele to the child. This finding does not *prove* that the accused man is the father; it says only that he cannot be excluded on the basis of this particular gene. (On the other hand, if a large enough number of genes are studied, and the man cannot be excluded by any of the genes, then it makes it more likely that he really is the father.) Case 2 in Figure 17.14 is an exclusion. The band denoted by the arrows is the band inherited by the child from the mother. The other band in the child does not match either of the bands in the accused, so the accused man could not be the biological father. (In theory, mutation could be invoked to explain the result, but this possibility is extremely unlikely.)

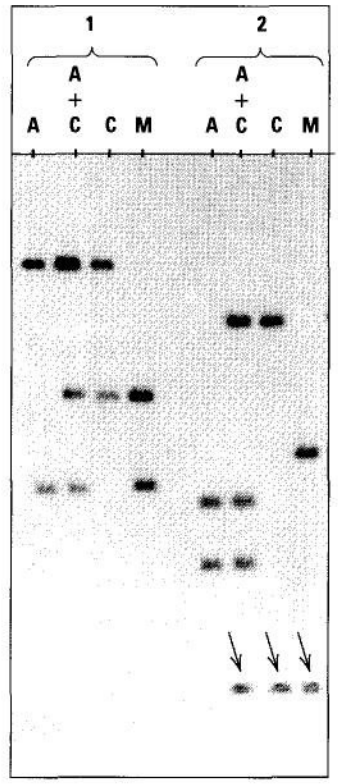


Figure 17.14 Use of DNA typing in paternity testing. The sets of lanes numbered 1 and 2 contain DNA samples from two different paternity cases. In each case, the lanes contain DNA fragments from the following sources: M, the mother; C, the child; A, the accused father. The lanes labeled A + C contain a mixture of DNA fragments from the accused father and the child. The arrows in case 2 point to bands of the same size that are present in lanes M, C, and A + C. Note that the male accused in case 2 could not be the father because neither of his bands is shared with the child. [Courtesy of R. W. Allen.]

Differences Among Populations

Most of the genetic variation in human populations is due to differences in genotype among individuals in the same subpopulation. Few alleles are very common in one ethnic group and very rare in others. Because human subpopulations have undergone genetic divergence relatively recently and have maintained a significant degree of intermixture, the genetic differences that exist among human subpopulations are largely differences in allele frequencies, not in the presence or absence of alleles. A typical example is shown in **Figure 17.15**. Each bar shows the range in frequency (minimum to maximum) of each allele of a VNTR (*D1S80*) observed among approximately 30 subpopulations throughout the world. There are clear differences in allele frequency, and for some alleles the range in frequency is quite large; for example, alleles with 21, 30, 31, or 34 repeats are relatively rare in some subpopulations but relatively common in others. Even the common alleles, with 18 and 24 repeats, can differ substantially in frequency from one subpopulation to the next. On the other hand, except for alleles that are relatively rare, most alleles are found in most subpopulations, and those that are relatively common in some subpopulations tend to be relatively common in most.

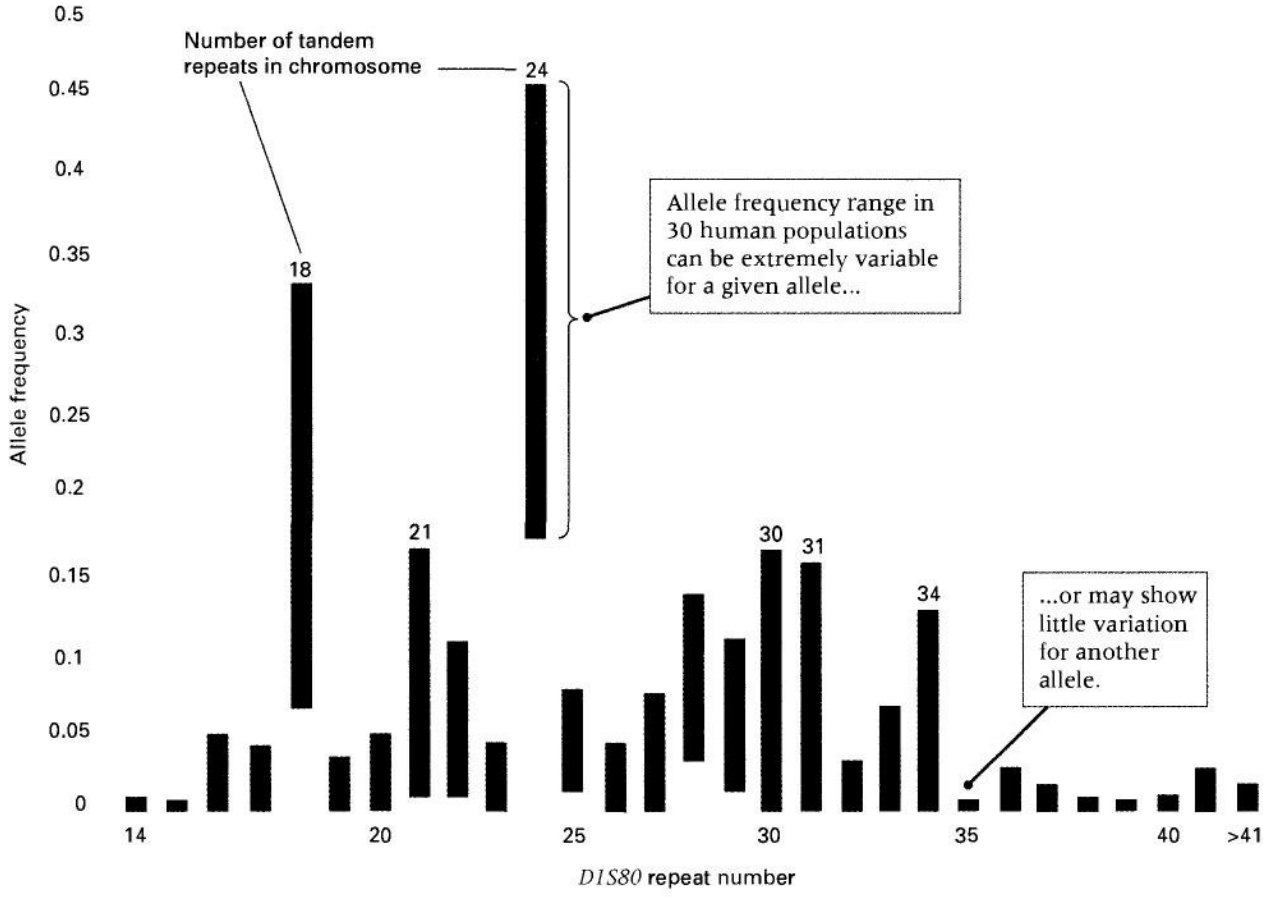


Figure 17.15 Range of allele frequencies found among human subpopulations for the VNTR *D1S80*. [Data from B. Budowle et al. *J. Forensic Science* 1995. 40:38.]

X-Linked Genes

If a gene with alleles *A* and *a* is X-linked, the result of random mating is illustrated in **Figure 17.16**. The principles are the same as those considered earlier, but male gametes carrying the X chromosome (Figure 17.16A) must be distinguished from those carrying the Y chromosome (Figure 17.16B). When the male gamete carries an X chromosome, the Punnett square is exactly the same as that for the two-allele autosomal gene in Figure 17.8, but all the offspring are female. Consequently, among female progeny, the genotype frequencies are

$$AA: p^2 \quad Aa: 2pq \quad aa: q^2$$

When the male gamete carries a Y chromosome, the outcome is quite different (Figure 17.16B). All the offspring are male, and each has only one X chromosome, which is inherited from the mother. Therefore, each

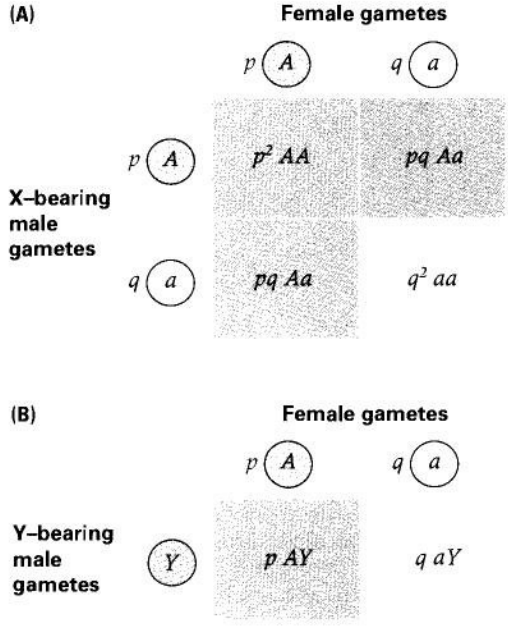


Figure 17.16 The results of random mating for an X-linked gene. (A) Genotype frequencies in females. (B) Genotype frequencies in males.

male offspring receives only one copy of each X-linked gene, and the genotype frequencies among males are the same as the allele frequencies: *A* males with frequency *p*, and *a* males with frequency *q*.

An important implication of Figure 17.16 is that if *a* is a rare recessive allele, then there will be many more males exhibiting the trait than females, because the frequency of affected females (q^2) will be much smaller than the frequency of affected males (q). This principle is illustrated in **Figure 17.17**. As the allele frequency of the recessive decreases toward 0, the frequencies of affected males and females both decrease, but the frequency of affected females decreases faster. The result is that the ratio of affected males to affected females increases. At an allele frequency of $q = 0.3$, for example, the ratio of affected males to affected females is 3.3, but for an allele frequency of $q = 0.1$, the ratio of affected males to affected females is 10.0. In general, the ratio of affected males to affected females is q/q^2, or $1/q$.

For an X-linked recessive trait, the frequency of affected males provides an estimate of the frequency of the recessive allele. A specific example is found in the common form of X-linked color blindness in human beings. This trait affects about 1 in 20 males among Caucasians, so $q = 1/20 = 0.05$. The expected frequency of color-blind females is therefore estimated as $q^2 = (0.05)^2 = 0.0025$, or about 1 in 400.

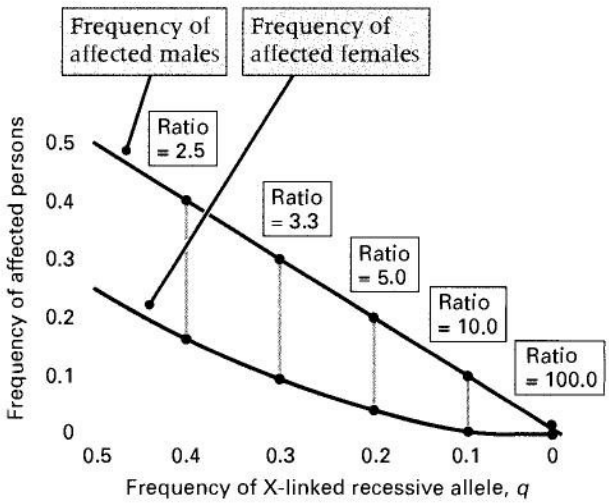


Figure 17.17 For a recessive X-linked allele, the upper curve gives the frequency of affected males (q) and the lower curve the frequency of affected females (q^2), for values of $q \leq 0.5$. Although both frequencies decrease as the recessive allele becomes rare, the frequency of affected females decreases faster. The result is that the ratio of affected males to affected females increases as the allele frequency decreases.

17.3 Inbreeding

Inbreeding means mating between relatives, such as first cousins. The principal consequence of inbreeding is that the frequency of heterozygous offspring is smaller than it is with random mating. This effect is seen most dramatically in repeated self-fertilization, which happens naturally in certain plants. Consider a hypothetical population consisting exclusively of *Aa* heterozygotes. With self-fertilization, each plant would produce offspring in the proportions 1/4 *AA*, 1/2 *Aa*, and 1/4 *aa*, which means that one generation of self-fertilization reduces the proportion of heterozygotes from 1 to 1/2. In the second generation, only the heterozygous plants can again produce heterozygous offspring, and only half of their offspring will again be heterozygous. Heterozygosity is therefore reduced to 1/4 of what it was originally. Three generations of self-fertilization reduces the heterozygosity to $1/4 \times 1/2 = 1/8$, and so on. The remainder of this section demonstrates how the reduction in heterozygosity because of inbreeding can be expressed in quantitative terms.

The Inbreeding Coefficient

Repeated self-fertilization is a particularly intense form of inbreeding, but weaker forms of inbreeding are qualitatively similar in that they also lead to a reduction in heterozygosity. A convenient measure of the effect of inbreeding is based on the reduction in heterozygosity. To put this into symbols, suppose that H_I is the frequency of heterozygous genotypes in a population of inbred individuals. The most widely used measure of inbreeding is called the **inbreeding coefficient,** symbolized *F*, which is defined as the proportionate reduction in H_I compared with the value of $2pq$ that would be expected with random

mating. Expressed as an equation, this definition implies that

$$F = (2pq - H_I)/2pq \qquad (3)$$

Equation (3) can be rearranged as

$$H_I = 2pq(1 - F)$$

The homozygous genotypes increase in frequency as the heterozygotes decrease in frequency, and overall, the genotype frequencies in the inbred population are given by

$$\begin{aligned} AA:&\ p^2(1 - F) + pF \\ Aa:&\ 2pq(1 - F) \\ aa:&\ q^2(1 - F) + qF \end{aligned} \qquad (4)$$

These expressions are modifications of the Hardy–Weinberg principle that take inbreeding into account. When $F = 0$ (no inbreeding), the genotype frequencies are the same as those given in the Hardy–Weinberg principle in Equation (1): p^2, $2pq$, and q^2. At the other extreme, when $F = 1$ (complete inbreeding), the inbred population consists entirely of *AA* and *aa* genotypes in the frequencies p and q, respectively. Inbreeding alone does not change the allele frequencies in a population. This you can verify for yourself by calculating that, for any value of the inbreeding coefficient F, the allele frequencies of *A* and *a* remain at p and q, respectively.

A graphical representation of the genotype frequencies in Equation (4) is shown in **Figure 17.18**. For any gene in a population of organisms, a proportion of the population equaling $1 - F$ may be considered as having, by chance, escaped the effects of inbreeding, which means that the genotype frequencies are those expected with random mating: p^2, $2pq$, and q^2. In another proportion, equaling F, the alleles may be considered to have become homozygous solely because of inbreeding, with the *AA* and *aa* genotype frequencies equal to p and q, respectively.

Figure 17.19 shows graphically the reduction in heterozygosity due to inbreeding. Each curve is given by an equation of the form $H_I = 2pq(1 - F)$, where the case $F = 0$ corresponds to random mating. The curves show that as the inbreeding coefficient increases, the frequency of heterozygous genotypes decreases proportionally until, when $F = 1$, there are no

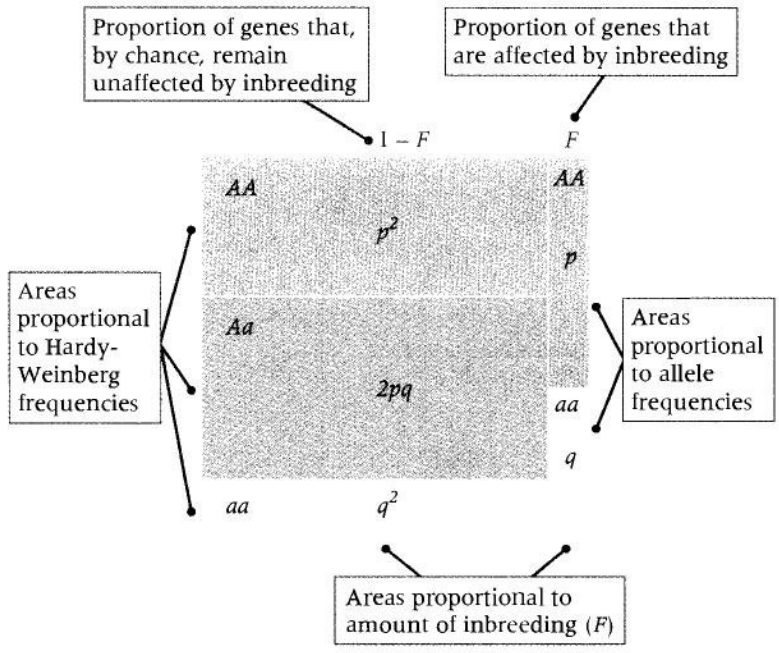


Figure 17.18 Effect of inbreeding on genotype frequencies. The large rectangles on the left pertain to alleles whose ancestries, by chance, are not affected by inbreeding and for which the genotype frequencies remain in Hardy–Weinberg proportions. The narrow rectangles on the right pertain to alleles whose ancestries are affected by the inbreeding, and in this case the genotype frequencies of *AA* and *aa* are related as $p : q$. (Note that there are no heterozygous genotypes in the latter case.)

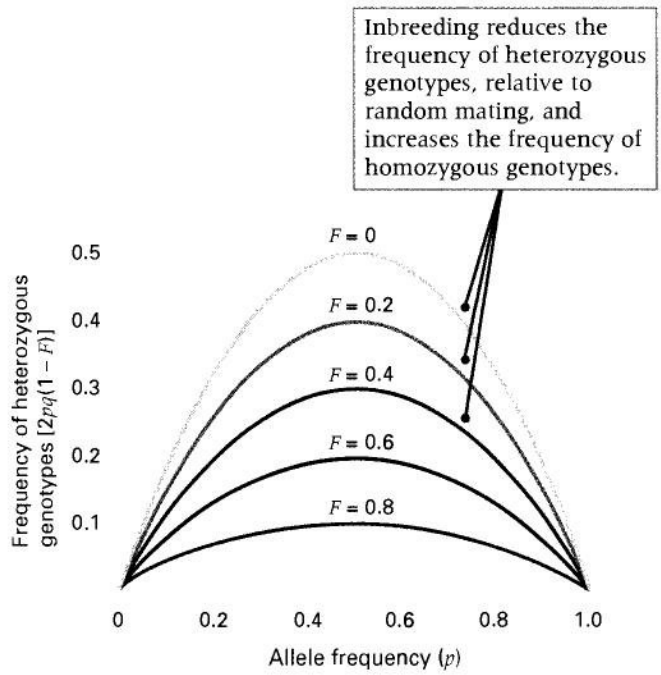


Figure 17.19 Frequency of heterozygous genotypes in an inbred population (y-axis) against allele frequency (x-axis). As the inbreeding becomes more intense (greater inbreeding coefficient F), the proportion of heterozygous genotypes decreases.

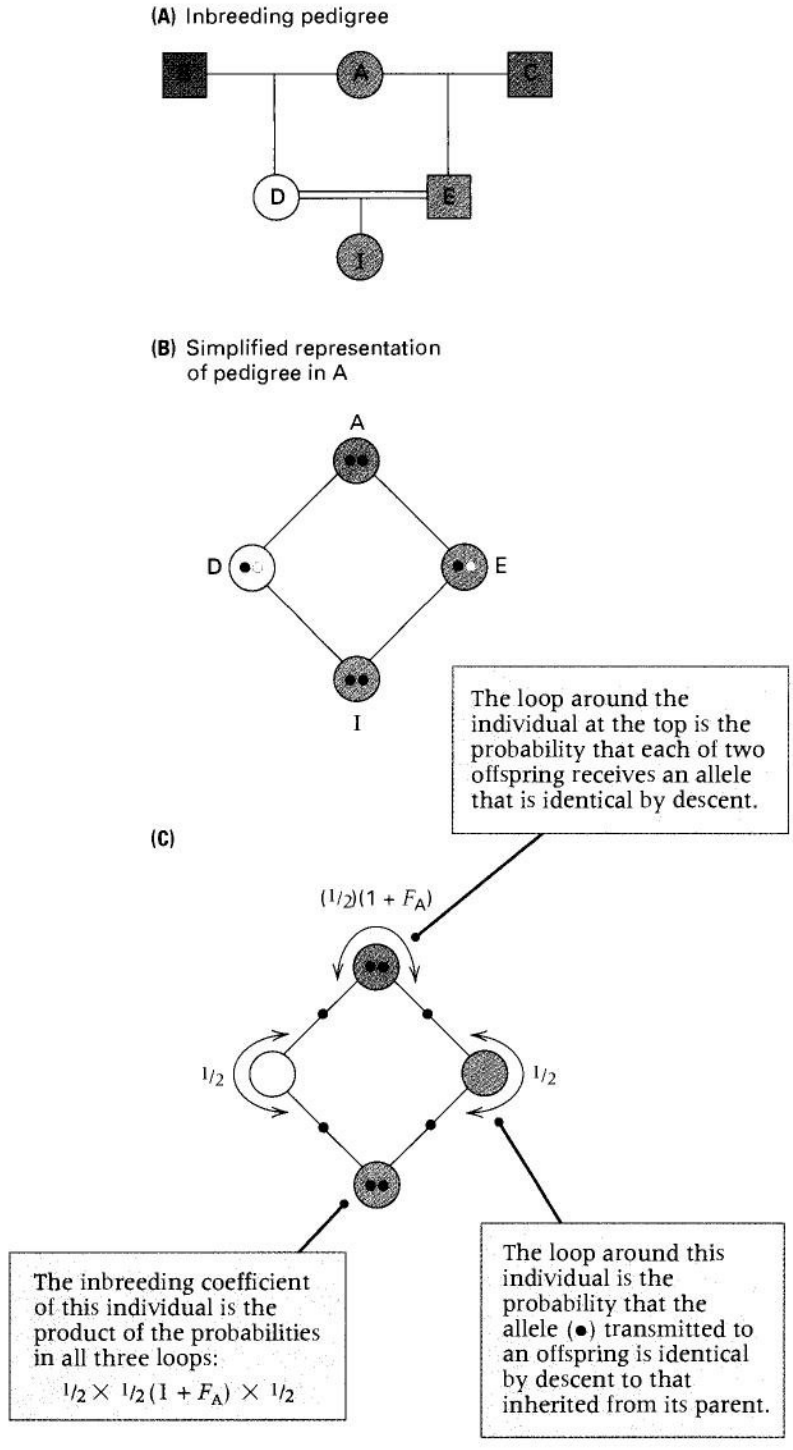


Figure 17.20 (A) An inbreeding pedigree in which individual I results from a mating between half siblings. (B) The same pedigree redrawn to show gametes in the closed loop leading to individual I. The colored dots represent alleles of a particular gene. (C) Calculation of the probability that the alleles indicated are identical by descent.

heterozygotes remaining in the inbred population. In other words, in a highly inbred population with $F = 1$, the genotypes consist either of *AA* or of *aa* in the relative proportions p and q, respectively.

A naturally self-fertilizing plant, *Phlox cuspidata*, provides a specific example of the effect of inbreeding. In this species, about 78 percent of the plants undergo self-fertilization, and it can be shown that the inbreeding coefficient corresponding to this level of self-fertilization is $F = 0.64$. We have seen earlier (Section 17.2) that the observed genotype counts—15 $Pgm^a\ Pgm^a$, 6 $Pgm^a\ Pgm^b$, and 14 $Pgm^b\ Pgm^b$—are not at all in agreement with the Hardy–Weinberg expectations for the allele frequencies of Pgm^a ($p = 0.51$) and Pgm^b ($q = 0.49$), because the numbers expected with random mating are 9.1, 17.5, and 8.4. Note that the largest discrepancy is that there are too few observed heterozygous genotypes.

Equation (4) tells us how to adjust the expected numbers to take the inbreeding into account. For the allele frequencies 0.51 and 0.49, and an inbreeding coefficient $F = 0.64$, the expected genotype frequencies are

$$Pgm^a\ Pgm^a: (0.51)^2(1 - 0.64) + (0.51)(0.64) = 0.420$$

$$Pgm^a\ Pgm^b: 2(0.51)(0.49) \times (1 - 0.64) = 0.180$$

$$Pgm^b\ Pgm^b: (0.49)^2(1 - 0.64) + (0.49)(0.64) = 0.400$$

In a sample size of 35, the expected numbers of the three genotypes are therefore $0.420 \times 35 = 14.7$, $0.180 \times 35 = 6.3$, and $0.400 \times 35 = 14$, compared with the observed numbers 15, 6, and 14. In this example, the agreement with the inbreeding model is excellent.

Allelic Identity by Descent

In human genetics, as well as in animal and plant breeding, a geneticist may wish to calculate the inbreeding coefficient of a particular individual whose pedigree is known in order to determine the probability of each possible genotype for the individual. This section introduces a concept that is particularly suited to pedigree calculations.

The diagnostic criterion of inbreeding is a closed loop in the pedigree. Such a closed

loop is often called a *path*. **Figure 17.20**A is an inbreeding pedigree in which individual I is the product of a mating between half siblings, and the path (closed loop) from I through the common ancestor A is evident. A streamlined version of this pedigree useful for inbreeding calculations is depicted in Figure 17.20B; individuals of both sexes are shown as circles, diagonal lines trace alleles transmitted from parent to offspring, and ancestors that did not contribute to the inbreeding of I (ancestors B and C) are left out altogether. The contrasting dots inside the circles represent particular alleles present in the corresponding individuals, and the depicted pattern of inheritance of the alleles represents the principal consequence of inbreeding and illustrates why closed loops are important. Specifically, individuals D and E both received the blue allele from A and received a white allele from B and C, respectively. The blue alleles in D and E have a very special relationship; they both originate by replication of the blue allele in A and therefore are said to show **identity by descent.** In the interval of the relatively few generations involved in most pedigrees, alleles that are identical by descent can be assumed to be identical in nucleotide sequence, because mutation can usually be ignored in so few generations. Individual I is shown as having received the blue allele from both D and E, and because the alleles in question are identical by descent, this individual must be homozygous for the allele inherited from the common ancestor. Without the closed loop in the pedigree, identity by descent of the alleles in I would be much less probable. The concept of identity by descent provides an alternative definition of the inbreeding coefficient, which can be shown to be completely equivalent to the definition in terms of heterozygosity given earlier. Specifically:

> The inbreeding coefficient of an individual is the probability that the individual carries alleles that are identical by descent.

Calculation of the Inbreeding Coefficient from Pedigrees

Figure 17.20C illustrates how the probability of identity by descent can be calculated in a pedigree, and the alleles are drawn as dots on the diagonal lines to emphasize their pattern of transmission. The double-headed arrows indicate which pairs of alleles must be identical by descent in order for I to have alleles that are identical by descent. The values of $1/2$ represent the probability that a diploid genotype will transmit the allele inherited from a particular parent; this value must be $1/2$ because of Mendelian segregation. The term $(1/2)(1 + F_A)$ can be explained by looking at the situation in **Figure 17.21**, which shows four possible pairs of gametes that A can transmit, each equally likely. In cases A and B, the transmitted alleles are certainly identical by descent because they are products of DNA replication of the same allele (blue or red). In cases C and D, the transmitted alleles have contrasting colors, but they can

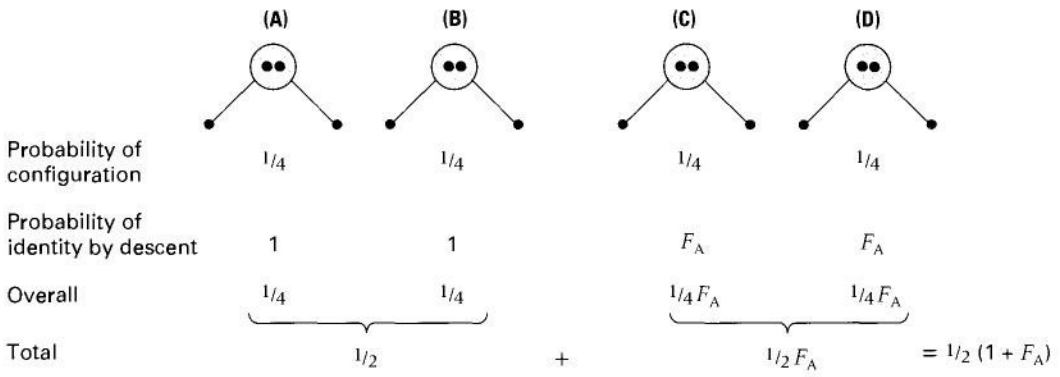


Figure 17.21 A diagram illustrating the reason why $(1/2)(1 + F_A)$ is the probability that two different gametes from the same individual, A, carry alleles that are identical by descent.

still be identical by descent if A is already inbred. However, the probability that A carries alleles that are identical by descent is, by definition, the inbreeding coefficient of A, and this is designated F_A. Altogether, the probability that A transmits two alleles that are identical by descent is the sum of the probabilities of the four eventualities in Figure 17.21, and this sum is equal to $(1/2)(1 + F_A)$. Matters are somewhat simpler when A is not itself inbred, because in this case $F_A = 0$, and the corresponding probability is simply 1/2.

Returning now to Figure 17.20C, the alleles in I can be identical by descent only if all three identities indicated by the loops are true, and this probability is the product of the three numbers shown. That is, the inbreeding coefficient of I, symbolized F_I, is

$$F_I = (1/2) \times (1/2)(1 + F_A) \times (1/2)$$
$$= (1/2)^3(1 + F_A)$$

For a more remote common ancestor, the formula becomes

$$F_I = (1/2)^n(1 + F_A) \qquad (5)$$

where n is the number of individuals in the path traced from one of the parents of I back through the common ancestor and forward again to the other parent of I.

Most pedigrees of interest are more complex than the one in Figure 17.20, but the inbreeding coefficient of any individual can still be calculated by using Equation (5). Application to complex pedigrees requires several additional rules:

1. If there are several distinct paths that can be traced through a common ancestor, then each path makes a contribution to the inbreeding coefficient calculated as in Equation (5).
2. If there are several common ancestors, then each is considered in turn in order to calculate its separate contribution to the inbreeding.
3. The overall inbreeding coefficient is then calculated as the sum of all these separate contributions of the common ancestors.

The extension of Equation (5) to complex pedigrees may be written as

$$F_I = \Sigma(1/2)^n(1 + F_A) \qquad (6)$$

where Σ denotes the summation of all paths through all common ancestors.

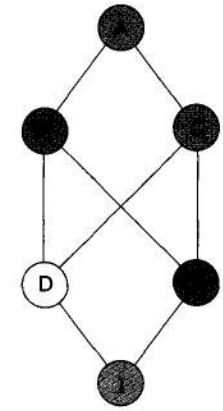


Figure 17.22 A complex pedigree in which the inbreeding of individual I results from several independent paths through three different common ancestors.

Application of Equation (6) can be illustrated in the pedigree in **Figure 17.22**. Each path and its contribution to the inbreeding coefficient of I is shown below (the red letter indicates the common ancestor).

EBD	$(1/2)^3(1 + F_B)$
ECD	$(1/2)^3(1 + F_C)$
ECABD	$(1/2)^5(1 + F_A)$
EBACD	$(1/2)^5(1 + F_A)$

The inbreeding coefficient of I is the sum of these contributions. If none of the common ancestors is inbred (that is, if $F_A = F_B = F_C = 0$), then

$$F_I = (1/2)^3 + (1/2)^3 + (1/2)^5 + (1/2)^5$$
$$= 5/16$$

The value 5/16 is to be interpreted to mean that the probability that individual I will be heterozygous at any particular locus is smaller, by the proportion 5/16, than the probability that a noninbred individual will be heterozygous at the same locus. Equivalently, the 5/16 could be interpreted to mean that individual I is heterozygous at 5/16 fewer loci in the whole genome than a noninbred individual.

Effects of Inbreeding

The effects of inbreeding differ according to the normal mating system of an organism. At one extreme, in regularly self-fertilizing plants, inbreeding is already so intense and the organisms are so highly homozygous

that additional inbreeding has virtually no effect. However:

> In most species, inbreeding is harmful, and much of the effect is due to rare recessive alleles that would not otherwise become homozygous.

Among human beings, inbreeding is usually uncommon because of social conventions and laws, although in small isolated populations (isolated villages, religious communities, aboriginal groups) it does occur, mainly through matings between relatives more distant than second cousins ("remote relatives"). The most common type of close inbreeding is between first cousins. The effect is always an increase in the frequency of genotypes that are homozygous for rare, usually harmful recessives. For example, among American whites, the frequency of albinism among offspring of matings between nonrelatives is approximately 1 in 20,000; among offspring of first-cousin matings, the frequency is approximately 1 in 2000. The reason for the 10-fold greater risk may be understood by comparing the genotype frequencies of homozygous recessives in Equation (4) for inbreeding and Equation (1) for random mating. Among the offspring of a mating between first cousins, $F = 1/16$, or 0.062. Therefore, the frequency of homozygous recessives produced by first-cousin mating will be

$$q^2(1 - 0.062) + q(0.062)$$

whereas among the offspring of nonrelatives, the frequency of homozygous recessives is q^2. For albinism, $q = 0.007$ (approximately), and the calculated frequencies are 5×10^{-4} for the offspring of first cousins and 5×10^{-5} for the offspring of nonrelatives. The increased frequency of rare homozygous recessive genotypes is the principal consequence of first-cousin mating (or any degree of inbreeding). The effect of first-cousin mating is illustrated in **Figure 17.23**, which shows a greater effect for rare alleles. For example, with an allele frequency $q = 0.05$, the relative risk of a homozygous offspring is $0.0335938/0.0025 = 13.4$, whereas with an allele frequency $q = 0.01$, the relative risk of a homozygous offspring is $0.00634375/0.0001 = 63.4$.

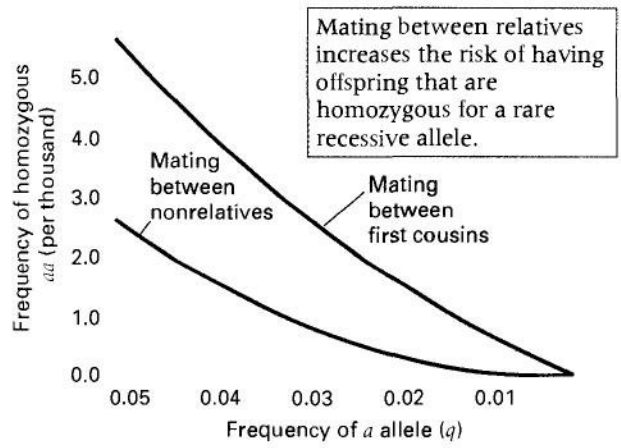


Figure 17.23 Effect of first-cousin mating on the frequency of offspring genotypes that are homozygous for a rare autosomal recessive allele. Although both curves decrease as the allele becomes rare, the curve for mating between nonrelatives decreases faster. As a consequence, the more rare the recessive allele, the greater the proportion of all affected individuals that result from first-cousin matings.

17.4 Genetics and Evolution

Evolution consists of progressive changes in the gene pool. Some authors prefer a stricter requirement that the changes must be associated with the progressive adaptation of a population to its environment. Whichever definition is adopted, evolution occurs because genetic variation exists in populations and because there is a natural selection favoring organisms that are best adapted to the environment. Genetic variation and natural selection are population phenomena, so they are conveniently discussed in terms of allele frequencies.

Four processes account for most of the changes in allele frequency in populations. They form the basis of cumulative change in the genetic characteristics of populations, leading to the descent with modification that characterizes the process of evolution. Although the point has yet to be proved, most evolutionary biologists believe that these same processes, when carried out continuously over geological time, can also account for the formation of new species and higher taxonomic categories. These process are

1. **Mutation,** the origin of new genetic capabilities in populations by means of spontaneous heritable changes in genes.

2. **Migration,** the movement of individuals among subpopulations within a larger population.
3. **Natural selection,** which results from the differing abilities of individuals to survive and reproduce in their environment.
4. **Random genetic drift,** the random, undirected changes in allele frequency that occur by chance in all populations, but particularly in small ones.

Some of the implications of these processes for population genetics are considered in the following sections.

17.5 Mutation and Migration

Mutation is an essential process in evolution because it is the ultimate source of genetic variation. But it is a relatively weak force for changing allele frequencies, primarily because typical mutation rates are so low. Moreover, most newly arising mutations are harmful to the organism and are eliminated from the population in a few generations. Although some mutations may be **selectively neutral,** which means they do not affect the ability of the organism to survive and reproduce, only a very few mutations are favorable for the organism and contribute to evolution. The low mutation rates that are observed are thought to have evolved as a compromise: The mutation rate is high enough to generate the favorable mutations that a species requires to continue evolving, but it is not so high that the species suffers excessive genetic damage from the preponderance of harmful mutations.

Migration is similar to mutation in that new alleles can be introduced into a local population, although the alleles derive from another subpopulation rather than from new mutations. In the absence of migration, the allele frequencies in each local population can change independently, so local populations can undergo considerable genetic differentiation. Genetic differentiation among subpopulations means that there are differing frequencies of common alleles among the local populations or that some local populations possess certain rare alleles not found in others. The accumulation of genetic differences among subpopulations can be minimized if the subpopulations exchange individuals (undergo migration). In fact, only a relatively small amount of migration among subpopulations (on the order of just a few immigrant individuals in each local population in each generation) is usually sufficient to prevent the accumulation of high levels of genetic differentiation. On the other hand, genetic differentiation can occur in spite of migration if other evolutionary forces, such as natural selection for adaptation to the local environment, are sufficiently strong.

Irreversible Mutation

To appreciate the magnitude of mutation as an evolutionary force, consider an allele A and a completely equivalent mutant allele a. (The mutational change in a could be, for example, a nucleotide substitution at an irrelevant site in an intron, or a change from one synonymous codon to another.) Suppose allele A undergoes mutation to a at a rate of μ per A allele per generation. What this rate means is that in any generation, the probability that a particular A allele mutates to a is μ. Equivalently, we could say that in any generation, the proportion of A alleles that do *not* mutate to a equals $1 - \mu$. If we let p and q represent the allele frequencies of A and a, respectively, then the allele frequency of A changes according to the rule that the allele frequency of A in any generation equals the allele frequency of A in the previous generation times the proportion of A alleles that failed to mutate in that generation. (Here we are assuming that reverse mutation of a to A can be neglected.) In symbols, the rule implies that

$$p_n = p_{n-1}(1 - \mu)$$

where the subscript n denotes any generation and $n - 1$ denotes the previous generation. Hence the allele frequency of A in generations 0, 1, 2, 3, and so forth changes according to

$$p_1 = p_0(1 - \mu)$$
$$p_2 = p_1(1 - \mu)$$
$$p_3 = p_2(1 - \mu)$$

Substituting p_1 from the first equation into the second, we obtain

$$p_2 = p_1(1 - \mu) = p_0(1 - \mu)^2$$

Substituting this into the equation for p_3 yields

$$p_3 = p_2(1 - \mu) = p_0(1 - \mu)^3$$

The general rule follows by successive substitutions:

$$p_n = p_0(1 - \mu)^n \qquad (7)$$

where p_n represents the allele frequency of A in the nth generation.

If we assume that allele A is fixed in the population at time $n = 0$, then $p_0 = 1$, and Equation (7) becomes

$$p_n = (1 - \mu)^n \qquad (8)$$

For realistic values of the mutation rate, Equation (8) implies a very slow rate of change in allele frequency. The pattern of change is illustrated by the lower curve in **Figure 17.24**, which assumes a mutation rate of $\mu = 1 \times 10^{-5}$ per generation. With this value of μ, the allele frequency of A decreases by half its present value every 69,314 generations. To put this rate in human terms, assuming 20 years per generation, it would require $69{,}314 \times 20 = 1{,}368{,}280$ years, or approximately 1.4 million years, to reduce the allele frequency of A by half. For example, if a human allele A were fixed at the time of *Homo erectus*, 1.4 million years ago, and A underwent mutation to a at a rate of 1×10^{-5} per generation, then 50 percent of the alleles present in the modern human population would still be of type A.

Reversible Mutation

When the possibility of reverse mutation is taken into account, the **forward mutation** of A to a is offset by **reverse mutation** of a to A. Eventually, an **equilibrium** of allele frequency is reached at which the allele frequencies become stable and no longer change, even though there is continuous mutation. The reason for the equilibrium is that at this point, each new A allele that is lost by forward mutation is replaced by a new A allele created by reverse mutation.

It is quite straightforward to deduce the equilibrium frequencies when mutation is reversible. As before, let A and a have allele frequencies p and q, respectively. Suppose that the rate of forward mutation ($A \rightarrow a$) is μ and that of reverse mutation ($a \rightarrow A$) is ν. In any generation, the number of A alleles lost to forward mutation is proportional to $p\mu$, and the number of new A alleles created by reverse mutation is proportional to $q\nu$. At equilibrium, the loss of A must equal the gain of a; hence $p\mu = q\nu$. Substituting $q = 1 - p$, we obtain

$$p\mu = (1 - p)\nu = \nu - p\nu$$

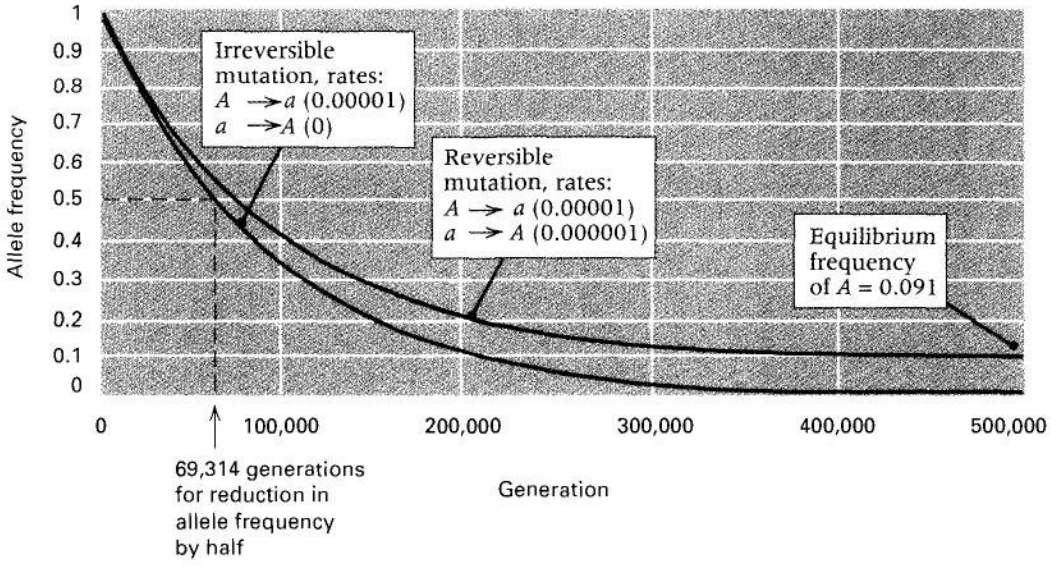


Figure 17.24 Effects of irreversible mutation (bottom curve) and reversible mutation (top curve) on allele frequency. In this example, the forward mutation rate is $\mu = 1 \times 10^{-5}$ and the reverse mutation rate is $\nu = 1 \times 10^{-6}$.

and therefore, at equilibrium,

$$\hat{p} = \frac{\nu}{\mu + \nu} \qquad (9)$$

where the symbol $\hat{p}$ is used, by convention, to mean the equilibrium value of p. As a check of the correctness of Equation (9), note that $\hat{q} = \mu/(\mu + \nu)$ and so $\hat{p}\mu = \hat{q}\nu$ as required.

In Figure 17.24, the upper curve gives the change in allele frequencies for reversible mutation with $\mu = 1 - 10^{-5}$ and $\nu = 1 \times 10^{-6}$. The equilibrium in this example is

$$\hat{p} = \frac{1 \times 10^{-6}}{1 \times 10^{-5} + 1 \times 10^{-6}} = \frac{1}{11}$$

The time required to reach equilibrium is again very long. When $\mu = 1 \times 10^{-5}$ and $\nu = 1 \times 10^{-6}$, it requires 63,013 generations for any specified values of p and q to go halfway to their equilibrium values. (The half-time of 69,314 generations derived earlier is different owing to the effect of reverse mutation.)

17.6 Natural Selection

The driving force of adaptive evolution is *natural selection*, which is a consequence of hereditary differences among organisms in their ability to survive and reproduce in the prevailing environment. Since it was first proposed by Charles Darwin in 1859, the concept of natural selection has been revised and extended, most notably by the incorporation of genetic concepts. In its modern formulation, the occurrence of natural selection rests on two premises, both of which are universally acknowledged to be true, and one deduction:

Premise 1: In all organisms, more offspring are produced than survive and reproduce. (Darwin borrowed this part of the theory from Thomas Malthus.)

Premise 2: Organisms differ in their ability to survive and reproduce, and some of these differences are due to genotype.

Deduction: In every generation, the genotypes that promote survival in the prevailing environment (favored genotypes) must be present in excess among individuals of reproductive age, and hence the favored genotypes will contribute disproportionately to the offspring of the next generation.

By means of this process, the alleles that enhance survival and reproduction increase in frequency from generation to generation, and the population becomes progressively better able to survive and reproduce in the prevailing environment. This progressive genetic "improvement" in populations constitutes the process of evolutionary **adaptation.**

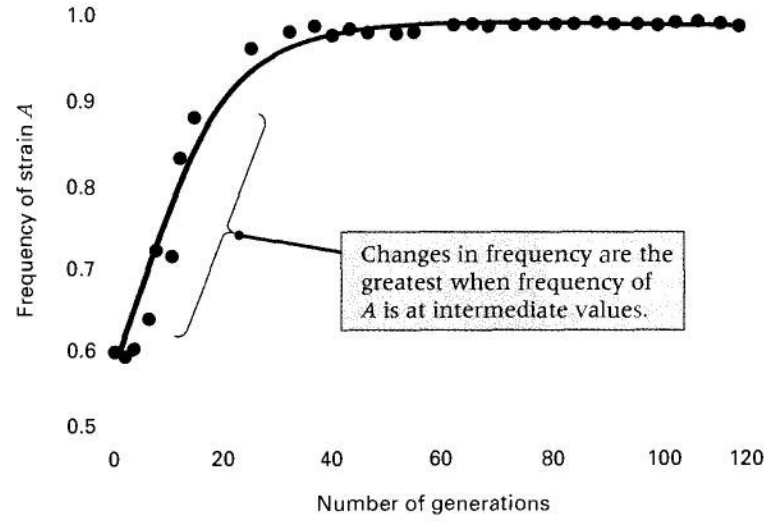


Figure 17.25 Increase in frequency of a favored strain of *E. coli* resulting from selection in a continuously growing population. The *y*-axis is the number of *A* cells at any time divided by the total number of cells ($A + B$). Note that the changes in frequency are the greatest when the frequency of the favored strain, *A*, is at intermediate values.

Selection in a Laboratory Experiment

Selection is easily studied experimentally in bacterial populations because of their short generation time (about 30 minutes). **Figure 17.25** shows the result of competition between two bacterial genotypes, *A* and *B*, under conditions in which *A* is the superior competitor. In the experiment, the competition was allowed to continue for 120 generations, in which time the proportion of *A* genotypes (p) increased from 0.60 to 0.995 and that of *B* genotypes decreased from 0.40 to 0.005. The data points give a satisfactory fit to an equation of the form

$$\frac{p_n}{q_n} = \left(\frac{p_0}{q_0}\right)\left(\frac{1}{w}\right)^n \qquad (10)$$

where p_0 and q_0 are the initial frequencies of *A* and *B* (in this case 0.6 and 0.4, respectively), p_n and q_n are the frequencies after n generations of competition (here, $n = 120$, $p_n = 0.995$, and $q_n = 0.005$), and w is a measure of the competitive ability of *B* when competing against *A* under the conditions of the experiment. Equation (10) can be derived step by step as follows: Because the relative rates of survival and reproduction of strains *A* and *B* are in the ratio $1 : w$, the ratio of frequencies in any generation, p_n/q_n, must equal the ratio of frequencies in the previous generation, p_{n-1}/q_{n-1}, times $1/w$. Therefore,

$$(p_1/q_1) = (p_0/q_0) \times (1/w)$$
$$(p_2/q_2) = (p_1/q_1) \times (1/w)$$
$$(p_3/q_3) = (p_2/q_2) \times (1/w)$$

and so on. Substituting successively from the first equation into the second, the second into the third, and so forth shows how Equation (10) is ultimately obtained.

The value of w is called the **relative fitness** of the *B* genotype relative to the *A* genotype under these particular conditions. The data in Figure 17.25 fit Equation (10) for a value of $w = 0.96$, which generates the smooth curve. Relative fitness measures the comparative contribution of each parental genotype to the pool of offspring genotypes produced in each generation. In this example, for each offspring cell produced by an *A* genotype, a *B* genotype produces an average of 0.96 offspring cell.

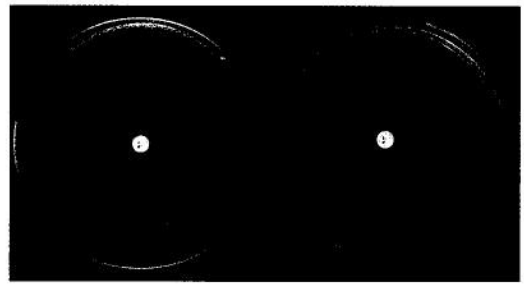
In population genetics, relative fitnesses are usually calculated by taking the most favored genotype (*A* in this case) as the standard with a fitness of 1.0. However, the selective disadvantage of a genotype is often of greater interest than its relative fitness, because some of the equations are simplified. The selective disadvantage of a disfavored genotype is called the **selection coefficient** against the genotype, and it is calculated as the difference between the fitness of the standard (taken as 1.0) and the relative fitness of the genotype in question. In the case at hand, the selection coefficient against *B*, denoted s, is

$$s = 1.000 - 0.96 = 0.04 \qquad (11)$$

In words, the selective disadvantage of strain *B* is 4 percent per generation. When the fitnesses are known, Equation (10) also enables us to predict the allele frequencies in any future generation, given the original frequencies. Alternatively, it can be used to calculate the number of generations required for selection to change the allele frequencies from any specified initial values to any later ones. For example, from the relative fitnesses of *A* and *B* just estimated, we can calculate the number of generations required to change the frequency of *A* from 0.1 to 0.8. In this example, $p_0/q_0 = 0.1/0.9$, $p_n/q_n = 0.8/0.2$, and $w = 0.96$. A little manipulation of Equation (10) gives

$$n = \frac{\log\left(\frac{0.1}{0.9}\right) - \log\left(\frac{0.8}{0.2}\right)}{\log(0.96)}$$

$$= 87.8 \text{ generations}$$



Widespread use of antibiotics has resulted in the natural selection of antibiotic resistance. These petri dishes show the growth of the bacterium *Staphylococcus aureus*. The small white disk contains a solution of penicillin, which diffuses out into the medium forming the halo around the disk. The bacterial strain on the left is sensitive. That on the right has a gene for a penicillinase protein and is resistant except for the highest doses nearest the disk. [John Durham/Photo Researchers, Inc.]

Selection in Diploid Organisms

Selection in diploids is analogous to that in haploids, but dominance and recessiveness create additional complications. **Figure 17.26** shows the change in allele frequencies for both a favored dominant and a favored recessive in a random-mating diploid population. The striking feature of the figure is that the frequency of the favored dominant allele changes very slowly when the allele is common, and the frequency of the favored recessive allele changes very slowly when the allele is rare. The reason is that with random mating, rare alleles are found much more frequently in heterozygotes than in homozygotes. With a favored dominant at high frequency, most of the recessive alleles are present in heterozygotes, and the heterozygotes are not exposed to selection and hence do not contribute to change in allele frequency. Conversely, with a favored recessive at low frequency, most of the favored alleles are in heterozygotes, and again the heterozygotes are not exposed to selection and do not contribute to change in allele frequency. The principle is quite general:

> Selection for or against a recessive allele is very inefficient when the recessive allele is rare.

The inefficiency of selection against rare recessive alleles has an important practical implication. There is a widely held belief that medical treatment to save the lives of persons with rare recessive disorders will cause a deterioration of the human gene pool because of the reproduction of persons who carry the harmful genes. This belief is unfounded. With rare alleles, the proportion of homozygous genotypes is so small that reproduction of the homozygous genotypes contributes a negligible amount to change in allele frequency. Considering their low frequency, it matters little whether homozygous genotypes reproduce or not. Similar reasoning applies to eugenic proposals to "cleanse" the human gene pool of harmful recessives by preventing the reproduction of affected persons. People with severe genetic disorders rarely reproduce anyway, and even when they do, they have essentially no effect on allele frequency. The underlying principle is that:

> The main reservoir of harmful recessive alleles in the human population is in the genomes of phenotypically normal carriers who are heterozygous.

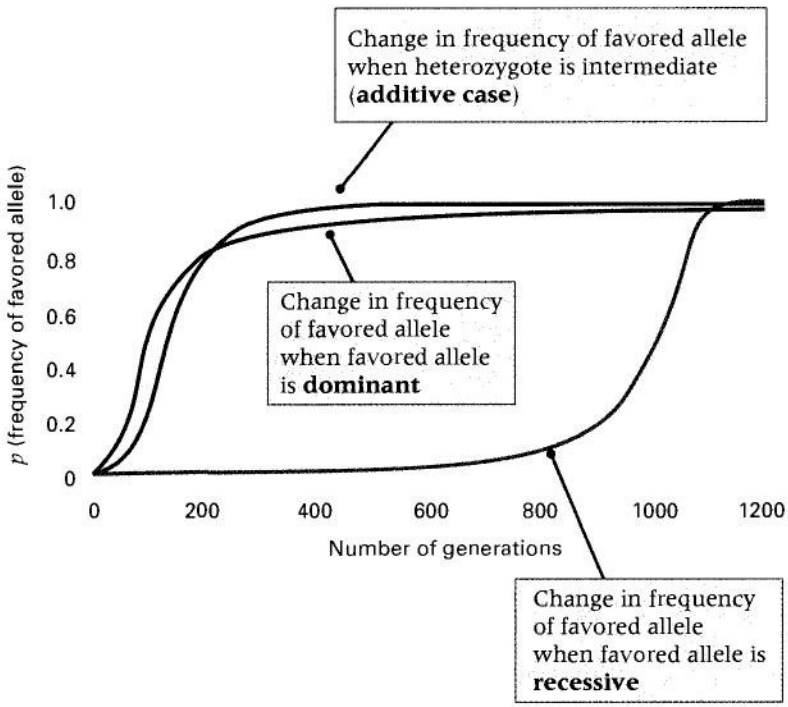


Figure 17.26 Theoretically expected change in frequency of an allele favored by selection in a diploid organism undergoing random mating when the favored allele is dominant, when the favored allele is recessive, and when the heterozygote is exactly intermediate in fitness (additive). In each case, the selection coefficient against the least fit homozygous genotype is 5 percent. The data are plotted in terms of *p*, the allele frequency of the beneficial allele. These curves demonstrate that selection for or against a recessive allele is very inefficient when the recessive allele is rare.

Components of Fitness

Data like those in Figure 17.25 are essentially unobtainable in higher organisms because of their long generation times. Observations extending over 120 generations would take 10 years with *Drosophila* and 2400 years with humans. How, then, is one to estimate fitness? The usual approach is to divide fitness into its component parts and estimate these separately. The two major components of the fitness of a genotype are usually considered to be

- The **viability,** which means the probability that a newly formed zygote of the specified genotype survives to reproductive age

- The **fertility,** which means the average number of offspring produced by an individual of the specified genotype during the reproductive period

Together, viability and fertility measure the contribution of each genotype to the offspring of the next generation.

Table 17.1 is an example of experiments carried out to estimate the viability of two genotypes coding for different electrophoretic forms of the enzyme alcohol dehydrogenase in *Drosophila*. The genotypes are homozygous for either the *F* or the *S* allele. The tabulated numbers are averages obtained in seven separate experiments in which larvae were placed together in competition in culture vials. The ratio of surviving larvae to those introduced originally is the probability of survival of each genotype. As noted at the bottom of the table, these numbers can be converted into relative viabilities in two ways, using either *SS* or *FF* as the standard. Because of the convention that the larger of the two relative values should be 1.0, the situation can be summarized either by saying that the viability of *FF* relative to *SS* is 0.94 or by saying that the selective disadvantage of *FF* relative to *SS* with respect to viability is 0.06.

Estimates of overall fitness based on viability and fertility components may be used for predicting changes in allele frequency that would be expected in experimental or natural populations. However, the fitness of a genotype may depend on the environment and on other factors, such as population density and the frequencies of other genotypes. Consequently, fitnesses estimated in a particular laboratory situation may have little relevance to predicting changes in allele frequency in natural populations unless other factors are comparable.

Selection–Mutation Balance

Natural selection usually acts to minimize the frequency of harmful alleles in a population. However, the harmful alleles can never be totally eliminated, because mutation of wildtype alleles continually creates new harmful mutations. With the forces of selection and mutation acting in opposite directions, the population eventually attains a state of equilibrium, or **selection–mutation balance,** at which new mutations exactly offset selective eliminations. In determining the allele frequencies at equilibrium, it makes a great deal of difference whether the harmful allele is completely recessive or has, instead, a small effect (partial dominance) on the fitness of the heterozygous carriers. To deduce the equilibrium frequencies, we will symbolize the wildtype allele as *A* and the harmful allele as *a* and represent the allele frequencies as *p* and *q*, respectively.

In the case of a complete recessive, the fitnesses of *AA*, *Aa*, and *aa* genotypes can be written as $1 : 1 : 1 - s$, where s represents the selection coefficient against the *aa* homozygote and measures the fraction of *aa* genotypes that fail to survive or reproduce. For a recessive lethal, for example, $s = 1$. When q is very small (as it will be near equilibrium), the number of *a* alleles eliminated by selection will be proportional to

Table 17.1 Relative viability of alcohol dehydrogenase genotypes in *Drosophila*

Experimental data and relative viability	Genotype *FF*	Genotype *SS*
Experimental data		
Number of larvae introduced per vial	200	100
Average number of survivors per vial	132.72	70.43
Probability of survival	0.6636	0.7043
Relative viability		
With *SS* as standard	0.6636/0.7043 = 0.94	0.7043/0.7043 = 1.0
With *FF* as standard	0.6636/0.6636 = 1.0	0.7043/0.6636 = 1.06

q^2s, assuming Hardy–Weinberg proportions. At the same time, the number of new a alleles introduced by mutation will be proportional to μ. At equilibrium, the selective eliminations must balance the new mutations, and so $q^2s = \mu$, or

$$\hat{q} = \sqrt{\frac{\mu}{s}} \qquad (12)$$

As before, $\hat{q}$ means the equilibrium value. To apply Equation (12) to a specific example, consider the recessive allele for Tay–Sachs disease, which in most non-Jewish populations has a frequency of $q = 0.001$. Because the condition is lethal, $s = 1$. Assuming that $q = 0.001$ represents the equilibrium frequency and that the allele has no effect on the fitness of the heterozygous carriers, the mutation rate required to account for the observed frequency would be

$$\mu = q^2s = (0.001)^2 \times 1.0 = 1 \times 10^{-6}$$

This example shows how considerations of selection–mutation balance can be used in estimating human mutation rates.

If a harmful allele shows partial dominance in having a small detrimental effect on the fitness of the heterozygous carriers, then the fitnesses of *AA*, *Aa*, and *aa* can be written as $1 : 1 - hs : 1 - s$, in which hs is the selection coefficient against the heterozygous carriers. The parameter h is called the **degree of dominance.** When $h = 1$, the harmful allele is completely dominant; when $h = 0$, it is completely recessive; and when $h = 1/2$, the fitness of the heterozygous genotype is exactly the average of those of the homozygous genotypes. For most harmful alleles that show partial dominance, the value of h is considerably smaller than $1/2$.

With random mating, when an allele is rare (as a harmful allele will be at equilibrium), the frequency of heterozygous genotypes greatly exceeds the frequency of those that are homozygous for the harmful recessive. This means that most of the alleles that are eliminated by selection are eliminated because of the selection against the heterozygous genotypes, because there are so many more of them. For each *Aa* genotype that fails to survive or reproduce, half of the alleles that are lost are *a*, so the number of alleles lost to selection in any generation is proportional to

$$2pq \times hs \times 1/2$$

On the other hand, the number of new *a* alleles created by $A \rightarrow a$ mutation equals μ. At equilibrium, the loss due to selection must balance the gain from mutation, so

$$2pq \times hs \times 1/2 = \mu,$$

or

$$\hat{p}\hat{q}hs = \mu$$

Because a harmful allele will be rare at equilibrium, especially if it shows partial dominance, $\hat{p}$ will be so close to 1 that it can be replaced with 1 in the above expression. Therefore, the equilibrium value for the frequency of a partially dominant harmful allele is

$$\hat{q} = \frac{\mu}{hs} \qquad (13)$$

Note that Equation (13) does not include a square root, which makes a substantial difference in the allele frequency. This effect can be seen by comparing the curves in **Figure 17.27** for a harmful allele that is lethal when homozygous ($s = 1$). The top curve is for a complete recessive ($h = 0$), and the other curves refer to various degrees of par-

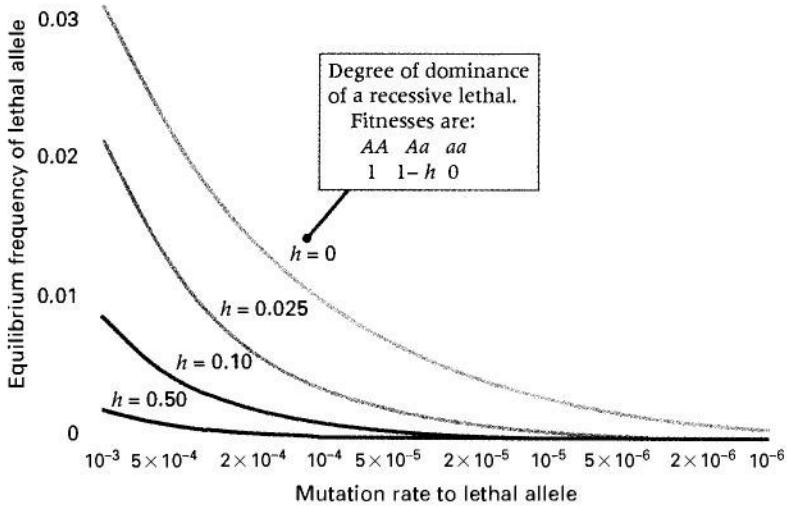


Figure 17.27 Effect of partial dominance on the equilibrium frequency of a harmful allele that is lethal when homozygous. The degree of dominance is measured by h. For example, a value of $h = 0.025$ means that the relative fitnesses of *AA* and *Aa* are 1 : 0.975.

tial dominance; for example, $h = 1/2$ means that the heterozygous genotype has a fitness exactly intermediate between the homozygous genotypes. Even a small amount of partial dominance has a substantial effect on reducing the equilibrium allele frequency.

Heterozygote Superiority

So far we have considered examples of selection in which the fitness of the heterozygote is intermediate between those of the homozygotes (or possibly equal to that of one homozygote). In these cases, the allele associated with the most fit homozygote eventually becomes fixed, unless the selection is opposed by mutation. In this section we consider the possibility that the fitness of the heterozygote is greater than that of both homozygotes; this case is called **overdominance** or **heterozygote superiority.**

When there is overdominance, neither allele can be eliminated by selection, because in each generation, the heterozygotes produce more offspring than the homozygotes. The selection in favor of heterozygous genotypes keeps both alleles in the population. Eventually there is an equilibrium in which the allele frequency no longer changes. With overdominance, the fitnesses of *AA*, *Aa*, and *aa* may be written as $1 - s : 1 : 1 - t$, where s and t are the selection coefficients against *AA* and *aa*, respectively, relative to *Aa*. Assuming random mating, in any generation, the proportion of *A* alleles eliminated by selection is $p^2s/p = ps$, and the proportion of *a* alleles eliminated by selection is $q^2t/q = qt$. At equilibrium, the selective eliminations of *A* must balance the selective eliminations of *a*, and hence $\hat{p}s = \hat{q}t$, or

$$\hat{p} = \frac{t}{s + t} \qquad (14)$$

With overdominance, the allele frequencies always go to equilibrium, but the rate of approach depends on the magnitudes of s and t. The equilibrium is attained most rapidly when there is strong selection against the homozygotes.

Overdominance does not appear to be a particularly common form of selection in natural populations. However, there are several well-established cases, the best known of which involves the sickle-cell hemoglobin mutation. The initial hint of a connection between sickle-cell anemia and malaria caused by the protozoan *Plasmodium falciparum* was the substantial overlap of the geographical distribution of sickle-cell anemia and that of malaria (**Figure 17.28**).

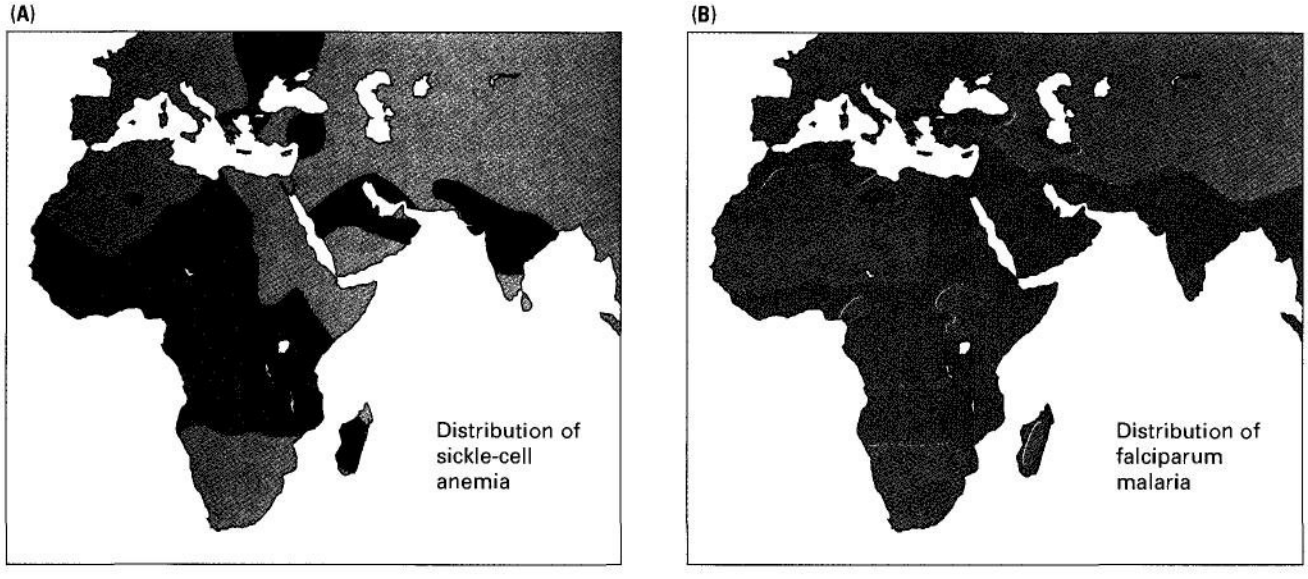


Figure 17.28 Geographic distribution of (A) sickle-cell anemia and (B) falciparum malaria in the 1920s, before extensive malaria-control programs were launched. Shades of brown indicate the areas in question.

CoNnection

Resistance in the Blood

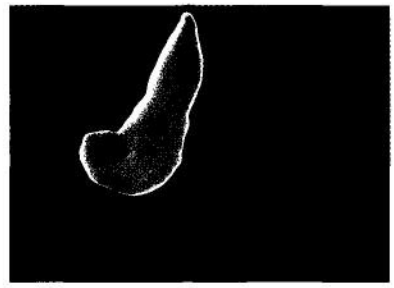
Anthony C. Allison 1954
Radcliffe Infirmary,
Oxford, England

Protection Afforded by Sickle-Cell Trait Against Subtertian Malarial Infection

Malaria is the most prevalent infectious disease in tropical and subtropical regions of the world, infecting up to 250 million people each year and causing as many as 2 million deaths. The disease is characterized by recurrent episodes of fever with alternating shivering and sweating. Patients suffer anemia due to destruction of red blood cells, as well as enlargement of the spleen, inflammation of the digestive system, bronchitis, and many other severe complications. The type of malaria caused by the protozoan parasite Plasmodium falciparum *is called "subtertian" malaria because there is less than a 3-day interval between bouts of fever. This parasite is spread through bites by the mosquito* Anopheles gambiae. *Upon transmission, the parasites multiply in the liver for about a week and then begin to infect and multiply in red blood cells (parasitemia), which are destroyed after a few days. In regions of Africa, the Middle East, the Mediterranean region, and India where falciparum malaria is endemic, there is also a relatively high frequency of the sickle-cell mutation affecting the amino acid sequence of the beta chain of hemoglobin. Heterozygous carriers have no severe clinical symptoms, but they have the so-called "sickle-cell trait," in which the red blood cells collapse into half-moon, or sickle, shapes after 1 to 3 days when sealed under a cover slip on a microscope slide. Homozygous affected persons have "sickle-cell anemia," in which many red blood cells sickle spontaneously while still in the bloodstream, causing severe complications and often death. Why would a genetic disease that is effectively lethal when homozygous be maintained at a frequency of 10 percent or more in a population? Allison noted the correlation between the sickle-cell mutation and malaria and speculated that the sickle-cell trait gives heterozygous carriers some protection against malaria. Key evidence supporting this hypothesis is presented here. Later work showed that the heterozygous carriers have an approximately 15 percent selective advantage as a result of their malaria resistance.*

It became imperative . . . to ascertain whether sickle cells can afford some degree of protection against malarial infection.

During the course of field work in Africa in 1949 I was led to question the view that the sickle-cell trait is neutral from the point of view of natural selection and to reconsider the possibility that it is as-



Unless well oxygenated, the mutant hemoglobin inside red blood cells tends to form needle-like crystals, deforming the cell from an elliptical shape into that of an elongated crescent or sickle. The sickled cells can clog capillary vessels throughout the body. [Dr. Gopal Murti/Science Photo Library/Photo Researchers, Inc.]

On the basis of the rates at which people of different genotypes are infected with malaria, and the mortality of infected people, the relative fitnesses of the genotypes *AA*, *AS*, and *SS* have been estimated as 0.85 : 1 : 0. (*S* represents the sickle-cell mutation.) In the absence of intensive medical treatment, the homozygous *SS* genotype is lethal as a result of severe anemia. On the other hand, there is overdominance because *AS* heterozygotes are more resistant to malarial infection than either of the homozygous genotypes.

From these fitnesses, we have $s = 0.15$ and $t = 1$, so the predicted equilibrium frequency of A from Equation (14) equals $1/(0.15 + 1) = 0.87$. The predicted equilibrium frequency of S is therefore $1 - 0.87 = 0.13$, which is a value reasonably close to that observed in many parts of West Africa.

17.7 Random Genetic Drift

The process of *random genetic drift* comes about because populations are not infinitely large (as we have been assuming all along) but rather finite, or limited, in size. The breeding individuals of any one generation produce a potentially infinite pool of ga-

sociated with a selective advantage. I noted that the incidence of the sickle-cell trait was higher in regions where malaria was prevalent than elsewhere. . . . It became imperative, then, to ascertain whether sickle cells can afford some degree of protection against malarial infection. . . . Children were chosen rather than adults as subjects for the observations so as to minimize the effect of acquired immunity to malaria. The recorded incidence of parasitemia in a group of 290 Ganda children [living near Kampala, Uganda] is presented in the accompanying Table. . . .

	With parasitemia	Without parasitemia	Total
Sicklers	12 (27.9%)	31 (72.1%)	43
Non-sicklers	113 (45.7%)	134 (54.3%)	247

It is apparent that the incidence of parasitemia is lower in the sickle-cell group than in the group without sickle cells. The difference is statistically significant ($\chi^2 = 4.8$ for 1 degree of freedom). . . . The parasite density in the two groups also differed: of 12 sicklers with malaria, 66.7% had only slight parasitemia while 33.3% had a moderate parasitemia. Of the 113 non-sicklers with malaria, 34% had slight parasitemia, the parasite density in the remainder being moderate or severe. . . . [Among a group of adult males who volunteered to be bitten by heavily infected *Anopheles gambiae* mosquitoes], an infection with *Plasmodium falciparum* was established in 14 out of 15 men without the sickle-cell trait, whereas in a comparable group of 15 men with the trait only 2 developed parasites. It is concluded that the abnormal erythrocytes of individuals with the sickle-cell trait are less easily parasitized by *P. falciparum* than are normal erythrocytes. Hence those who are heterozygous for the sickle-cell gene will have a selective advantage in regions where malaria is hyperendemic. This fact may explain why the sickle-cell gene remains common in these areas in spite of the elimination of genes in patients dying of sickle-cell anemia.

Source: British Medical Journal 1: 290–294

metes. In the absence of fertility differences, the allele frequencies among gametes will equal the allele frequencies among adults. However, because of the limited size of the population, only a few of these gametes will participate in fertilization and be represented among zygotes of the next generation. In other words, a process of *random sampling* takes place in going from one generation to the next. Because there is variation among samples as a result of chance, the allele frequencies among gametes and zygotes may differ as a result of chance. Changes in allele frequency that come about because of the generation-to-generation sampling in finite populations are the cause of random genetic drift.

To look at this process more generally, consider a population consisting of exactly N diploid individuals in each generation. At each autosomal locus, there will be exactly $2N$ alleles. Suppose that i of these are of type A and the remaining $2N - i$ are of type a. The allele frequency of A, designated p, therefore equals $i/2N$. The number of A alleles in the next generation cannot be specified with certainty, because it is governed by the chance sampling process of random genetic drift. On the other hand, it is possible to specify a *probability* for each possible number of A alleles in the next generation. This probability is given by the terms of the binomial distribution (Chapter 4). In particular, the probability $P(k|i)$ that there will be exactly k copies of A in the next generation, given that there are exactly i copies among the parents, is given by

$$P(k|i) = \frac{2N!}{k!\,(2N - k)!}\, p^k q^{2N-k} \qquad (15)$$

In Equation (15), $p = i/2N$, $q = 1 - p$, and k can equal 0, 1, 2, and so forth, up to and including $2N$. The part of the equation with factorials is the number of possible orders in which the k copies of allele A can be chosen. (The A alleles could be the first k chosen, the last k chosen, alternating with a, and so on.) The powers of p and q

represent the probability that any particular order will be realized. For any specified values of N and i, one can use Equation (15) along with a table of random numbers (or, better yet, a personal computer) to calculate an actual value for k.

Let's consider a concrete example to illustrate the essential features of random genetic drift. **Figure 17.29**A shows the result of random genetic drift in 12 subpopulations, each consisting of 8 diploid individuals, and initially containing equal numbers of A and a alleles. Within each subpopulation, mating is random. A computer program generating values of k with Equation (15) was used to calculate the number of A alleles in each subpopulation in each successive generation. These numbers are shown on the y axis, and the dispersion of allele frequencies resulting from random genetic drift is apparent. These changes in allele frequency would be less pronounced, and would require more generations, in larger populations than in the very small populations illustrated here, but the overall effect would be the same. Because of the prominent role of $2N$ in Equation (15), the dispersion of allele frequency resulting from random genetic drift depends on population size; the smaller the population, the greater the dispersion and the more rapidly it takes place.

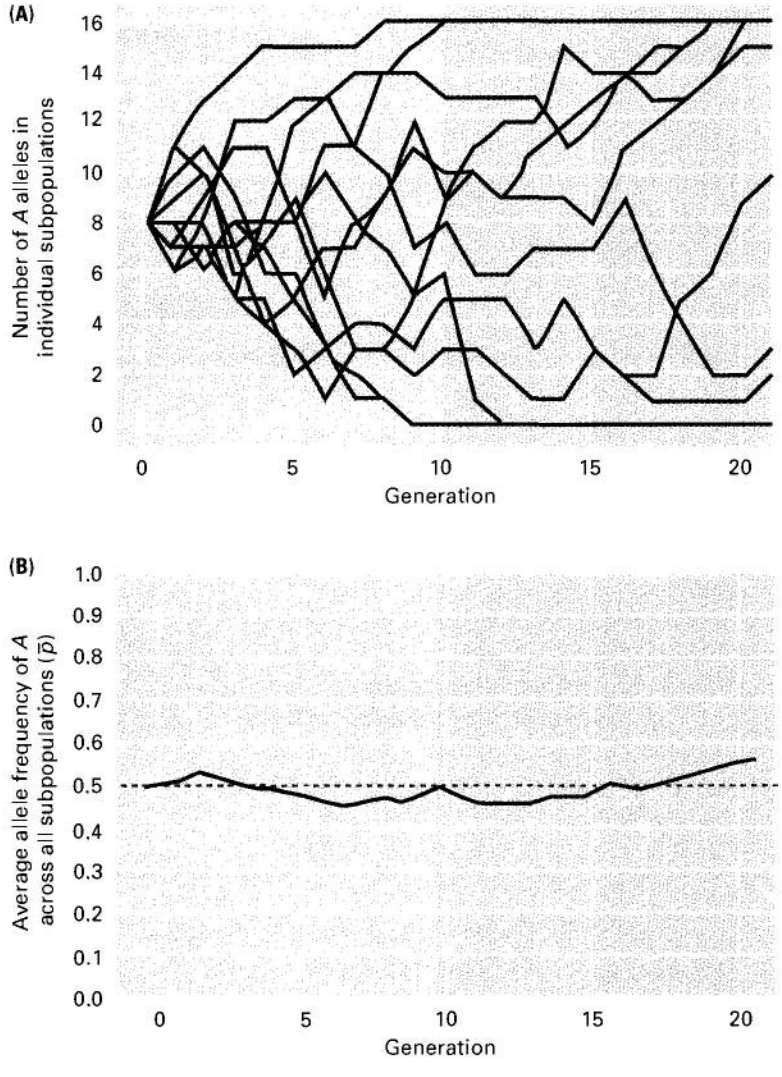


Figure 17.29 (A) Random genetic drift in 12 hypothetical subpopulations of 8 diploid oganisms. (B) Average allele frequency among the subpopulations in part A.

In Figure 17.29A, the principal effect of random genetic drift is evident even in the first few generations: The allele frequencies begin to spread out over a wider range. By the seventh generation, the spreading is extreme, and the number of A alleles ranges from 1 to 15. This spreading out means that the allele frequencies among the subpopulations become progressively more different. In general:

> Random genetic drift causes differences in allele frequency among subpopulations; it can be a major cause of genetic differentiation among subpopulations.

Although allele frequencies among subpopulations spread out over a wide range because of random genetic drift, the *average* allele frequency among subpopulations remains approximately constant. This point is illustrated in Figure 17.29B. The average allele frequency stays close to 0.5, its initial value. If an infinite number of subpopulations were being considered instead of only 12 subpopulatons, then the average allele frequency would be exactly 0.5 in every generation. This principle implies that in spite of the random drift of allele frequency in individual subpopulations, the average allele frequency among a large number of subpopulations remains constant and equal to the average allele frequency among the original subpopulations.

After a sufficient number of generations of random genetic drift, some of the subpopulations become fixed for A and others for a. Because we are excluding the occurrence of mutation, a population that be-

comes fixed for an allele remains fixed thereafter. After 21 generations in Figure 17.29A, only four of the populations are still segregating; eventually, these too will become fixed. Because the average allele frequency of A remains constant, it follows that a fraction p_0 of the populations will ultimately become fixed for A and a fraction $1 - p_0$ will become fixed for a. (The symbol p_0 represents the allele frequency of A in the initial generation.) Therefore:

> With random genetic drift, the probability of ultimate fixation of a particular allele is equal to the frequency of the allele in the original population.

In Figure 17.29A, five of the fixed populations become fixed for A and three for a, which is not very different from the equal numbers expected theoretically with an infinite number of subpopulations.

Figure 17.30A shows an actual example of random genetic drift in small experimental populations of *Drosophila*. The figure is based on 107 subpopulations, each initiated with 8 bw^{75}/bw females (bw = brown eyes) and 8 bw^{75}/bw males and maintained at a constant size of 16 by randomly choosing 8 males and 8 females from among the progeny of each generation. Note how the allele frequencies among subpopulations spread out because of random genetic drift and how subpopulations soon begin to be fixed for either bw^{75} or bw. Although the data are somewhat rough because there are only 107 subpopulations, the overall pattern of genetic differentiation has a reasonable resemblance to that expected from the theory based on the binomial distribution, which is shown in Figure 17.30B.

If random genetic drift were the only force at work, then all alleles would eventually become either fixed or lost and there would be no polymorphism. On the other hand, many factors can act to retard or prevent the effects of random genetic drift, of which the following are the most important: (1) large population size; (2) mutation and migration, which impede fixation because alleles lost by random genetic drift can be reintroduced by either process; and (3) natural selection, particularly those modes of selection that tend to maintain genetic diversity, such as heterozygote superiority.

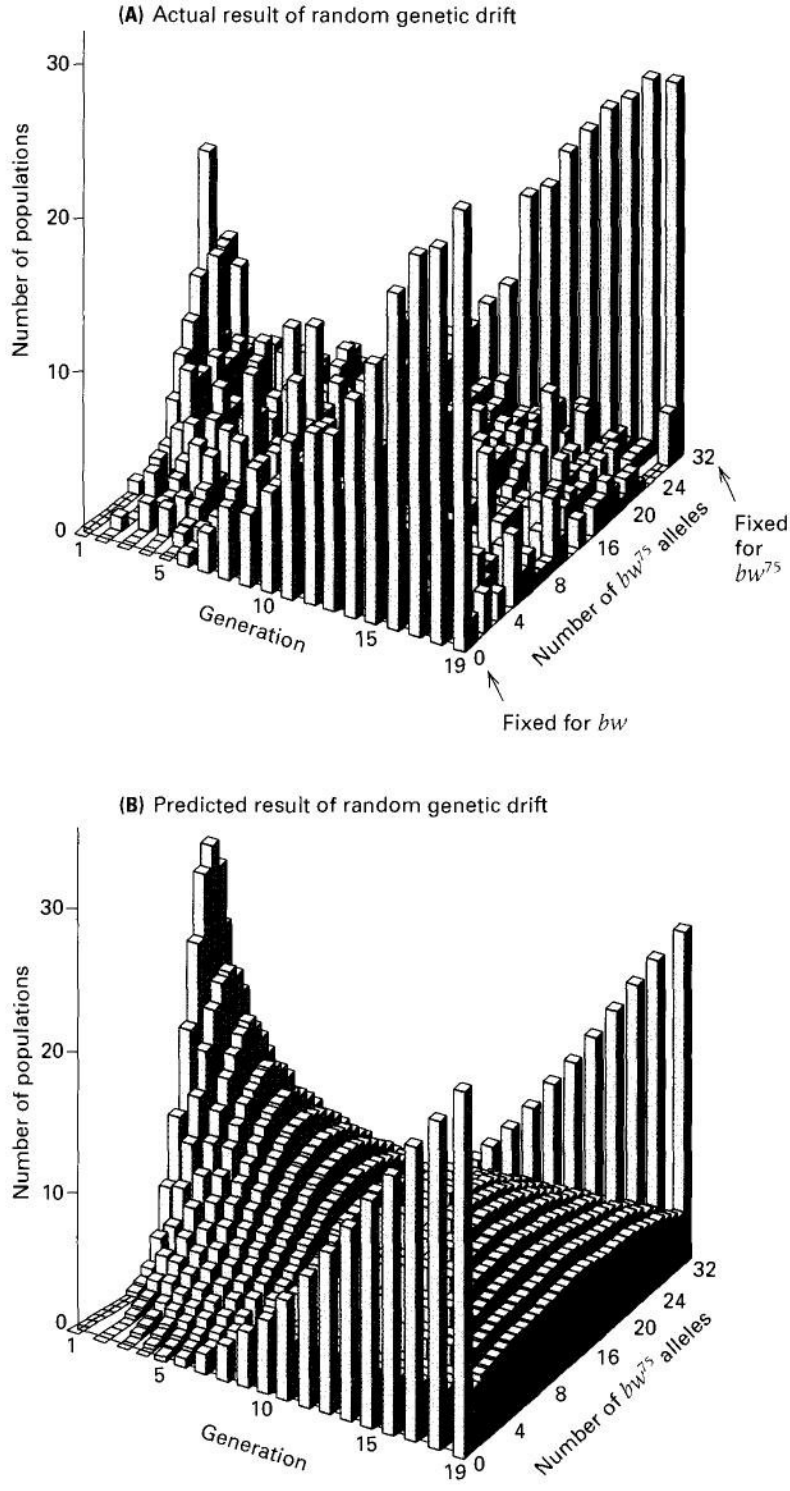


Figure 17.30 (A) Random genetic drift in 107 experimental populations of *Drosophila melanogaster*, each consisting of 8 females and 8 males. (B) Theoretical expectation of the same situation, calculated from the binomial distribution. [Data in part A from P. Buri, 1956, *Evolution* 10:367. Graphs from D. L. Hartl and A. G. Clark, 1997, *Principles of Population Genetics*, 3d ed. (Sunderland, MA: Sinauer Associates).]

Chapter Summary

Population genetics is the application of the principles of genetics to populations of organisms. A subpopulation, or local population, is a group of organisms of the same species living within a geographical region of such size that most matings are between members of the group. In most natural populations, many genes are polymorphic in that they have two or more common alleles. One of the goals of population genetics is to determine the nature and causes of genetic variation in natural populations.

The relationship between the relative proportions of particular alleles (allele frequencies) and genotypes (genotype frequencies) is determined in part by the frequencies with which particular genotypes form mating pairs. In random mating, mating pairs are independent of genotype. When a population undergoes random mating for an autosomal gene with two alleles, the frequencies of the genotypes are given by the Hardy–Weinberg principle. If the alleles of the gene are A and a, and their allele frequencies are p and q, respectively, then the Hardy–Weinberg principle states that the genotype frequencies with random mating are AA, p^2; Aa, $2pq$; and aa, q^2. These are often good approximations for genotype frequencies within subpopulations. An important implication of the Hardy–Weinberg principle is that rare alleles are found much more frequently in heterozygotes than in homozygotes ($2pq$ versus q^2).

Inbreeding means mating between relatives, and the extent of inbreeding is measured by the inbreeding coefficient, F. The main consequence of inbreeding is that an allele present in a common ancestor may be transmitted to both parents of an inbred individual in a later generation and become homozygous in the inbred offspring. The replicas of an ancestral allele are said to be identical by descent. The inbreeding coefficient of an inbred organism can be deduced directly from the pedigree of inbreeding by calculating the probability of identity by descent along every possible path of descent through each common ancestor. Among inbred individuals, the frequency of heterozygous genotypes is smaller, and that of homozygous genotypes greater, than it would be with random mating.

Evolution is cumulative genetic change in a population through time. Adaptation is a progressive increase in the degree to which a species becomes genetically well suited to its environment. A principal mechanism of adaptation is natural selection, in which individuals superior in survival or reproductive ability in the prevailing environment contribute a disproportionate share of genes to future generations, thereby gradually increasing the frequency of the favorable alleles in the whole population. At least three other processes also can change allele frequency: mutation (heritable change in a gene), migration (movement of individuals among subpopulations), and random genetic drift (which results from chance sampling due to restricted population size). Spontaneous mutation rates are generally so low that the effect of mutation on changing allele frequency is minor, except for rare alleles. Migration can have significant effects on allele frequency because migration rates may be very large. The main effect of migration is the tendency to equalize allele frequencies among the local populations that exchange migrants. Selection occurs through differences in viability (the probability of survival of a genotype) and in fertility (the probability of successful reproduction).

Populations maintain harmful alleles at low frequencies as a result of a balance between selection, which tends to eliminate the alleles, and mutation, which tends to increase their frequencies. When there is selection–mutation balance, the allele frequency at equilibrium is usually greater if the allele is completely recessive than if it is partially dominant. This difference arises because selection is quite ineffective at influencing the frequency of a completely recessive allele when the allele is rare, because of the almost exclusive appearance of the allele in heterozygotes.

A few examples are known in which the heterozygous genotype has a greater fitness than either of the homozygous genotypes. Heterozygote superiority results in an equilibrium in which both alleles are maintained in the population. An example is sickle-cell anemia in regions of the world where falciparum malaria is endemic. Heterozygous persons have an increased resistance to malaria and only a mild anemia, a circumstance that results in greater fitness than that of either homozygote.

Random genetic drift is a statistical process of change in allele frequency in small populations, resulting from the inability of every individual to contribute equally to the offspring of successive generations. In a subdivided population, random genetic drift results in differences in allele frequency among the subpopulations. In an isolated population, barring mutation, an allele will ultimately become either fixed or lost because of random genetic drift. The probability of ultimate fixation of an allele as a result of random genetic drift is equal to the frequency of the allele in the initial population.

Key Terms

adaptation
allele frequency
assortative mating (positive and negative)
consanguineous mating
degree of dominance
DNA typing
equilibrium
evolution
fertility
fixed allele
forward mutation

gene pool
genotype frequency
Hardy–Weinberg principle
heterozygote superiority
identity by descent
inbreeding
inbreeding coefficient
lost allele
mating system
migration
mutation
natural selection
overdominance
population genetics
random genetic drift
random mating
relative fitness
reverse mutation
selection coefficient
selection–mutation balance
selectively neutral mutation
subpopulation
viability

Review the Basics

- With regard to allele frequency, what does it mean to say that an allele is fixed? that an allele is lost?
- Distinguish between inbreeding and assortative mating.
- What does the Hardy–Weinberg principle imply about the relative frequencies of heterozygous carriers and homozygous affected organisms for a rare, harmful recessive allele?
- Traits due to recessive alleles in the X chromosome are usually much more prevalent in males than in females. Explain why this discrepancy is expected with random mating.
- Why are the effects of inbreeding more easily observed with a rare recessive allele than with a rare dominant allele?
- Name four evolutionary processes that can change allele frequencies in natural populations. How do allele frequencies change in the absence of these processes?
- What is the role of mutation in the evolutionary process? Why is mutation a weak force for changing allele frequencies?
- What do population geneticists mean by *fitness*? What is the fitness of an organism that dies before the age of reproduction? What is the fitness of an organism that is sterile?
- Why, in a random-mating population, does the allele frequency of a rare recessive allele change slowly under selection?
- Many recessive alleles are extremely harmful when homozogous, so there is selection in every generation that tends to reduce the allele frequency. Yet harmful recessive alleles are maintained at a low frequency for almost every gene. What process prevents harmful recessive alleles from being completely eliminated?
- Heterozygote superiority of the type observed with sickle-cell anemia is sometimes called balancing selection. Do you think this is an appropriate term? Explain why or why not.
- What is random genetic drift and why does it occur? Explain why this process implies that, in the absence of other forces, the ancestry of all alleles present at a locus in a population can eventually be traced back to a single allele present in some ancestral population.

Guide to Problem Solving

Problem 1 A sample of 300 plants from a population is examined for the electrophoretic mobility of an enzyme that varies according to the genotype determined by two alleles, *F* and *S*, of a single gene. The results are 7 plants with genotype *FF*, 106 with genotype *FS*, and 187 with genotype *SS*. What are the allele frequencies of *F* and *S*? What are the expected numbers of the three genotypes, assuming random mating?

Answer The allele frequencies are determined by counting the alleles. The 7 *FF* plants represent 14 *F* alleles, the 106 *FS* plants represent 106 *F* and 106 *S* alleles, and the 187 *SS* plants represent 374 *S* alleles, for a total of $300 \times 2 = 600$ alleles altogether. The allele frequency p of *F* is $(14 + 106)/600 = 0.2$, and the allele frequency q of *S* is $(106 + 374)/600 = 0.8$. As a check on the calculations, note that the allele frequencies sum to unity, as they should. For the second part of the question, the expected genotype frequencies with random mating are p^2 *FF*, $2pq$ *FS*, and q^2 *SS*, so the expected numbers are

$$\begin{aligned} FF:&\quad (0.2)^2 \times 300 = 12 \\ FS:&\quad 2(0.2)(0.8) \times 300 = 96 \\ SS:&\quad (0.8)^2 \times 300 = 192 \end{aligned}$$

As a check on these calculations, note that $12 + 96 + 192 = 300$.

Problem 2 Xeroderma pigmentosum (XP) is an often-fatal skin cancer resulting from a recessive mutant allele that affects DNA repair. In the United States, the frequency of homozygous-recessive affected people is approximately 1 in 250,000. (The mutant allele can be in any one of about eight different genes, but for the purposes of this problem it makes no difference.)

(a) What is the expected frequency of XP among the offspring of first-cousin matings?

(b) What is the ratio of XP among the offspring of first-cousin matings to that among the offspring of nonrelatives?

Answer First we must calculate the frequency of the recessive allele, q, using the information that the frequency of homozygous recessives is 1 in 250,000. Assuming random mating frequencies, $q = \sqrt{(1/250{,}000)} = 0.002$. **(a)** Among the offspring of first-cousin matings, the expected frequency of XP equals $q^2(1 - F) + qF$, where $F = 1/16$ for the offspring of first cousins. In this case the formula yields $(0.002)^2(15/16) + (0.002)(1/16) = 0.00012875$, or about 1 in 7767. **(b)** The ratio is given in the following formula, where $q = 0.002$ and $F = 1/16$.

$$\text{Ratio} = \frac{q^2(1 - F) + qF}{q^2} = \frac{(1/7{,}767)}{} = 32.2$$

In other words, the offspring of first cousins have more than a 32-fold greater risk of being homozygous for a recessive mutant allele causing XP.

Problem 3 Warfarin is a rat killer that acts by hindering blood coagulation. Continued use of the poison has resulted in the evolution of resistance in most rat populations because of selection favoring a resistance mutation R. The normal, sensitive allele may be denoted S. In the absence of warfarin, the relative fitnesses of SS, RS, and RR are in the ratio 1.00 : 0.77 : 0.46, and in the presence of warfarin, the ratio of fitnesses is 0.68 : 1.00 : 0.37.

(a) With regard to persistence or loss of the R and S alleles, what is the expected result of the continued use of the chemical?

(b) What would happen if warfarin were no longer used?

(c) Under what circumstances might the R allele become fixed?

Answer

(a) In the presence of warfarin, there is heterozygote superiority, so continued use would be expected to result in a stable equilibrium in which both R and S persisted in the population.

(b) In the absence of warfarin, the genotype SS is favored and RS is intermediate in fitness, so allele S would become fixed and R would become lost.

(c) The allele R could become fixed in a small population where, because of chance fluctuations in allele frequency, the S allele became lost. (However, small populations would also increase the chance that the R allele might be lost.)

Analysis and Applications

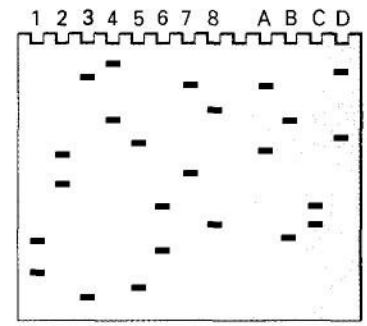
17.1 If the genotype AA is an embryonic lethal and the genotype aa is fully viable but sterile, what genotype frequencies would be found in adults in an equilibrium population containing the A and a alleles? Is it necessary to assume random mating?

17.2 Allozyme phenotypes of alcohol dehydrogenase in the flowering plant *Phlox drummondii* are determined by codominant alleles of a single gene. In one sample of 35 plants, the following data were obtained:

Genotype	AA	AB	BB	BC	CC	AC
Number	2	5	12	10	5	1

What are the frequencies of the alleles A, B, and C in this sample?

17.3 The accompanying illustration shows the gel patterns observed with a probe for a VNTR (variable number of tandem repeats) locus among four pairs of parents (1–8) and four children (A–D). One child comes from each pair of parents.

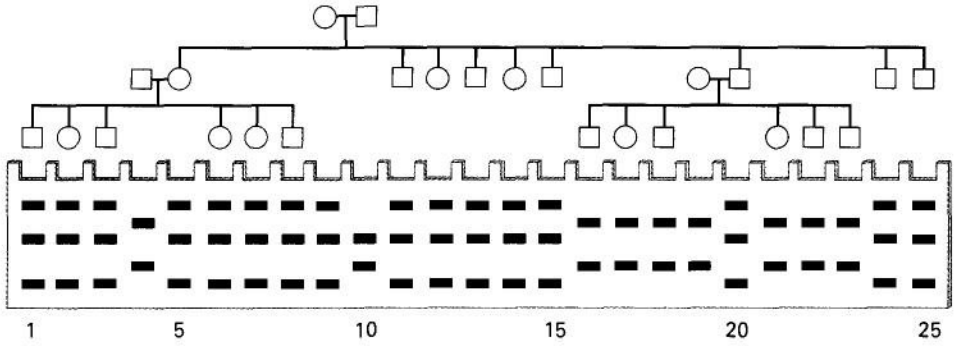


(a) Why does each person have two bands?

(b) How would it be possible for a person to have only one band?

(c) What type of dominance is illustrated by the VNTR alleles?

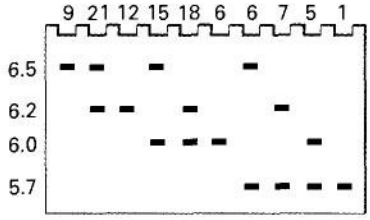
(d) Which pairs of people who have the DNA in lanes 1–8 are the parents of each child?



Problem 17.4

17.4 The illustration above shows a human pedigree, along with the pattern of restriction fragments observed in a DNA sample from each of the individuals when hybridized with a labeled probe. Examine the pedigree and the band patterns, and suggest a mode of inheritance. Is the Hardy–Weinberg principle relevant to this situation? Explain why or why not.

17.5 DNA from 100 unrelated people was digested with the restriction enzyme *Hin*dIII, and the resulting fragments were separated and probed with a sequence for a particular gene. Four fragment lengths that hybridized with the probe were observed—namely 5.7, 6.0, 6.2, and 6.5 kb—where each fragment defines a different restriction-fragment allele. The accompanying illustration shows the gel patterns observed; the number of individuals with each gel pattern is shown across the top. Estimate the allele frequencies of the four restriction-fragment alleles.

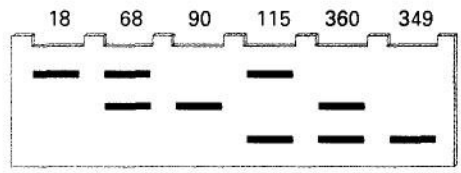


17.6 If the frequencies of genotypes *AA*, *Aa*, and *aa* are denoted *P*, *Q*, and *R*, respectively, show that the Hardy–Weinberg principle implies that $Q^2/4 = PR$.

17.7 Which of the following genotype frequencies of *AA*, *Aa*, and *aa*, respectively, satisfy the Hardy–Weinberg principle?

(a) 0.25, 0.50, 0.25
(b) 0.36, 0.55, 0.09
(c) 0.49, 0.42, 0.09
(d) 0.64, 0.27, 0.09
(e) 0.29, 0.42, 0.29

17.8 A population geneticist obseved the RFLP phenotypes shown in the diagram in the next column, where the number above each well is the number of individuals exhibiting the corresponding phenotype. Test whether the observed phenotype frequencies are consistent with those expected from a single locus with three alleles in a random-mating population, using the chi-square criterion. Explain why, in this test, the appropriate number of degrees of freedom is 3.



17.9 Hartnup disease is an autosomal-recessive disorder of intestinal and renal transport of amino acids. The frequency of affected newborn infants is about 1 in 14,000. Assuming random mating, what is the frequency of heterozygotes?

17.10 A randomly mating population of dairy cattle contains an autosomal-recessive allele causing dwarfism. If the frequency of dwarf calves is 10 percent, what is the frequency of heterozygous carriers of the allele in the entire herd? What is the frequency of heterozygotes among nondwarf cattle?

17.11 In certain grasses, the ability to grow in soil contaminated with the toxic metal nickel is determined by a dominant allele.

(a) If 60 percent of the seeds in a randomly mating population are able to germinate in contaminated soil, what is the frequency of the resistance allele?
(b) Among plants that germinate, what proportion are homozygous?

17.12 In a Pygmy group in Central Africa, the frequencies of alleles determining the ABO blood groups were estimated as 0.74 for I^O, 0.16 for I^A, and 0.10 for I^B. Assuming random mating, what are the expected frequencies of ABO genotypes and phenotypes?

GeNETics on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes,* choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- A colorful introduction to the use of VNTRs in **DNA typing** is located at this keyword site. Another method that makes use of **microsatellite markers** can be found at this other keyword site. Although both of these approaches rely on Southern blotting and DNA hybridization to visualize the DNA fragments used to discriminate among genotypes, alternative methods that employ DNA amplification require much less DNA to start with—as little as that contained in cells and saliva on a cigarette butt, under the fingernails of a victim who scratched the assailant, or in a single hair root.

- This **Hardy–Weinberg** site has a clever demonstration of the principle in a game that makes use of several decks of playing cards. For a one-locus, two-allele situation, the red suits (hearts and diamonds) are regarded as one allele and the black suits (spades and clubs) as the alternative allele. Each person in the game gets two cards, defining his or her genotype, and then a simulated round of random mating takes place. Best of all, no algebra is required!

- At the keyword site **CFTR** you can learn about geographical variation in common cystic fibrosis mutations in the human population. The high frequency of the disease has been difficult to explain in view of the seriousness of the condition in affected individuals. Supposing that mutant alleles are completely recessive, and that the frequency of homozygous recessives (1/1700) represents an equilibrium between mutation and selection, then the rate of mutation to new alleles would have to be 6×10^{-4} per generation, which is much higher than can easily be explained. Another alternative is that heterozygotes have a selective advantage of some kind, so the mechanism of selection is actually overdominance. Recent data support the hypothesis that heterozygotes are more resistant to typhoid fever. The bacterium *Salmonella typhi*, the pathogen for typhoid fever, invades the gastrointestinal system through the CFTR membrane channel encoded in the *CFTR* gene. Cells expressing wildtype CFTR internalize more *S. typhi* than mutant cells, and transgenic mice that are heterozygous for *CFTR* mutations internalize 86 percent fewer bacteria than do wildtype homozygotes. The resistance hypothesis is also consistent with the high frequency of particular *CFTR* mutations that are found in patients.

- The occasionally extreme effects of random genetic drift are dramatically illustrated by **Pingelap disease**, a form of total color blindness combined with extreme nearsightedness and cataract, which affects 4–10 percent of the Pingelapese people of the eastern Caroline Islands in the Pacific; victims are blind from infancy. Consult this keyword site to learn the molecular and cellular basis of this condition and why its frequency is so high in this particular human population.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 17 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 17.



[Courtesy TIGR, www.tigr.org.]

17.13 Among 35 individuals of the flowering plant *Phlox roemariana,* the following genotypes were observed for a gene determining electrophoretic forms of the enzyme phosphoglucose isomerase: 2 *AA,* 13 *AB,* 20 *BB.*

(a) What are the frequencies of the alleles *A* and *B*?
(b) Assuming random mating, what are the expected numbers of the genotypes?

17.14 If an X-linked recessive trait is present in 2 percent of the males in a population with random mating, what is the frequency of the trait in females? What is the frequency of carrier females?

17.15 In a population of *Drosophila*, an X-linked recessive allele causing yellow body color is present in genotypes at frequencies typical of random mating; the frequency of the recessive allele is 0.20. Among 1000 females and 1000 males, what are the expected numbers of the yellow and wildtype phenotypes in each sex?

17.16 How does the frequency of heterozygotes in an inbred population compare with that in a randomly mating population with the same allele frequencies?

17.17 Which of the following genotype frequencies of *AA*, *Aa*, and *aa*, respectively, are suggestive of inbreeding?

(a) 0.25, 0.50, 0.25
(b) 0.36, 0.55, 0.09
(c) 0.49, 0.42, 0.09
(d) 0.64, 0.27, 0.09
(e) 0.29, 0.42, 0.29

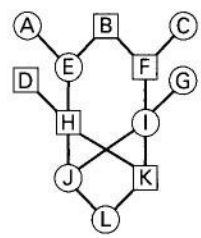
17.18 Galactosemia is an autosomal-recessive condition associated with liver enlargement, cataracts, and mental retardation. Among the offspring of unrelated individuals, the frequency of galactosemia is 8.5×10^{-6}. What is the expected frequency among the offspring of first cousins ($F = 1/16$) and among the offspring of second cousins ($F = 1/64$)?

17.19 Suppose the frequency of a recessive genetic disorder among the children of nonrelatives in a population is 1 in 2.6 million.

(a) What is the estimated frequency of the harmful recessive allele?
(b) What is the expected frequency of homozygous recessives among the offspring of first cousins?
(c) If first-cousin matings make up 1 percent of all matings in the population, and all other matings are random, what is the total frequency of homozygous recessive genotypes?
(d) What proportion of all homozygous recessive genotypes come from the first-cousin matings?

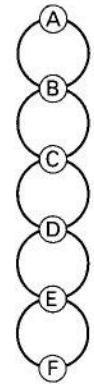
17.20 Consider the accompanying pedigree.

(a) Identify all of the common ancestors.
(b) Identify all paths through each common ancestor, underlining the common ancestor in each path.
(c) Calculate the inbreeding coefficient of individual L, assuming that F_B is unknown.
(d) Calculate the inbreeding coefficient of individual L, assuming that $F_B = 0$.



17.21 The accompanying pedigree shows five consecutive generations of self-fertilization. Assume that $F_A = 0$.

(a) Calculate the inbreeding coefficients of individuals B through F.
(b) Deduce a general formula $F(t)$ for the inbreeding coefficient after t generations of self-fertilization.



17.22 Self-fertilization in the annual plant *Phlox cuspidata* results in an average inbreeding coefficient of $F = 0.66$.

(a) What frequencies of the genotypes for the enzyme phosphoglucose isomerase would be expected in a population with alleles *A* and *B* at frequencies 0.43 and 0.57, respectively?
(b) What frequencies of the genotypes would be expected with random mating?

17.23 Two strains of *Escherichia coli*, *A* and *B*, are inoculated into a chemostat in equal frequencies and undergo competition. After 40 generations, the frequency of the *B* strain is 35 percent. What is the fitness of strain *B* relative to strain *A*, under the particular experimental conditions, and what is the selection coefficient against strain *B*?

17.24 A strain of *Drosophila* has an insertion of the transposable element *mariner* in the wildtype w^+ allele of the gene *white* in the X chromosome. The insertion mutation results in peach-colored eyes instead of white. Suppose this transposable element undergoes spontaneous excision from its site of insertion with a frequency of 5 percent per generation:

(a) What is the frequency of the mutant allele in X chromosomes after 1 generation?
(b) After 10 generations?
(c) After 20 generations?
(d) After how many generations will the frequency of mutant X chromosomes be below 0.01?

17.25 An allele *A* is a hotspot of mutation and undergoes mutation to *a* at a frequency of 10^{-4} per generation. If the frequency of reverse mutation from *a* to *A* is 10^{-7} per generation, what is the expected equilibrium allele frequency of *a*?

17.26 An allele *A* mutates to a nonfunctional allele *a* at a rate 10^{-6} per generation. Reverse mutation from *a* to *A* can be ignored. Determine the expected equilibrium frequency of allele *a* if:

(a) The allele *a* is a recessive lethal with no effect in heterozygous genotypes.

(b) The allele *a* is a "recessive" lethal that results in a 1 percent selective disadvantage in heterozygous genotypes.
(c) By what percentage does the small amount of dominance reduce the equilibrium frequency of *a*?

17.27 Huntington disease is a form of neuromuscular degeneration, usually beginning in middle age, that results from a dominant mutation in a gene encoding a protein called huntingtin. The early onset reduces the reproductive fitness of affected individuals by about 20 percent. A large study in Michigan estimated that the frequency of the dominant allele is about 0.00005. Assume that this population undergoes random mating and is in equilibrium.
(a) What is the expected frequency of individuals who carry the dominant allele?
(b) What is the estimated mutation rate to the dominant deleterious allele?

17.28 Determine the ultimate fate of the *A* allele with random mating when the relative fitnesses of *AA*, *Aa*, and *aa* are
(a) 1.00, 0.98, 0.96
(b) 1.00, 1.02, 1.02
(c) 1.00, 1.02, 0.98
(d) 1.00, 1.01, 1.02

17.29 Which allele frequency will change faster? Explain why.
(a) A rare beneficial allele that is dominant or a rare beneficial allele that is recessive?
(b) A common deleterious allele that is dominant or a common deleterious allele that is recessive?

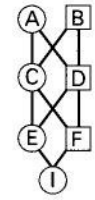
17.30 An overdominant locus has relative fitnesses of *AA*, *Aa*, and *aa* given by 0.90 : 1.00 : 0.50.
(a) What is the expected equilibrium frequency of the *a* allele?
(b) What is the expected frequency of heterozygous genotypes at equilibrium?
(c) If homozygous *aa* were lethal, how would this change the equilibrium allele frequency of *a*?
(d) What is the expected frequency of heterozygous genotypes at the new equilibrium?

Challenge Problems

17.31 In a randomly mating population that contains *n* equally frequent alleles, what is the expected frequency of heterozygotes?

17.32 Human genetic polymorphisms in which there are a variable number of tandem repeats (VNTR) are often used in DNA typing. The table below shows the alleles of four VNTR loci and their estimated average frequencies in genetically heterogeneous Caucasian populations. (The allele 9.3 of the *HUMTHO1* locus has a partial copy of one of the tandem repeats.) Using Equation (2), estimate
(a) the probability that an individual will be of genotype (6, 9.3) for *HUMTHO1*, (10, 11) for *HUMFES*, (6, 6) for *D12S67*, and (18, 24) for *D1S80*.
(b) The probability that an individual will be of genotype (8, 10) for *HUMTHO1*, (8, 13) for *HUMFES*, (1, 2) for *D12S67*, and (19, 19) for *D1S80*.

17.33 In the pedigree shown here, individual I is the result of two successive generations of brother–sister mating. Assume that $F_A = F_B = 0$.
(a) Calculate the inbreeding coefficients of individuals E and F (who result from one generation of brother–sister mating).
(b) Calculate the inbreeding coefficient of I.



HUMTHO1		*HUMFES*		*D12S67*		*D1S80*	
Repeats	**Freq.**	**Repeats**	**Freq.**	**Repeats**	**Freq.**	**Repeats**	**Freq.**
6	0.230	8	0.007	1	0.015	18	0.293
7	0.160	10	0.321	2	0.005	19	0.011
8	0.105	11	0.373	3	0.058	20	0.021
9	0.193	12	0.233	4	0.078	21	0.032
9.3	0.310	13	0.066	5	0.118	22	0.043
10	0.002			6	0.324	23	0.016
				7	0.196	24	0.335
				8	0.127	25	0.037
				9	0.059	26	0.016
				10	0.020	28	0.078
						29	0.059
						30	0.016
						31	0.043

Problem 17.32—VNTR locus

Further Reading

Allison, A. C. 1956. Sickle cells and evolution. *Scientific American*, August.

Ayala, F. 1978. The mechanisms of evolution. *Scientific American*, September.

Bittles, A. H., W. M. Mason, J. Greene, and N. A. Rao. 1991. Reproductive behavior and health in consanguineous marriages. *Science* 252: 789.

Bodmer, W. F., and L. L. Cavalli-Sforza. 1976. *Genetics, Evolution, and Man*. New York: Freeman.

Crow, J. F., and M. Kimura. 1970. *An Introduction to Population Genetics Theory*. New York: Harper & Row.

Dobzhansky, T. 1941. *Genetics and the Origin of the Species*. New York: Columbia University Press.

Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction of Quantitative Genetics*. 4th ed. Essex, England: Longman.

Graur, D., and W.-H. Li. 2000. *Fundamentals of Molecular Evolution*. 2d ed. Sunderland, MA: Sinauer.

Hartl, D. L. 2000. *A Primer of Population Genetics*. 3d ed. Sunderland, MA: Sinauer.

Hartl, D. L., and A. G. Clark. 1997. *Principles of Population Genetics*. 3rd ed. Sunderland, MA: Sinauer.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge, England: Cambridge University Press.

Li, W.-H. 1997. *Molecular Evolution*. Sunderland, MA: Sinauer.

Mitton, J. F. 1997. *Selection in Natural Populations*. New York, NY: Oxford University Press.

Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M., and S. Kumar 2000 *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.

Neufeld, D. J., and N. Colman. 1990. When science takes the witness stand. *Scientific American*, May.

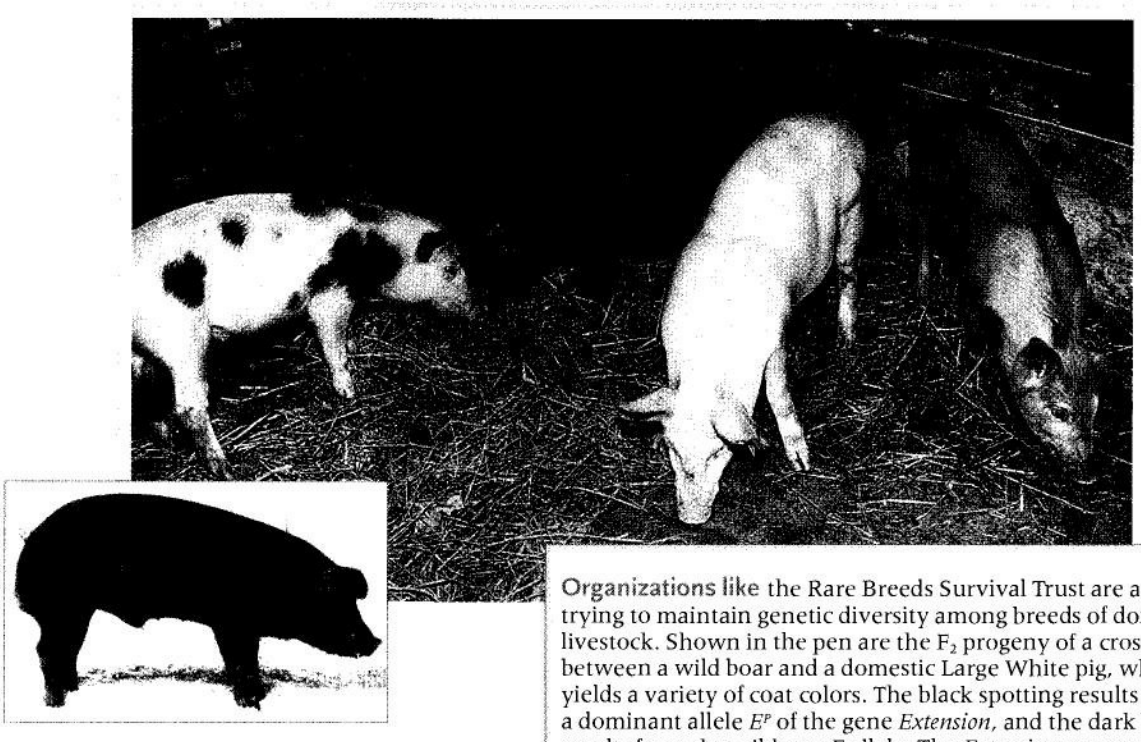
Weir, B. S. 1996. *Genetic Data Analysis II*. Sunderland, MA: Sinauer.

White, R., and J.-M. Lalouel. 1988. Chromosome mapping with DNA markers. *Scientific American*, February.

Wright, S. 1978. *Evolution and the Genetics of Populations* (4 volumes). Chicago: University of Chicago Press

CHAPTER 18

The Genetic Basis of Complex Inheritance



Organizations like the Rare Breeds Survival Trust are actively trying to maintain genetic diversity among breeds of domestic livestock. Shown in the pen are the F_2 progeny of a cross between a wild boar and a domestic Large White pig, which yields a variety of coat colors. The black spotting results from a dominant allele E^p of the gene *Extension*, and the dark brown results from the wildtype *E* allele. The *Extension* gene encodes a receptor on melanocyte cells that regulates the production of different types of melanin. The red pig shown at left is homozygous for a recessive allele of the *E* gene; mutation in the homologous gene in humans results in red hair and fair skin. [Pigs: courtesy of Leif Andersson and The Rare Breeds Survival Trust, UK.]

CHAPTER OUTLINE

18.1 Complex Traits
- Continuous, Meristic, and Threshold Traits
- The Normal Distribution

18.2 Causes of Variation
- Genotypic Variation
- Environmental Variation
- Genetics and Environment Combined
- Genotype–Environment Interaction and Association

18.3 Genetic Analysis of Complex Traits
- The Number of Genes Affecting Complex Traits
- Broad-Sense Heritability
- Twin Studies

18.4 Artificial Selection
- Narrow-Sense Heritability
- Phenotypic Change with Individual Selection: A Prediction Equation
- Long-Term Artificial Selection
- Inbreeding Depression and Heterosis

18.5 Correlation Between Relatives
- Covariance and Correlation
- The Geometrical Meaning of a Correlation
- Estimation of Narrow-Sense Heritability

18.6 Heritabilities of Threshold Traits

18.7 Identification of Genes Affecting Complex Traits
- Linkage Analysis in the Genetic Mapping of Quantitative Trait Loci
- Candidate Genes for Complex Traits

PRINCIPLES

- Complex traits are determined by multiple genetic and environmental factors acting together.
- The relative contributions of genotype and environment to a trait are measured by the variance due to genotype (genotypic variance) and the variance due to environment (environmental variance).
- Correlations between relatives are used to estimate various components of variation, such as genotypic variance, additive variance, and dominance variance.
- Additive variance accounts for the parent–offspring correlation; dominance variance accounts for the sib–sib correlation over and above that expected from the additive variance.
- Narrow-sense heritability is the ratio of additive (transmissible) variance to the total phenotypic variance; it is widely used in plant and animal breeding to predict the response to artificial selection.
- Genes that affect quantitative traits can be identified by genetic mapping using highly polymorphic DNA markers, or by specifying candidate genes on the basis of knowledge of the physiology and development of the trait.

CONNECTIONS

A Maize'n Grass
George W. Beadle 1972
The Mystery of Maize

Win, Place, or Show?
Hugh E. Montgomery and 18 other investigators 1998
Human Gene for Physical Performance

Earlier chapters have emphasized traits in which differences in phenotype result from alternative genotypes of a single gene. These traits are particularly suited for genetic analysis through the study of pedigrees, not only because of the small number of genotypes and phenotypes, but also because of the simple correspondence between genotype and phenotype. However, many traits that are important in medical genetics, plant breeding, and animal breeding are influenced by *multiple* genes as well as by the environment. These are known as **complex traits** or **multifactorial traits** because of the multiple genetic and environmental factors implicated in their causation, and they are said to show *complex inheritance*. The inheritance is complex because a single genotype can have many possible phenotypes (depending on the environment), and a single phenotype can include many possible genotypes. Genetic analysis of complex inheritance requires special concepts and methods, which are introduced in this chapter.

18.1 Complex Traits

Complex traits are typically influenced not only by the alleles of two or more genes but also by the environment. The phenotype of an organism is therefore potentially influenced by

- ***Genetic factors*** in the form of alternative genotypes of one or more genes.
- ***Environmental factors*** in the form of conditions that are favorable or unfavorable for the development of the trait.

Examples of environmental factors include the effect of nutrition on the growth rate of animals and the effects of fertilizer, rainfall, and planting density on the yield of crop plants.

With some complex traits, differences in phenotype result largely from differences in genotype, and the environment plays a minor role. With others, differences in phenotype result largely from the effects of environment, and genetic factors play a minor role. However, most complex traits fall between these extremes, and both genotype and environment must be taken into account in their analysis.

In a genetically heterogeneous population, many genotypes are formed by the processes of segregation and recombination. Variation in genotype can be eliminated by studying inbred lines, which are homozygous for most genes, or the F_1 progeny from a cross of inbred lines, which are uniformly heterozygous for all genes in which the parental inbreds differ. In contrast, complete elimination of environmental variation is impossible, no matter how hard the experimenter may try to render the environment identical for all members of a population. With plants, for example, small variations in soil quality or exposure to the sun will produce slightly different environments, sometimes even for adjacent plants. Similarly, highly inbred *Drosophila* still show variation in phenotype (for example, in body size) brought about by environmental differences among animals within the same culture bottle. Therefore, traits that are susceptible to small environmental effects will never be uniform, even in inbred lines.

Most traits that are important in plant and animal breeding are complex traits. In agricultural production, one economically important complex trait is yield—for example, the size of the harvest of corn, tomatoes, soybeans, or grapes per unit area. In domestic animals, important complex traits include meat quality, milk production per cow, egg laying per hen, fleece weight per sheep, and litter size per sow. Important complex traits in human genetics include infant growth rate, adult weight, blood pressure, serum cholesterol, and length of life. In evolutionary studies, fitness is the preeminent complex trait.

Most complex traits cannot be studied by means of the usual pedigree methods, because the effects of segregation of alleles of one gene may be concealed by effects of other genes, and environmental effects may cause identical genotypes to have different phenotypes. Therefore, individual pedigrees do not fit any simple pattern of dominance, recessiveness, or X linkage. Nevertheless, genetic effects on the traits can be assessed by comparing the phenotypes of relatives who, because of their familial relationship, must have a certain proportion of their genes in common. Such studies utilize many of the concepts of population genetics discussed in Chapter 17.

Three categories of traits are frequently found to have complex inheritance. They are described in the following section.

Continuous, Meristic, and Threshold Traits

Most phenotypic variation in populations is not manifested in a few easily distinguished categories. Instead, the traits vary continuously from one phenotypic extreme to the other, with no clear-cut breaks in between. Height and weight are prime examples. Other examples include milk production in cattle, growth rate in poultry, yield in corn, and blood pressure in human beings. Such traits are called **continuous traits** or **quantitative traits,** because there is a continuous or quantitative gradation from one phenotype to the next.

Two other types of complex traits are not continuous:

Meristic traits are traits in which the phenotype is determined by counting. Some examples are number of skin ridges forming the fingerprints, number of kernels on an ear of corn, number of eggs laid by a hen, number of bristles on the abdomen of a fly, and number of puppies in a litter.

Threshold traits are traits that have only two, or a few, phenotypic classes, but their inheritance is determined by the effects of multiple genes acting together with the environment. Examples of threshold traits are twinning in cattle and parthenogenesis (development of unfertilized eggs) in turkeys. In a threshold trait, each organism has an underlying risk or *liability* to express the trait. Even though the underlying liability is not directly observable, if it is high enough (above a "threshold"), the trait will be expressed (for example, a pregnancy results in twins); otherwise, the trait is not expressed (the pregnancy results in a single birth). In many threshold-trait disorders, the phenotypic classes are "affected" and "not affected." Examples of threshold-trait disorders in human beings include adult-onset diabetes, schizophrenia, and many congenital abnormalities, such as spina bifida. Threshold traits can be interpreted as continuous traits by imagining that the underlying *liability* is a continuous variable.



French soccer fans at a match in St. Denis illustrate continuous variation among people, in height, weight, skin color, facial features, and many other traits. [Associated Press/ Michel Euler.]

The Normal Distribution

The **distribution** of a trait in a population is a description of the population in terms of the proportion of individuals that have each of the possible phenotypes. To create such a description with continuous traits, it is usually convenient to begin by grouping similar phenotypes into classes. Data for an example pertaining to the distribution of height among 4995 British women are given in **Table 18.1** and in **Figure 18.1**. One can imagine each bar in the graph in Figure 18.1 being built, step by step, as each of the women's height is measured, by placing a small square along the x-axis at the location corresponding to the height of that woman. As sampling proceeds, the squares begin to pile up in certain places, leading ultimately to the histogram shown.

Displaying a distribution completely, in either tabular form (Table 18.1) or graphical form (Figure 18.1), is always helpful but often unnecessary; it often happens that a description of the distribution in terms of two major features—the *mean* and the *variance*—is sufficient. To discuss the mean and the variance in quantitative terms, we will

Table 18.1 Distribution of height among British women

Interval number (i)	Height interval (inches)	Midpoint (x_i)	Number of women (f_i)
1	53–55	54	5
2	55–57	56	33
3	57–59	58	254
4	59–61	60	813
5	61–63	62	1340
6	63–65	64	1454
7	65–67	66	750
8	67–69	68	275
9	69–71	70	56
10	71–73	72	11
11	73–75	74	4
			Total N = 4995

need some symbols. In Table 18.1, the height intervals are numbered from 1 (53–55 inches) to 11 (73–75 inches). The symbol x_i designates the midpoint of the height interval numbered i; for example, x_1 = 54 inches, x_2 = 56 inches, and so on.

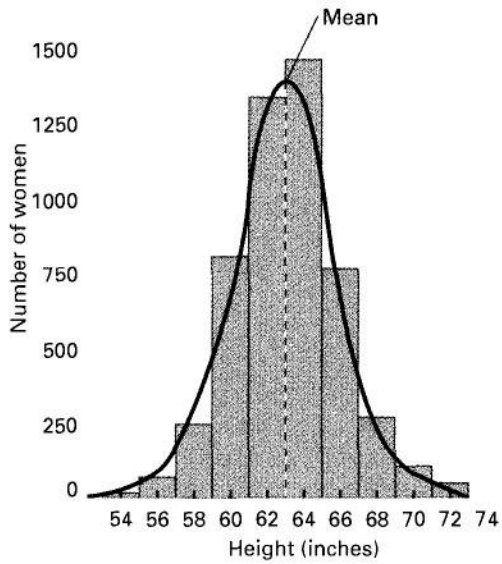


Figure 18.1 Distribution of height among 4995 British women and the smooth normal distribution that approximates the data.

The number of women in height interval i is designated f_i; for example, f_1 = 5 women, f_2 = 33 women, and so forth. The total size of the sample, in this case 4995, is denoted by N. The mean and variance serve to characterize the distribution of height among these women as well as the distribution of many other quantitative traits.

The **mean,** or average, is the peak of the distribution. The mean of a population is estimated from a sample of individuals from the population, as follows:

$$\bar{x} = \frac{\Sigma f_i x_i}{N} \qquad (1)$$

where $\bar{x}$ is the estimate of the mean and Σ symbolizes summation over all classes of data (in this example, summation over all 11 height intervals). In Table 18.1, the mean height in the sample of women is 63.1 inches.

The **variance** is a measure of the spread of the distribution and is estimated in terms of the squared *deviation* (difference) of each observation from the mean. The variance is estimated from a sample of individuals as follows:

$$s^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{N - 1} \qquad (2)$$

where s^2 is the estimated variance and x_i, f_i, and N are as in Table 18.1. The difference $(x_i - \bar{x})$ is the deviation of each height category from the mean height in the entire sample. The variance describes the extent to which the phenotypes are clustered around the mean, as shown in **Figure 18.2.** A large value implies that the distribution is spread out, and a small value implies that it is clustered near the mean. From the data in Table 18.1, the variance of the population of British women is estimated as s^2 = 7.24 in^2. (Note that the units of the variance are squared.)

A quantity closely related to the variance—the **standard deviation** of the distribution—is defined as the square root of the variance. For the data in Table 18.1, the estimated standard deviation s is obtained from Equation (2) as follows:

$$s = \sqrt{s^2} = \sqrt{7.24 \text{ in}^2} = 2.69 \text{ inches}$$

The standard deviation has the useful feature of having the same units of dimension as the mean—in this example, inches. Just

as the mean of the sample, $\bar{x}$, is used to estimate the unknown mean of the population μ, the standard deviation of the sample, s, is used to estimate the unknown standard deviation of the populaion, σ, and s^2 used to estimate σ^2.

When the data are symmetrical, or approximately symmetrical, the distribution of a trait can often be approximated by the smooth, arching curve known as the **normal distribution**, which is shown for the height example in Figure 18.1. Because the normal distribution is symmetrical, half of its area is determined by points that have values greater than the mean and half by points with values less than the mean, and thus the proportion of phenotypes that exceed the mean is 1/2. One important characteristic of the normal distribution is that it is completely determined by the value of the mean and the variance.

The mean and the standard deviation (square root of the variance) of a normal distribution provide a great deal of information about the distribution of phenotypes in a population, as is illustrated in **Figure 18.3**. Specifically, for a normal distribution:

1. Approximately 68 percent of the population have a phenotype within 1 standard deviation of the mean (in the symbols of Figure 18.3, between $\mu - \sigma$ and $\mu + \sigma$).
2. Approximately 95 percent lie within 2 standard deviations of the mean (between $\mu - 2\sigma$ and $\mu + 2\sigma$).
3. Approximately 99.7 percent lie within 3 standard deviations of the mean (between $\mu - 3\sigma$ and $\mu + 3\sigma$).

Applying these rules to the data in Figure 18.1, in which the mean and standard deviation are 63.1 and 2.69 inches, reveals that approximately 68 percent of the women are expected to have heights in the range from 63.1 − 2(2.69) inches to 63.1 + 2.69 inches (that is, 60.4–65.8), and approximately 95 percent are expected to have heights in the range from 63.1 − 2(2.69) inches to 63.1 + 2(2.69) inches (that is, 57.7–68.5).

Real data frequently conform to the normal distribution. Normal distributions are usually the rule when the phenotype is determined by the cumulative effect of many individually small independent factors. This is the case for many multifactorial traits.

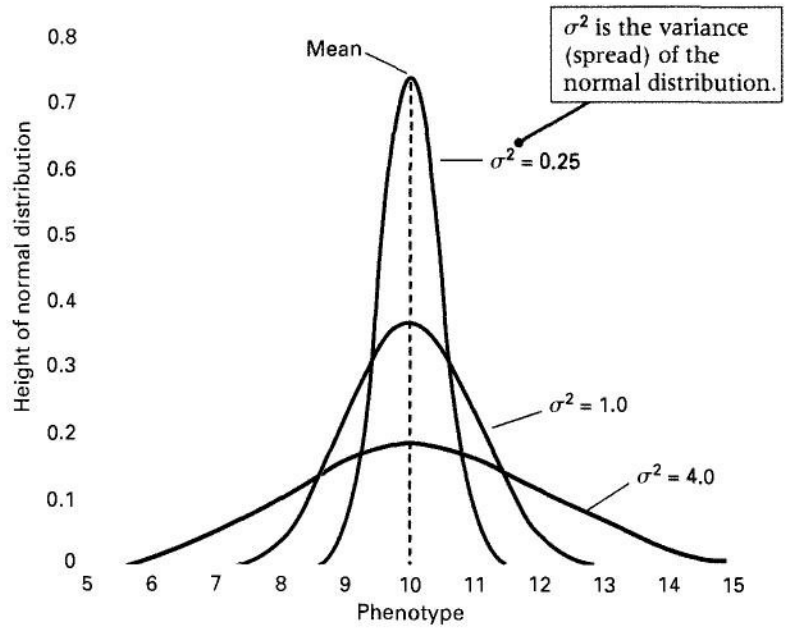


Figure 18.2 Graphs showing that the variance of a distribution measures the spread of the distribution around the mean. The area under each curve covering any range of phenotypes equals the proportion of individuals having phenotypes within that range.

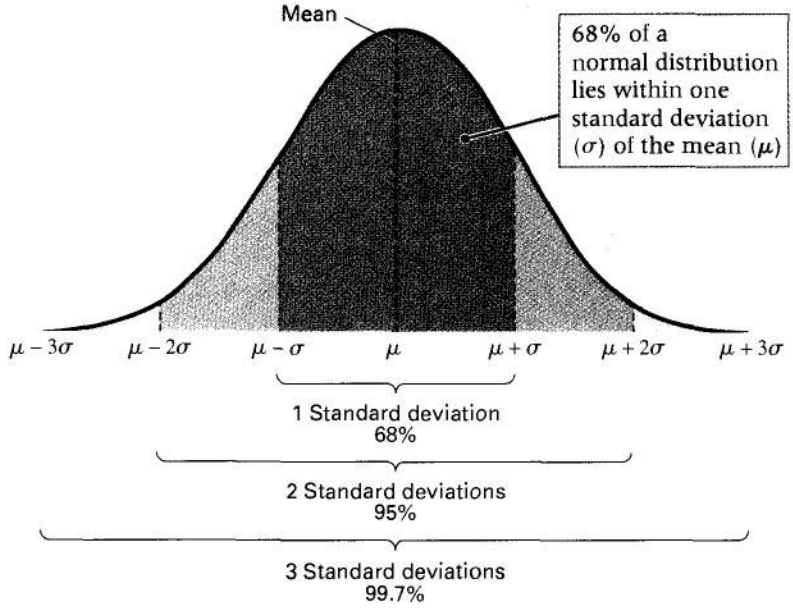


Figure 18.3 Features of a normal distribution. The proportions of individuals lying within 1, 2, and 3 standard deviations of the mean are approximately 68 percent, 95 percent, and 99.7 percent, respectively. In this normal distribution, μ represents the mean and σ the standard deviation.

Connection

A Maize'n Grass

George W. Beadle 1972
University of Chicago
The Mystery of Maize

Cultivated corn (Zea mays) is a wonder crop. It can be boiled, baked, popped, and chopped; it can be ground, fermented, fried, and dried. It is fodder for livestock and raw material for industry. Domesticated 5 to 10 thousand years ago by native American Indians in the Balsas River drainage, about 100 miles southwest of Mexico City, maize became a staple for many people in the Middle Americas. The native Americans developed 200–300 varieties, essentially all that we have today. Accomplished without knowledge of genetics, their achievement in plant breeding has been called the most extraordinary in all of human existence. But the ancestral source of maize was for a long time very obscure. A few early geneticists suspected teosinte because they could imagine a handful of major gene differences affecting growth and development that might convert teosinte into something resembling maize. Genetic evidence for this hypothesis is presented in this piece by George Beadle. Best known for his work in Neurospora *and for the "one gene, one enzyme" concept (see the Connection on page 452), Beadle had begun his career studying corn, and when he retired as president of the University of Chicago in 1968, he immediately resumed his maize experiments in a half-acre lot off Ellis Avenue near 55th Street in the middle of the city, carrying out crosses that he had first conceived as a graduate student at Cornell 40 years earlier. Sure enough, the major differences between the species could be accounted for by mutations in a few key genes. A number of these have since been cloned and sequenced and their functions identified. One is a gene that in maize is known as* teosinte branched, *whose mutant phenotype is a highly branched plant similar in growth habit to teosinte.*

It was a plant that natives in parts of Mexico called teocentli, from the Aztec language, meaning "support of the gods."

For most cultivated plants, wild relatives are known from which they could reasonably have been derived. . . . In the

18.2 Causes of Variation

In considering the genetics of complex traits, an important objective is to assess the relative importance of genotype and environment. In some cases in experimental organisms, it is possible to separate genotype and environment with respect to their effects on the mean phenotype. For example, a plant breeder may study the yield of a series of inbred lines grown in a group of environments that differ in planting density or amount of fertilizer. It would then be possible:

- To compare yields of the same genotype grown in different environments, and thereby rank the *environments* relative to their effects on yield
- To compare yields of different genotypes grown in the same environment and thereby rank the *genotypes* relative to their effects on yield

Such a fine discrimination between genetic and environmental effects is not usually possible, particularly in human genetics. For example, with regard to the height of the women in Figure 18.1, environment could be considered favorable or unfavorable for tall stature only in comparison with the mean height of a genetically identical population reared in a different environment. This reference population does not exist. Likewise, the genetic composition of the population could be judged as favorable or unfavorable for tall stature only in comparison with the mean of a genetically different population reared in an identical environment. This reference population does not exist either.

Without such standards of comparison, it is impossible to distinguish genetic from environmental effects on the mean phenotype. However, it is still possible to assess genetic versus environmental contributions to the *variance* in phenotype, because instead of comparing the means of two or more populations, we can compare the phenotypes of individuals within the *same* population. Some of the differences in phe-

case of corn, early plant explorers had great difficulty finding any wild relative that could conceivably have been its ancestor. . . . Finally they found what looked as if it might be a relative. It was a plant that natives in parts of Mexico called *teocentli,* from the Aztec language, meaning "support of the gods." This is now anglicized to "teosinte." . . . I well remember discussions with Rollins A. Emerson [at Cornell University in about 1930] in which he pointed out that just two mutations could make teosinte an easily usable food plant—one to a nonshattering rachis [the elongated axis of the ear], so the fruits would not be scattered and lost as food; the other to a soft fruitcase, so the kernels could be threshed free of them. . . . To jump ahead now forty years, . . . I decided that on retirement [as president of the University of Chicago] I would return to a study of the relation of teosinte to corn. . . . Mendel's laws say that if original parents differ from each other by only one gene, in the second generation of descendants [from the hybrid cross] each original parental type will appear with a statistical frequency of one in four. With a difference of two genes, the frequency with each parental type will be reproduced in one in sixteen; and so on for more genes. . . . I resolved to grow up to 50,000 second-generation plants, if necessary. . . . With this year's planting, I estimate that we will have grown and examined just about 50,000. . . . What have been the results? In the second-generation cross of Chapalote corn and Chalco teosinte we found good parent corn types and good parent teosinte types with a frequency of about 1 in each 500 plants. These frequencies are intermediate between those expected with four and those expected with five gene differences. Of course, the genetic complexities are greater than this kind of arithmetic implies, but I will not go into them here. . . . It seems clear that the genetic differences between corn and teosinte cannot be so great as to [exclude] an ancestral relationship of teosinte to corn. And it seems reasonable to assume that [native American Indians] could have selected and preserved the relatively few mutations required to produce a useful plant from teosinte—early corn.

Source: Field Museum Natural History Bulletin 43: 2–11

notype result from differences in genotype and others from differences in environment, and it is often possible to separate these effects.

In any distribution of phenotypes, such as the one in Figure 18.1, four sources contribute to phenotypic variation:

1. Genotypic variation
2. Environmental variation
3. Variation due to genotype–environment interaction
4. Variation due to genotype–environment association

These sources of variation are discussed in the following sections.



An ear of wild teosinte (left) looks dramatically different from an ear of modern cultivated maize (right), but the organisms are very closely related, producing fertile hybrids in both reciprocal crosses (middle). Early native Americans first domesticated teosinte and, through many generations, selected the mutant phenotypes most favorable for cultivation, leading ultimately to modern maize. [© John Doebley, Plimoth Plantation Agriculture.]

Genotypic Variation

The variation in phenotype caused by differences in genotype among individuals is termed **genotypic variance.** Figure 18.4 illustrates the genetic variation expected among the F_2 generation from a cross of two inbred lines that differ in genotype for

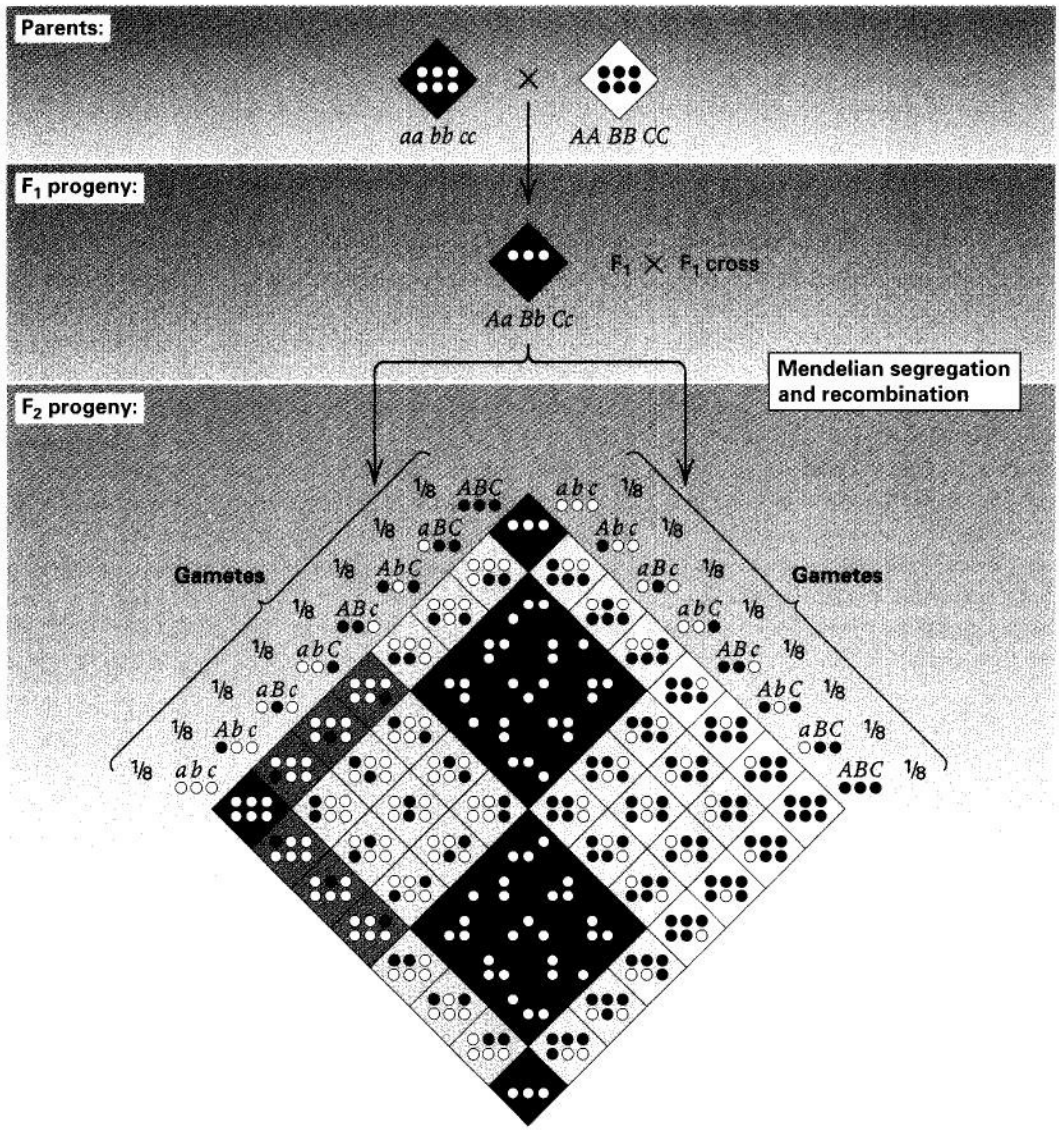


Figure 18.4 Segregation of three independent genes affecting a quantitative trait. Each uppercase allele in a genotype contributes one unit to the phenotype.

three unlinked genes. The alleles of the three genes are represented as *Aa*, *Bb*, and *Cc*, and the genetic variation in the F_2 caused by segregation and recombination is evident in the differences in color. Relative to a meristic trait (a trait whose phenotype is determined by counting, such as ears per stalk in corn), if we assume that each uppercase allele is favorable for the expression of the trait and adds one unit to the phenotype, whereas each lowercase allele is without effect, then the *aa bb cc* genotype has a phenotype of 0, and the *AA BB CC* genotype has a phenotype of 6. There are seven possible phenotypes (0–6) in the F_2 generation. The distribution of phenotypes in the F_2 generation is shown in the colored bar graph in **Figure 18.5**. The normal distribution approximating the data has a mean of 3 and a variance of 1.5. In this case, we are assuming that *all* of the variation in phenotype in the population results from differences in genotype among the individuals.

Figure 18.5 also includes a bar graph with diagonal lines that represent the theoretical distribution when the trait is deter-

mined by 30 unlinked genes segregating in a randomly mating population, grouped into the same number of phenotypic classes as the 3-gene case. We assume that 15 of the genes are nearly fixed for the favorable allele and 15 for the unfavorable allele. The contribution of each favorable allele to the phenotype has been chosen to make the mean of the distribution equal to 3 and the variance equal to 1.5. Note that the distribution with 30 genes is virtually identical to that with 3 genes and that both are approximated by the same normal distribution. If such distributions were encountered in actual research, the experimenter would not be able to distinguish between them. The key point is that:

> Even in the absence of environmental variation, the distribution of phenotypes, by itself, provides no information about the number of genes influencing a trait and no information about the dominance relations of the alleles.

However, the number of genes influencing a complex trait is important in determining the potential for long-term genetic improvement of a population by means of artificial selection. For example, in the 3-gene case in Figure 18.5, the best possible genotype would have a phenotype of 6, but in the 30-gene case, the best possible genotype (homozygous for the favorable allele of all 30 genes) would have a phenotype of 30.

Later in this chapter, some methods for estimating how many genes affect a quantitative trait will be presented. All the methods depend on comparing the phenotypic distributions in the F_1 and F_2 generations of crosses between nearly or completely homozygous lines.

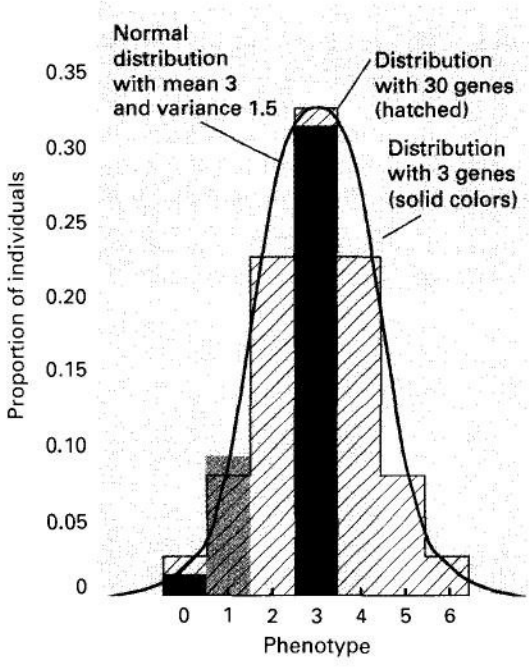


Figure 18.5 The colored bar graph is the distribution of phenotypes determined by the segregation of the 3 genes illustrated in Figure 18.4. The bar graph with the diagonal lines is the theoretical distribution expected from the segregation of 30 independent genes. Both distributions are approximated by the same normal distribution (black curve).

Environmental Variation

The variation in phenotype caused by differences in environment among individuals is termed **environmental variance.** Figure 18.6 is an example showing the distribution of seed weight in edible beans. The mean of the distribution is 500 mg, and the standard deviation is 95 mg. All of the beans in this population are genetically identical and homozygous because they are highly inbred. Therefore, in this population, *all* of the phenotypic variation in seed weight results

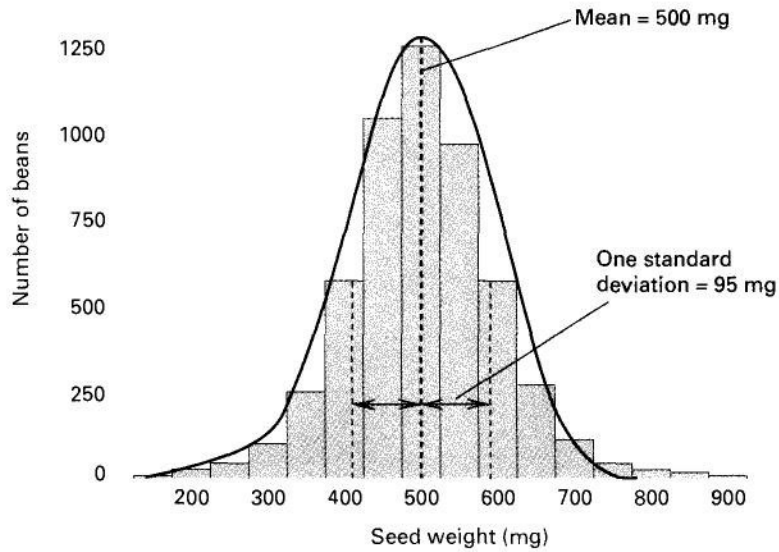


Figure 18.6 Distribution of seed weight in a homozygous line of edible beans. All variation in phenotype among individuals results from environmental differences.

from environmental variance. A comparison of Figures 18.5 and 18.6 demonstrates the following principle:

> The distribution of a trait in a population provides no information about the relative importance of genotype and environment. Variation in the trait can be entirely genetic, it can be entirely environmental, or it can reflect a combination of both influences.

Genetics and Environment Combined

Genotypic and environmental variance are seldom separated as clearly as in Figures 18.5 and 18.6, because usually they work together. Their combined effects are illustrated for a simple hypothetical case in **Figure 18.7**. Figure 18.7A is the distribution of phenotypes for three genotypes, ignoring (for the moment) environmental effects. As

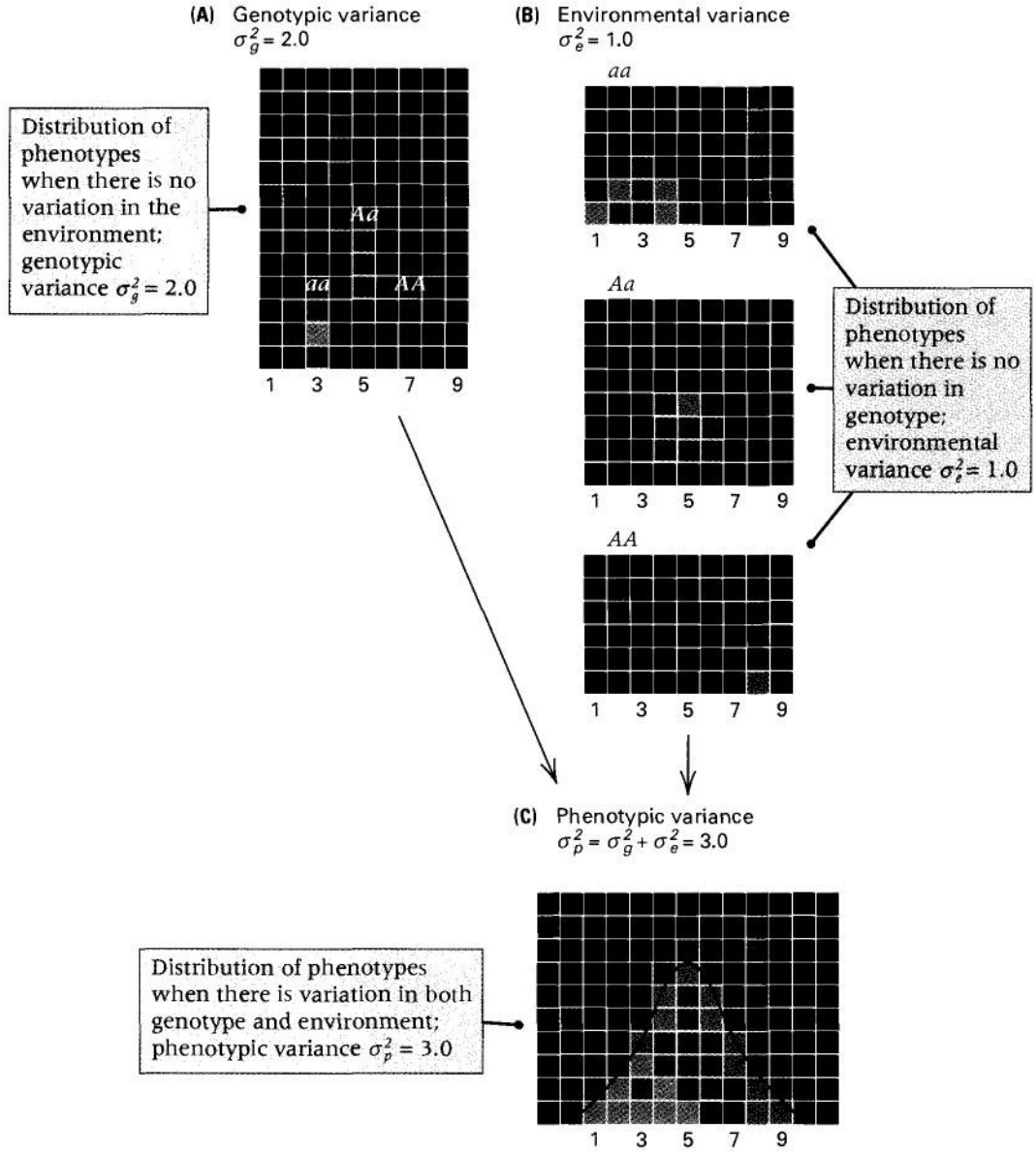


Figure 18.7 The combined effects of genotypic and environmental variance. (A) Population affected only by genotypic variance, σ_g^2. (B) Populations of each genotype affected only by environmental variance, σ_e^2. (C) Population affected by both genotypic and environmental variance, in which the phenotypic variance, σ_p^2, equals the sum of σ_g^2 and σ_e^2.

depicted, the trait can have one of three distinct and nonoverlapping phenotypes determined by the effects of two additive alleles. The genotypes are in random-mating proportions for an allele frequency of 1/2, and the distribution of phenotypes has mean 5 and variance 2. Because it results solely from differences in genotype, this variance is *genotypic variance*, which is symbolized σ_g^2. Figure 18.7B is the distribution of phenotypes in the presence of environmental variation, illustrated for each genotype separately. Each distribution of environmental effects has a mean of 0 and a variance of 1. Because this variance results solely from differences in environment, it is *environmental variance*, which is symbolized σ_e^2. When the effects of genotype and environment are combined in the same population, each genotype is affected by environmental variation, and the distribution shown in Figure18.7C of the figure results. The variance of this distribution is the **phenotypic variance,** which is symbolized σ_p^2. Because we are assuming that genotype and environment have separate and independent effects on phenotype, we expect σ_p^2 to be greater than either σ_g^2 or σ_e^2 alone. In fact,

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2 \qquad (3)$$

In words, Equation (3) states that:

> When genetic and environmental effects contribute independently to phenotype, the total variance equals the sum of the genotypic variance and the environmental variance.

Equation (3) is one of the most important relations in the analysis of complex inheritance. How it can be used to analyze data will be explained shortly. Although the equation serves as an excellent approximation in very many cases, it is valid in an exact sense only when genotype and environment are independent in their effects on phenotype. The two most important departures from independence are discussed in the next section.

Genotype–Environment Interaction and Association

In the simplest cases, environmental effects on phenotype are additive, and each environment adds (or detracts) the same amount to (or from) the phenotype, independent of the genotype. When this is not true, environmental effects on phenotype differ according to genotype, and a **genotype–environment interaction (G–E interaction)** is said to be present. In some cases, G–E interaction can even change the relative ranking of the genotypes, so a genotype that is superior in one environment may become inferior in another.

An example of genotype–environment interaction in maize is illustrated in **Figure 18.8.** The two strains of corn are hybrids formed by crossing different pairs of inbred lines, and their overall means, averaged across all of the environments, are approximately the same. However, the strain designated A clearly outperforms B in the stressful (negative) environments, whereas the performance is reversed when the environment is of high quality. (Environmental quality is judged on the basis of soil fertility, moisture, and other factors.) In some organisms, particularly plants, experiments like those illustrated in Figure 18.8 can be

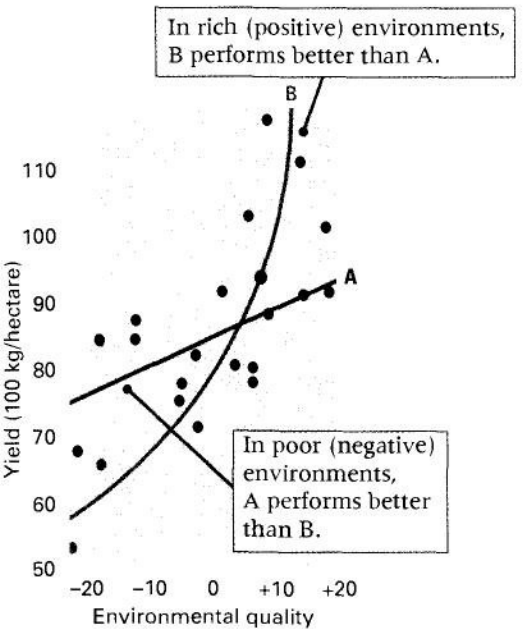


Figure 18.8 Genotype–environment interaction in maize. Strain A is superior when environmental quality is low (negative numbers), but strain B is superior when environmental quality is high. [Data from W. A. Russell. 1974. *Annual Corn & Sorghum Research Conference*, 29: 81.]

carried out to determine the contribution of G–E interaction to the total observed variation in phenotype. In other organisms, particularly human beings, the effect cannot be evaluated separately.

Interaction of genotype and environment is common and is very important in both plants and animals. Because of this interaction, no one plant variety will outperform all others in all types of soil and climate, and therefore plant breeders must develop special varieties that are suited to each growing area.

When the different genotypes are not distributed at random in all the possible environments, there is **genotype-environment association (G–E association).** In these circumstances, certain genotypes are preferentially associated with certain environments, which may either increase or decrease the phenotype of these genotypes compared with what would result in the absence of G–E association. An example of deliberate genotype–environment association can be found in dairy husbandry, in which some farmers feed each of their cows in proportion to its level of milk production. Because of this practice, cows with superior genotypes with respect to milk production also have a superior environment in that they receive more feed. In plant breeding, genotype–environment association can often be eliminated or minimized by appropriate randomization of genotypes within the experimental plots. In other cases, human genetics again being a prime example, the possibility of G–E association cannot usually be controlled.

18.3 Genetic Analysis of Complex Traits

Equation (3) can be used to separate the effects of genotype and environment on the total phenotypic variance. Two types of data are required:

1. The phenotypic variance of a genetically heterogeneous population, which provides an estimate of $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$
2. The phenotypic variance of a genetically uniform population, which provides an estimate of σ_e^2 because a genetically uniform population has a value of $\sigma_g^2 = 0$

An example of a genetically uniform population is the F_1 generation from a cross between two highly homozygous strains, such as inbred lines. An example of a genetically heterogeneous population is the F_2 generation from the same cross. If the environments of both populations are the same, and if there is no G−E interaction, then the estimates may be combined to deduce the value of σ_g^2.

As an illustration of this approach, we will consider the Mexican cave-dwelling fish *Astyanax fasciatus*, one of about 80 species of cave fishes that are known worldwide. Among other genetic changes accompanying evolution in perpetual darkness, these species have all evolved reduced eyes and body pigment (**Figure 18.9**). To estimate the number of genes contributing to reduced eye size, an inbred line from a cave population was mated with an inbred line from a surface population, and the offspring were reared in the same environment. The variances in eye diameter in the F_1 and F_2 generations were estimated as 0.057 and 0.563, respectively. Written in terms of the components of variance, these are

$$F_2 \text{ variance: } \sigma_p^2 = \sigma_g^2 + \sigma_e^2 = 0.563$$
$$F_1 \text{ variance: } \sigma_e^2 = 0.057$$

The estimate of genotypic variance, σ_g^2, is obtained by subtracting the second equation from the first; that is,

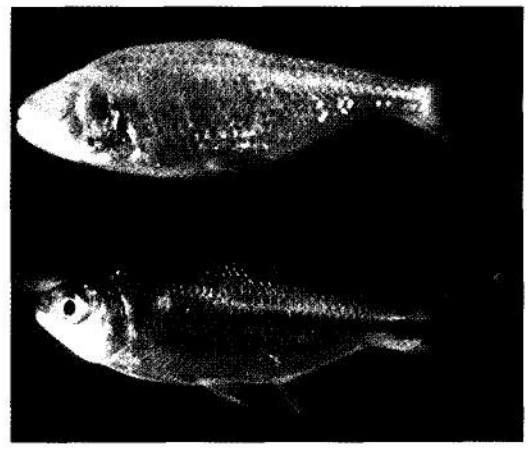


Figure 18.9 Reduced eye size and pigmentation in a cave-dwelling *Astyanax* (above) and a surface-dwelling relative (below). [Courtesy of Richard Borowsky and Horst Wilkens.]

$$(\sigma_g^2 + \sigma_e^2) - \sigma_e^2 = \sigma_g^2$$

and therefore

$$0.563 - 0.057 = 0.506 = \sigma_g^2$$

Hence the estimate of σ_g^2 is 0.506, whereas that of σ_e^2 is 0.057. In this example, the genotypic variance is much greater than the environmental variance, but this is not always the case.

The next section shows what other information can be obtained from an estimate of the genotypic variance.

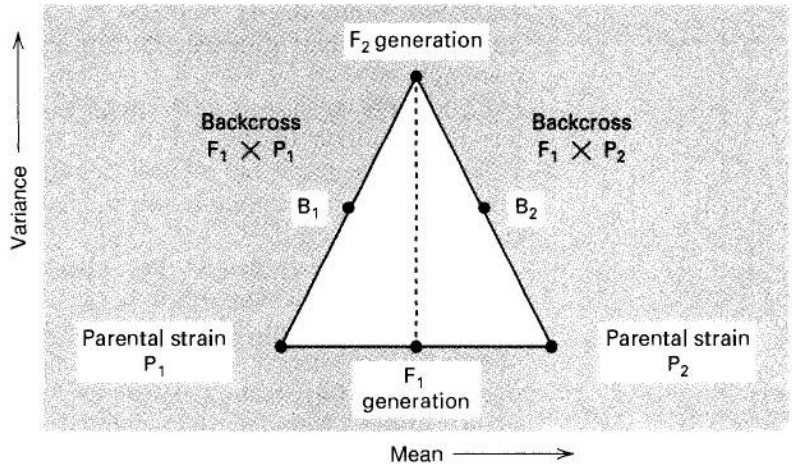


Figure 18.10 Means and variances of parents (P), backcross (B_1 and B_2), and hybrid (F_1 and F_2) progeny of inbred lines for an ideal quantitative trait affected by unlinked and completely additive genes. [After R. Lande. 1981. *Genetics* 99: 541.]

The Number of Genes Affecting Complex Traits

When the number of genes influencing a complex trait is not too large, knowledge of the genotypic variance can be used to estimate the number of genes. The required information consists of the means and variances of two phenotypically divergent strains, as well as the variances of the F_1, F_2, and backcross generations. In ideal cases, the data appear as in **Figure 18.10**, in which P_1 and P_2 represent the divergent parental strains (for example, inbred lines). The points lie on a triangle, with increasing variance according to the increasing genetic heterogeneity (genotypic variance) of the populations. If the F_1 and backcross means lie exactly between their parental means, then these means will lie at the midpoints along the sides of the triangle, as shown in Figure 18.10. This finding implies that the alleles affecting the trait are *additive;* that is, for each gene, the phenotype of the heterozygote is the average of the phenotypes of the corresponding homozygotes. In such a simple situation, it can be shown that the number, n, of genes contributing to the trait is

$$n = \frac{D^2}{8\sigma_g^2} \qquad (4)$$

where D represents the difference between the means of the original parental strains, P_1 and P_2. This equation can be verified by applying it to the ideal case of 3 genes in Figure 18.4. The parental strains are the homozygous genotypes with mean phenotypes of 0 and 6. Consequently, $D = 6 - 0 = 6$. The genotypic variance is given in Figure 18.5 as $\sigma_g^2 = 1.5$. Substituting D and σ_g^2 into Equation (4), we obtain $n = (6)^2/(8 \times 1.5) = 3$, which is correct because there are three independent and equivalent genes that affect the trait.

Applied to actual data, Equation (4) requires several assumptions that are not necessarily valid. In addition to the assumption that all generations are reared in the same environment, the theory also makes the genetic assumptions that (1) the alleles of each gene are additive, (2) the genes contribute equally to the trait, (3) the genes are unlinked, and (4) the original parental strains are homozygous for alternative alleles of each gene. However, when the assumptions are invalid, the calculated n is smaller than the actual number of genes affecting the trait. The estimated number is a minimum because almost any departure from the genetic assumptions leads to a smaller genotypic variance in the F_2 generation and so, for the same value of D, would yield a larger value of n in Equation (4). This is why the estimated n is the *minimum* number of genes that can account for the data. For the cave-dwelling *Astyanax* fish discussed in the preceding section, the parental strains had average phenotypes of 7.05 and 2.10, giving $D = 4.95$. The

estimated value of $\sigma_g^2 = 0.506$, so the minimum number of genes affecting eye diameter is $n = (4.95)^2/(8 \times 0.506) = 6.0$. Therefore, at least six different genes affect the diameter of the eye of the fish. Three of these genes have been identified using the genetic mapping methods discussed in Section 18.7.

The number of genes that affect a quantitative trait is important because it influences the amount by which a population can be genetically improved by selective breeding. With traits determined by a small number of genes, the potential for change in a trait is relatively small, and a population consisting of the best possible genotypes may have a mean value that is only 2 or 3 standard deviations above the mean of the original population. However, traits determined by a large number of genes have a large potential for improvement. For example, after a population of the flour beetle *Tribolium* was selected for increased pupa weight for many generations, the mean value for pupa weight was found to be 17 standard deviations above the mean of the original population. Determination of traits by a large number of genes implies that:

> Selective breeding can create an improved population in which the value of *every* individual greatly exceeds that of the *best* individuals that existed in the original population.

This principle at first seems paradoxical, because in a large enough population, every possible genotype should be created at some low frequency. The resolution of the paradox is that real populations subjected to selective breeding typically consist of a few hundred organisms (at most), and hence in the real world many of the theoretically possible genotypes are never formed because their expected frequencies are much too low. However, as selection takes place, and the allele frequencies change, these genotypes become more common and allow the selection of superior organisms in future generations.

Broad-Sense Heritability

Estimates of the number of genes that determine quantitative traits are frequently unavailable because the necessary experiments are impractical or have not been carried out. Another attribute of quantitative traits, which requires less data to evaluate, makes use of the ratio of the genotypic variance to the total phenotypic variance. This ratio of σ_g^2 to σ_p^2 is called the **broad-sense heritability,** symbolized as H^2, and it measures the importance of genetic variation, relative to environmental variation, in causing variation in the phenotype of a trait of interest. Broad-sense heritability is a ratio of variances, specifically

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2} = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)} \qquad (5)$$

Substitution of the data for eye diameter in *Astyanax*, in which $\sigma_g^2 = 0.506$ and $\sigma_g^2 + \sigma_e^2 = 0.563$, into Equation (5) yields $H^2 = 0.506/0.563 = 0.90$ for the estimate of broad-sense heritability. This value implies that 90 percent of the variation in eye diameter in this population results from differences in genotype among the fish.

Knowledge of heritability is useful in the context of plant and animal breeding, because heritability can be used to predict the magnitude and speed of improvement in the population. The broad-sense heritability defined in Equation (5) is used in predicting the outcome of selection practiced among clones, inbred lines, or varieties. Analogous predictions for randomly bred populations utilize another type of heritability, different from H^2, which we will discuss shortly. Broad-sense heritability measures how much of the total variance in phenotype results from differences in genotype. For this reason, H^2 is often of interest in regard to human quantitative traits.

Twin Studies

In human beings, twins would seem to be ideal subjects for separating genotypic and environmental variance, because **identical twins,** which arise from the splitting of a single fertilized egg, are genetically identical and are often strikingly similar in such traits as facial features and body build. **Fraternal twins,** which arise from two fertilized eggs, have the same genetic relationship as ordinary siblings, and therefore only half of the genes in either twin are identical with those in the other. Theoretically, the variance between members of an identical-twin pair

would be equivalent to σ^2_e, because the twins are genetically identical, whereas the variance between members of a fraternal-twin pair would include not only σ^2_e but also part of the genotypic variance (approximately $\sigma^2_g/2$, because of the identity of half of the genes in fraternal twins). Consequently, both σ^2_g and σ^2_e could be estimated from twin data and combined as in Equation (5) to estimate H^2. **Table 18.2** summarizes estimates of H^2 based on twin studies of several traits.

Unfortunately, twin studies are subject to several important sources of error, most of which increase the similarity of identical twins, so the numbers in Table 18.2 should be considered very approximate and probably too high. Four of the potential sources of error are:

1. Genotype–environment interaction, which increases the variance in fraternal twins but not in identical twins
2. Frequent sharing of embryonic membranes between identical twins, resulting in more similar intrauterine environments than those of fraternal twins
3. Greater similarity in the treatment of identical twins by parents, teachers, and peers, resulting in a decreased environmental variance in identical twins
4. Different sexes in half of the pairs of fraternal twins, in contrast with the same sex of identical twins

These pitfalls and others imply that data from human twin studies should be interpreted with caution and reservation.

Table 18.2 Broad-sense heritability, in percent, based on twin studies

Trait	Heritability (H^2)
Longevity	29
Height	85
Weight	63
Amino acid excretion	72
Serum lipid levels	44
Maximum blood lactate	34
Maximum heart rate	84
Verbal ability	63
Numerical ability	76
Memory	47
Sociability index	66
Masculinity index	12
Temperament index	58

18.4 Artificial Selection

Artificial selection is the breeders' practice of choosing a select group of organisms from a population to become the parents of the next generation. When artificial selection is carried out either by choosing the best organisms in a species that reproduces asexually or by choosing the best subpopulation among a series of inbred subpopulations, the broad-sense heritability is used to predict how rapidly progress can be achieved. Broad-sense heritability is important in this context, because with clones or inbred lines, superior genotypes can be perpetuated without disruption of favorable gene combinations by segregation and recombination. An example is the selection of superior varieties of plants that are propagated asexually by means of cuttings or grafts. Because there is no sexual reproduction, each plant has exactly the same genotype as its parent.

In sexually reproducing populations that are genetically heterogeneous, broad-sense heritability is not relevant in predicting progress resulting from artificial selection, because superior genotypes must necessarily be broken up by the processes of segregation and recombination. For example, if the best genotype is heterozygous for each of two unlinked loci, *Aa Bb*, then because of segregation and independent assortment, among the progeny of a cross between parents with the best genotypes—*Aa Bb* × *Aa Bb*—only 1/4 will have the same favorable *Aa Bb* genotype as the parents. The rest of the progeny will be genetically inferior to the parents. For this reason, to the extent that high genetic merit may depend on particular combinations of alleles, each generation of artificial selection results in a slight setback in that the offspring of superior parents are generally not

quite so good as the parents themselves. Progress under selection can still be predicted, but the prediction must make use of another type of heritability, the narrow-sense heritability, which is discussed in the next section.

Narrow-Sense Heritability

Figure 18.11 illustrates a typical form of artificial selection and its result. The organism is *Nicotiana longiflora* (tobacco), and the trait is the length of the corolla tube (the corolla is a collective term for all the petals of a flower). Figure 18.11A shows the distribution of phenotypes in the parental generation, and Figure 18.11B shows the distribution of phenotypes in the offspring generation. The parental generation is the population from which the parents were chosen for breeding. The type of selection is called **individual selection,** because each member of the population to be selected is evaluated according to its own individual phenotype. The selection is practiced by choosing some arbitrary level of phenotype, called the **truncation point,** that determines which individuals will be saved for breeding purposes. All individuals with a phenotype above the threshold are randomly mated among themselves to produce the next generation.

Phenotypic Change with Individual Selection: A Prediction Equation

In evaluating progress through individual selection, three distinct phenotypic means are important. These means are symbolized in Figure 18.11 as M, M^* and M' and are defined as follows:

1. M is the mean phenotype of the entire population in the parental generation, including both the selected and the nonselected individuals.
2. M^* is the mean phenotype among those individuals selected as parents (those with a phenotype above the truncation point).
3. M' is the mean phenotype among the progeny of selected parents.

The relationship among these three means is given by

$$M' = M + h^2(M^* - M) \qquad (6)$$

where the symbol h^2 is the **narrow-sense heritability** of the trait in question.

Later in this chapter, a method for estimating narrow-sense heritability from the similarity in phenotype among relatives will be explained. In Figure 18.11, h^2 is the only unknown quantity, so it can be estimated from the data themselves. Rearranging Equation (6) and substituting the values for the means from Figure 18.11, we find that

$$h^2 = \frac{M' - M}{M^* - M} = \frac{77 - 70}{81 - 70}$$
$$= 0.636$$

In a manner analogous to the way in which total phenotypic variance can be split into the sum of the genotypic variance and the environmental variance (Equation 3), the genotypic variance can be split into parts resulting from the additive effects of alleles, dominance effects, and effects of interaction between alleles of different genes. The difference between the broad-sense heritability, H^2, and the narrow-sense heritability, h^2, is that H^2 includes all of these genetic contributions to variation, whereas h^2 includes only the additive effects of alleles. From the standpoint of animal or plant improvement, h^2 is the heritability of interest, because:

> The narrow-sense heritability, h^2, is the proportion of the variance in phenotype that can be used to predict changes in population mean with individual selection according to Equation (6).

The distinction between the broad-sense heritability and the narrow-sense heritability can be appreciated intuitively by considering a population in which there is a rare recessive gene. In such a case, most homozygous recessive genotypes come from matings between heterozygous carriers. Some matings have more than one affected offspring. Hence, affected siblings can resemble each other more than they resemble their parents. For example, if a is a recessive allele, the mating $Aa \times Aa$, in which both parents show the dominant phenotype, may yield two offspring that are aa, which both show the recessive phenotype; in this case, the offspring are more similar to each other than to either of the parents. It is the dominance of the wildtype allele that causes this paradox, contributing to the broad-sense heritability of the trait but not to the narrow-sense heritability.

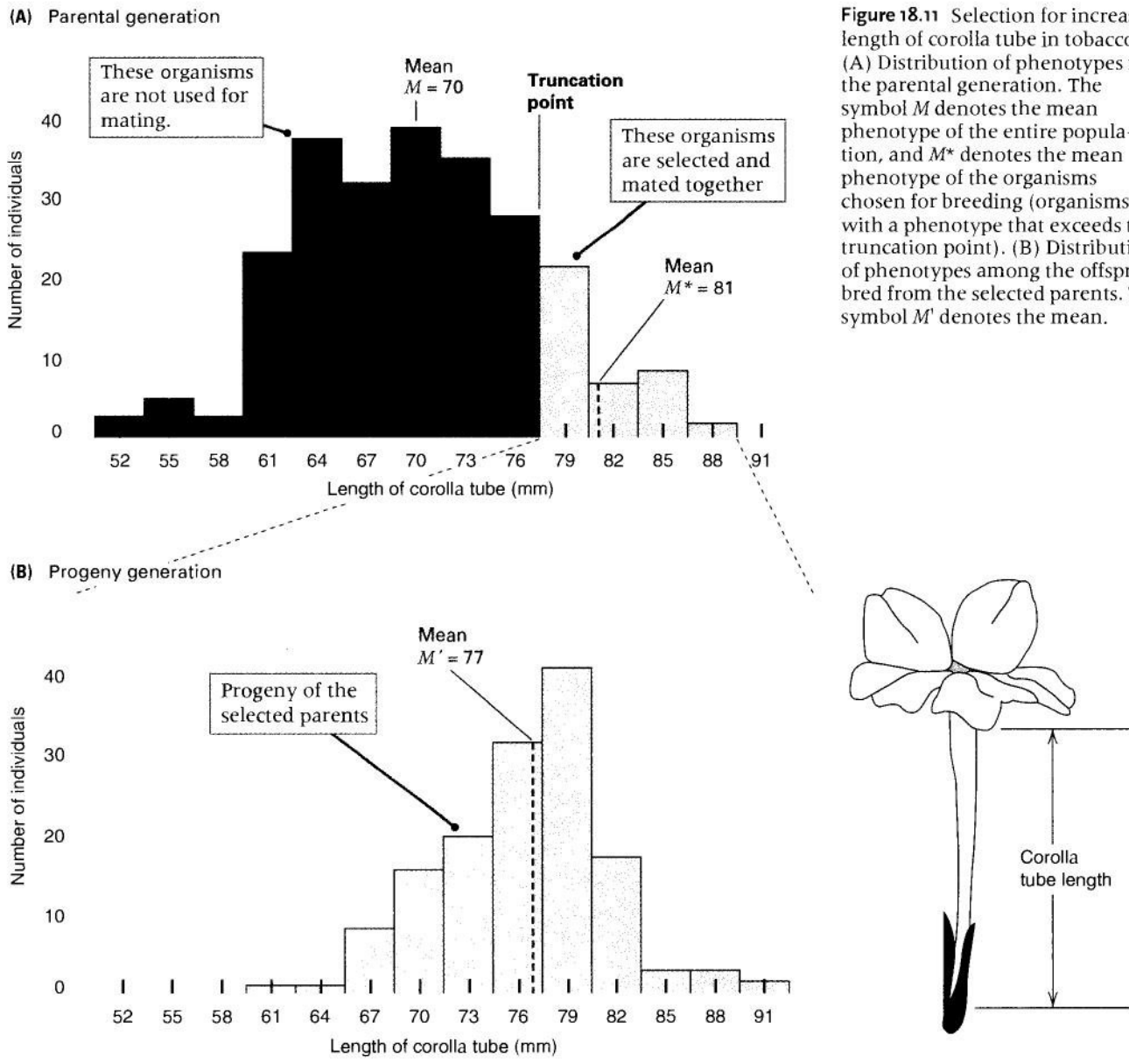


Figure 18.11 Selection for increased length of corolla tube in tobacco. (A) Distribution of phenotypes in the parental generation. The symbol M denotes the mean phenotype of the entire population, and M^* denotes the mean phenotype of the organisms chosen for breeding (organisms with a phenotype that exceeds the truncation point). (B) Distribution of phenotypes among the offspring bred from the selected parents. The symbol M' denotes the mean.

The narrow-sense heritability includes only those genetic effects that contribute to the resemblance between parents and their offspring, because narrow-sense heritability measures how similar offspring are to their parents.

In general, the narrow-sense heritability of a trait is smaller than the broad-sense heritability. For example, in the parental generation in Figure 18.11, the broad-sense heritability of corolla tube length is $H^2 = 0.82$. The two types of heritability are equal only when the alleles affecting the trait are additive in their effects; with additive effects, each heterozygous genotype shows a phenotype that is exactly intermediate between the phenotypes of the respective homozygous genotypes.

Equation (6) is of fundamental importance in quantitative genetics because of its predictive value. This can be seen in the following example. The selection in Figure 18.11 was carried out for several generations. After two generations, the mean of the population was 83, and parents having a mean of 90 were selected. By use of the

estimate $h^2 = 0.636$, the mean in the next generation can be predicted. The information provided is that $M = 83$ and $M^* = 90$. Therefore, Equation (6) implies that the predicted mean is

$$M' = 83 + (0.636)(90 - 83) = 87.4$$

This value is in good agreement with the observed value of 87.9.

Long-Term Artificial Selection

Artificial selection is analogous to natural selection in that both types of selection cause an increase in the frequency of alleles that improve the selected trait (or traits). The principles of natural selection discussed in Chapter 17 also apply to artificial selection. For example, artificial selection is most effective in changing the frequency of alleles that are in an intermediate range of frequency ($0.2 < p < 0.8$). Selection is less effective for alleles with frequencies outside this range and is least effective for rare recessive alleles. For complex traits, including fitness, the total selection is shared among all the genes that influence the trait, and the selection coefficient affecting each allele is determined by (1) the magnitude of the effect of the allele, (2) the frequency of the allele, (3) the total number of genes affecting the trait, (4) the narrow-sense heritability of the trait, and (5) the proportion of the population that is selected for breeding.

The value of heritability is determined both by the magnitude of effects and by the frequency of alleles. If all favorable alleles were fixed ($p = 1$) or lost ($p = 0$), then the heritability of the trait would be 0. Therefore, the heritability of a quantitative trait is expected to decrease over many generations of artificial selection as a result of favorable alleles becoming nearly fixed. For example, ten generations of selection for less fat in a population of Duroc pigs decreased the heritability of fatness from 73 to 30 percent because of changes in allele frequency that resulted from the selection.

Population improvement by means of artificial selection cannot continue indefinitely. A population may respond to selection until its mean is many standard deviations different from the mean of the original population, but eventually the population reaches a **selection limit** at which successive generations show no further improvement. Progress may stop because all alleles affecting the trait are either fixed or lost, and so the narrow-sense heritability of the trait becomes 0. However, it is more common for a selection limit to be reached because natural selection counteracts artificial selection. Many genes that respond to artificial selection as a result of their favorable effect on a selected trait also have indirect harmful effects on fitness. For example, selection for increased size of eggs in poultry results in a decrease in the number of eggs, and selection for extreme body size (large or small) in most animals results in a decrease in fertility. When one trait (for example, number of eggs) changes in the course of selection for a different trait (for example, size of eggs), then the unselected trait is said to have undergone a **correlated response** to selection. Correlated response of fitness is typical in long-term artificial selection. Each increment of progress in the selected trait is partially offset by a decrease in fitness because of a correlated response; eventually, artificial selection for the trait of interest is exactly balanced by natural selection against the trait. Thus a selection limit is reached, and no further progress is possible without a change in the strategy of selection.

Inbreeding Depression and Heterosis

Inbreeding can have harmful effects on economically important traits such as yield of grain and egg production. This decline in performance is called **inbreeding depression,** and it results principally from rare harmful recessive alleles becoming homozygous because of inbreeding. The degree of inbreeding is measured by the inbreeding coefficient F discussed in Chapter 17. **Figure 18.12** is an example of inbreeding depression of yield in corn, in which the yield decreases linearly as the inbreeding coefficient increases.

Most highly inbred strains suffer from many genetic defects, as might be expected from the uncovering of deleterious recessive alleles. One would also expect that if two different inbred strains were crossed, then the F_1 would show improved features, because a harmful recessive allele inherited

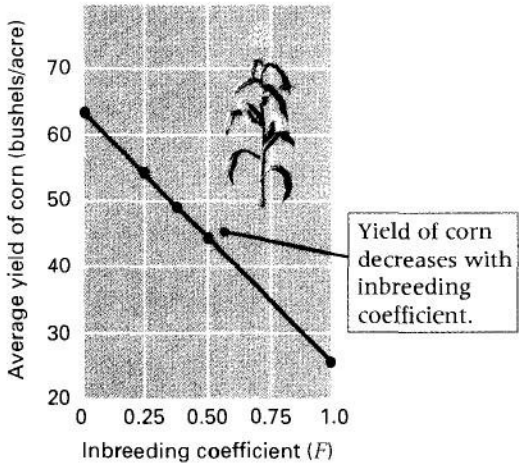


Figure 18.12 Inbreeding depression for yield in corn. [Data from N. Neal. 1935. *J. Amer. Soc. Agronomy* 27: 666]

from one parent would be likely to be covered up by a normal dominant allele from the other parent. In many organisms, the F_1 generation of a cross between inbred lines is superior to the parental lines, and the phenomenon is called **hybrid vigor** or **heterosis.** The phenomenon, which is widely used in the production of corn and other agricultural products, yields genetically identical hybrid plants with traits that are sometimes more favorable than those of the ancestral plants from which the inbreds were derived. The features that most commonly distinguish hybrid plants from their inbred parents are their rapid growth, larger size, and greater yield. Furthermore, the F_1 plants have a fairly uniform phenotype (because $\sigma_g^2 = 0$). Genetically heterogeneous crops with high yields or certain other desirable features can also be produced by traditional plant-breeding programs, but growers often prefer hybrid plants because of their uniformity. For example, uniform height and time of maturity facilitate machine harvesting, and plants that all bear fruit at the same time accommodate picking and shipping schedules.

Hybrid varieties of corn are used almost exclusively in the United States for commercial crops. A farmer cannot plant the seeds from his crop because the F_2 generation consists of a variety of genotypes, most of which do not show hybrid vigor. The production of hybrid seeds is a major industry in corn-growing regions of the United States.

18.5 Correlation Between Relatives

Studies of complex inheritance rely extensively on similarity among relatives to assess the importance of genetic factors. Particularly in the study of complex behavioral traits in human beings, genetic interpretation of familial resemblance is not always straightforward because of the possibility of nongenetic, but nevertheless familial, sources of resemblance. However, the situation is usually less complex in plant and animal breeding, because genotypes and environments are under experimental control.

Covariance and Correlation

Genetic data about families are frequently pairs of numbers: pairs of parents, pairs of twins, or pairs consisting of a single parent and a single offspring. An important issue in quantitative genetics is the degree to which the numbers in each pair are associated. The usual way to measure this association is to calculate a statistical quantity called the *correlation coefficient* between the members of each pair.

The correlation coefficient among relatives is based on the covariance in phenotype among them. Much as the variance describes the tendency of a set of measurements to vary (Equation 2), the covariance describes the tendency of pairs of numbers to vary together (co-vary). Calculation of the covariance is similar to calculation of the variance in Equation (2) except that the squared deviation term $(x_i - \bar{x})^2$ is replaced with the product of the deviations of the pairs of measurements from their respective means—that is, $(x_i - \bar{x})(y_i - \bar{y})$.

For example, $(x_i - \bar{x})$ could be the deviation of a particular father's height from the overall "father mean," and $(y_i - \bar{y})$ could be the deviation of his son's height from the overall "son mean." In symbols, let f_i be the number of pairs of relatives with phenotypic measurements x_i and y_i. Then the estimated **covariance** (*Cov*) of the trait among the relatives is

$$Cov(x,y) = \frac{\Sigma f_i (x_i - \bar{x})(y_i - \bar{y})}{N - 1} \qquad (7)$$

where N is the total number of pairs of relatives studied.

The **correlation coefficient** (r) of the trait between the relatives is calculated from the covariance as follows:

$$r = \frac{Cov(x, y)}{s_x s_y} \qquad (8)$$

where s_x and s_y are the standard deviations of the measurements, estimated from Equation (2). The correlation coefficient can range from -1.0 to $+1.0$. A value of $+1.0$ means perfect association. When $r = 0$, x and y are not associated.

The Geometrical Meaning of a Correlation

Figure 18.13 shows a geometrical interpretation of the correlation coefficient. Just as the distribution of a single variable can be built up, step by step, by placing small squares corresponding to each individual along the x-axis, so too can the joint distribution of two variables (for example, height of fathers and height of sons) be built up, step by step, by placing small cubes on the x-y plane at a position determined by each pair of measurements. As these cubes pile up, the joint distribution assumes the form of an inverted bowl, cross sections of which are themselves normal distributions. When there is no correlation between the variables, the surface is completely symmetrical (Figure 18.13A), which means that the distribution of y is completely independent of that of x. However, when the distributions are correlated, the distribution of y varies according to the particular value of x, and the joint distribution becomes asymmetrical with a ridge built up in the x-y plane (Figure 18.13B).

There is usually not enough data to form a full three-dimensional joint distribution of x and y as depicted in Figure 18.13. Instead, the data are in the form of individ-

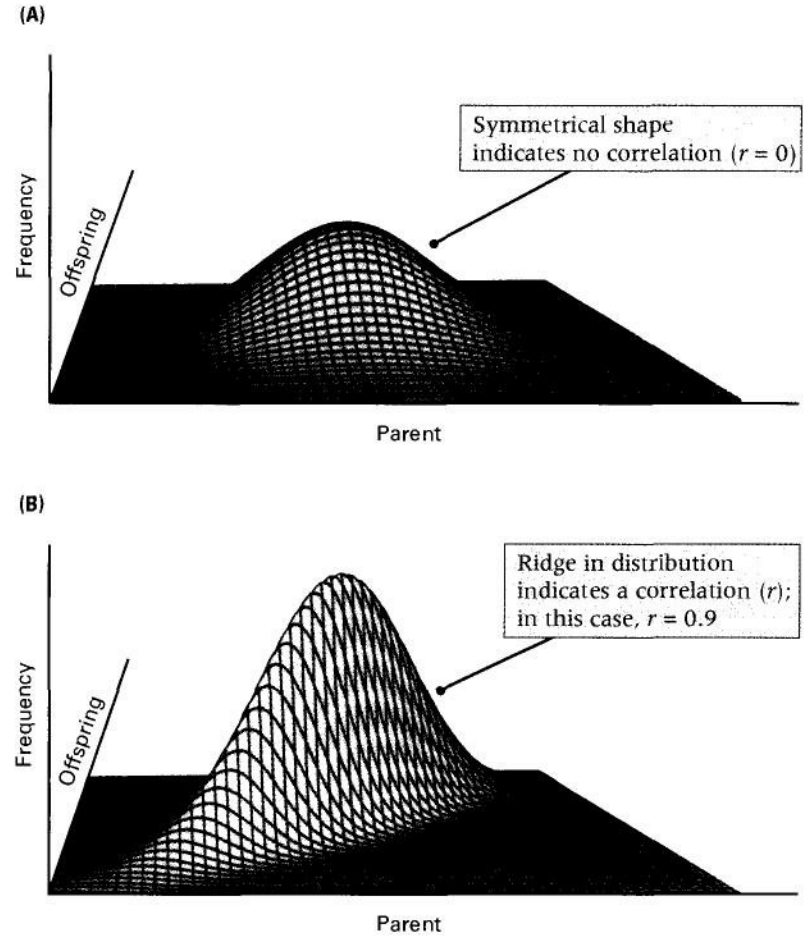


Figure 18.13 Distribution of a quantitative trait in parents and offspring when there is no correlation (A) or a high correlation (B) between them. [From R. R. Sokal and F. J. Rohlf. 1969. *Biometry.* San Francisco: Freeman.]

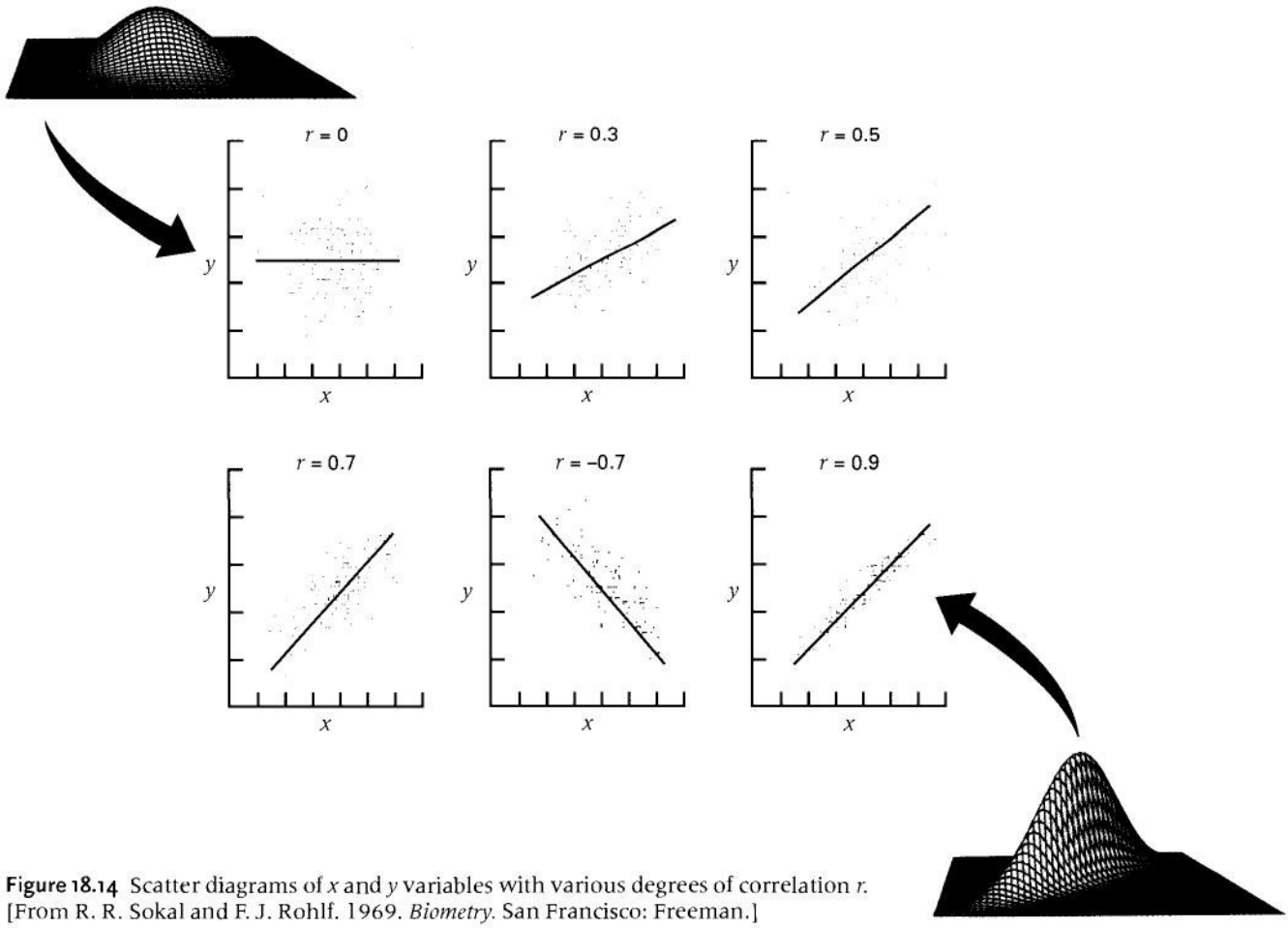


Figure 18.14 Scatter diagrams of x and y variables with various degrees of correlation r. [From R. R. Sokal and F. J. Rohlf. 1969. *Biometry.* San Francisco: Freeman.]

ual points, sparse enough to be plotted on a scatter diagram like those in **Figure 18.14**. In these diagrams, the clustering of points around each line corresponds to the ridge of points in Figure 18.13. The higher the ridge, the more clustered the points are around the line.

Estimation of Narrow-Sense Heritability

Covariance and correlation are important in quantitative genetics, because the correlation coefficient of a trait between individuals with various degrees of genetic relationship is related fairly simply to the narrow-sense or broad-sense heritability, as shown in **Table 18.3**. The table gives the theoretical values of the correlation coefficient for various pairs of relatives; h^2 represents the narrow-sense heritability and H^2 the broad-sense heritability. For parent–offspring, half-sibling, or first-cousin pairs, narrow-sense heritability can be estimated directly by multiplication. Specifically, h^2 can be estimated as twice the parent–offspring correlation, four times the half-sibling correlation, or eight times the first-cousin correlation.

Table 18.3 Theoretical correlation coefficient in phenotype between relatives

Degree of relationship	Correlation coefficient*
Offspring and one parent	$h^2/2$
Offspring and average of parents	$h^2/2$
Half siblings	$h^2/4$
First cousins	$h^2/8$
Monozygotic twins	H^2
Full siblings	$\sim H^2/2$

*Contributions from interactions among alleles of different genes have been ignored. For this and other reasons, H^2 correlations are approximate

With full siblings and identical twins, the correlation coefficient is related to the broad-sense heritability, H^2, because phenotypic resemblance depends not only on additive effects but also on dominance. In these relatives, dominance contributes to resemblance because the relatives can share *both* of their alleles as a result of their common ancestry, whereas parents and offspring, half siblings, and first cousins can share at most a single allele of any gene because of common ancestry. Therefore, to the extent that phenotype depends on dominance effects, full siblings can resemble each other more than they resemble their parents.

The potentially greater resemblance between siblings than between parents and offspring may again be understood by considering an autosomal recessive trait caused by a rare recessive allele. In a randomly mating population, the probability that an offspring will be affected, if the mother is affected, is q (the allele frequency), which corresponds to the random-mating probability that the sperm giving rise to the offspring carries the recessive allele. In contrast, the probability that both of two siblings will be affected, if one of them is affected, is 1/4, because most of the matings that produce affected individuals will be between heterozygous parents. When the trait is rare, the parent–offspring resemblance will be very small (as q approaches 0), whereas the sibling–sibling resemblance will remain large. This discrepancy is entirely a result of dominance, and it arises only because, at any particular locus, full siblings may have the same genotype.

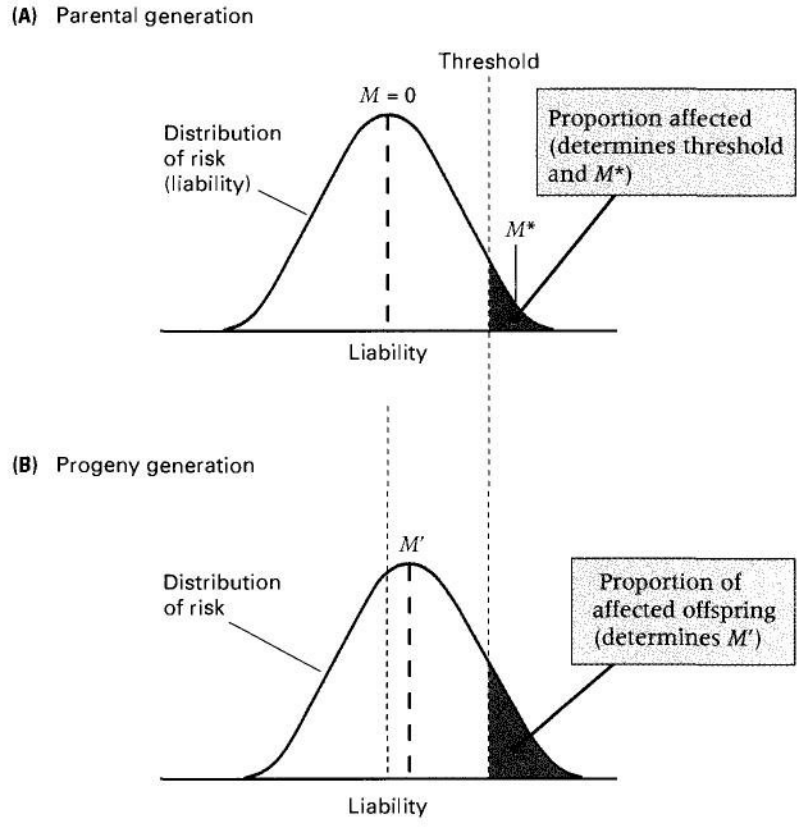


Figure 18.15 Interpretation of a threshold trait in terms of an underlying continuous distribution of liability. Individuals with a liability greater than the threshold are affected. M (arbitrarily set equal to 0), M^*, and M' are the mean liabilities of the parental generation, of affected individuals, and of the offspring of affected individuals, respectively.

18.6 Heritabilities of Threshold Traits

Application of the concepts of individual selection can be used to understand the meaning of heritability in the context of threshold traits. The analogy is illustrated in **Figure 18.15**. Threshold traits depend on an underlying quantitative trait, called the **liability,** which refers to the degree of genetic risk or predisposition to the trait. Only those individuals with a liability greater than a certain threshold develop the trait. Figure 18.15A shows the distribution of liability in a population, where liability is measured on an arbitrary scale so that the mean liability (M) equals 0 and the variance in liability equals 1. Affected individuals are in the right-hand area of the distribution. Using various characteristics of the normal distribution, we can use the observed proportion of affected individuals to calculate the position of the threshold and the value of M^*, which in this context is the mean liability of affected individuals.

The distribution of liability among offspring of affected individuals is shown in Figure 18.15B. It is shifted slightly to the right, which implies that, relative to the population as a whole, a greater proportion of individuals with affected parents have liabilities above the threshold and so will be affected. The observed proportion of affected individuals in the offspring generation can be used to calculate M'. The

important point is that M^* and M' can be obtained from the observed proportions corresponding to the shaded red areas in Figure 18.15. With $M = 0$, Equation (6) can be used to calculate the narrow-sense heritability of liability of the trait in question.

Examples of narrow-sense heritability of the liability to important congenital abnormalities are given in Table 18.4. How these heritabilities translate into risk is illustrated in Figure 18.16, along with observed data for the most common congenital abnormalities in Caucasians. The theoretical curves correspond to the relationships among the incidence of the trait in the general population, the type of inheritance, and the risk among first-degree relatives of affected individuals. (First-degree relatives of an affected individual include parents,

Table 18.4 Narrow-sense heritability of liability for congenital abnormalities

Trait	Description	Population frequency (%)	Narrow-sense heritability (%)
Cleft lip	Upper lip not completely formed	0.08	65
Spina bifida	Exposed spinal cord	0.11	65
Club foot	Abnormally formed foot	0.08	80
Pyloric stenosis	Obstructed stomach	0.40	85
Dislocation of hip	Hip joint mispositioned	0.06	95
Hydrocephalus	Fluid accumulation around brain	0.12	40
Celiac disease	Inability to digest fats, starches, and sugars	0.04	80
Hypospadias	Abnormally formed penis	0.30	75
Atrial septal defect	Hole between upper chambers of heart	0.07	70
Patent ductus arteriosus	Hole between aorta and pulmonary artery; blood bypasses lungs	0.08	75

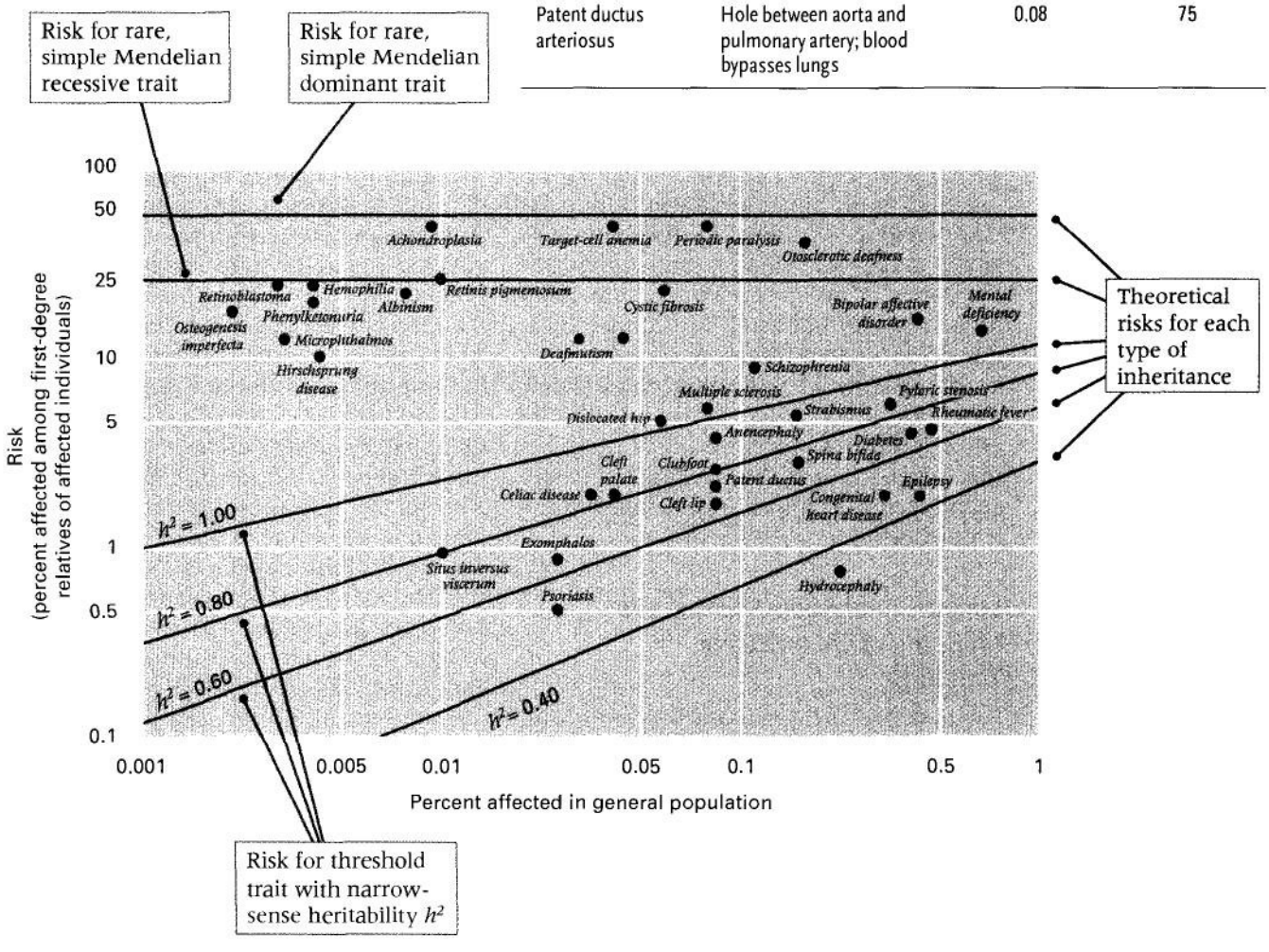


Figure 18.16 Recurrence risks of common abnormalities. Diagonal lines are the theoretical risks for threshold traits with the indicated values of the narrow-sense heritability of liability. Horizontal lines are the theoretical risks for simple dominant or recessive traits.

offspring, and siblings.) Simple Mendelian inheritance produces the highest risks, but these traits, as a group, tend to be rare. The most common traits are threshold traits, and with a few exceptions, their risks tend to be moderate or low, corresponding to the heritabilities indicated.

18.7 Identification of Genes Affecting Complex Traits

A gene that affects a complex trait is a **quantitative-trait locus (QTL)**. The term derives from the fact that many complex traits of medical or commercial importance are quantitative traits. QTLs cannot usually be identified in pedigrees, because their individual effects are obscured by the segregation of other genes and by environmental variation. Even so, genes that affect the traits can be localized if they are genetically linked with polymorphic DNA markers, such as those discussed in Chapter 2, because the effects of the genotype affecting the complex trait will be correlated with the genotype of a linked genetic marker. Locating QTLs in the genome is important to the manipulation of genes in breeding programs and to the cloning and studying of the genes in order to identify their functions. Another approach to the identification of genes that influence complex traits is through the identification of candidate genes that are known to be involved in physiological or developmental processes that affect the trait. These complementary approaches to complex inheritance are examined below.

Linkage Analysis in the Genetic Mapping of Quantitative Trait Loci

Polymorphic DNA markers are abundant, are distributed throughout the genome, and often have multiple codominant alleles; thus they are ideally suited for linkage studies of quantitative traits. In QTL studies, as many widely scattered markers as possible are monitored, along with the complex trait, in successive generations of a genetically heterogeneous population. Statistical studies are then carried out to identify which DNA markers are associated with the complex trait in the sense that their presence in a genotype is consistently accompanied by the phenotype of interest. These DNA markers identify regions of the genome that contain one or more QTLs that have important effects on the trait, and the DNA markers can be used to trace the segregation of the important regions in breeding programs and even as starting points for cloning genes with particularly large effects.

An example of genetic mapping of QTLs for several quantitative traits in tomatoes is illustrated in **Figure 18.17**. More than 300 highly polymorphic genetic markers have been mapped in the tomato genome, with an average spacing between markers of 5 map units. The chromosome maps in Figure 18.17 show a subset of 67 markers that were segregating in crosses between the domestic tomato and a wild South American relative. The average spacing between the markers is 20 map units. Backcross progeny of the cross $F_1 \times$ domestic tomato were tested for the genetic markers, and the fruits of the backcross progeny were assayed for three quantitative traits—fruit weight, content of soluble solids, and acidity. Statistical analysis of the data was carried out in order to detect marker alleles that were associated with phenotypic differences in any of the traits; a significant association indicates genetic linkage between the marker gene and one or more QTLs affecting the trait. A total of six QTLs affecting fruit weight were detected (green bars), as well as four QTLs affecting soluble solids (blue bars) and five QTLs affecting acidity (dark red bars). Although additional QTLs with smaller effects undoubtedly remained undetected in these types of experiments, the effects of the mapped QTLs are substantial: The mapped QTLs account for 58 percent of the total phenotypic variance in fruit weight, 44 percent of the phenotypic variance in soluble solids, and 48 percent of the phenotypic variance in acidity. The genetic markers linked to the QTLs with substantial effects make it possible to trace the transmission of the QTLs in pedigrees and to manipulate the QTLs in breeding programs by following the transmission of the linked marker genes. Figure 18.17 also indicates a number of chromosomal regions containing QTLs for two or more of the traits—for example, the QTL regions on chromosomes 6 and 7, which affect all three traits. Be-

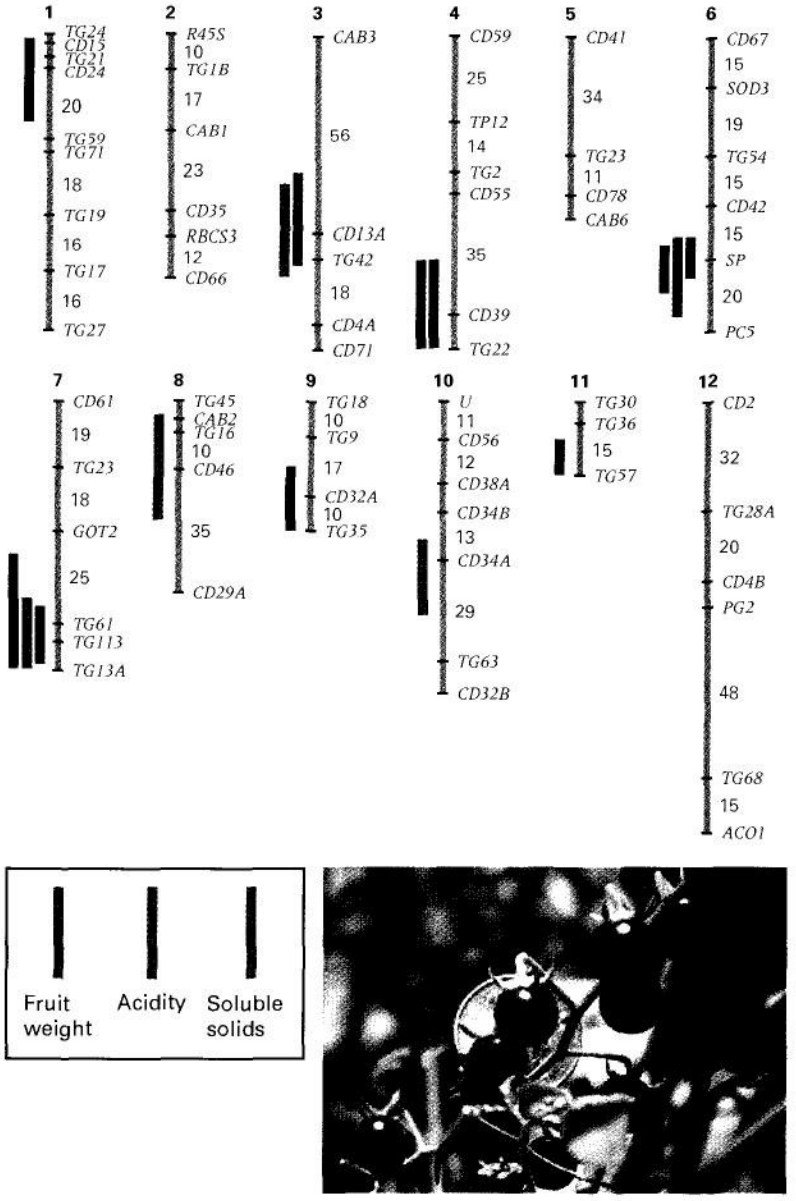


Figure 18.17 Location of QTLs for several quantitative traits in the tomato genome. The genetic markers are shown for each of the 12 chromosomes. The numbers in red are distances in map units between adjacent markers, but only map distances of 10 or greater are indicated. The regions in which the QTLs are located are indicated by the bars: green bars, QTLs for fruit weight; blue bars, QTLs for content of soluble solids; and dark red bars, QTLs for acidity (pH). The data are from crosses between the domestic tomato (*Lycopersicon esculentum*) and a wild South American relative with small, green fruit (*Lycopersicon chmielewskii*). The photograph is of fruits of wild tomato. Note the small size in comparison with the coin (a U.S. quarter). The F_1 generation was backcrossed with the domestic tomato, and fruits from the progeny were assayed for the genetic markers and each of the quantitative traits. [Data from A. H. Paterson et al. 1988. *Nature* 335: 721. Photograph courtesy of Steven D. Tanksley.]

cause the locations of the QTLs can be specified only to within 20–30 map units, it is unclear whether the coincidences result from (1) pleiotropy, in which a single gene affects several traits simultaneously, or (2) independent effects of multiple, tightly linked genes. However, the locations of QTLs for different traits coincide frequently enough so that, in many cases, pleiotropy is likely to be the explanation.

CoNection

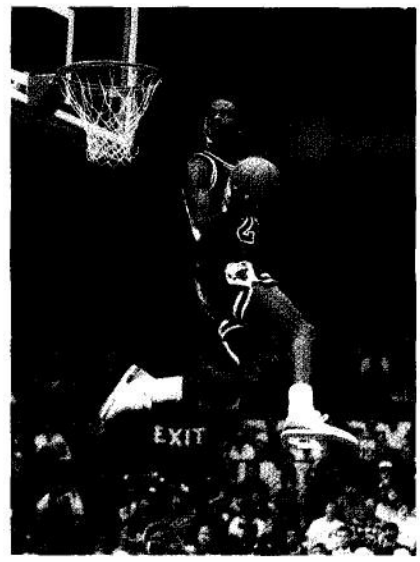
Win, Place, or Show?

Hugh E. Montgomery and 18 other investigators 1998
University College, London,
and 6 other research institutions
Human Gene for Physical Performance

A professional basketball executive once said of Chicago Bulls superstar Michael Jordan, "Gene for gene, DNA for DNA, nobody can replace him." The implication that Jordan's phenomenal success was not due exclusively to his work ethic, practice, commitment, consistency, and character is interesting. But nobody doubts that Jordan's success in basketball was associated in part with his being tall. And height is a multifactorial trait partly under genetic control. About 50 percent of the variation in height among people is due to genetic differences. But might invisible, metabolic differences also influence physical performance? Some people find this possibility unsettling, because it potentially raises all sorts of issues about the value of effort and practice, the fairness of competition, genetic testing for entry into elite physical training programs, and so forth. The extent to which genotype influences physical performance is not known, but some genetic differences may be important. In this paper, a polymorphism in the angiotensinconverting enzyme ACE is reported to have a major effect on physical endurance. It is not difficult to see how ACE could play a role. The enzyme degrades a class of vasodilator proteins and also converts angiotensin I into a vasoconstrictor. Note, however, that the authors are cautious in their conclusions and are careful not to speculate that the gene for ACE is a QTL (quantitative-trait locus) affecting physical performance. It is still too early to reach firm conclusions.

A specific genetic factor that strongly influences human physical performance has not so far been reported, but here we show that a polymorphism in the gene encoding angiotensin-converting enzyme does just that. An "insertion" allele of the gene is associated with elite endurance performance among high-altitude mountaineers. Also, after physical training, repetitive weight-lifting is improved eleven-fold in individuals homozygous for the "insertion" allele compared with those homozygous for the "deletion" allele. The endocrine reninangiotensin system is important in controlling the circulation. . . . A polymorphism in the human *ACE* gene for angiotensin-converting enzyme has been described in which the deletion (*D*) rather than insertion (*I*) allele is associated with higher activity by tissue ACE. . . . Our initial studies suggested that the *I* allele was associated with improved endurance performance. . . . High-altitude mountaineers perform extreme-endurance exercise. [The *ACE* genotypes of 25] elite

We show that a polymorphism in the gene encoding angiotensin-converting enzyme [influences human physical performance].



Michael "Air" Jordan: Possibly the greatest basketball player of all time. [Corbis/Bettmann-UPI.]

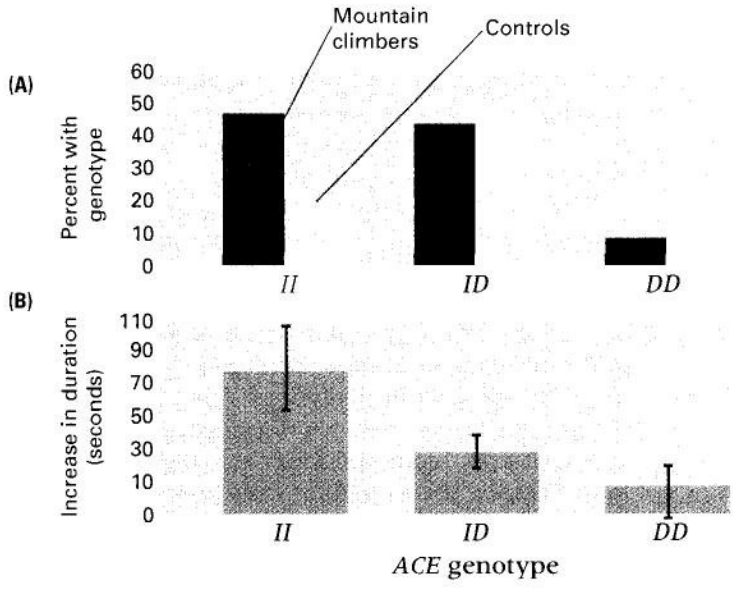
Candidate Genes for Complex Traits

A **candidate gene** for a complex trait is a gene for which there is some *a priori* basis for suspecting that the gene may affect the trait. In human behavioral genetics, for example, if a pharmacological agent is known to affect a personality trait, and the molecular target of the drug is known, then the gene that codes for the target molecule and any genes whose products interact with the drug or with the target molecule are all candidate genes for affecting the personality trait.

One example of the use of candidate genes in the study of human behavioral genetics is in the identification of a

unrelated male British mountaineers . . . were compared with those of 1906 British males. . . . Both the genotype distribution and allele frequency differed significantly between climbers and controls (Figure A).

In a second study, *ACE* genotype was determined in 123 Caucasian males recruited into the United Kingdom army consecutively. . . . The maximum duration (in seconds) for which they could perform repetitive elbow flexion while holding a 15-kg barbell was assessed both before and after a 10-week physical training period. . . . Duration of exercise improved significantly for those of *II* and *ID* genotypes, but not for those of *DD* genotype (Figure B). . . . Increased performance is likely to be due to an improvement in the endurance characteristics of the tested muscles. . . . Further work will be needed to determine whether this correlation holds beyond the limited group studied here.



Source: Nature 393: 221

naturally occurring genetic polymorphism that affects neuroticism and, in particular, the anxiety component of neuroticism. The neurotransmitter substance serotonin (5-hydroxytryptamine) is known to influence a variety of psychiatric conditions, such as anxiety and depression. Among the important components in serotonin action is the serotonin transporter protein SLC6A4 (SLC stands for *solute carrier*), encoded in a gene in chromosomal region *17q11.1–q12*. Neurons that release serotonin to stimulate other neurons also take it up again through the SLC6A4 transporter. The uptake terminates the stimulation and also recycles the molecule for later use.

The serotonin transporter became a strong candidate gene for anxiety-related personality traits when it was discovered that the transporter is the target of a class of antidepressants known as selective serotonin reuptake inhibitors. The widely prescribed antidepressant Prozac is an example of such a drug.

Motivated by the strong suggestion that the serotonin transporter might be involved in anxiety-related traits, researchers looked for evidence of genetic polymorphisms affecting the SLC6A4 transporter gene in human populations. Such a polymorphism was found in the promoter region of the transporter. About 1 kb upstream from the transcription start site is a series of 16 tandem repeats of a nearly identical DNA sequence of about 15 base pairs. This is the *L* (long) allele, which has an allele frequency

of 57 percent among Caucasians. There is also an *S* (*short*) allele, in which three of the repeated sequences are not present. The *S* allele frequency is 43 percent. The genotypes *LL*, *LS*, and *SS* are found in the Hardy–Weinberg proportions of 32 percent, 49 percent, and 19 percent. Further studies revealed that the polymorphism does have a physiological effect. In cells grown in culture, *LL* cells had approximately 50 percent more mRNA for the transporter than *LS* or *SS* cells, and *LL* cells had approximately 35 percent more membrane-bound transporter protein than *LS* or *SS* cells.

On the basis of these results, the researchers predicted that the serotonin transporter polymorphism would be found to be associated with anxiety-related personality traits. The predicted result was observed in a study of 505 people who were genotyped for the transporter polymorphism and whose personality traits were classified by means of a self-report questionnaire. Significant associations were found between the transporter genotype and the overall neuroticism score, and the highest correlations were with the anxiety-related traits "tension" and "harm avoidance." A comparison between the genotypes with respect to neuroticism score is shown in **Figure 18.18**, where the green bars represent scores for *LL* genotypes and the beige bars those for *LS* and *SS* genotypes (symbolized *S*−).

Although there is a significant association, it is in the opposite direction from what one might expect from the activity of selective serotonin uptake inhibitors. The effectiveness of these inhibitors might suggest that because *S*− genotypes have reduced serotonin transporter, they should have reduced anxiety, but the observed result is the other way around. There is also a great deal of overlap in the *LL* and *S*− distributions. The overlap means that the SLC6A4 transporter polymorphism accounts for a relatively small proportion of the total variance in neuroticism score in the whole population. In quantitative terms, the transporter polymorphism accounts for 3–4 percent of the total variance in neuroticism score and for 7–9 percent of the genotypic variance in neuroticism score. If all genes affecting neuroticism had the same magnitude of effect as the serotonin transporter polymorphism, approximately 10–15 genes would be implicated in the 45 percent of the variance that can be attributed to genetic factors.

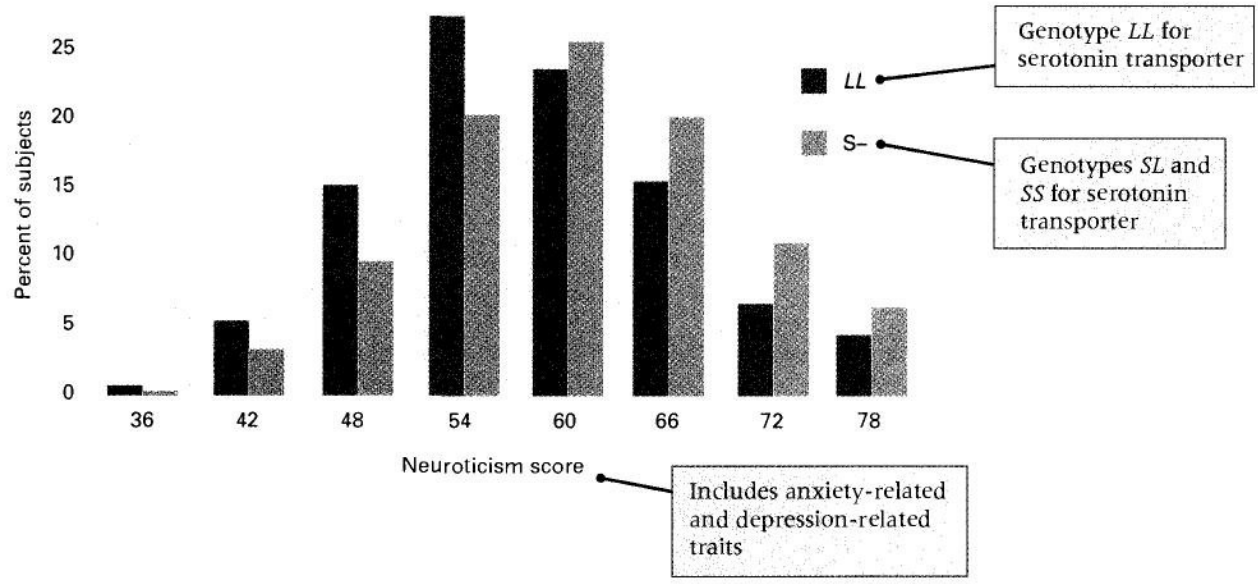


Figure 18.18 Distribution of neuroticism scores among different genotypes for the *SLC6A4* serotonin transporter polymorphism. The genotype *S*− includes both *SL* and *SS*. The total is 163 subjects for the *LL* genotype and 342 for the *S*− genotypes. [After K.-P. Lesch, et al. 1996. *Science* 274: 1527.]

Chapter Summary

Many traits that are important in agriculture and human genetics are determined by the effects of multiple genes and by the environment. Such traits are known as complex traits, and their inheritance is called complex inheritance. There are three types of complex traits: continuous, meristic, and threshold traits. Continuous traits (also called quantitative traits) are traits expressed according to a continuous scale of measurement, such as height. Meristic traits are traits expressed in whole numbers, such as the number of grains on an ear of corn. Threshold traits have an underlying risk and are either expressed or not expressed in each individual; an example is diabetes. The genes affecting complex traits are no different from those affecting simple Mendelian traits, and the genes can have multiple alleles, partial dominance, and so forth. When several genes affect a complex trait, the pattern of genetic transmission need not fit a simple Mendelian pattern, because the effects of one gene can be obscured by other genes or the environment. However, the number of genes can be estimated, and many of them can be mapped using linkage to polymorphic DNA markers.

Many continuous and meristic traits have a distribution that approximates the bell-shaped curve of a normal distribution. A normal distribution can be completely described by two quantities: the mean and the variance. The standard deviation of a distribution is the square root of the variance. In a normal distribution, approximately 68 percent of the individuals have a phenotype within 1 standard deviation of the mean, and approximately 95 percent of the individuals have a phenotype within 2 standard deviations of the mean.

Variation in phenotype of multifactorial traits among individuals in a population derives from four principal sources: (1) variation in genotype, which is measured by the genotypic variance; (2) variation in environment, which is measured by the environmental variance; (3) variation resulting from the interaction between genotype and environment (G–E interaction); and (4) variation resulting from nonrandom association of genotypes and environments (G–E association). The ratio of genotypic variance to the phenotypic variance of a trait is called the broad-sense heritability; this quantity is useful in predicting the outcome of artificial selection among clones or inbred lines. When artificial selection is carried out in a randomly mating population, the narrow-sense heritability is used for prediction. The value of the narrow-sense heritability can be determined from the correlation in phenotype among relatives.

One common type of artificial selection is called individual truncation selection, in which only those individuals that have a phenotype above a certain value (the truncation point) are saved for breeding the next generation. Artificial selection usually results in improvement of the selected population. The general principle is that the deviation of the progeny mean from the mean of the previous generation equals the narrow-sense heritability times the deviation of the parental mean from the mean of the parental generation: $M' - M = h^2(M^* - M)$. When selection is carried out for many generations, progress often slows or ceases because (1) some of the favorable genes become nearly fixed in the population, which decreases the narrow-sense heritability, and (2) natural selection may counteract the artificial selection.

A gene that affects a complex trait is a quantitative-trait locus, or QTL. The locations of QTLs in the genome can be determined by genetic mapping with respect to polymorphic DNA markers. The mapping of QTLs aids in the manipulation of genes in breeding programs and in the identification of the genes. QTLs for complex traits can also be identified by testing candidate genes known to contribute to the development or physiology of the trait.

Key Terms

artificial selection
broad-sense heritability
candidate gene
complex trait
continuous trait
correlated response
correlation coefficient
covariance
distribution
environmental variance
fraternal twins
genotype–environment association
genotype–environment interaction
genotypic variance
heterosis
hybrid vigor
identical twins
inbreeding depression
individual selection
liability
mean
meristic trait
multifactorial trait
narrow-sense heritability
normal distribution
phenotypic variance
quantitative trait
quantitative-trait locus (QTL)
selection limit
standard deviation
threshold trait
truncation point
variance

Review the Basics

- Give an example of a trait in human beings that is affected both by environmental factors and by genetic factors. Specify some of the environmental factors that affect the trait.
- Does the distribution of phenotypes of a trait in a population tell one anything about the relative importance of genes and environment in causing differences in phenotype among individuals? Does it tell one anything about the number of genes that may affect the trait? Explain.
- What is the genotypic variance of a quantitative trait? What is the environmental variance?
- Distinguish between the variance due to genotype–enviroment interaction and the variance due to genotype–environment association. Which type of variance is more easily controlled by the experimenter?
- In regard to the genotypic variance, what is special about an inbred line or about the F_1 progeny of a cross between two inbred lines?
- Distinguish between broad-sense heritability and narrow-sense heritability. Which type of heritability is relevant to individual truncation selection?
- The distribution of bristle number on one of the abdominal segments in a population of *Drosophila* ranges from 12 to 26. The narrow-sense heritability of the trait is approximately 50 percent. Do you think that it would be possible, by practicing artificial selection over a number of generations, to obtain a population in which the *mean* bristle number was greater than 26? a population in which the *minimum* bristle number was greater than 26? Explain why or why not.
- Why are correlations between relatives of interest to the quantitative geneticist?
- In the context of multifactorial inheritance, what is a QTL? Does a QTL differ from any other kind of gene? How can QTLs be detected?
- In the context of multifactorial inheritance, what is a candidate gene? Would you regard the human *PAH* gene for phenylalanine hydroxylase as a natural candidate gene for mental retardation? Why or why not?

Guide to Problem Solving

Problem 1 A genetically heterogeneous population of wheat has a variance in the number of days to maturation of 40, whereas two inbred populations derived from it have a variance in the number of days to maturation of 10.

(a) What is the genotypic variance, σ_g^2, the environmental variance, σ_e^2, and the broad-sense heritability, H^2, of days to maturation in this population?

(b) If the inbred lines were crossed, what would be the predicted variance in days to maturation of the F_1 generation?

Answer

(a) The phenotypic variance, σ_p^2, in the genetically heterogeneous population is the sum $\sigma_g^2 + \sigma_e^2 = 40$. The variance in the inbred lines equals the environmental variance (because $\sigma_g^2 = 0$ in a genetically homogeneous population); hence $\sigma_e^2 = 10$. Therefore, σ_g^2 in the heterogeneous population equals $40 - 10 = 30$. The broad-sense heritability H^2 equals $\sigma_g^2/(\sigma_g^2 + \sigma_e^2) = 30/40 = 75$ percent.

(b) If the inbred lines were crossed, the F_1 generation would be genetically uniform, and consequently the predicted variance would be $\sigma_e^2 = 10$.

Problem 2 Artificial selection for fruit weight is carried out in a tomato population. From each plant, all the fruits are weighed, and the average weight per tomato is regarded as the phenotype of the plant. In the population as a whole, the average fruit weight per plant is 75 grams. Plants whose fruit weight averages 100 grams per tomato are selected and mated at random to produce the next generation. The narrow-sense heritability of average fruit weight in this population is estimated to be 20 percent.

(a) What is the expected average fruit weight among plants after one generation of selection?

(b) What is the expected average fruit weight among plants after five consecutive generations of selection when, in each generation, the chosen parents have an average fruit weight 25 grams above the popula-

tion average in that generation? Assume that the narrow-sense heritability remains constant during this time.

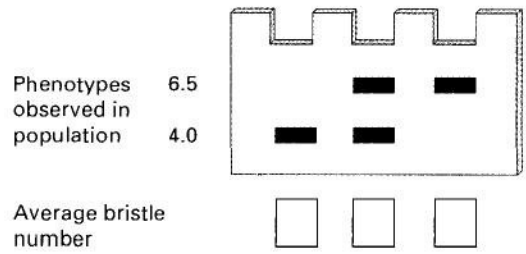
Answer In the first generation of selection, the mean of the population is $M = 75$ grams, that of the selected parents is $M^* = 100$ grams, and the narrow-sense heritability h^2 is given as 0.20. **(a)** The formula below gives the prediction equation for this case, and the predicted average fruit weight is 80 grams per tomato. **(b)** It is clear from the formula that if $M^* - M = 25$ in each generation, and h^2 remains at 0.20, then the expected gain per generation is 5 grams. After five generations the expected gain is 25 grams, yielding an expected average fruit weight after five generations of 100 grams per tomato. In the supermarket, this represents a decrease from about 6 tomatoes per pound to about 4 to 5 tomatoes per pound.

$$M' = M + h^2(M^* - M)$$
$$= 75 + (0.20)(100 - 75) = 75 + 5$$
$$= 80 \text{ grams}$$

Problem 3 A restriction fragment polymorphism (RFLP) in *Drosophila* is linked to a QTL (quantitative trait locus) affecting the number of abdominal bristles. A probe for the RFLP hybridizes with a 6.5-kb restriction fragment, with a 4.0-kb restriction fragment, or with both (in heterozygotes). The QTL genotypes may be designated *QQ*, *Qq*, and *qq*. The average abdominal bristle numbers in these genotypes are 20, 18, and 16, respectively. The typical standard deviation in bristle number is about 2, so the effect of the QTL is equal to the phenotypic standard deviation. This would usually be regarded as a rather large effect. A mating is carried out as shown, and each of the progeny is examined for bristle number and then subjected to electrophoresis and DNA hybridization to classify the RFLP genotype.

$$\frac{6.5\ Q}{4.0\ q}\,♀♀ \times \frac{6.5\ Q}{4.0\ q}\,♂♂$$

As expected, the progeny phenotypes with respect to the RFLP fall into three classes: those with the 4.0 band only (homozygous 4.0), those with both bands (4.0/6.5 heterozygotes), and those with the 6.5 band only (homozygous 6.5). In the boxes under the lanes in the gel shown above, write the average bristle number for each of the three RFLP phenotypes, assuming that the frequency of recombination between the RFLP locus and the QTL is 10 percent. Remember that there is no recombination in *Drosophila* males.



Problem 3

Answer Solving this kind of problem absolutely requires a Punnett square so that all the progeny genotypes and phenotypes can be displayed with their respective frequencies. The appropriate format for this problem is shown in the table below, where the frequencies of the female gametes result from linkage with a recombination frequency of 0.1, and those in males result from random segregation with no recombination. The average bristle number in each class of progeny is shown in bold type. From this tabulation it is clear that among flies with the 4.0 band only, the ratio of $Qq : qq$ genotypes is 1 : 9, so the average bristle number in this class is $(1/10) \times 18 + (9/10) \times 16 = 16.2$. Among flies with the 4.0 and the 6.5 band, the ratio of $QQ : Qq : qq$ genotypes is 1 : 18 : 1, so the average bristle number in this class is $(1/20) \times 20 + (18/20) \times 18 + (1/20) \times 16 = 18.0$. Finally, among flies with the 6.5 band only, the ratio of $QQ : Qq$ genotypes is 9 : 1, so the average bristle number in this class is $(9/10) \times 20 + (1/10) \times 18 = 19.8$. Note that the difference in mean phenotype between the RFLP genotypes is less than 1 phenotypic standard deviation, which means that even a difference this large would be very difficult to detect and that doing so would require large sample sizes.

Female gamete	Frequency	Male gamete (each with frequency 0.5) 6.5 *Q* 0.5			4.0 *q* 0.5		
6.5 *Q*	(1 − 0.1)/2 = 0.45	0.225	$\frac{6.5\ Q}{6.5\ Q}$	**20**	0.225	$\frac{6.5\ Q}{4.0\ q}$	**18**
6.5 *q*	0.1/2 = 0.05	0.025	$\frac{6.5\ q}{6.5\ Q}$	**18**	0.025	$\frac{6.5\ q}{4.0\ q}$	**16**
4.0 *Q*	0.1/2 = 0.05	0.025	$\frac{4.0\ Q}{6.5\ Q}$	**20**	0.025	$\frac{4.0\ Q}{4.0\ q}$	**18**
4.0 *q*	(1 − 0.1)/2 = 0.45	0.225	$\frac{4.0\ q}{6.5\ Q}$	**18**	0.225	$\frac{4.0\ q}{4.0\ q}$	**16**

Problem 3

Analysis and Applications

18.1 Two varieties of corn, A and B, are field-tested in Indiana and North Carolina. Strain A is more productive in Indiana, but strain B is more productive in North Carolina. What phenomenon in quantitative genetics does this example illustrate?

18.2 A distribution has the feature that the standard deviation is equal to the variance. What are the possible values for the variance?

18.3 The following questions pertain to a normal distribution.

(a) What term applies to the value along the x-axis that corresponds to the peak of the distribution?

(b) If two normal distributions have the same mean but different variances, which is the broader?

(c) What proportion of the population is expected to lie within one standard deviation of the mean? Within two standard deviations?

18.4 Distinguish between the broad-sense heritability of a quantitative trait and the narrow-sense heritability. If a population is fixed for all genes that affect a particular quantitative trait, what are the values of the narrow-sense and broad-sense heritabilities?

18.5 When we compare a quantitative trait in the F_1 and F_2 generations obtained by crossing two highly inbred strains, which set of progeny provides an estimate of the environmental variance? What determines the variance of the other set of progeny?

18.6 For the difference between the domestic tomato, *Lycopersicon esculentum*, and the wild South American relative, *Lycopersicon chmielewskii*, the environmental variance σ_e^2 accounts for 13 percent of the total phenotypic variance σ_p^2 of fruit weight, 9 percent of σ_p^2 of soluble-solid content, and 11 percent of σ_p^2 of acidity. What are the broad-sense heritabilities of these traits?

18.7 Ten female mice had the following numbers of live-born offspring in their first litters: 11, 9, 13, 10, 9, 8, 10, 11, 10, 13. Considering these females as representative of the total population from which they came, estimate the mean, variance, and standard deviation of size of the first litter in the entire population.

18.8 Suppose a quantitative trait is determined by three independently segregating genes in a randomly mating population. Alleles *A*, *B*, and *C* are dominant to their recessive counterparts *a*, *b*, and *c*. Genotype *aa bb cc* has a phenotype of 0, and two units in phenotype are added for each locus that is heterozygous or homozygous for the favorable allele.

(a) What is the expected distribution of phenotypes in a population in which each favorable allele has a frequency of 1/2?

(b) Calculate the mean, variance, and standard deviation of the phenotype in this population. (In problems such as this, the variance is most easily calculated as the mean of the squares of the phenotypic values minus the square of the mean phenotypic value: in symbols, $s^2 = Mean(x^2) - [Mean(x)]^2$.)

(c) How do the answers to parts **(a)** and **(b)** compare to those for a trait affected by additive alleles (in which each favorable allele present in a genotype adds one unit to the phenotype)?

18.9 Values of IQ score are distributed approximately according to a normal distribution with mean 100 and standard deviation 15. What proportion of the population has a value above 130? Below 85? Above 85?

18.10 The data in the adjoining table pertain to milk production over an 8-month lactation among 304 two-year-old Jersey cows.

Pounds of milk produced	Number of cows
1000–1500	1
1500–2000	0
2000–2500	5
2500–3000	23
3000–3500	60
3500–4000	58
4000–4500	67
4500–5000	54
5000–5500	23
5500–6000	11
6000–6500	2

The data have been grouped by intervals, but for purposes of computation, each cow may be treated as though the milk production were equal to the midpoint of the interval (for example, 1250 for cows in the 1000–1500 interval).

(a) Estimate the mean, variance, and standard deviation in milk yield.

(b) What range of yield would be expected to include 68 percent of the cows?

(c) Round the limits of this range to the nearest 500 pounds, and compare the observed with the expected numbers of animals.

(d) Do the same calculations as in parts **(b)** and **(c)** for the range expected to include 95 percent of the animals.

18.11 In the F_2 generation of a cross of two cultivated varieties of tobacco, the number of leaves per plant was distributed according to a normal distribution with mean 18 and standard deviation 3. What proportion of the population is expected to have the following phenotypes?

(a) between 15 and 21 leaves

(b) between 12 and 24 leaves

(c) fewer than 15 leaves

(d) more than 24 leaves

(e) between 21 and 24 leaves

18.12 Two highly homozygous inbred strains of mice are crossed and the 6-week weight of the F_1 progeny determined.

(a) What is the magnitude of the genotypic variance in the F_1 generation?

(b) If all alleles affecting 6-week weight were additive, what would be the expected mean phenotype of the F_1 generation compared with the average 6-week weight of the original inbred strains?

18.13 In a cross between two cultivated inbred varieties of tobacco, the variance in leaf number per plant is 1.46 in the F_1 generation and 5.97 in the F_2 generation. What are the genotypic and environmental variances? What is the broad-sense heritability in leaf number?

18.14 Two inbred lines of *Drosophila* are crossed, and the F_1 population has a mean number of abdominal bristles of 20 and a standard deviation of 2. The F_2 generation has a mean of 20 and a standard deviation of 3. What are the environmental variance, the genetic variance, and the broad-sense heritability of bristle number in this population?

18.15 Maternal effects are nongenetic influences on offspring phenotype that derive from the phenotype of the mother. For example, in many mammals, larger mothers have larger offspring, in part because of a maternal effect on birth weight. What result would a maternal effect have on the correlation in birth weight between mothers and their offspring compared with that between fathers and their offspring? How would such a maternal effect influence the estimate of narrow-sense heritability?

18.16 In terms of the narrow-sense heritability, what is the theoretical correlation coefficient in phenotype between first cousins who are the offspring of monozygotic twins?

18.17 If the correlation coefficient of a trait between first cousins is 0.09, what is the estimated narrow-sense heritability of the trait?

18.18 The asexual unicellular protozoan *Difflugia* has a number of well-formed and easily counted teeth encircling the region of the cell that functions as the mouth. The cells also have a variable number of spiny projections. The correlation coefficients between a parent and its offspring in a genetically heterogeneous population are 0.956 and 0.287 for tooth number and length of longest spine, respectively.

(a) Estimate the broad-sense heritabilities of these traits.

(b) In view of the fact that parents and offspring in asexual organisms are related as identical twins in sexual organisms, how can the large difference in broad-sense heritability be explained?

18.19 In an experiment with weight gain between ages 3 and 6 weeks in mice, the difference in mean phenotype between two strains was 17.6 grams and the genotypic variance was estimated as 0.88. Estimate the minimum number of genes affecting this trait.

18.20 Estimate the minimum number of genes affecting fruit weight in a population of the domestic tomato produced by crossing two inbred strains. When average fruit weight is expressed as the logarithm of fruit weight in grams, the inbred lines have average fruit weights of -0.137 and 1.689. The F_1 generation had a variance of 0.0144, and the F_2 generation had a variance of 0.0570.

18.21 A flock of broiler chickens has a mean weight gain of 700 g between ages 5 and 9 weeks, and the narrow-sense heritability of weight gain in this flock is 0.80. Selection for increased weight gain is carried out for five consecutive generations, and in each generation the average of the parents is 50 g greater than the average of the population from which the parents were derived. Assuming that the heritability of the trait remains constant at 80 percent, what is the expected mean weight gain after the five generations?

18.22 In a selection experiment for increased plasma cholesterol levels in mice, parents with a mean level of 2.37 units were selected from a population with a mean of 2.26 units, and the progeny of the selected parents had an average level of 2.33 units. Estimate the narrow-sense heritability of this trait from these data.

18.23 A mouse population has an average weight gain between ages 3 and 6 weeks of 12 g, and the narrow-sense heritability of the weight gain between 3 and 6 weeks is 20 percent.

(a) What average weight gain would be expected among the offspring of parents whose average weight gain was 16 g?

(b) What average weight gain would be expected among the offspring of parents whose average weight gain was 8 g?

18.24 To estimate the heritability of maze-learning ability in rats, a selection experiment was carried out. From a population in which the average number of trials necessary to learn the maze was 10.8, with a variance of 4.0, animals were selected that managed to learn the maze in an average of 5.8 trials. Their offspring required an average of 8.8 trials to learn the maze. What is the estimated narrow-sense heritability of maze-learning ability in this population?

18.25 A replicate of the population in the preceding problem was reared in another laboratory under rather different conditions of handling and other stimulation. The mean number of trials required to learn the maze was still 10.8, but the variance was increased to 9.0. Animals with a mean learning time of 5.8 trials were again selected, and the mean learning time of the offspring was 9.9.

(a) What is the heritability of the trait under these conditions?

(b) Is this result consistent with that in Problem 18.24 and how can the difference be explained?

18.26 A representative sample of lamb weights at the time of weaning in a large flock is shown below. If the narrow-sense heritability of weaning weight is 20 percent, and the half of the flock consisting of the heaviest lambs is saved for breeding for the next generation, what is the best estimate of the average weaning weight of the progeny? (*Note*: If a normal distribution has mean μ and standard deviation σ, then the mean of the upper half of the distribution is given by $\mu + 0.8\sigma$.)

81	81	83	101	86
65	68	77	66	92
94	85	105	60	90
94	90	81	63	58

18.27 A quantitative trait is affected by 20 genetically unlinked, additive loci at which, for each locus, the presence of 1 or 2 favorable alleles adds 1 or 2 units to the phenotype, respectively. A population is segregating for all 20 loci with allele frequencies equal to 1/2 for both the favorable allele and the unfavorable allele at each locus.

(a) What is the mean phenotype in this population?

(b) If artificial selection were carried out until the favorable allele at each locus was fixed, what would the phenotype of the selected population be?

(c) In the original population, what is the expected frequency of the genotype that is homozygous for the favorable allele at every locus?

(d) How many individuals would have to be present in the original population so that one individual would be expected that is homozygous for all favorable alleles? (This calculation shows why artificial selection can result in a selected population whose mean phenotypic value is much greater than that of any genotype found in the original population.)

18.28 For a phenotype determined by a single, completely penetrant recessive allele at frequency q in a random-mating population, it can be shown that the narrow-sense heritability is $2q\ (1 + q)$.

(a) Calculate the narrow-sense heritability for $q = 1.0$, 0.5, 0.1, 0.05, 0.01, 0.005, and 0.001.

(b) Note that the narrow-sense heritability goes to 0 as q goes to 0, yet the phenotype is completely determined by heredity. How can this be explained?

18.29 In artificial selection, the deviation of the mean of the selected parents from the mean of the population $(M^* - M)$ is often called the *selection differential* and is symbolized S. If artificial selection is carried out with a constant selection differential of S per generation, and h^2 remains constant, show that after t generations of selection, the mean of the selected population is given by $M_t = M_0 + th^2S$, where M_0 is the mean of the original population.

18.30 With reference to the selection differential S defined in the previous problem, what is the corresponding formula for M_t if the selection differential can change and in generation i is symbolized as S_i?

Challenge Problems

18.31 A quantitative trait is determined by three independently segregating, completely additive genes in a randomly mating population. Alleles *A*, *B*, and *C* are favorable for the trait. Genotype *aa bb cc* has a phenotype of 0, and in every other genotype one unit in phenotype is added for each favorable allele that is present. Consider a random-mating population in which the allele frequencies are 0.5 for the favorable allele and the unfavorable allele at each locus.

(a) What is the mean phenotype in the population?

(b) If artificial selection is carried out such that the selected parents have a phenotype of either 5 or 6, what is the mean phenotype among the selected parents?

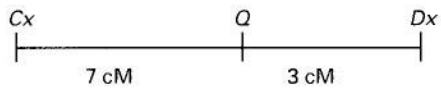
(c) What are the allele frequencies among the selected parents?

(d) If the selected parents are mated at random, what is the expected mean phenotype among the offspring?

(e) Do these results conform to the formula $M' = M + h^2(M^* - M)$?

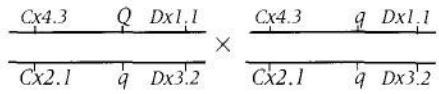
(f) What is the value of h^2? Is this value unexpected?

18.32 Two varieties of chickens differ in the average number of eggs laid per hen per year. One variety yields an average of 300 eggs per year, the other an average of 270 eggs per year. The difference is due to a single quantitative-trait locus (QTL) located between two restriction fragment length polymorphisms. The genetic map is



In this diagram, *Cx* and *Dx* are the locations of the restriction fragment length polymorphisms, and *Q* is the location of the QTL affecting egg production. The map distances are given in centimorgans (cM). The *Cx* alleles are denoted *Cx4.3* and *Cx2.1* according to the size of the restriction fragment produced by digestion with a particular restriction enzyme, and the *Dx* alleles are denoted *Dx3.2* and *Dx1.1*. The allele *Q* is associated with the higher egg production and *q* with the lower egg production, with the *Qq* genotype having an annual egg production equal

to the average of the *QQ* (300 eggs) and *qq* (270 eggs) genotypes. A cross of the following type is made:



Female offspring are classified into the following four types:

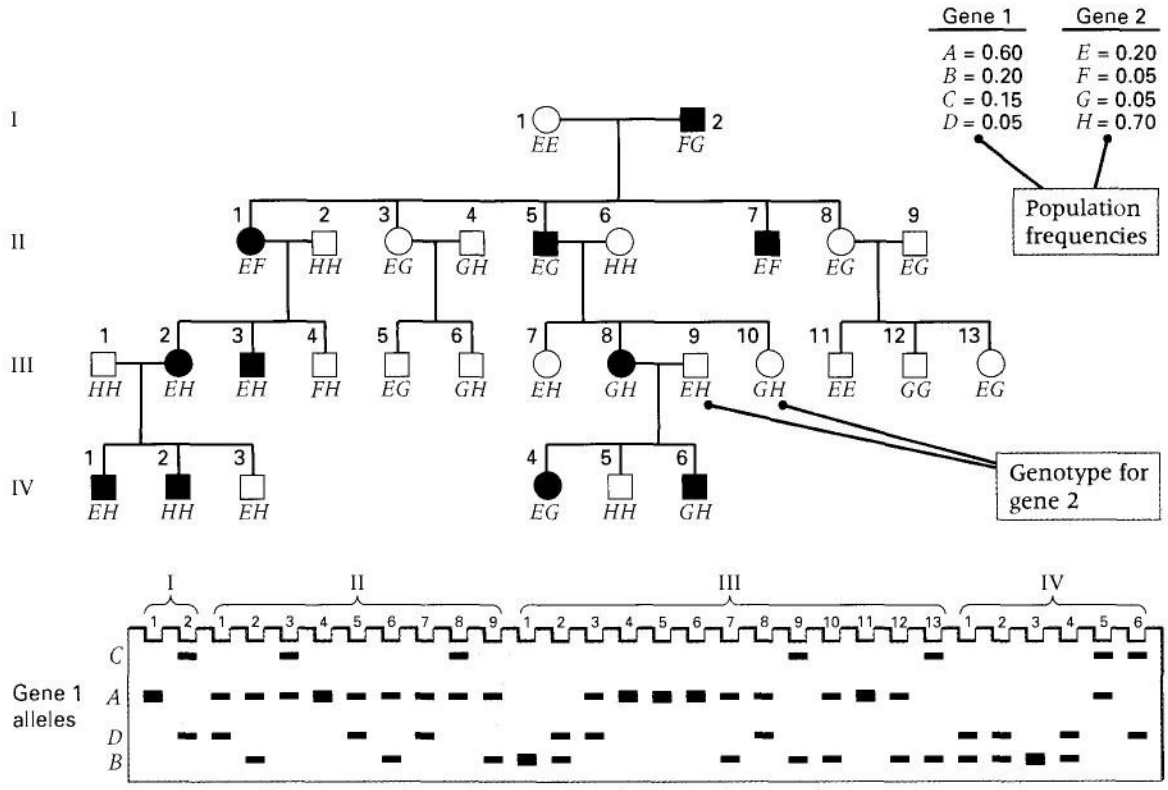
(a) *Cx4.3 Dx1.1/Cx2.1 Dx3.2*
(b) *Cx4.3 Dx3.2/Cx2.1 Dx3.2*
(c) *Cx2.1 Dx1.1/Cx2.1 Dx3.2*
(d) *Cx2.1 Dx3.2/Cx2.1 Dx3.2*

What is the expected average egg production of each of these genotypes? You may assume complete chromosome interference across the region.

18.33 The pedigree shown below includes information about the alleles of three genes that are segregating.

- Gene 1 has four alleles denoted *A*, *B*, *C*, and *D*. Each allele is identified according to the electrophoretic mobility of a DNA restriction fragment that hybridizes with a probe for gene 1. The pattern of bands observed in each individual in the pedigree is shown in the diagram of the gel below the pedigree. When an allele is homozygous, the band is twice as thick as otherwise, because homozygous genotypes yield twice as many restriction fragments corresponding to the allele than do homozygotes. The small table at the upper right of the pedigree gives the frequencies of the gene 1 alleles in the population as a whole.
- Gene 2 has four alleles denoted *E*, *F*, *G*, and *H*. These also are identified by the electrophoretic mobility of restriction fragments. For each person in the pedigree, the genotype with respect to gene 2 is indicated. The population frequencies of the gene-2 alleles are also given in the table at the upper right.
- Gene 3 is associated with a rare genetic disorder. Affected people are indicated by the red symbols in the pedigree. The genotype associated with the disease has complete penetrance. There are two alleles of gene 3, which are denoted *Q* and *q*.

(a) What genetic hypothesis can explain the pattern of affected and nonaffected individuals in the pedigree?
(b) If the male IV-2 fathers a daughter, what is the probability that the daughter will be affected?
(c) Assuming random mating, if the male IV-2 fathers a daughter, what is the probability that the daughter will have the genotype *HF* for gene 2?
(d) Identify the genotype of each person in the pedigree with respect to genes 1, 2, and 3. Wherever possible, specify the linkage phase (coupling or repulsion) of the alleles present in each individual.
(e) Examine the data for evidence of linkage. Are any of the pairs of genes linked? On the basis of this pedigree, what is the estimated recombination frequency between gene 1 and gene 2? between gene 2 and gene 3?



Electrophoresis pattern when DNA from each individual is probed for gene 1

Problem 18.33

GeNETics on the web

GeNETics on the Web will introduce you to some of the most important sites for finding genetic information on the Internet. To explore these sites, visit the Jones and Bartlett home page at

http://www.jbpub.com/genetics

For the book *Genetics: Analysis of Genes and Genomes*, choose the link that says **Enter GeNETics on the Web**. You will be presented with a chapter-by -chapter list of highlighted keywords. Select any highlighted keyword and you will be linked to a Web site containing genetic information related to the keyword.

- This keyword will open a table of frequencies of the most common types of **birth defects** in the U. S. population. The overall frequency of these conditions is about 1 in 25 births. Most of the conditions that have genetic risk factors are complex traits affected by multiple genes; these include disorders of the heart and circulation, muscles and skeleton, genital and urinary tract, and nervous system and eye, which together have an incidence of 1 in 35.

- Older genetics textbooks often reported that blue **eye color** versus brown eye color in human beings has a simple Mendelian mode of inheritance, with blue eyes recessive to brown. This hypothesis is still widely disseminated in elementary and high school textbooks. For the latest information on the inheritance of eye color, log on to this keyword site .

- Many human disorders are complex, multifactorial traits with significant genetic components as well as multiple environmental factors in their causation. Identifying the specific genetic and environmental causes of such conditions is extremely difficult. A case in point is **schizophrenia**, a relatively common mental disease with an estimated incidence of 1–1.5 percent in the U. S. population. Though serious, schizophrenia is usually treatable with drug therapy. Schizophrenia appears to have a significant genetic component, and there are probably multiple forms of the disease. At this keyword site you can learn more about this condition and the meaning of the statement "People do not cause schizophrenia; they merely blame each other for doing so."

- The keyword **serotonin** will link you to a site offering a three-dimensional display of over 200,000 small biological molecules, including Prozac® and serotonin. Using the mouse, you can tilt and rotate the structures to analyze the spatial relationships among the atoms. Use this site to compare the three-dimensional structures of 5-hydroxytryptamine (serotonin) and fluoxetine (Prozac®), which can be found in the indices for biomolecules and for pharmaceuticals–prescription drugs, respectively. It is also interesting to compare melatonin with these structures. Melatonin is involved in maintaining our circadian rhythms.

- The Mutable Site changes frequently. Each new update includes a different site that highlights genetics resources available on the World Wide Web. Select the **Mutable Site** for Chapter 18 and you will be linked automatically.

- The Pic Site showcases some of the most visually appealing genetics sites on the World Wide Web. To visit the genetics Web site pictured below, select the **PIC Site** for Chapter 18.



Further Reading

Black, D. M., and E. Solomon. 1993. The search for the familial breast/ovarian cancer gene. *Trends in Genetics* 9: 22.

Bouchard, T. J., Jr., D. T. Lykken, M. McGue, N. L. Segal, and A. Tellegen. 1990. Sources of human psychological differences: The Minnesota study of twins reared apart. *Science* 250: 223.

Comprehensive Human Genetic Linkage Center. 1994. A comprehensive human linkage map with centimorgan density. *Science* 265: 2049.

Crow, J. F. 1993. Francis Galton: Count and measure, measure and count. *Genetics* 135: 1.

Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction of Quantitative Genetics*, 4th ed. Essex, England: Longman.

Feldman, M. W., and R. C. Lewontin. 1975. The heritability hang-up. *Science* 190: 1163.
Haley, C. S. 1996. Livestock QTLs: Bringing home the bacon. *Trends in Genetics* 11: 488.
Hartl, D. L. 2000. *A Primer of Populaton Genetics*. 3d ed. Sunderland MA: Sinauer.
Hartl, D. L., and A. G. Clark. 1997. *Principles of Population Genetics*. 3rd ed. Sunderland, MA: Sinauer.
Lander, E. S., and N. J. Schork. 1994. Genetic dissection of complex traits. *Science* 265: 2037.
MacGregor, A. J., H. Snieder, N. J. Schork, and T. D. Spector. 2000. Twins: Novel uses to study complex traits and genetic diseases. *Trends in Genetics* 16: 131.
Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln, and S. D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 335: 721.
Pirchner, F. 1983. *Population Genetics in Animal Breeding*. 2d ed. New York: Plenum.
Plomin, R., and J. C. Defries. 1998. The genetics of cognitive abilities and disabilities. *Scientific American*, May.
Risch, N. J. 2000 Searching for genetic determinants in the new millenium. *Nature* 405: 847.
Risch, N., and H. Zhang. 1995. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science* 268: 1584.
Stigler, S. M. 1995. Galton and identification by fingerprints. *Genetics* 140: 857.
Stuber, C. W. 1996. Mapping and manipulating quantitative traits in maize. *Trends in Genetics* 11: 477.
Tanksley, S. D., and S. R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277: 1063.

Answers to Even-Numbered Problems

Chapter 1

1.2 They noted that the strands were used as a template for the formation of a new daughter strand having a complementary sequence of bases. They also noted that genetic information could be coded by the sequence of bases along the DNA molecule, analogous to letters of the alphabet printed on a strip of paper. Finally, they noted that changes in genetic information could result from errors in replication, and the altered nucleotide sequence could then be perpetuated.

1.4 Replication results in two daughter DNA duplexes, each identical in base sequence to the parental molecule (except for possible mutations in the sequence). Each of the daughter molecules contains one of the original intact parental strands.

1.6 A mixture of heat-killed S cells and living R cells causes pneumonia in mice, but neither heat-killed S cells nor living R cells alone do so.

1.8 Phenol would not destroy the transforming activity; strong alkali would.

1.10 Because the mature T2 phage contains only DNA and protein; the labeled RNA was left behind in material released by the burst cells.

1.12 Because the amount of A equals that of T, and the amount of G equals that of C, it can be inferred that the DNA in this bacteriophage is double-stranded; 56 percent of the base pairs are A–T pairs, and 44 percent are G–C pairs.

1.14 Because A pairs with T, and G pairs with C, the base composition of the other strand must be 24 percent T, 28 percent A, 22 percent C, and 26 percent G.

1.16 Because of A–T and G–C base pairing, the complementary strand has the sequence

3'-TAGCATACGTGAAATGGGCC-5'

Note that the paired strands have opposite 5'-to-3' polarity.

1.18 The probability that four particular bases have the sequence 5'-GGCC-3' is $(1/4)^4 = 1/256$, so 256 base pairs is the average spacing between consecutive occurrences; similarly, the average spacing for the sequence 5'-GAATTC-3' is 4096 base pairs.

1.20 The top strand is the RNA because it contains U instead of T; the mismatched base pair is the sixth from the right.

1.22 The complementary sequence is 3'-UAUGCUAU-5'.

1.24 The codon for leucine must be 5'-UUA-3'.

1.26 Six. Either DNA strand could be transcribed, and each transcript could be translated in any one of three reading frames.

1.28 The result means that an mRNA is translated in nonoverlapping groups of three nucleotides: the genetic code is a triplet code.

1.30 The codon 5'-UGG-3' codes for Trp, and in this random polymer the Trp codon is expected with a frequency of $1/4 \times 3/4 \times 3/4 = 9/64$. The amino acid Val could be specified by either 5'-GUU-3' or 5'-GUG-3'; the former has an expected frequency of $3/4 \times 1/4 \times 1/4 = 3/64$, the latter of $3/4 \times 1/4 \times 3/4 = 9/64$, totaling 12/64. The amino acid Phe could be specified only by 5'-UUU-3' in this random polymer, so Phe would have an expected frequency of $(1/4)^3 = 1/64$.

Chapter 2

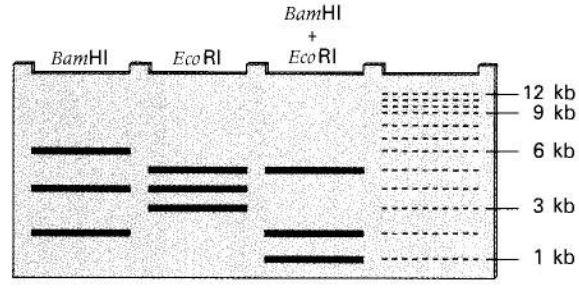
2.2 A DNA palindrome reads the same forward and backward, but along opposite strands.

(a) Yes, partner strand reads 3'-TTAA-5'.
(b) No, partner strand reads 3'-TTTT-5'.
(c) Yes, partner strand reads 3'-TTNAA-5', where the two N's must be Watson–Crick base-paired.
(d) No, partner strand reads 3'-TT(T/A)TT-5'.
(e) Yes, partner strand reads 3'-TT(C/G)AA-5'.

2.4

(a) *Hae* III produces blunt ends.
(b) *Mae*I produces a 5' overhang.
(c) *Cfo*I produces a 3' overhang.

2.6



2.8

(a) B_5 = T.
(b) B_6 = G.
(c) B_7 = pyrimidine.
(d) B_8 = T or A.

2.10 The probability of a restriction site is $1/4^n$, where n is the number of nucleotides in the site.

(a) $1/4^4 = 0.0039$.
(b) $1/4^6 = 0.00024$.
(c) $1/4^8 = 0.000015$.

2.12

(a) $3 \times 10^9/256 \approx 12$ million.

(b) $3 \times 10^9/4096 \approx 730{,}000$.

(c) $3 \times 10^9/65536 \approx 46{,}000$.

2.14 There are two *Bam*HI sites in the original molecule. Cleavage produces the following fragments:

```
5'-ATTG-3'
3'-TAACCTAG-5'

5'-GATCCAAACCCCAAAG-3'
     3'-GTTTGGGGTTTCCTAG-5'

5'-GATCCTTA-3'
     3'-GAAT-5'
```

If the fragments come back together and are sealed, the middle fragment can be present in either the original orientation or reversed, because the ends are complementary. Hence, the possible products are

```
5'-ATTGGATCCAAACCCCAAAGGATCCTTA-3'
3'-TAACCTAGGTTTGGGGTTTCCTAGGAAT-5'
```

or

```
5'-ATTGGATCCTTTGGGGTTTGGATCCTTA-3'
3'-TAACCTAGGAAACCCCAAACCTAGGAAT-5'
```

2.16

(a)
```
5'-ATGCCCTGCA-3'
3'-TACGGG-5'

     5'-GTACCATGA-3'
3'-ACGTCATGGTACTGCGC-5'

5'-CGCGTTACGCAGCTGATCGAAA-3'
     3'-AATGCGTCGACTAGCTTTGCGC-5'

5'-CGCGTATATATGCC-3'
     3'-ATATATACGG-5'
```

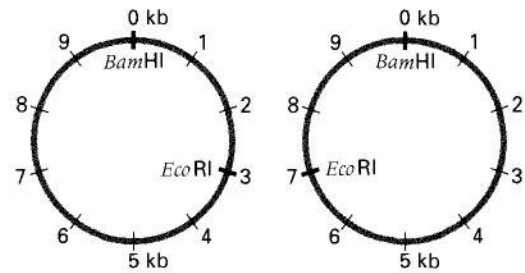
(b)
```
5'-ATGCCCTGCAGTACCATGA-3'
3'-TACGGGACGTCATGGTACTGCGC-5'

          5'-CGCGTTACGCAG-3'
     3'-AATGCGTC-5'

5'-CTGATCGAAA-3'
          3'-GACTAGCTTTGCGC-5'

5'-CGCGTATATATGCC-3'
     3'-ATATATACGG-5'
```

2.18 The results imply that the circle contains one *Bam*HI site and one *Eco*RI site separated by a distance of 3 kb. Either of the diagrams below fulfills this requirement. The numbers are the coordinates in kb, arbitrarily assigning the *Bam*HI site to position 0.



2.20 There are three possible genotypes: A_1A_1 (homozygous), A_1A_2 (heterozygous), and A_2A_2 (homozygous). Genomic DNA from homozygous A_1A_1 yields a single band of 4 kb, that from heterozygous A_1A_2 yields bands of 4 kb and 6kb, and that from A_2A_2 yields a single band of 6 kb.

2.22 Two alleles and three genotypes. The top two DNA molecules are indistinguishable using the probe shown, because both yield a 3-kb restriction fragment that hybridizes with the probe. The bottom molecule yields a 6-kb fragment. Thus the possible genotypes are "*3/3*" (homozygous), "*3/6*" (heterozygous), and "*6/6*" (homozygous).

2.24 The priming sites are shown in red.

```
5'-GATTACCGGTAAATGCCGGA-
   TTAACCCGGGTTATCAGGCCACGTACAACTGGAGTCC-3'

3'-CTAATGGCCATTTACGGCCT-
   AATTGGGCCCAATAGTCCGGTGCATGTTGACCTCAGG-5'
```

Hence, the amplified segment has the sequence

```
5'-AATGCCGGATTAACCCGGGTTATCAGGCCACGTACA-3'
3'-TTACGGCCTAATTGGGCCCAATAGTCCGGTGCATGT-5'
```

2.26

(a) 3 kb/3×10^6 kb = 0.0001 percent.

(b) $(3 \times 2^{10}$ kb$)/(3 \times 2^{10}$ kb + 3×10^6 kb) = 0.1 percent.

(c) $(3 \times 2^{20}$ kb$)/(3 \times 2^{20}$ kb + 3×10^6 kb) = 51.2 percent.

(d) $(3 \times 2^{30}$ kb$)/(3 \times 2^{30}$ kb + 3×10^6 kb) = 99.9 percent.

2.28 Because the evidence is based on total genomic DNA from an individual, all bands found in the evidence must match bands in the perpetrator. Anyone with one or more bands that do not match those in the evidence can be excluded. To find the matching bands in the suspects, for each locus draw straight horizontal lines through the bands of X. For locus 1, suspects A, C, D, and G are excluded, but B, E, and F not excluded; for locus 2, suspects A, B, C, D, and G are excluded, but E and F not excluded; and for locus 3, suspects A, C, D, F, and G are excluded, but B and E not excluded. Taking all three probes together, all except suspect E can be excluded. This does not necessarily imply that E is the robber, only that E cannot be excluded. Nevertheless, a match of three highly polymorphic DNA markers is highly suggestive.

2.30 After cleavage, phosphate 2 goes with B1, and phosphate 4 goes with B3, so the B1 and B3 mononucleotides will be radioactive.

Chapter 3

3.2 Multiply the number of possible gametes formed for each gene (one for each homozygous gene and two for each heterozygous gene). In the case of *AA Bb Cc Dd Ee*, there are a total of $1 \times 2 \times 2 \times 2 \times 2 = 16$ possible gametes. With n homozygous genes and m heterozygous genes, the number of possible gametes is $1^n \times 2^m$, which equals 2^m.

3.4 For the child to be affected, both III-1 and III-2 must be heterozygous *Aa*. For III-1 to be *Aa*, individual II-2 must be *Aa* and must transmit the recessive allele to III-1. The probability that II-2 is *Aa* equals 2/3, and then the probability of transmitting the *a* allele to III-1 is 1/2. Altogether, the

probability that III-1 is a carrier equals $(2/3) \times (1/2) = 1/3$. This is also the probability that III-2 is a carrier. Given that both III-1 and III-2 have genotype *Aa*, the probability of an *aa* offspring is 1/4. Overall, the probability that both III-1 and III-2 are carriers and that they have an *aa* child is $(1/3) \times (1/3) \times (1/4) = 1/36$. The numbers are multiplied because each of the events is independent. (The reason for the 2/3 is as follows: Because II-3 is affected, the genotypes of I-1 and I-2 must be *Aa*. Among nonaffected individuals from this mating (individuals II-2 and II-4), the ratios of *AA* and *Aa* are 1/4 : 2/4, so the probability of *Aa* is 2/3.)

3.6 1/2; 1/2. The probability argument is that each birth is independent of all the previous ones, so the number of girls and boys already born has no influence on the sexes of future children.

3.8

(a) The parent with the dominant phenotype must carry one copy of the recessive allele and hence must be heterozygous.

(b) Because some of the progeny are homozygous recessive, both parents must be heterozygous.

(c) The parent with the dominant phenotype could be either homozygous dominant or heterozygous; the occurrence of no homozygous recessive offspring encourages the suspicion that the parent may be homozygous dominant, but because there are only two offspring, heterozygosity cannot be ruled out.

3.10

(a) Two phenotypic classes are expected for the *A*, *a* pair of alleles (*A*− and *aa*), two for the *B*, *b* pair (*B*− and *bb*), and three for the *R*, *r* pair (*RR*, *Rr*, and *rr*), yielding a total number of phenotypic classes of $2 \times 2 \times 3 = 12$.

(b) The probability of an *aa bb* RR offspring is 1/4 (*aa*) × 1/4 (*bb*) × 1/4 (*RR*) = 1/64.

(c) Homozygosity may occur for either allele of each of the three genes, yielding such combinations as *AA BB rr*, *aa bb RR*, *AA bb rr*, and so forth. Because the probability of homozygosity for either allele is 1/2 for each gene, the proportion expected to be homozygous for all three genes is $(1/2) \times (1/2) \times (1/2) = 1/8$.

3.12 Because both parents have solid coats but produce some spotted offspring, they must be *Ss*. With respect to the *A*, *a* pair of alleles, the female parent (tan) is *aa*, and because there are some tan offspring, the genotype of the black male parent must be *Aa*. Thus the parental genotypes are *Ss aa* (solid tan female) and *Ss Aa* (solid black male).

3.14

(a) Because the trait is rare, it is reasonable to assume that the affected father is heterozygous *HD*/*hd*, where *hd* represents the normal allele. Half of his gametes contain the *HD* allele, so the probability is 1/2 that the son received the allele and will later develop the disorder.

(b) We do not know whether the son is heterozygous *HD*/*hd*, but the probability is 1/2 that he is; if he is heterozygous, half of his gametes will contain the *HD* allele. Therefore, the overall probability that his child has the *HD* allele is $(1/2) \times (1/2) = 1/4$.

3.16 Both parents must be heterozygous (*Aa*) because each had an albino (*aa*) parent. Therefore, the probability of an albino child is 1/4, and the probability of two homozygous recessive children is $(1/4) \times (1/4) = 1/16$. The probability that at least one child is an albino is the probability that exactly one is albino plus the probability that both are albinos. The probability that exactly one is an albino is $2 \times (3/4) \times (1/4) = 6/16$, in which the factor of 2 comes from the fact that there may be two birth orders: normal-albino or albino-normal. Therefore, the overall probability of at least one albino child equals $1/16 + 6/16 = 7/16$. Alternatively, the probability of at least one albino child may be calculated as 1 minus the probability that both are nonalbino, or $1 - (3/4)^2 = 7/16$.

3.18 The mother is heterozygous for alleles yielding bands of 6 and 9 kb, and the father is heterozygous for alleles yielding bands of 3 and 12 kb. Four genotypes of progeny are possible (*"6/3"*, *"6/12"*, *"9/3"*, and *"9/12"*) and the phenotypes are, respectively, bands of 3 and 6 kb, bands of 6 and 12 kb, bands of 3 and 9 kb, and bands of 9 and 12 kb.

3.20 Because the genes segregate independently, they can be considered separately. For the *Cr*, *cr* pair of alleles, the genotypes of the zygotes are *Cp Cp*, *Cp cp*, and *cp cp* in the proportions 1/4, 1/2, and 1/4, respectively. However, the *Cp Cp* zygotes do not survive, so among the survivors the genotypes are *Cp cp* (creeper) and *cp cp* (noncreeper), in the proportions 2/3 and 1/3, respectively. For the *W*, *w* pair of alleles, the progeny genotypes are W− (white) and *ww* (yellow), in the proportions 3/4 and 1/4, respectively. Altogether, the phenotypes and proportions of the surviving offspring can be obtained by multiplying (2/3 creeper +1/3 noncreeper) × (3/4 white + 1/4 yellow), which yields 6/12 creeper, white + 2/12 creeper, yellow + 3/12 noncreeper, white + 1/12 noncreeper, yellow. Note that the proportions sum to 12/12 = 1, which serves as a check.

3.22 The 9 : 3 : 3 : 1 ratio is that of the genotypes *A*− *B*−, *A*− *bb*, *aa B*−, and *aa bb*. The modified 9 : 7 ratio implies that all of the last three genotypes have the same phenotype. The F_1 testcross is between the genotypes *Aa Bb* × *aa bb*, and the progeny are expected in the proportions 1/4 *Aa Bb*, 1/4 *Aa bb*, 1/4 *aa Bb*, and 1/4 *aa bb*. Again, the last three genotypes have the same phenotype, so the ratio of phenotypes among progeny of the testcross is 1 : 3.

3.24

(a) $(1/2)^6 = 1/64$.

(b) $6 \times (1/6)^6 = 6/64$.

(c) $15 \times (1/6)^6 = 15/64$.

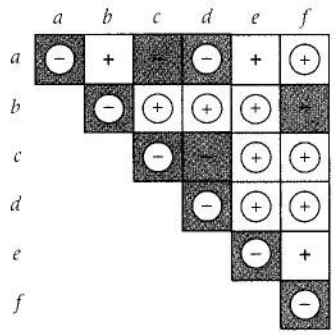
(d) $20 \times (1/6)^6 = 20/64$.

(e) $1 - 1/64 - 6/64 - 15/64 = 42/64$.

3.26 The probability that one or more offspring are *aa* equals 1 minus the probability that all are *A*−. In the mating *Aa* × *Aa*, the probability that all of *n* offspring will be *Aa* equals $(3/4)^n$, and the question asks for the value of *n* such that $1 - (3/4)^n \geq 0.95$. Solving this inequality yields $n \geq \log(1 - 0.95)/\log(3/4) = 10.4$. Therefore, 11 is the smallest number of offspring for which there is a greater than 95 percent chance that at least one will be *aa*. As a check, note that $(3/4)^{10} = 0.056$ and $(3/4)^{11} = 0.042$, so with 10 offspring, the chance of one or more *aa* equals 0.944, and with 11 offspring, the chance of one or more *aa* equals 0.958.

3.28 Let *A* be the allele associated with the 6-kb band and *B* that associated with the 9-kb band. The alternative alleles (which yield no amplification with PCR) may be denoted *a* and *b*. The genotype of parent A is therefore *AA bb*, and that of parent B is *aa BB*. Each RAPD polymorphism is dominant, so the F_1 progeny (*Aa Bb*) have both bands. With independent assortment, the F_2 progeny are 9/16 *A*−*B*− (bands of 6 and 9 kb), 3/16 *A*− *bb* (band of 6 kb), 3/16 *aa B*− (band of 9 kb), and *aa bb* (no bands).

3.30 The existing data allow the mutants to be grouped into three complementation groups as follows: {*a, c, d*}, {*b, f*}, and {*e*}. The missing entries are shown in the accompaying table.



Chapter 4

4.2 Illustration B is anaphase of mitosis, because the homologous chromosomes are not paired. Illustration A is anaphase I of meiosis, because the homologous chromosomes are paired. Illustration C is anaphase II of meiosis, because the chromosome number has been reduced by half.

4.4 In leptotene, the chromosomes first become visible as *thin threads*. In zygotene, they become *paired threads* because of synapsis. The chromosomes continue to condense to become *thick threads* in pachytene. In diplotene, the sister chromatids become evident, and each chromosome is seen to be a *paired thread* consisting of the sister chromatids. (Early signs of the bipartite nature of each chromosome are evident in pachytene.) In diakinesis, the homologous chromosomes repulse each other and *move apart;* they would fall apart entirely were they not held together by the chiasmata.

4.6

(a) 20 chromosomes and 40 chromatids.
(d) 20 chromosomes and 40 chromatids.
(c) 10 chromosomes and 20 chromatids.

4.8 The wheat parent produces gametes with 14 chromosomes and the rye parent gametes with 7, so the hybrid plants have 21 chromosomes.

4.10 Because there is an affected son, the phenotypically normal mother must be heterozygous. This implies that the daughter has a probability of 1/2 of being heterozygous.

4.12 The *Bar* mutation is an X-linked dominant. Because the sexes are affected unequally, some association with the sex chromosomes is suggested. Mating (a) provides an important clue. Because a male receives his X chromosome from his mother, the wildtype phenotype of the sons suggests that the *Bar* gene is on the X chromosome. The fact that all daughters are affected is also consistent with X-linkage, provided that the *Bar* mutation is dominant. Mating (b) confirms the hypothesis, because the females from mating (a) would have the genotype *Bar*/+ and so would produce the indicated progeny.

4.14 All the females will be v/v^+; *bw/bw* and will have brown eyes; all the males will be *v*/Y; *bw/bw* and will have white eyes.

4.16

(a) The mother is a heterozygous carrier of the mutation, so the probability that a daughter is a carrier is 1/2. The probability that both daughters are carriers is $(1/2)^2 = 1/4$.
(b) If the daughter is not heterozygous, the probability of an affected son is 0, and if the daughter is heterozygous, the probability of an affected son is 1/2; the overall probability is therefore 1/2 (that is, the chance that the daughter is a carrier) × 1/2 (that is, the probability of an affected son if the daughter is a carrier) = 1/4.

4.18 20, because in females there are 5 homozygous and 10 heterozygous genotypes, and in males there are 5 hemizygous genotypes.

4.20 The genotype of the female is *Cc*, where *c* denotes the allele for color blindness. In meiosis, the *c*-bearing chromatids undergo nondisjunction and produce an XX-bearing egg of genotype *cc*. Fertilization by a normal Y-bearing sperm results in an XXY zygote with genotype *cc*, which results in color blindness.

4.22

(a) II-2 must be heterozygous, because she has an affected son, which means that half of her sons will be affected. The probability of a nonaffected child (son or daughter) is therefore 3/4, and the probability of two nonaffected children is $(3/4)^2 = 9/16$.
(b) Because the mother of II-4 is heterozygous, the probability that II-4 is heterozygous is 1/2. If she is heterozygous, half of her sons will be affected. Therefore, the overall probability of an affected child (son or daughter) is 1/2 × 1/4 = 1/8.

4.24 $[(8!/0!8!) + (8!/2!6!) + (8!/4!4!) + (8!/6!2!) + (8!/8!0!)](1/2)^8 = 128/2^8 = 0.500$.

4.26 1/4; 1/8

4.28 With 1 : 1 segregation, the expected numbers are 125 in each class, so the χ^2 value equals $(140 - 125)^2/125 + (110 - 125)^2/125 = 3.6$, and there is 1 degree of freedom because there are two classes of data. The associated *P* value is approximately 6 percent, which means that a fit at least as bad would be expected 6 percent of the time even with Mendelian segregation. Therefore, on the basis of these data, there is no justification for rejecting the hypothesis of 1 : 1 segregation. (On the other hand, the observed *P* value is close to 5 percent, which is the conventional level for rejecting the hypothesis, so the result should not inspire great confidence.)

4.30 The χ^2 value equals $(234 - 200)^2/200 + (203 - 200)^2/186 + (175 - 200)^2/200 + (188 - 200)^2/200 = 9.67$. It has 3 degrees of freedom (four classes of data), and the P value is

about 0.025. We therefore conclude that these data are not in satisfactory agreement with the hypothesis of 1 : 1 : 1 : 1 segregation.

Chapter 5

5.2 There is no crossing-over in male *Drosophila*, so the gametes will be *A B* and *a b* in equal proportions. In the female, each of the nonrecombinant gametes *A B* and *a b* is expected with a frequency of $(1 - 0.05)/2 = 0.475$, and each of the recombinant gametes *A b* and *a B* is expected with a frequency of $0.05/2 = 0.025$.

5.4 Independent assortment takes place when the genes always have one or more crossovers between them, provided that the chromatids participate in the crossovers at random.

5.6 Each meiotic cell produces four products: namely, *A B* and *a b*, which contain the chromatids that do not participate in the crossover, and *A b* and *a B*, which contain the chromatids that do take part in the crossover. Therefore, the frequency of recombinant gametes is 50 percent. Because one crossover occurs in the region between the genes in every cell, the occurrence of multiple crossing-over does not affect the recombination frequency as long as the chromatids participating in each crossover are chosen at random.

5.8

(a) With the exception of the persons who carry recombinant chromosomes listed in part **(d)** of the answer, all affected people are $HD\ A_2/hd\ A_1$, and all unaffected persons are $hd\ A_1/hd\ A_1$.

(b) Evidently, the A_2 allele is rare enough that, by chance, a homozygous A_2A_2 person does not occur in this sample. (Alternatively, the A_2 allele could be very strongly associated with *HD*.)

(c) There is evidence for linkage between the genes, because most persons who inherit *HD* also inherit the A_2 allele associated with the dominant allele in this pedigree.

(d) The recombinant chromosomes appear in sample numbers 8, 10, 16, 24, and 30.

(e) There are 25 persons in whom a recombinant chromosome could have been detected, of whom 5 are carrying a recombinant chromosome. Hence the estimated frequency of recombination from this pedigree is $5/25 = 20\%$.

5.10

(a) The coat color gene cannot be linked to the gene with alleles A_1 and A_2, because this gene is autosomal and the coat color gene is X-linked. (One indication that the A_1, A_2 allele pair is autosomal is that some males exhibit both bands and so are heterozygous.)

(b) The coat color gene is linked to the gene with alleles B_3, and B_4, because most progeny from a doubly heterozygous mother carry a nonrecombinant X chromosome.

(c) There are three recombinant genotypes: numbers 2, 11, and 16.

(d) There are a total of 30 offspring in whom recombination in the mother could be detected, so the estimated frequency of recombination is $3/30 = 10\%$.

5.12

(a) No, they are not alleles. If they were alleles, they would undergo segregation, and the progeny would consist equally of either DDT-resistant or Dieldrin-resistant animals. There would be no progeny resistant to both insecticides and no progeny sensitive to both.

(b) The test for linkage is whether the parental and nonparental chromosomes occur in equal proportions, so the reciprocal classes must be pooled. This yields 177 parental and 205 nonparental. The chi-square = 2.05 with 1 df, for which *P* is approximately 0.15. The genes are therefore unlinked (they show independent assortment), even though they are in the same chromosome.

(c) The conclusion is that *DDT* and *DI* are so far apart in the chromosome that one or more crossovers always occur between them.

5.14 The most frequent gametes from the double heterozygous parent are the nonrecombinant gametes, in this case *Gl ra* and *gl Ra*. The genotype of the parent was therefore *Gl ra/gl Ra*. The recombinant gametes are *Gl Ra* and *gl ra*, and the frequency of recombination is calculated as the ratio $(6 + 3)/(88 + 103 + 6 + 3) = 4.5$ percent.

5.16 The most frequent classes of gametes are the nonrecombinants, and so the parental genotype was *A B c/a b C*. The least frequent gametes are the double recombinants, and the allele that distinguishes them from the nonrecombinants (in this case, *C* and *c*) is in the middle. Therefore, the gene order is *A*–*C*–*B*.

5.18

(a) The nonrecombinant gametes (the most frequent classes) are $y\ v^+\ sn$ and $y^+\ v\ sn^+$, and the double recombinants (the least-frequent classes) indicate that *sn* is in the middle. Therefore, the correct order of the genes is *y*–*sn*–*v*, and the parental genotype was $y\ sn\ v^+/y^+\ sn^+\ v$. The total progeny is 1000. The recombination frequency in the *y*–*sn* interval is $(108 + 5 + 95 + 3)/1000 = 21.1$ percent, and the recombination frequency in the *sn*–*v* interval is $(53 + 5 + 63 + 3)/1000 = 12.4$ percent. The genetic map is *y*–21.1–*sn*–12.4–*v*.

(b) On the assumption of independence, the expected frequency of double crossovers is $0.211 \times 0.124 = 0.0262$, whereas the observed frequency is $(5 + 3)/1000 = 0.008$. The coincidence is the ratio $0.008/0.0262 = 0.31$, so the interference is $1 - 0.31 = 0.69$, or 69 percent.

5.20 Let x be the observed frequency of double crossovers. Then single crossovers in the *a*–*b* interval occur with frequency $0.15 - x$, and single crossovers in the *b*–*c* interval occur with frequency $0.20 - x$. The observed recombination frequency between *a* and *c* equals the sum of the single-crossover frequencies (because double crossovers are undetected), which implies that $0.15 - x + 0.20 - x = 0.31$, or $x = 0.02$. The expected frequency of double crossovers equals $0.15 \times 0.20 = 0.03$, and the coincidence is therefore $0.02/0.03 = 0.67$. The interference equals $1 - 0.67 = 33$ percent.

5.22 The frequency of *dpy-21 unc-34* recombinants is expected to be $0.24/2 = 0.12$ among both eggs and sperm, so

the expected frequency of *dpy-21 unc-34* *dpy-21 unc-34* zygotes is $0.12^2 = 1.44$ percent.

5.24 Consider row 1: there are − entries for 3 and 7, and so 1, 3, and 7 form one complementation group. Row 2: all +, and so 2 is the only representative of a second complementation group. Row 3: mutation already classified. Row 4: there is a − for 6 only and no other − in column 4, and so 4 and 6 make up a third complementation group. Row 5: − only for 9 and no other − in column 5, and so 5 and 9 form a fourth complementation group. Rows 6 and 7: mutations already classified. Row 8: all +, and so 8 is the only member of a fifth complementation group. In summary, there are five complementation groups as follows: group 1 (mutations 1, 3, and 7); group 2 (mutation 2); group 3 (mutations 4 and 6); group 4 (mutations 5 and 9); group 5 (mutation 8).

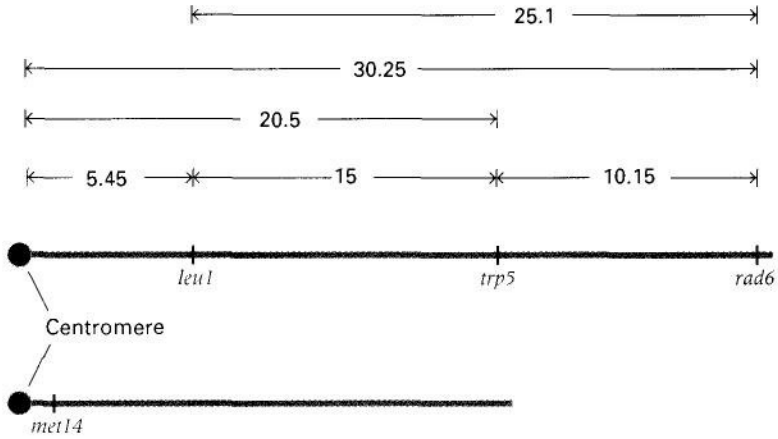
5.26 Considered individually, gene *c* yields 90 first-division segregation asci and 10 (1 + 9) second-division segregation asci, so the frequency of recombination between *c* and the centromere is (10 100) × 1 2 = 5 percent. Gene *v* yields 79 first-division and 21 (20 + 1) second-division asci, so the frequency of recombination between *v* and the centromere is (21/100) × 1 2 = 10.5 percent. If the genes are taken together, the asci include 34 PD, 36 NPD, and 30 (20 + 1 + 9) TT. Because NPD = PD, the genes are unlinked. In summary, *c* and *v* are in different chromosomes, *c* being 5 map units from the centromere and *v* being 10.5 map units from the centromere.

5.28 For each pair of genes, classify the tetrads as PD, NPD, or TT, and tabulate the results as follows:

leu2-trp1	PD 230 + 235	= 465
	NPD 215 + 220	= 435
	TT 54 + 46	= 100
leu2-met14	PD 235 + 220	= 455
	NPD 230 + 215	= 445
	TT 54 + 46	= 100
trp1-met14	PD 235 + 215 + 46	= 496
	NPD 230 + 220 + 54	= 504
	TT	0

For *leu2-trp1*, PD = NPD; therefore, *leu2* and *trp1* are unlinked. Similarly, for *leu2-met14*, PD = NPD; these genes are also unlinked. Likewise, *trp1* and *met14* are unlinked because PD = NPD. However, for the *trp1-met14* gene pair, there are no tetratypes, which means that each gene is closely linked to its own centromere. Because they do not recombine with their centromeres, the segregation of *trp1* and *met14* can be used as genetic markers of centromere segregation; their presence marks the sister spores created by segregation in the first meiotic division. Using the *trp1* and *met14* markers in this fashion allows the remaining *leu2* gene to be mapped with respect to its centromere; all tetratype tetrads will result from crossing-over between *leu2* and its centromere because neither of the other two genes recombines with its centromere. The map distance between *leu2* and its centromere is therefore (1/2)(100/1000) × 100 = 5 centimorgans.

5.30 For the *rad6-trp5* gene pair, PD = 188 + 206 = 140 + 154 + 154 + 109 = 797, NPD = 0, and TT = 105 + 92 + 3 + 2 + 1 = 203. Because PD >> NPD, the genes are linked. In this case, TT > 0 and NPD = 0, so the map distance is calculated as (1/2)(203/1000) × 100 = 10.15 centimorgans. For the *trp5-leu1* gene pair, PD = 188 + 206 + 105 + 92 + 109 = 700, NPD = 0, and TT = 140 + 154 + 3 + 2 + 1 = 300; here again, PD >> NPD, TT > 0, and NPD = 0, so the map distance is calculated as (1 2)(300/1000) × 100 = 15 centimorgans. For the *rad6-leu1* gene pair, PD = 188 + 206 + 109 = 503, NPD = 1, and TT = 105 + 92 + 140 + 154 + 3 + 2 = 496. In this case, PD >> NPD, but TT > 0 and NPD > 0. The map distance is calculated as (1 2)[(496 + 6 × 1)/1000] × 100 = 25.1 centimorgans. For the *leu1* − *met14* gene pair, PD = 188 + 105 + 140 + 2 = 435, NPD = 206 + 92 + 154 + 3 + 1 = 456, and TT = 109. In this case PD = NPD, so *leu1* is not linked to *met14*. However, we know from the previous problem that *met14* is tightly linked with its centromere, and this information can be used to calculate the map distance between *leu1* and its centromere. Because TT = 109 and also TT < 2/3, the map distance between *leu1* and its centromere is (1 2)(109 1000) × 100 = 5.45 centimorgans. For the *rad6-met14* gene pair, PD = 188 + 1 = 189, NPD = 206, and TT = 105 + 92 + 140 + 154 + 109 + 3 + 2 = 605. Here again, PD = NPD, which indicates that there is no linkage between the genes. Because *met14* is tightly linked to its centromere, and TT = 605 but TT < 2/3, the map distance



Problem 5.30—Genetic Map

between *rad6* and its centromere is $(1/2)(605/1000) \times 100 = 30.25$ centimorgans. For the *trp5* – *met14* gene pair, PD = 188 + 105 = 293, NPD = 206 + 92 = 298, and TT = 140 + 154 + 109 + 3 + 2 + 1 = 409. Once again, PD = NPD, indicating no linkage between the genes. As before, we use *met14* as a centromere marker to calculate the distance between *trp5* and its centromere as $(1/2)(409/1000) \times 100 = 20.5$ centimorgans. The overall conclusion is that *rad6*, *trp5*, and *leu1* are all linked to each other; *met14* is tightly linked to its centromere but not to the other three genes. The genetic map based on these data is shown on page 791.

Chapter 6

6.2 A replication fork has an asymmetrical structure because of the inability of the DNA polymerase to carry out synthesis in the 3'-to-5' direction, and because of the antiparallel orientation of the strands in the DNA double helix. Both daughter strands are synthesized in the 3'-to-5' direction; therefore, one strand (the leading strand) is synthesized continuously, whereas the other strand (the lagging strand) requires the transient existence of the short pieces of DNA known as precursor fragments that serve as replication intermediates. The precursor fragments are joined together by DNA ligase.

6.4 Because new nucleotides are added only to the 3' hydroxyl of the growing strand, and a dideoxynucleotide does not have a 3' hydroxyl group.

6.6 DNA polymerase joins a 5'-triphosphate with a 3'-OH group, and the outermost two phosphates are released; the resulting phosphodiester bond contains one phosphate. DNA ligase joins a single 5'-P group with a 3'-OH group.

6.8 Neither RNA polymerases nor primases require a primer to get synthesis started.

6.10 The c–d and e–f strands are leading, the a–b and g–h strands are lagging; a, d, f, and g are 5' ends; and b, c, e, and h are 3' ends.

6.12 Samples 1 and 2 are made up of deoxyribonucleotides and therefore are DNA samples. Sample 3 is composed of ribonucleotides and is therefore a sample of RNA. Because of Watson–Crick base pairing, double-stranded DNA must have equal proportions of dAMP and dTMP as well as of dGMP and dCMP. Only sample 1 has this feature and is therefore likely to be double-stranded DNA. Sample 2 is single-stranded DNA, and sample 3 is single-stranded RNA.

6.14

(a) Because 18 percent is adenine, 18 percent is also thymine; so [T] = 18 percent.

(b) Because [A] + [T] = 36 percent, [G] + [C] equals 64 percent, or [G] = [C] = 32 percent each. Overall, the base composition is [A] = 18 percent, [T] = 18 percent, [G] = 32 percent, and [C] = 32 percent; the [G] + [C] content is 64 percent.

6.16

(a) 1/2 "light" : 1/2 "heavy."

(b) 3/4 "light" : 1/4 "heavy."

(c) 7/8 "light" : 1/8 "heavy."

The observed results were:

(a) all "intermediate."

(b) 1/2 "light" : 1/2 "intermediate."

(c) 3/4 "light" : 1/4 "intermediate."

6.18 If the Holliday junction is resolved with no recombination of outside markers, one molecule should be "light" and the other "heavy." If the Holiday junction is resolved with recombination of outside markers, one molecule should be "light" on the left-hand side and "heavy" on the right-hand side, and the other the reverse; in this case the density of both molecules would be intermediate.

6.20 Remember that the chemical group at the growing end of the leading strand is always a 3'-OH group, because DNA polymerases can add nucleotides only to such a group. Because DNA is antiparallel, the opposite end of the leading strand is a 5'-P group. Therefore,

(a) 3'-OH.

(b) 3'-OH.

(c) 5'-P.

6.22

(a) The first base in the DNA that is copied is a T, so the first base in the RNA is an A. RNA grows by addition to the 3'-OH group, so the 5'-P end remains free. Therefore, the sequence of the eight-nucleotide primer is

5'-AGUCAUGC-3'

(b) The C at the 3' end has a free -OH; the A at the 5' end has a free *tri*phosphate.

(c) Right to left, because this would require discontinuous replication of the strand shown.

6.24 In the chromosome on the left, the DNA in each chromatid should have one solid line (labeled strand) and one dashed line (unlabeled strand). In the chromosomes on the right, the darker chromatids should have DNA with both lines solid (both strands labeled), and the lighter chromatid should have one line solid (labeled strand) and one dashed (unlabeled strand).

6.26 For each site, the probability equals the product of the probability of each of the nucleotide pairs in turn, 1/4 for any of the specified pairs and 1/2 for R-Y. The average spacing is the reciprocal of the probability. Hence for *Taq*I, the average distance between sites is $4^4 = 256$ base pairs; for *Bam*HI, it is $4^6 = 4096$ base pairs; and for *Hae*II, it is $4^4\,2^2 = 1024$ base pairs.

6.28

(a) C → T mismatch repair in one heteroduplex, G → A in the other.

(b) T → C mismatch repair in one heteroduplex, A → G in the other.

(c) either a T → C mismatch repair in one heteroduplex and a G → A in the other, or a C → T mismatch repair in one heteroduplex and a A → G in the other.

6.30 The 27th nucleotide from the 5' end is a C in **(a)** but an A in **(b)**.

Chapter 7

7.2 Causes of a nonreverting mutation include (1) deletion of the gene, (2) multiple nucleotide substitutions in the gene, and (3) complex DNA rearrangement with more than one breakpoint in the gene. Deletions do not revert because the genetic information in the gene is missing, and the others do not revert because the nucleotide substitutions or DNA rearrangements are unlikely to be reversed precisely in one step.

7.4 The serine codon must be UCN; otherwise, it could not change into an alanine in one step. The alanine codon must be GCN. Hence the amino acid change results from a U → G change in the codon or an A → C base change in the transcribed strand of the *psbA* gene. It is therefore a transversion.

7.6 If the +1 frameshift mutation shifts the reading frame to produce a nonsense codon, no −1 frameshift beyond the nonsense codon can prevent the termination of translation.

7.8 Any single nucleotide substitution creates a new codon with a random nucleotide sequence, so the probability that it is a chain-termination codon (UAA, UAG, or UGA) is 3/64, or approximately 5 percent.

7.10 No. Dominance and recessiveness are characteristics of the phenotype. They depend on the function of the gene and on which attributes of the phenotype are examined.

7.12 Because there are many ways in which an amino acid replacement can impair the activity of a wildtype protein, but very few ways in which mutant protein can regain wildtype activity other than restoration of the original amino acid.

7.14 Because each base pair contains one purine, the number of spontaneous depurinations per diploid cell per minute is $(3 \times 10^{-9})(6 \times 10^{9}) = 18$, or $18 \times 60 \times 24 = 26{,}000$ per diploid cell per day.

7.16

(a) 1 : 1 because the mutation is recessive and because males get their X chromosome from the mother.

(b) All females, because at 29°C the males (who in this case get their X chromosome from their father) die.

7.18 The mutation frequency in the Lac^- culture is at least $0.90 \times 10^{-8} = 9 \times 10^{-9}$. Because the mutation rate to a nonsense suppressor is independent of the presence of a suppressible mutation (in this case, in the Lac^- mutant), the mutation rate in the original Lac^+ culture is also at least 9×10^{-9}. The qualifier *at least* is important because not all nonsense suppressors may suppress the mutation in the particular Lac^- strain (that is, some suppressors may insert an amino acid that still results in a nonfunctional enzyme.)

7.20
```
5'-ATTAGTAGGG...AT...GGGAAAAA-
   AAAAAAAAAACCC...AT...CCCATTAGTA-3'

3'-TAATCATCCC...TA...CCCTTTTT-
   TTTTTTTTTTGGG...TA...GGGTAATCAT-5'
```

7.22 Amino acid replacements at many positions in the A protein do not eliminate its ability to function; only amino acid replacements at critical positions that do eliminate tryptophan synthase activity are detectable as Trp^- mutants.

7.24 Six amino acid replacements are possible because the codons UAY (Y stands for either pyrimidine, C or U) can mutate in a single step to codons for Phe (UUY), Ser (UCY), Cys (UGY), His (CAY), Asn (AAY), or Asp (GAY).

7.26 Single nucleotide substitutions that can account for the amino acid replacements are (1) Met (AUG) to Leu (UUG), (2) Met (AUG) to Lys (AAG), (3) Leu (CUN) to Pro (CCN), (4) Pro (CCN) to Thr (ACN), and (5) Thr (ACR) to Arg (AGR), where R stands for any purine (A or G) and N for any nucleotide. Substitutions (1), (2), (4), and (5) are transversion, and substitution (3) is a transition. Because 5-bromouracil primarily induces transitions, the Leu → Pro replacement is expected to be the most frequent.

7.28 The *a* and *b* gene products may be subunits of a multimeric protein. A few proteins with compensatory amino acid replacements can interact to give wildtype activity, but most mutant proteins cannot interact in this way with each other or with wildtype.

7.30

(a) All progeny would be able to grow in the absence of arginine.

(b) The genotype of the diploid is $Su^+/Su^-\ arg^+/arg^-$, where Su^+ indicates the ability to suppress, and Su^- indicates inability to suppress. The resulting haploid spores are 1/4 $Su^+\ arg^+$, 1/4 $Su^+\ arg^-$, 1/4 $Su^-\ arg^+$, and 1/4 $Su^-\ arg^-$. The first three are Arg^+ and the last Arg^-, so the ratio is 3 Arg^+ : 1 Arg^-.

(c) The diploid genotype is $Su^+\ arg^-/Su^-\ arg^+$. The haploid spores are in the proportions 0.05 $Su^+\ arg^+$, 0.45 $Su^+\ arg^-$, 0.45 $Su^-\ arg^+$, and 0.05 $Su^-\ arg^-$. Therefore, 95 percent of the progeny are Arg^+ and 5 percent are Arg^-.

Chapter 8

8.2 A nucleosome core particle contains 2 molecules each of histones 2A, 2B, 3, and 4, plus 145 bp of DNA; a nucleosome consists of a core particle plus one molecule of histone H1 and an additional 55 bp of "linker" DNA.

8.4 Matching of chromomeres, which are locally folded regions of chromatin. Polytene chromosomes do not divide because the cells in which they exist do not divide.

8.6

(a) 120×10^6 bp/5000 bands = 24 kb/band.

(b) 13,600 genes/5000 bands = 2.7 genes/band.

8.8 Because DNA replication cannot be initiated at the extreme 3' end of a DNA strand; a short region at each 3' end remains unreplicated.

8.10 The internal sequence gives a sharp band because all the fragments containing the sequence were produced by cleavage of each chromosomal DNA molecule at the nearest restriction site on each side; the fragments are therefore all of the same size. The subtelomeric sequence gives a diffuse band because only the unique restriction site is common to all the fragments; the other end terminates in the telomeric sequence, which can differ in length from one cell to the next.

8.12 The repeated sequences are located at the ends of the element. They may be in direct or inverted orientation, depending on the particular transposable element.

8.14 The supercoiled form is in equilibrium with an underwound form having many unpaired bases. At the instant that a segment is single-stranded, S1 can attack it. Because a nicked circle lacks the strain of supercoiling, there is no tendency to have single-stranded regions, and so a nicked circle is resistant to S1 cleavage.

8.16

(a) Renaturation is concentration-dependent because it requires complementary molecules to collide by chance before the formation of base pairs can occur.

(b) Here, the lengths of the GC tracts are important. Molecule 2 has a long GC segment, which will be the last region to separate. Hence molecule 1 has the lower temperature for strand separation.

8.18 All such mutations are probably lethal. The fact that the amino acid sequence of histones is virtually the same in all organisms suggests that functional histones are extremely intolerant of amino acid changes.

8.20 To relieve the stress of overwinding the direction of supercoiling is positive, and there are three superhelical turns.

8.22 No, because there is a dynamic equilibrium between supercoiled molecules with all bases paired and an underwound form with some bases unpaired, and all degrees in between.

8.24 With the repeating sequence, any two molecules can form a stable duplex, whereas each random sequence can anneal only with other fragments having the same, or a very similar, sequence.

8.26

(a) $g = 30$, $T_m = 81.3°C$.

(b) $g = 50$, $T_m = 89.5°C$.

(c) $g = 70$, $T_m = 97.7°C$.

8.28 In the first experiment, the nucleosomes form at random positions in different molecules. In the second experiment, the protein binds to a unique sequence in all the molecules, and the nucleosomes form by sequential addition of histone octamers on both sides of the protein-binding site. Because nuclease attacks only in the linker regions, the cuts are made in small regions that are highly localized.

Chapter 9

9.2 45, X.

9.4 2; they represent inactivated X chromosomes.

9.6 The chromosomes underwent replication with no cell division (endoreduplication), resulting in an autotetraploid.

9.8 The recessive allele is X-linked, and the pattern of presence or absence of sweat glands in mosaic females results from random inactivation of either of the X chromosomes in females that are heterozygous.

9.10 The mother has a Robertsonian translocation that includes chromosome 21. The child with Down syndrome has 46 chromosomes, including two copies of the normal chromosome 21 plus an additional copy attached to another chromosome (the Robertsonian translocation). This differs from the usual situation, in which Down syndrome children have trisomy 21 (karyotype 47,+21).

9.12 The order is *a e d f c b* or the other way around.

9.14 The result is one normal chromosome and one inverted chromosome (these derive from chromatids not involved in either crossover), a normal chromosome carrying the alleles *A B c D E*, and an inverted chromosome carrying the alleles *a b C d e*. Note that the two-strand double crossover has transferred an allele between the normal and the inverted chromosomes.

9.16 Starting with chromosome (c), compare the chromosomes pairwise to find those that differ by a single inversion. In this manner, you can deduce that inversion of the *i-d-c* region in (c) gave rise to (d), inversion of the *h-g-c-d* region in (d) gave rise to (a), and inversion of the *c-d-e-f* region in (a) gave rise to (b). Because you were told that (c) is the ancestral sequence, the evolutionary ancestry is (c) → (d) → (a) → (b).

9.18 Translocation heterozygotes are semisterile because adjacent segregation from the four synapsed chromosomes in meiosis produces aneuploid gametes, which have large parts of chromosomes missing or present in excess. The aneuploid gametes fail to function in plants, and in animals they result in zygote lethality. The translocation homozygote is fully fertile because the translocated chromosomes undergo synapsis in pairs, and segregation is completely regular. In the cross of translocation homozygote × normal homozygote, all progeny are expected to be semisterile.

9.20 15 altogether: 12 are permutations of the general form $[(AA \times BB) \times CC] \times DD$, where first an allotetraploid is created, then an allohexaploid, and then the allooctoploid; and 3 are permutations of the general form $[(AA \times BB) \times (CC \times DD)]$, where two allotetraploids undergo hybridization and endoreduplication.

9.22 An autopolyploid series is formed by the combination of identical sets of chromosomes. Therefore, the chromosome numbers must be even multiples of the monoploid number, or 10 (diploid), 20 (tetraploid), 30, 40, and 50.

9.24 In this male, there was a reciprocal translocation between the *Curly*-bearing chromosome and the Y chromosome. Because there is no crossing-over in the male, all Y-bearing sperm that give rise to viable progeny contain *Cy*, and all X-bearing sperm that give rise to viable progeny contain Cy^+.

9.26 Viable progeny are expected to result only from adjacent segregation, in which the Y chromosme segregates with the normal chromosome 2, and the two parts of the X–2

translocation segregate together. The expected result is 50 percent ruby, dumpy males and 50 percent nonruby, nondumpy females.

9.28 The parental classes are wildtype and *brachytic, fine-stripe* and the recombinant classes are *brachytic, fine-stripe*$^+$ and *brachytic*$^+$, *fine-stripe*. The recombination frequency between *brachytic* and *fine-stripe* is therefore equal to $(17 + 1 + 6 + 8)/682 = 0.047$.

Because semisterility is associated with the presence of the translocation, determining the recombination frequency between each of the mutations and the translocation breakpoint follows the same logic as mapping any other kind of genetic marker. For the parental classes, the mutations and the translocation are on different chromosomes. Members of the recombinant classes are those plants that show both the mutation in question and semisterility or neither semisterility nor the mutation in question. The recombination frequency between *brachytic* and the translocation breakpoint is therefore $(17 + 25 + 19 + 8)/682 = 0.102$; the recombination frequency between *fine-stripe* and the translocation breakpoint is $(19 + 6 + 1 + 25)/682 = 0.075$.

9.30 The wildtype yeast strain is disomic for chromosome 2, and the result of the cross is a trisomic. Segregation of markers on this chromosome, like *his7*, is aberrant, but segregation of markers on the other chromosomes is completely normal.

Chapter 10

10.2 The final dilution should contain 500 viable cells/ml, which is a 10^4-fold dilution of the original suspension. You could do any combination of 100-fold and 10-fold dilutions that multiply to 10^4; for example, two 100-fold dilutions, four 10-fold dilutions, or one 100-fold dilution and two 10-fold dilutions.

10.4 Minimal medium containing galactose and biotin; the galactose would require the presence of *gal*$^+$, and the biotin would allow both *bio*$^+$ and *bio*$^-$ to grow.

10.6 Infect an *E. coli* λ lysogen with P1. The λ lysogen contains λ prophage, so some of the resulting P1 phage will include the part of the chromosome that contains the prophage.

10.8 The specialized-transducing particles originate from aberrant prophage excision.

10.10 Plate on minimal medium that lacks leucine (selects for Leu$^+$) but contains streptomycin (selects for Str-r). The *leu*$^+$ allele is the selected marker, and *str-r* the counterselected marker.

10.12 The mean number of phage per cell equals 1. If you know the Poisson distribution, then it is apparent that the proportion of uninfected cells is $e^{-1} = 0.37$. If you do not know the Poisson distribution, then the answer can be calculated as follows. The probability that a bacterial cell escapes infection by a particular phage is $1 - 10^{-6} = 0.999999$, and the probability that it escapes infection by all one million phages is $P_0 = (0.999999)^{1000000}$. Therefore, $\ln P_0 = 1000000 \times \ln(0.999999) = -1$, so $P_0 = e^{-1} = 0.37$.

10.14

(a) Medium containing galactose, tryptophan, histidine, and streptomycin.

(b) Medium containing glucose (or any sugar other than galactose), histidine, and streptomycin.

(c) Medium containing glucose (or any sugar other than galactose), tryptophan, and streptomycin.

(d) Medium containing glucose (or any sugar other than galactose) and streptomycin.

10.16 All receive the *a*$^+$ allele, but whether or not a particular *b*$^+$ *str-r* recombinant contains *a*$^+$ depends on the positions of the genetic exchanges.

10.18 The first selection is for Met$^+$ and so *lac*$^+$ must have been transferred. The probability that a marker that is transferred is incorporated into the recombinants is about 50 percent, so about 50 percent of the recombinants would be expected to be *lac*$^+$. The second selection is for Lac$^+$, and *met*$^+$ is a late marker. The great majority of mating pairs will spontaneously break apart before transfer of *met*$^+$, so the frequency of *met*$^+$ recombinants in this experiment will be close to zero.

10.20

(a) The map order and distances from the *E. coli* cross are *pro–14–met–2–arg–6–ile;* those from the *S. enterica* cross are *ile–14–arg–4–met–25–pro*. These are the same except for being written in the reverse order.

(b) The *E. coli* Hrf transfers *pro* first, the *S. enterica* Hfr transfers *ile* first, and the directions of transfer are opposite.

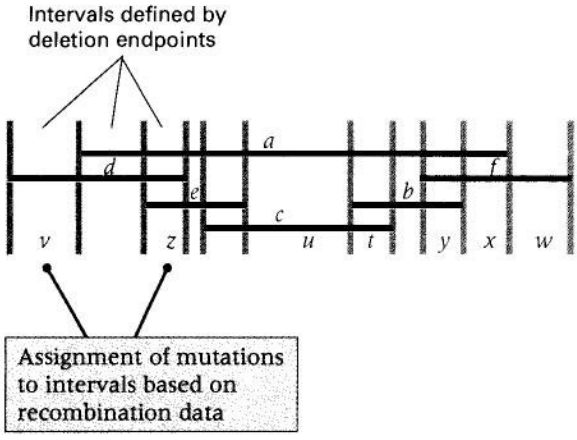
(c) The total region is 22 minutes in *E. coli* but 43 minutes in *S. enterica;* hence the *S. enterica* Hfr transfers the chromosome only half as fast as the *E. coli* Hfr.

10.22 The three possible orders are (1) *pur–pro–his*, (2) *pur–his–pro*, and (3) *pro–pur–his*. The predictions of the three orders follow. (1) Virtually all *pur*$^+$ *his*$^-$ transductants should be *pro*$^-$, but this is not true. (2) Virtually all *pur*$^+$ *pro*$^-$ transductants should be *his*$^-$, but this is not true. (3) Some *pur*$^+$ *pro*$^-$ transductants will be *his*$^-$, and some *pur*$^+$ *his*$^-$ transductants will be *pro*$^-$ (depending on where the exchanges occur). Order (3) is the only one that is not contradicted by the data.

10.24 It is probable that the *amp-r* gene in the natural isolate was present in a transposable element. Transposition into λ occurs in a small proportion of cases, and when these infect the laboratory strain, the transposable element can transpose into the chromosome before the infecting λ is lost.

10.26 Both the order of times of entry and the level of the plateaus imply that the order of gene transfer is *a b c d*. The times of entry are obtained by plotting the number of recombinants of each type versus time and extrapolating back to the time axis. The values are *a*, 10 minutes; *b*, 15 minutes; *c*, 20 minutes; *d*, 30 minutes. The low plateau value for the *d* gene probably results from close linkage with *str-s*, because in this case, the *d* allele in the Hfr strain would tend to be inherited along with the *str-r* allele.

10.28 The genetic map shows the genetic intervals defined by the deletion endpoints and the locations of the mutations with respect to these intervals.



10.30 The mean number of P2 phage per cell is 3, and the probability of a cell remaining uninfected by P2 is $e^{-3} = 0.050$; the mean number of P4 phage per cell is 2, and the probability of a cell remaining uninfected by P4 is $e^{-2} = 0.135$.

(a) $0.050 \times 0.135 \times 10^8 = 6.75 \times 10^5$.

(b) $6.75 \times 10^5 + 0.050 \times (1 - 0.135) \times 10^8 = 5.00 \times 10^6$.

(c) $(1 - 0.050) \times 0.135 \times 10^8 = 1.28 \times 10^7$.

(d) $(1 - 0.050) \times (1 - 0.135) \times 10^8 = 8.22 \times 10^7$.

Chapter 11

11.2

(a) 27.

(b) 27.

(c) 18.

11.4 UAR and UGA.

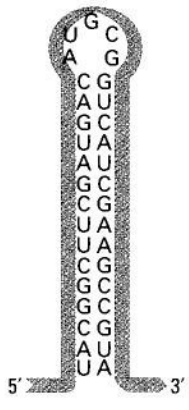
11.6 Only Met (AUG) and Trp (W).

11.8 The mRNA sequence is

5'-AAAAUGCCCUUAAUCUCAGCGUCCUAC-3'

and translation begins with the first AUG, which codes for methionine. The amino acid sequence coded by this region is Met-Pro-Leu-Ile-Ser-Ala-Ser-Tyr.

11.10 5'-GUCAUCGAAGCCGUA-3'. The stem and loop:



11.12 Codons are read from the 5' end of the mRNA, and the first amino acid in a polypeptide chain is at the amino terminus. If the G were at the 5' end, the first codon would be GAA, coding for glutamic acid at the amino terminus of the polypeptide. If the G were at the 3' end, the last codon would be AAG, coding for arginine at the carboxyl terminus. Thus the G was added at the 5' end.

11.14 The artificial mRNA can be read in any one of three reading frames, depending on where translation starts. The reading frames are

GUC GUC GUC . . ., UCG UCG UCG . . ., and CGU CGU CGU . . .

The first codes for polyvaline, the second for polyserine, and the third for polyarginine.

11.16 Methionine has 1 codon, histidine 2, and threonine 4, so the total number is $1 \times 2 \times 4 = 8$. The sequence AUGCAYACN encompasses them all. Because arginine has six codons, the possible number of sequences coding for Met-Arg-Thr is $1 \times 6 \times 4 = 24$. The sequences are AUGCGNACN (16 possibilities) and AUGAGRACN (8 possibilities).

11.18 The final four nucleotides in the open reading frame must be NUAA, NUAG, or NUGA. The last amino acid must therefore correspond to NUA (Leu, Ile, or Val, depending on the identity of N), NUA (also Leu, Ile, or Val), or NUG (Leu, Met, or Val). Hence, every polypeptide would have to terminate with Leu, Ile, Val, or Met.

11.20 The terminal nucleotides would have to be NNUAA, NNUAG, or NNUGA, with the last amino acid encoded by NNU. This formula rules out Trp, Gln, Met, Lys, or Glu as the carboxyl amino acid.

11.22 Because the pairing of codon and anticodon is antiparallel, it is convenient to write the anticodon as 3'-UAI-5'. The first two codon positions are therefore 5'-AU-3', and the third is either A, U, or C (because each can pair with I). The possible codons are 5'-AUA-3', 5'-AUU-3', and 5'-AUC-3', all of which code for isoleucine.

11.24

(a) The deletion must have fused the amino-coding terminus of the *B* gene with the carboxyl-coding terminus of the *A* gene. The nontranscribed strand must therefore be oriented 5'-*B*-*A*-3'.

(b) The number of bases deleted must be a multiple of 3; otherwise, the carboxyl terminus would not have the correct reading frame.

11.26 The probability of a correct amino acid at a particular site is 0.999, so the probability of all 300 amino acids being correctly translated is $(0.999)^{300} = 0.74$, or approximately 3/4.

11.28 If transcribed from left to right, the transcribed strand is the bottom one, and the mRNA sequence is

AGA CUU CAG GCU CAA CGU GGU

which codes for Arg-Leu-Gln-Ala-Gln-Arg-Gly. To invert the molecule, the strands must be interchanged in order to preserve the correct polarity, and so the sequence of the inverted molecule is

5'-AGAACGTTGAGCCTGAAGGGT-3'
3'-TCTTGCAACTCGGACTTCCCA-5'

This codes for an mRNA with sequence

AGA ACG UUG AGC CUG AAG GGU

which codes for Arg-Thr-Leu-Ser-Leu-Lys-Gly.

11.30

(a) 5'-UACCAUGUCAAACAGCGUAUGGUAGCAGUG-3'.

(b) Met–Ser–Asn–Ser–Val–Trp.

Chapter 12

12.2 The hemoglobin mRNA has a very long lifetime before being degraded.

12.4 The repressor gene need not be near because the repressor is diffusible. The *trp* repressor is one example in which the repressor gene is located quite far from the structural genes.

12.6 The mutant gene should bind more of the activator protein or bind it more efficiently. Therefore, the mutant gene should be induced with lower levels of the activator protein or expressed at higher levels than the wildtype gene, or both.

12.8

(a) β-galactosidase (*lacZ* product), lactose permease (*lacY* product), and the transacetylase (*lacA* product).

(b) The promoter region becomes inaccessible to RNA polymerase.

(c) Repressor synthesis is constitutive.

12.10 Inducers combine with repressors and prevent them from binding with the operator regions of the respective operons.

12.12 No; there is often a polar effect in which the relative amount of each protein synthesized decreases toward the 3' end of the mRNA.

12.14

(a) A missense mutation in A could be a missense mutation in B, a synonymous mutation in B (not changing the amino acid sequence), or a nonsense mutation.

(b) A deletion in one reading frame is also a deletion in the other.

(c) A frameshift mutation in one reading frame is also a frameshift mutation in the other.

12.16

(a) When proline is at a sufficiently low level that the aporepressor is no longer bound and transcription is derepressed.

(b) When the level of charged proline tRNA is high enough that translation proceeds through the leader sequence.

(c) When the level of charged proline tRNA is low enough that the translation stalls at the proline codons in the leader sequence.

12.18 It would cause constitutive attenuation; the tryptophan operan could never be transcribed completely, and tryptophan would be required for growth.

12.20 The α cell would be unable to switch to the **a** mating type.

12.22 The mutant diploid *MAT***a**/*MAT*α has the phenotype of an α haploid. The genotype lacks an active **a**1-α2 complex to repress the haploid-specific genes as well as α1, and the α1 gene product activates the α-specific genes.

12.24

(a) The strain is constitutive and has either the geno-type *lacI*$^-$ or *lacO*^{c}.

(b) The second mutant must be *lacZ*$^-$ because no β-galactosidase is made; and because permease synthesis is induced by lactose, the mutant must have a normal repressor-operator system. Therefore, the genotype is *lacI*$^+$ *lacO*$^+$ *lacZ*$^-$ *lacY*$^+$.

(c) Because the partial diploid is regulated, the operator in the operon with a functional *lacZ* gene must be wild-type. Hence the mutant in part **(a)** must be *lacI*$^-$.

12.26 Amino acids with codons are one step away from the nonsense codon, because their tRNA genes are the most likely to mutate to nonsense suppressors. For UGA, these are UGC (Cys), UGU (Cys), UGG (Trp), UUA (Leu), UCA (Ser), AGA (Arg), CGA (Arg), and GGA (Gly).

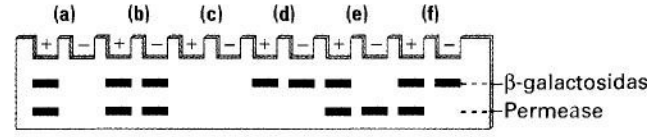
12.28

(a) Species A, all sites methylated in egg DNA, sites 1 and 3 methylated in sperm DNA; species B, sites 2 and 3 methylated in both egg DNA and sperm DNA; species C, no sites methylated in egg DNA, sites 1 and 2 methylated in sperm DNA.

(b) Species A and C show differential methylation in the gametes.

(c) Species A shows greater methylation in egg DNA.

12.30 The *lacZ*$^-$ and *lacY*$^-$ mutant alleles still produce normal-length polypeptide chains, whereas the deletions are said to produce no polypeptide product. Hence the expected pattern of bands in the Western blot are as shown here.

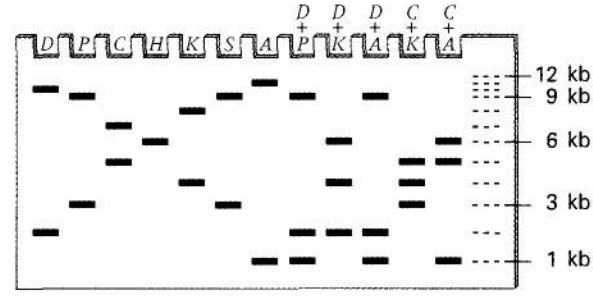


Chapter 13

13.2 64 possible sites; the chance that two different sites have compatible ends is 1/4, because the 5' overhang of one must be a Watson–Crick pairing partner with the 5' overhang of the other.

13.4 The "R" of one site must be complementary to the "Y" of the other site, so the probability is 1/2.

13.6



13.8 Genes encoding proteins that are used in the synthesis of amino acids, nucleosides and nucleotides, vitamins, and other small molecules that are not present in minimal medium but are present in complete medium.

13.10 The ends may be blunt or they may be cohesive (sticky) with either a 3' overhang or a 5' overhang.

13.12 No. The complement of 5'-GGCC-3' is 3'-CCGG-5'.

13.14

(a) The probability of a *Taq*I site is $1/6 \times 1/3 \times 1/3 \times 1/6 = 1/324$, so *Taq*I cleavage is expected every 324 base pairs; the probability of a *Mae*III site is $1/3 \times 1/6 \times 1 \times 1/6 \times 1/3 = 1/324$, so *Mae*III cleavage is expected every 324 base pairs.

(b) The answers are the same as in part **(a)**, because for both *Taq*I and *Mae*III sites, the number of A + T = the number of G + C.

13.16 Bacterial cells do not normally recognize eukaryotic promoters, and if transcription does occur, the transcript usually contains introns that the bacterial cell cannot remove. If these problems are overcome, the protein may still not be produced because the mRNA lacks a bacterial ribosome-binding site, because the protein might require post-translational processing, or because the protein might be unstable in bacterial cells and subject to degradation.

13.18 The finding is that the 3.6-kb and 5.3-kb fragments in the free phage are joined to form an 8.9-kb fragment in the intracellular form. This suggests that the the 3.6-kb and 5.3-kb fragments are terminal fragments of a linear molecule and that the intracellular form is circular.

13.20 A base substitution may destroy a restriction site and prevent two potential fragments from being separated, or a deletion may eliminate a restriction site.

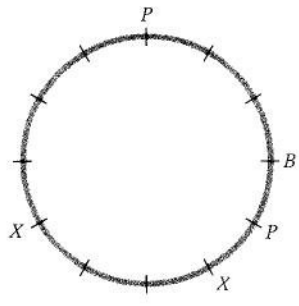
13.22

(a) The *tet-r* gene is not cleaved with *Bgl*I, so addition of tetracycline to the medium requires that the colonies be tetracycline-resistant (Tet-r) and hence contain the plasmid.

(b) Cells with the phenotype Tet-r Kan-r or Tet-r Kan-s will form colonies.

(c) Colonies with the phenotype Tet-r Kan-s contain inserts within the cleaved *kan* gene.

13.24 Each segment in the restriction map shown here is 1 kb.

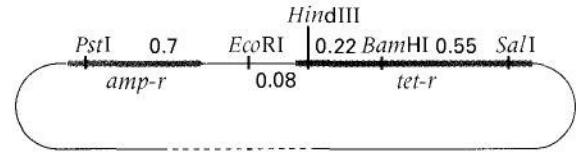


13.26 The *Pst*I fragment can be inserted in two possible orientations, only one of which will place the promotor in the correct orientation to allow transcription of the coding sequence. The two orientations are expected in equal proportions, so if only one clone is studied, it has a 50 percent chance of having the fragment inserted in the wrong orientation.

13.28 The restriction map is shown here. This type of clone would be expected to transcribe the gene of interest, because the *Pst*I fragment containing the coding sequence is in the correct orientation with respect to the promoter flanking the MCS.



13.30 The restriction map is shown in the accompanying figure.



Rest of plasmid between *Pst*I and *Sal*I = 2.45 kb

Chapter 14

14.2 Lineage mutations affect the developmental fates of one or more cells within a group of cells related by descent that normally undergo a defined pattern of development. Lineage mutations are detected by the occurrence of abnormal types or numbers of cells at characteristic positions within the embryo.

14.4 A female that is homozygous for a maternal-effect lethal produces defective eggs resulting in inviable embryos. However, the homozygous genotype itself is viable.

14.6 Cells normally destined to die adopt some other developmental fate, with the result that there are supernumerary copies of one or more cell types.

14.8 Yes, because mutations in these genes result in the wildtype floral organ identity being replaced by a different floral organ identity.

14.10 Heterochronic mutations affect the timing of developmental events relative to each other; hence a mutation that uniformly slows or accelerates development is not a heterochronic mutation.

14.12 In a loss-of-function mutation, the genetic information in a gene is not expressed in some or all cells. In a gain-of-function mutation, the genetic information is expressed at inappropriate times or in inappropriate cells. Both mutations can occur in the same gene: Some alleles may be loss-of-function and others gain-of-function. However, a particular allele must be one or the other (or neither).

14.14 Because the gene is sufficient, a gain-of-function mutation will induce the developmental fate to occur. The allele will be dominant because a normal allele in the homologous chromosome will not prevent the developmental pathway from occurring.

14.16 The loss-of-function mutation results in the floral phenotype sepal–sepal–carpel–carpel; the gain-of-function mutation results in the floral phenotype petal–petal–stamen–stamen. The results imply that Ap3/Pi is necessary for petal development in whorl 2 and for stamen development in whorl 3. In a wildtype genetic background, Ap3/Pi is sufficient for petal development in whorl 1 and for stamen development in whorl 4.

14.18 Petal–petal–stamen–carpel.

14.20 The material required for rescue is either mRNA transcribed from the maternal o^+ gene or a protein product of the gene that is localized in the nucleus.

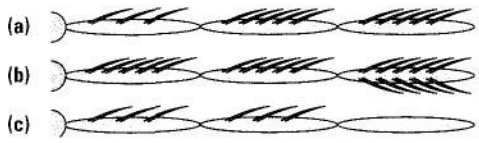
14.22 Look for embryos with a phenotype similar to *bicoid* mutants, lacking the head and thorax.

14.24 Whorls 1, 2, 3 ,and 4 yield sepal, petal, stamen, and leaf, respectively.

14.26 The phenotype of whorl 1 would probably be leaf, because that is the default state, but the phenotype of whorl 2 is not predicable, because the result of *ap3* and *pi* expression alone is not evident from the model.

14.28 a^+ is necessary for development of the medial and distal segments, b^+ is necessary for the development of the proximal segment, and c^+ is necessary for the development of the proximal and medial segments. When c^+ is missing from the medial segment, but a^+ is present in this segment, the segment is transformed to proximal, which implies that it is the interaction between a^+ and c^+ that yields the medial segment.

14.30



Chapter 15

15.2 At the G_1/S checkpoint, because of the unrepaired double-stranded DNA breaks caused by the x rays.

15.4 It binds to and sequesters the transcription factor E2F, which is needed for transcription of the genes required for DNA synthesis.

15.6 A DNA damage checkpoint, which responds to double-stranded breaks or other damage to DNA; a centrosome duplication checkpoint, which is activated unless the centrosome has duplicated and the products have moved into position; and a spindle checkpoint, which is activated when the spindle is not completely assembled or when one or more chromosomes are unattached to the spindle or improperly aligned on the spindle.

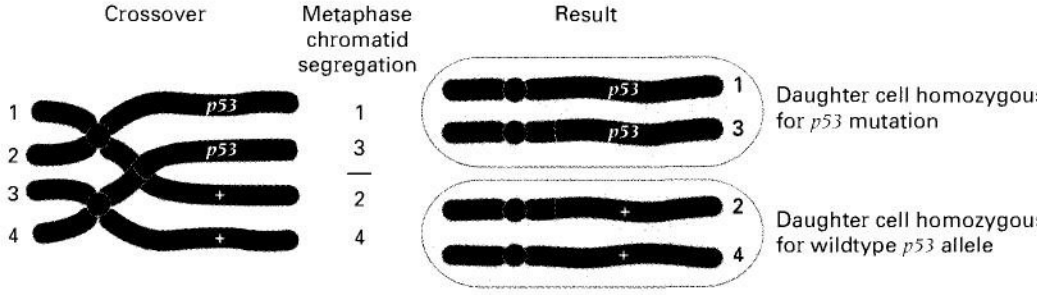
15.8 A decrease in processivity means that the polymerase complex detaches from the template after a smaller number of nucleotides have been replaced. More frequent reinitiation of DNA synthesis is therefore required, which prolongs the time spent in DNA replication. The prolongation allows greater time for repair enzymes to repair damaged DNA.

15.10 It is a "genetic disease" in that genetic mutations take place in cells and make them undergo uncontrolled growth and division. However, most cancers are caused by somatic mutations in the cells of affected individuals; hence they are not inherited from the parents (familial) but, rather, are due to new somatic mutations.

15.12 The tumors are sporadic in origin, so the DNA of the normal cells is not informative, because the changes occurred clonally in somatic cells. In patient 1, there is no indication of a change in DNA structure. The mutation is likely to be a missense or nonsense mutation. In patient 2, the DNA fragment that hybridizes to *p53* cDNA in tumor cells is smaller than the wildtype fragment; a partial deletion of the *p53* gene has occurred. In patient 3, there is no DNA present in tumor cells that is homologous to the *p53* gene; the entire gene seems to have been deleted. In patient 4, the DNA homologous to the *p53* gene is present in two fragments rather than one. A chromosomal aberration that disrupts the function of p53 must have occurred. The most likely candidates are translocation and inversion. One breakpoint in the aberration must have occurred within the *p53* gene.

15.14 Because p53 plays a key role in monitoring DNA damage through transcriptional activation of *p21*, *GADD*, and *14-3-3σ*, and in activating the pathway of apoptosis through transcriptional activation of *Bax*.

15.16 The diagram at the bottom of the page shows how mitotic recombination can cause loss of heterozygosity. The diagram also shows that all genetic markers distal to (toward the telomere side of) the site of recombination also lose heterozygosity.



Problem 15.16 Mitotic recombination

15.18 3 tumors/(4×10^6 cells) = $0.75 \times 10^{-6} = 7.5 \times 10^{-7}$ per cell. This is not exactly a "mutation" rate, because it includes such events as chromosome loss, mitotic recombination, and so forth, which are not, strictly speaking, "mutations."

15.20

(a) The smaller band should be associated with the mutant *RB1* allele, the larger with the nonmutant allele.

(b) The loss of heterozygosity is independent in different tumors; hence the mechanism need not be the same in both cases.

(c) In this case, the wildtype allele has undergone a new mutation (or perhaps a small deletion or insertion), which inactivates the gene but does not detectably change the size of the band.

15.22 The son's genotype is *AL p53*$^-$ *BL/AS p53*$^+$ *BS*.

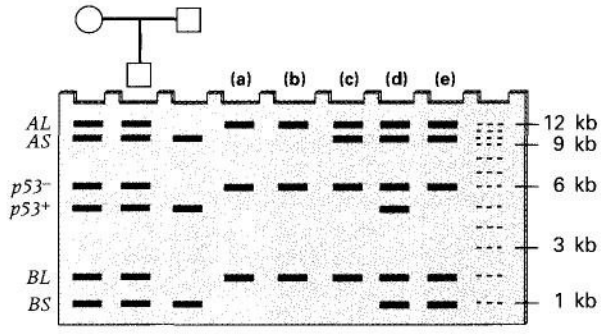
(a) The bands *AS p53*$^+$ and *BS* are lost.

(b) *AL p53*$^-$ and *BL* all become homozygous.

(c) *p53*$^-$ and *BL*, but *AL/AS* remains heterozygous.

(d) No apparent change in banding pattern from noncancerous cells.

(e) *p53*$^-$ has replaced *p53*$^+$. The expected band patterns are shown in the illustration.



15.24

(a) The cells should stop dividing at 36°C because they would no longer sense and propagate a growth-stimulatory signal (probably the presence of nutrients rather than growth factors in this example).

(b) Recessive. The Ras produced from the wildtype gene in the same cell would still propagate the signal.

(c) Such a mutant cell will continue to divide even when nutrients are exhausted. Such a mutant will die when starved, rather than cease growing.

(d) The mutation would be dominant. The mutant Ras will remain in the GTP-bound form regardless of what the wildtype Ras is doing and will ceaselessly propagate a signal to grow.

15.26 Effectively, HPV removes all the brakes on cellular proliferation, turns on all the stimuli, and removes the infected cell's ability to induce apoptosis to kill itself. The cell will effectively be *p53*$^-$, because E6 is present to ensure that levels of p53 do not rise. This means that synthesis of Bax will not be induced, and cells will not undergo apoptosis. In addition, synthesis of p21, GADD45, and 14-3-3σ will not be accelerated, so the cells will not arrest in the cell cycle. Because E7 disables RB, E2F will be active at all times, resulting in new rounds of DNA synthesis and continued cellular proliferation.

15.28 I-1 or I-2, III-3.

15.30 *RAD9* is probably not involved in the G_1/S DNA checkpoint, because the *cdc28, cdc6,* and *cdc8 rad9* double mutants survive incubation at 36°C, implying that the G_1/S checkpoint is intact in the *rad9* mutant. *RAD9* is also probably not involved in the centrosome duplication checkpoint, because the *cdc7 rad9* double mutant survives incubation at 36°C. *RAD9* is involved in the S phase DNA checkpoint, because the *cdc2, cdc13,* and *cdc17 rad9* (as well as the *cdc9 rad9*) double mutants die at 36°C. *RAD9* is probably not involved in the spindle or in the exit from mitosis checkpoints, because the *cdc16, cdc20, cdc23,* and *cdc15 rad9* double mutants survive incubation at 36°C.

Chapter 16

16.2 Because the severity depends on the proportion of defective mitochondria that are present in the affected individual.

16.4 UGA should not be used for Trp, and AUA should not be used for Ile. To encode a polypeptide of the same length, neither AGA nor AGG should be used for termination; otherwise, these would be translated in the cytoplasm as Arg and would yield a longer polypeptide chain.

16.6 All transition mutations in the third position are synonymous. Among 128 possible transversion mutations, 56, or 56/128 = 43.8 percent, are synonymous.

16.8 Pollen is not totally devoid of cytoplasm, and occasionally it contains a chloroplast that is transmitted to the zygote.

16.10 This trait is maternally inherited, so the phenotype is that of the mother. The progeny are

(a) green.

(b) white.

(c) variegated.

(d) green.

16.12 If the daughter chloroplasts segregate at random, both of them will go to the same pole half the time. For all four products to contain a descendant chloroplast, the segregation has to be perfect in three divisions, which occurs with probability $(1/2)^3 = 1/8$. Therefore, the probability that at least one cell lacks a descendant chloroplast is $1 - 1/8 = 7/8$.

16.14 The inheritance is not Mendelian because there is no evidence of segregation. The bands corresponding to the two largest DNA fragments are mutually exclusive and show maternal inheritance; the bands corresponding to the two smallest DNA fragments are mutually exclusive and show paternal inheritance. One possibility is that the smaller bands derive from mitochondrial DNA and the larger from chloroplast DNA; another is that the smaller bands derive from chloroplast DNA and the larger from mitochondrial DNA. In any event, the larger and smaller bands are maternally and paternally derived, respectively.

16.16

(a) All *cpl-r* because the chloroplast comes from the mt^+ parent.

(b) All *mit-s* because the mitochondria come from the mt^- parent.

(c) Resistant : sensitive in a ratio of 2 : 2 because a nuclear gene undergoes Mendelian segregation.

16.18 The *klv-r* versus *klv-s* difference segregates as 1 : 1, consistent with their being alleles of a gene in the nucleus. The association with mating type can be explained by linkage: The *klv* and *mt* loci are genetically linked with a frequency of recombination of $r = 20/200 = 10$ percent (10 map units).

16.20 Petite, because neither parent contributes functional mitochondria.

16.22 In autogamy, one product of meiosis becomes a homozygous diploid nucleus, so half of the *Aa* cells that undergo autogamy become *AA* and half become *aa*.

16.24 Begin by considering the situation without cytoplasmic exchange. In this case the *KK* killer exconjugant becomes *Kk* (kappa), where (kappa) means that the kappa bacterium is present in the cytoplasm, and the *kk* nonkiller exconjugant becomes *Kk* (no kappa). Immediately after autogamy, the ratios become 1/4 *KK* (kappa), 1/4 *kk* (kappa), 1/4 *KK* (no kappa), and 1/4 *kk* (no kappa). The kappa bacterium cannot be maintained in the *kk* (kappa) strain, so the overall ratio of killer : nonkiller is that of (kappa) : (no kappa), which is 1 : 3. On the other hand, with cytoplasmic exchange, the 1/4 *KK* (no kappa) and 1/4 *kk* (no kappa) become 1/4 *KK* (kappa) and 1/4 *kk* (kappa). The latter cannot maintain kappa, but the former can, so the overall ratio of killer : nonkiller with cytoplasmic exchange is 1 : 1. The cytoplasmic exchange is necessary to obtain the 1 : 1 ratio because, although the *K* allele is necessary to maintain kappa, it cannot create kappa. The kappa bacterium must be transferred from the killer partner.

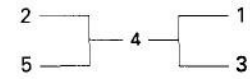
16.26 The only functional pollen is *a B*. All the progeny of the *AA bb* plants are *Aa Bb*, and those of the *aa BB* plants are *aa BB*.

16.28 The result excludes hypothesis (1) but does not distinguish between hypotheses (2) and (3).

16.30 The accompanying table shows the percent of restriction sites shared by each pair of mitochondrial DNA molecules.

	1	2	3	4	5
1	100	0	80	10	0
2		100	0	10	100
3			100	11	0
4				100	10

The entries in the table may be explained by example. Molecules 3 and 4 together have 9 restriction sites, of which 1 is shared, so the entry in the table is $1/9 = 11$ percent. The diagonal entries of 100 mean that each molecule shares 100 percent of the restriction sites with itself. The two most closely related molecules are 2 and 5 (in fact, they are identical), so these must be grouped together. The two next closest are 1 and 3, so these also must be grouped together. The 2–5 group shares an average of 10 percent restriction sites with molecule 4, and the 1–3 group shares an average of 10.5 percent restriction sites [that is, $(10 + 11)/2$] with molecule 4. Therefore, the 2–5 group and the 1–3 group are about equally distant from molecule 4. The diagram shows the inferred relationships among the species. Data of this type do not indicate where the "root" should be positioned in the tree diagram.



Chapter 17

17.2 The numbers of alleles of each type are *A*, $2 + 2 + 5 + 1 = 10$; *B*, $5 + 12 + 12 + 10 = 39$; *C*, $10 + 5 + 5 + 1 = 21$. The total number is 70, so the allele frequencies are *A*, $10/70 = 0.14$; *B*, $39/70 = 0.56$; and *C*, $21/70 = 0.30$.

17.4 Each individual has a pattern of bands exactly like his or her mother, so the most likely explanation is mitochondrial inheritance. The Hardy–Weinberg principle is not relevant to mitochondrial inheritance, because the genotype of the progeny is determined not by the genotypes of both parents but only by the genotype of the mother.

17.6 Hardy–Weinberg frequencies imply that $P = p^2$, $Q = 2pq$, and $R = q^2$. Substituting, $Q^2/4 = p^2q^2$ and $PR = p^2q^2$ also; hence $Q^2/4 = PR$.

17.8 The total number of individuals equals 1000. If we let A_1, A_2, and A_3 be the alleles associated with the single DNA fragments, then from largest to smallest, the allele frequencies are $p_1 = (2 \times 18 + 68 + 115)/2000 = 0.1095$, $p_2 = (2 \times 90 + 68 + 360)/2000 = 0.3040$, and $p_3 = (2 \times 349 + 115 + 360)/2000 = 0.5865$. The expected genotype frequences (given in the same left-to-right order as shown in the gel) are 11.99, 66.58, 92.42, 128.44, 356.59, and 343.98. The chi-square value works out to 4.62. There are 3 degrees of freedom because there are six classes of data, and two allele frequencies have been estimated from the data (only two, because the third could have been deduced by subtracting the others from 1), and $6 - 1 - 2 = 3$. The *P* value is about 0.25, so the fit to the genetic hypothesis is satisfactory.

17.10 Because the frequency of homozygous recessives is 0.10, the frequency of the recessive allele *q* is $(0.10)^{1/2} = 0.316$. This means $p = 1 - 0.316 = 0.684$, so the frequency of heterozygotes is $2 \times 0.316 \times 0.684 = 0.43$. Because the proportion 0.90 of the individuals are nondwarfs, the frequency of heterozygotes among nondwarfs is $0.43/0.90 = 0.48$.

17.12 I^AI^A, $(0.16)^2 = 0.026$; I^AI^O, $2(0.16)(0.74) = 0.236$; I^BI^B, $(0.10)^2 = 0.01$; I^BI^O, $2(0.10)(0.74) = 0.148$; I^OI^O, $(0.74)^2 = 0.548$; I^AI^B, $2(0.16)(0.10) = 0.032$. The phenotype frequencies are O, 0.548; A, $0.026 + 0.236 = 0.262$; B, $0.01 + 0.148 = 0.158$; AB, 0.032.

17.14 The frequency in males implies that $q = 0.2$. The expected frequency of the trait in females is $q^2 = 0.0004$. The expected frequency of heterozygous females is $2pq = 2(0.98)(0.02) = 0.0392$.

17.16 The frequency of heterozygotes is smaller by the amount $2pqF$, in which F is the inbreeding coefficient.

17.18 The allele frequency is the square root of the frequency of affected offspring among unrelated individuals, or $(8.5 \times 10^{-6})^{1/2} = 2.9 \times 10^{-3}$. With inbreeding, the expected frequency of homozygous recessives is $q^2(1 - F) + qF$, in which F is the inbreeding coefficient. When $F = 1/16$, the expected frequency of homozygous recessives is $(2.9 \times 10^{-3})^2(1 - 1/16) + (2.9 \times 10^{-3})(1/16) = 1.9 \times 10^{-4}$; when $F = 1/64$, the value is 5.3×10^{-5}.

17.20

(a) B, H, and I.

(b) JHEBFIK, JIFBEHK, JHK, JIK.

(c) Respective contributions to the inbreeding coefficient of L are $(1/2)^7(1 + F_B) + (1/2)^7(1 + F_B) + (1/2)^3 + (1/2)^3 = (1/2)^6(1 + F_B) + (1/2)^2$.

(d) If $F_B = 0$, this reduces to $17/64 = 0.266$.

17.22

(a) Here $F = 0.66$, $p = 0.43$, and $q = 0.57$. The expected genotype frequencies are AA, $(0.43)^2(1 - 0.66) + (0.43)(0.66) = 0.347$; AB, $2(0.43)(0.57)(1 - 0.66) = 0.167$; BB, $(0.57)^2(1 - 0.66) + (0.57)(0.66) = 0.487$.

(b) With random mating, the expected genotype frequencies are AA, $(0.43)^2 = 0.185$; AB, $2(0.43)(0.57) = 0.490$; BB, $(0.57)^2 = 0.325$.

17.24 Let p_t be the frequency of the mutant X chromosomes in generation t, with $p_0 = 1$.

(a) $p_1 = p_0(1 - 0.05) = 0.95$.

(b) $p_{10} = p_0(1 - 0.05)^{10} = 0.599$.

(c) $p_{10} = p_0(1 - 0.05)^{20} = 0.358$.

(d) Solve $p_t = p_0(1 - 0.05)^t = 0.01$; hence $t = 90$ generations.

17.26

(a) $q = \sqrt{(\mu/s)}$, where $\mu = 10^{-6}$ and $s = 1$; $q = 0.001$.

(b) $q = \mu/hs$, where μ is as before and $hs = 0.01$; in this case, $q = 0.0001$.

(c) The equilibrium frequency of a is reduced by $(0.001 - 0.0001)/0.001 = 90$ percent.

17.28

(a) A fixed.

(b) A lost.

(c) stable equilibrium.

(d) A lost.

17.30

(a) $q = 0.10/(0.50 + 0.10) = 0.167$.

(b) $2pq = 28.8$ percent.

(c) $q = 0.10/(1.00 + 0.10) = 0.091$.

(d) 16.5 percent.

Chapter 18

18.2 Since the variance equals the square of the standard deviation, the only possibilities are that the variance equals either 1 or 0.

18.4 Broad-sense heritability is the proportion of the phenotypic variance attributable to differences in genotype, or $H^2 = \sigma_g^2/\sigma_p^2$, which includes dominance effects and interactions between alleles. Narrow-sense heritability is the proportion of the phenotypic variance due only to additive effects, or σ_a^2/σ_p^2, which is used to predict the resemblance between parents and offspring in artificial selection. If all allele frequencies equal 1 or 0, both heritabilities equal zero.

18.6 For fruit weight, $\sigma_e^2/\sigma_p^2 = 0.13$ is given. Because $\sigma_p^2 = \sigma_g^2 + \sigma_e^2$, then $H^2 = \sigma_g^2/\sigma_p^2 = 1 - \sigma_e^2/\sigma_p^2 = 1 - 0.13 = 0.87$. Therefore, 87 percent is the broad-sense heritability of fruit weight. The values of H^2 for soluble-solid content and acidity are 91 percent and 89 percent, respectively. Mean 3.00, variance 1.50, standard deviation 1.22.

18.8

(a) Phenotypes 0, 2, 4, and 6 with frequencies 0.015625, 0.140625, 0.421875, and 0.421875, respectively.

(b) Mean 4.50, variance 2.25, standard deviation 1.50.

(c) Phenotypes 0, 1, 2, 3, 4, 5, and 6 with frequencies 0.0156, 0.0938, 0.2344, 0.3125, 0.2344, 0.0938, and 0.0156, respectively. Mean 3.00, variance 1.50, standard deviation 1.22

18.10

(a) To determine the mean, multiply each value for the pounds of milk produced (using the midpoint value) by the number of cows, and divide by 304 (the total number of cows) to obtain a value of the mean of 4032.9. The estimated variance is 693,633.8, and the estimated standard deviation, which is the square root of the variance, is 832.8.

(b) The range that includes 68 percent of the cows is the mean minus the standard deviation to the mean plus the standard deviation, or $4033 - 833 = 3200$ to $4033 + 833 = 4866$.

(c) Rounding to the nearest 500 yields 3000–5000. The observed number in this range is 239; the expected number is $0.68 \times 304 = 207$.

(d) For 95 percent of the animals, the range is within two standard deviations, or 2367–5698. Rounding off to the nearest 500 yields 2500–6000. The observed number is 296, which compares well to the expected number of $0.95 \times 304 = 289$.

18.12

(a) The genotypic variance is 0 because all F_1 mice have the same genotype.

(b) With additive alleles, the mean of the F_1 generation will equal the average of the parental strains.

18.14 From the F_1 data, $\sigma_e^2 = (2)^2 = 4$; from the F_2 data, $\sigma_g^2 + \sigma_e^2 = (3)^2 = 9$; hence $\sigma_g^2 = 9 - 4 = 5$, and so $H^2 = 5/9 = 56$ percent.

18.16 The offspring are genetically related as half-siblings, and Table 18.3 says that the theoretical correlation coefficient is $h^2/4$.

18.18

(a) Because *Difflugia* is an asexual organism, parent and offspring have the same genetic relationship as monozygotic twins; hence the broad-sense heritabilities are 0.956 and 0.287, respectively.

(b) The environmental component of the variance for length of longest spine is larger than that for tooth number.

18.20 $\sigma_g^2 = 0.0570 - 0.0144 = 0.0426$ and $D = 1.689 - (-0.137) = 1.826$. The minimum number of genes is estimated as $n = (1.826)^2/(8 \times 0.0426) = 9.8$.

18.22 $h^2 = (2.33 - 2.26)/(2.37 - 2.26) = 63.6$ percent.

18.24 Use Equation (6) with $M = 10.8$, $M^* = 5.8$, and $M' = 8.8$. The narrow-sense heritability is estimated as $h^2 = (8.8 - 10.8)/(5.8 - 10.8) = 0.40$.

18.26 The mean weaning weight of the 20 individuals is 81 pounds, and the standard deviation is 13.8 pounds. The mean of the upper half of the distribution can be calculated by using the relationship given in the problem, which says that the mean of the upper half of the population is $81 + 0.8 \times 13.8 = 92$ pounds. Using Equation (6), we find that the expected weaning weight of the offspring is $81 + 0.2(92 - 81) = 83.2$ pounds.

18.28

(a) 1.0, 0.67, 0.18, 0.10, 0.02, 0.01, and 0.002, respectively.

(b) With rare recessive alleles, most homozygous offspring come from matings between heterozygous genotypes, so there is essentially no parent–offspring correlation in phenotype.

18.30 Successive substitutions in this case yield $M_t = M_0 + h^2(S_1 + S_2 + S_3 + \cdots + S_{t-1})$.

Concise Dictionary of Genetics

A

aberrant 4 : 4 segregation The presence of equal numbers of alleles among the spores in a single ascus, yet with a spore pair in which the two spores have different genotypes.

acentric chromosome A chromosome with no centromere.

acridine A chemical mutagen that intercalates between the bases of a DNA molecule, causing single-base insertions or deletions.

acrocentric chromosome A chromosome with the centromere near one end.

active site The part of an enzyme at which substrate molecules bind and are converted into their reaction products.

acylated tRNA A tRNA molecule to which an amino acid is linked.

adaptation Any characteristic of an organism that improves its chance of survival and reproduction in its environment; the evolutionary process by which a species undergoes progressive modification favoring its survival and reproduction in a given environment.

addition rule The principle that the probability that any one of a set of mutually exclusive events is realized equals the sum of the probabilities of the separate events.

additive variance The magnitude of the genetic variance that results from the additive action of genes and that accounts for the genetic component of the parent-offspring correlation; the value that the genetic variance would assume if there were no dominance or interaction of alleles affecting the trait.

adenine (A) A nitrogenous purine base found in DNA and RNA.

adenyl cyclase The enzyme that catalyzes the synthesis of cyclic AMP.

adjacent segregation Segregation of a heterozygous reciprocal translocation in which a translocated chromosome and a normal chromosome segregate together, producing an aneuploid gamete.

adjacent-1 segregation Segregation from a heterozygous reciprocal translocation in which homologous centromeres go to opposite poles of the first-division spindle.

adjacent-2 segregation Segregation from a heterozygous reciprocal translocation in which homologous centromeres go to the same pole of the first-division spindle.

agarose A component of agar used as a gelling agent in gel electrophoresis; its value is that few molecules bind to it, so it does not interfere with electrophoretic movement.

agglutination The clumping or aggregation, as of viruses or blood cells, caused by an antigen-antibody interaction.

albinism Absence of melanin pigment in the iris, skin, and hair of an animal; absence of chlorophyll in plants.

alkaptonuria A recessively inherited metabolic disorder in which a defect in the breakdown of tyrosine leads to excretion of homogentisic acid (alkapton) in the urine.

alkylating agent An organic compound capable of transferring an alkyl group to other molecules.

allele Any of the alternative forms of a given gene.

allele frequency The relative proportion of all alleles of a gene that are of a designated type.

allopolyploid A polyploid formed by hybridization between two different species.

allosteric protein Any protein whose activity is altered by a change of shape induced by binding a small molecule.

allozyme Any of the alternative electrophoretic forms of a protein coded by different alleles of a single gene.

α helix A fundamental unit of protein folding in which successive amino acids form a right-handed helical structure held together by hydrogen bonding between the amino and carboxyl components of the peptide bonds in successive loops of the helix.

alpha satellite Highly repetitive DNA sequences associated with mammalian centromeres.

alternate segregation Segregation from a heterozygous reciprocal translocation in which both parts of the reciprocal translocation separate from both nontranslocated chromosomes in the first meiotic division.

amber codon Common jargon for the UAG stop codon; an amber mutation is a mutation in which a sense codon has been altered to UAG.

Ames test A bacterial test for mutagenicity; also used to screen for potential carcinogens.

amino acid Any one of a class of organic molecules that have an amino group and a carboxyl group; 20 different amino acids are the usual components of proteins.

amino acid attachment site The 3′ terminus of a tRNA molecule at which an amino acid is attached.

aminoacylated tRNA A tRNA covalently attached to its amino acid; charged tRNA.

aminoacyl site The tRNA-binding site on the ribosome to which each incoming charged tRNA is initially bound.

A site *See* **aminoacyl site.**

aminoacyl tRNA synthetase The enzyme that attaches the correct amino acid to a tRNA molecule.

amino terminus The end of a polypeptide chain at which the amino acid bears a free amino group ($-NH_2$).

amniocentesis A procedure for obtaining fetal cells from the amniotic fluid for the diagnosis of genetic abnormalities.

amorph A mutation in which the function of the gene product is completely abolished.

amplified fragment length polymorphism (AFLP) A type of DNA marker assayed by PCR amplification of genomic restriction fragments onto whose ends specific oligonucleotide sequences have been ligated to serve as primer-binding sites.

analog *See* **base analog.**

anaphase The stage of mitosis or meiosis in which chromosomes move to opposite ends of the spindle. In anaphase I of meiosis, homologous centromeres separate; in anaphase II, sister centromeres separate.

anaphase-promoting complex (APC/C) A ubiquitin-protein ligase that targets proteins whose destruction enables a cell to transition from metaphase into anaphase.

aneuploid A cell or organism in which the chromosome number is not an exact multiple of the haploid number; more generally, aneuploidy refers to a condition in which particular genes or chromosomal regions are present in extra or fewer copies compared with wildtype.

annealing The coming together of two strands of nucleic acid that have complementary base sequences to form a duplex molecule.

antibiotic-resistant mutant A cell or organism that carries a mutation conferring resistance to an antibiotic.

antibiotic-resistant mutation A mutation conferring resistance to one or more antibiotics.

antibody A blood protein produced in response to a specific antigen and capable of binding with the antigen.

anticipation A phenomenon observed in certain trinucleotide-repeat diseases in which, in each successive generation, the mean age of onset of the disease is less than in the previous generation.

anticodon The three bases in a tRNA molecule that are complementary to the three-base codon in mRNA.

antigen A substance able to stimulate the production of antibodies.

antigen presentation An immune process in which a cell carrying an antigen bound to an MHC protein stimulates a responsive cell.

antigen-presenting cell Any cell, such as a macrophage, expressing certain products of the major histocompatibility complex and capable of stimulating T cells that have appropriate antigen receptors.

antimorph A mutation whose phenotypic effects are opposite to those of the wildtype allele.

antiparallel The chemical orientation of the two strands of a double-stranded nucleic acid molecule; the 5'-to-3' orientations of the two strands are opposite one another.

AP endonuclease An endonuclease that cleaves a DNA strand at any site at which the deoxyribose lacks a base.

AP repair Repair of duplex DNA in which one or more deoxyribose sugars has lost its base and become an apurinic or apyrimidinic nucleotide.

apoptosis Genetically programmed cell death, especially in embryonic development.

aporepressor A protein converted into a repressor by binding with a particular molecule.

Archaea One of the three major classes of organisms; previously called archaebacteria, they are unicellular microorganisms, usually found in extreme environments, that differ as much from bacteria as either group differs from eukaryotes. An archaeon is a single cell or species of archaea. *See also* **Bacteria.**

artificial selection Selection, imposed by a breeder, in which organisms of only certain phenotypes are allowed to breed.

ascospore *See* **ascus.**

ascus A sac containing the spores (ascospores) produced by meiosis in certain groups of fungi, including *Neurospora* and yeast.

assortative mating Nonrandom selection of mating partners with respect to one or more traits; it is positive when like phenotypes mate more frequently than would be expected by chance and is negative when the reverse occurs.

ATP Adenosine triphosphate, the primary molecule for storing chemical energy in a living cell.

attached-X chromosome A chromosome in which two X chromosomes are joined to a common centromere; also called a compound-X chromosome.

attachment site The base sequence in a bacterial chromosome at which bacteriophage DNA can integrate to form a prophage; either of the two attachment sites that flank an integrated prophage.

attenuation *See* **attenuator.**

attenuator A regulatory base sequence near the beginning of an mRNA molecule at which transcription can be terminated; when an attenuator is present, it precedes the coding sequences.

attractant A substance that draws organisms to it through chemotaxis.

AUG The usual initiation codon for polypeptide synthesis as well as the internal codon for methionine.

autogamy Reproductive process in ciliates in which the genotype is reconstituted from the fusion of two identical haploid nuclei resulting from mitosis of a single product of meiosis; the result is homozygosity for all genes.

autonomous determination Cellular differentiation determined intrinsically and not dependent on external signals or interactions with other cells.

autopolyploidy Type of polyploidy in which there are more than two sets of homologous chromosomes.

autoradiography A process for the production of a photographic image of the distribution of a radioactive substance in a cell or large cellular molecule; the image is produced on a photographic emulsion by decay emission from the radioactive material.

autoregulation Regulation of gene expression by the product of the gene itself.

autosomes All chromosomes other than the sex chromosomes.

auxotroph A mutant microorganism unable to synthesize a compound required for its growth but able to grow if the compound is provided.

B

backcross The cross of an F_1 heterozygote with a partner that has the same genotype as one of its parents.

Bacteria One of the major kingdoms of living things; includes most bacteria. *See also* **Archaea.**

bacterial artificial chromosome (BAC) A recombinant DNA plasmid in which a large fragment of DNA has been ligated into a suitable vector, making possible its replication and segregation in bacterial cells.

bacterial attachment site *See* **attachment site.**

bacterial transformation *See* **transformation.**

bacteriophage A virus that infects bacterial cells; commonly called a phage.

band In gel electrophoresis, a compact region of heavy staining due to the accumulation of molecules of a given size; in *Drosophila* genetics, any of the horizontal striations found in the giant polytene chromosomes in the salivary glands of third in-star larvae; in metaphase chromosomes, any of the horizontal striations revealed by special staining procedures.

Barr body A darkly staining body found in the interphase nucleus of certain cells of female mammals; consists of the condensed, inactivated X chromosome.

basal transcription factors Transcription factors that are associated with transcription of a wide variety of genes.

base Single-ring (pyrimidine) or double-ring (purine) component of a nucleic acid.

base analog A chemical so similar to one of the normal bases that it can be incorporated into DNA.

base composition The relative proportions of the bases in a nucleic acid or of A + T versus G + C in duplex DNA.

base pair A pair of nitrogenous bases, most commonly one purine and one pyrimidine, held together by hydrogen bonds in a double-stranded region of a nucleic acid molecule; commonly abbreviated bp, the term is often used interchangeably with the term *nucleotide pair*. The normal base pairs in DNA are AT and GC.

base pairing Specific hydrogen bonding of A with T or of G with C in duplex DNA.

base stacking The tendency of the hydrophobic faces of nucleotides to come together in solution to minimize contact with the water molecules.

base-substitution mutation Incorporation of an incorrect base into a DNA duplex.

B cells A class of white blood cells, derived from bone marrow, with the potential for producing antibodies.

becquerel A unit of radiation equal to 1 disintegration/second, abbreviated Bq; equal to 2.7×10^{-11} curie.

behavior The observed response of an organism to stimulation or environmental change.

B-form DNA The right-handed helical structure of DNA proposed by Watson and Crick.

β-galactosidase An enzyme that cleaves lactose into its glucose and galactose constituents; produced by a gene in the *lac* operon.

β sheet A fundamental structure formed in protein folding in which two polypeptide segments are held together in an antiparallel array by hydrogen bonds.

bidirectional replication DNA replication proceeding in opposite directions from an origin of replication. *See also* **θ replication.**

binding site A DNA or RNA base sequence that serves as the target for binding by a protein; the site on the protein that does the binding.

biochemical pathway A diagram showing the order in which intermediate molecules are produced in the synthesis or degradation of a metabolite in a cell.

bivalent A pair of homologous chromosomes, each consisting of two chromatids, associated in meiosis I.

blastoderm Structure formed in the early development of an insect larva; the syncytial blastoderm is formed from repeated cleavage of the zygote nucleus without cytoplasmic division; the cellular blastoderm is formed by migration of the nuclei to the surface and their inclusion in separate cell membranes.

blastula A hollow sphere of cells formed early in development.

block In a biochemical pathway, a defective step at which the normal biochemical reaction cannot be carried out, often because of a defective enzyme encoded by a mutant gene.

blood-group system A set of antigens on red blood cells resulting from the action of a series of alleles of a single gene, such as the ABO, MN, or Rh blood-group systems.

blunt ends Ends of a DNA molecule in which all terminal bases are paired; the term usually refers to termini formed by a restriction enzyme that does not produce single-stranded ends.

branched pathway A metabolic pathway in which an intermediate serves as a precursor for more than one product.

branch migration In a DNA duplex invaded by a single strand from another molecule, the process in which the size of the heteroduplex region increases by progressive displacement of the original strand.

broad-sense heritability The ratio of genotypic variance to total phenotypic variance.

C

cAMP-CRP complex A regulatory complex consisting of cyclic AMP (cAMP) and the CAP protein; the complex is needed for transcription of certain operons.

cancer Any of a large number of diseases characterized by the uncontrolled proliferation of cells.

candidate gene A gene proposed to be involved in the genetic determination of a trait because of the role of the gene product in the cell or organism.

cap A complex structure at the 5' termini of most eukaryotic mRNA molecules, having a 5'-5' linkage instead of the usual 3'-5' linkage.

CAP protein Acronym for catabolite activator protein, the protein that binds cAMP and that regulates the activity of inducible operons in prokaryotes; also called CRP, for cAMP receptor protein.

carbon-source mutant A cell or organism that carries a mutation preventing the use of a particular molecule or class of molecules as a source of carbon.

carboxyl terminus The end of a polypeptide chain at which the amino acid has a free carboxyl group (−COOH).

carcinogen A physical agent or chemical reagent that causes cancer.

carrier A heterozygote for a recessive allele.

cascade A series of enzyme activations serving to amplify a weak chemical signal.

cassette In bacterial genetics, a circular molecule of duplex DNA containing a target sequence for a site-specific recombinase (integrase) and usually one or more protein-coding sequences. In yeast, either of two sets of inactive mating-type genes that can become active by relocating to the *MAT* locus.

catabolite activator protein *See* **CAP protein.**

***cdc* mutants** Mutants in which the cell division cycle is defective.

cDNA *See* **complementary DNA.**

cell cycle The growth cycle of a cell; in eukaryotes, it is subdivided into G_1 (gap 1), S (DNA synthesis), G_2 (gap 2), and M (mitosis).

cell cycle arrest A blockage in the progression of the cell cycle.

cell fate The pathway of differentiation that a cell normally undergoes.

cell hybrid Product of fusion of two cells in culture, often from different species.

cell lineage The ancestor-descendant relationships of a group of cells in development.

cell senescence A normal process in which mammalian cells in culture cease dividing after about 50 doublings.

cellular oncogene A gene coding for a cellular growth factor whose abnormal expression predisposes to malignancy. *See also* **oncogene.**

centimorgan A unit of distance in the genetic map equal to 1 percent recombination; also called a map unit.

central dogma The concept that genetic information is transferred from the nucleotide sequence in DNA to the nucleotide sequence in an RNA transcript to the amino acid sequence of a polypeptide chain.

centriole In animal cells, one of a pair of particulate structures composed of an array of microtubules around which the spindle is organized.

centromere The region of the chromosome that is associated with spindle fibers and that participates in normal chromosome movement during mitosis and meiosis.

centrosome A localized region of clear cytoplasm found near the nucleus of nondividing cells, which in dividing cells duplicates to form the centers around which the spindle is organized. *See also* **centriole.**

centrosome duplication checkpoint A mechanism that arrests the cell cycle while the centrosome remains undivided.

chain elongation The process of addition of successive amino acids to the growing end of a polypeptide chain.

chain initiation The process by which polypeptide synthesis is begun.

chain termination The process of ending polypeptide synthesis and releasing the polypeptide from the ribosome; a chain-termination mutation creates a new stop codon, resulting in premature termination of synthesis of the polypeptide chain.

chaperone A protein that assists in the three-dimensional folding of another protein.

Chargaff's rules In double-stranded DNA, the amount of A equals that of T, and the amount of G equals that of C.

charged tRNA A tRNA molecule to which an amino acid is linked; acylated tRNA.

checkpoint Any mechanism that arrests the cell cycle until one or more essential processes are completed.

chemoreceptor Any molecule, usually a protein, that binds with other molecules and stimulates a behavioral response, such as chemotaxis, sensation, etc. *See also* **chemosensor.**

chemosensor In bacteria, a class of proteins that bind with particular attractants or repellents. Chemosensors interact with chemoreceptors to stimulate chemotaxis. *See also* **chemotaxis.**

chemotaxis Behavioral response in which organisms move toward chemicals that attract them and away from chemicals that repel them.

chiasma The cytological manifestation of crossing-over; the cross-shaped exchange configuration between nonsister chromatids of homologous chromosomes that is visible in prophase I of meiosis. The plural can be either *chiasmata* or *chiasmas*.

chimeric gene A gene produced by recombination, or by genetic engineering, that is a mosaic of DNA sequences from two or more different genes.

chi-square (χ^2) A statistical quantity calculated to assess the goodness of fit between a set of observed numbers and the theoretically expected numbers.

chromatid Either of the longitudinal subunits produced by chromosome replication.

chromatid interference In meiosis, the effect that crossing-over between one pair of nonsister chromatids may have on the probability that a second crossing-over in the same chromosome will involve the same or different chromatids; chromatid interference does not generally occur.

chromatin The aggregate of DNA and histone proteins that makes up a eukaryotic chromosome.

chromatin-remodeling complex Any of a number of complex protein aggregates that reorganizes the nucleosomes of chromatin in preparation for transcription.

chromocenter The aggregate of centromeres and adjacent heterochromatin in nuclei of *Drosophila* larval salivary gland cells.

chromomere A tightly coiled, bead-like region of a chromosome most readily seen during pachytene of meiosis; the beads are in register in a polytene chromosome, resulting in the banded appearance of the chromosome.

chromosome In eukaryotes, a DNA molecule that contains genes in linear order to which numerous proteins are bound and that has a telomere at each end and a centromere; in prokaryotes, the DNA is associated with fewer proteins, lacks telomeres and a centromere, and is often circular; in viruses, the chromosome is DNA or RNA, single-stranded or double-stranded, linear or circular, and often free of bound proteins.

chromosome complement The set of chromosomes in a cell or organism.

chromosome interference In meiosis, the phenomenon by which a crossing-over at one position inhibits the occurrence of another crossing-over at a nearby position.

chromosome map A diagram showing the locations and relative spacing of genes along a chromosome.

chromosome painting Use of differentially labeled, chromosome-specific DNA strands for hybridization with chromosomes to label each chromosome with a different color.

chromosome theory of heredity The theory that chromosomes are the cellular objects that contain the genes.

circadian rhythm A biological cycle with an approximate 24-hour period synchronized with the cycle of day and night.

circular permutation A permutation of a group of elements in which the elements are in the same order but the beginning of the sequence differs.

***cis* configuration** The arrangement of linked genes in a double heterozygote in which both mutations are present in the same chromosome—for example, $a_1\ a_2/++$; also called coupling.

***cis*-dominant mutation** A mutation that affects the expression of only those genes on the same DNA molecule as that containing the mutation.

***cis* heterozygote** *See* ***cis* configuration.**

***cis–trans* test** *See* **complementation test.**

cistron A DNA sequence specifying a single genetic function as defined by a complementation test; a nucleotide sequence coding for a single polypeptide.

***ClB* method** A genetic procedure used to detect X-linked recessive lethal mutations in *Drosophila melanogaster;* so named because one X chromosome in the female parent is marked with an inversion (*C*), a recessive lethal allele (*l*), and the dominant allele for Bar eyes (*B*).

cleavage division Mitosis in the early embryo.

clone A collection of organisms derived from a single parent and, except for new mutations, genetically identical to that parent; in genetic engineering, the linking of a specific gene or DNA fragment to a replicable DNA molecule, such as a plasmid or phage DNA.

cloned DNA A DNA sequence incorporated into a vector molecule capable of replication in the same or a different organism.

cloning The process of producing cloned genes.

coding region The part of a DNA sequence that codes for the amino acids in a protein.

coding sequence A region of a DNA strand with the same sequence as is found in the coding region of a messenger RNA, except that T is present in DNA instead of U.

coding strand In a transcribed region of a DNA duplex, the strand that is not transcribed.

codominance The expression of both alleles in a heterozygote.

codon A sequence of three adjacent nucleotides in an mRNA molecule, specifying either an amino acid or a stop signal in protein synthesis.

coefficient of coincidence An experimental value obtained by dividing the observed number of double recombinants by the expected number calculated under the assumption that the two events take place independently.

cohesive end A single-stranded region at the end of a double-stranded DNA molecule that can adhere to a complementary single-stranded sequence at the other end or in another molecule.

cointegrate A DNA molecule, usually circular and formed by recombination, that joins two replicons.

colchicine A chemical that prevents formation of the spindle in nuclear division.

colinearity The linear correspondence between the order of amino acids in a polypeptide chain and the corresponding sequence of nucleotides in the DNA molecule.

colony A visible cluster of cells formed on a solid growth medium by repeated division of a single parental cell and its daughter cells.

colony hybridization assay A technique for identifying colonies containing a particular cloned gene; many colonies are transferred to a filter, lysed, and exposed to radioactive DNA or RNA complementary to the DNA sequence of interest, after which colonies that contain a sequence complementary to the probe are located by autoradiography.

color blindness In human beings, the usual form of color blindness is X-linked red-green color blindness. Unequal crossing-over between the adjacent red and green opsin pigment genes results in chimeric opsin genes causing mild or severe green-vision defects (deuteranomaly or deuteranopia, respectively) and mild or severe red-vision defects (protanomaly or protanopia, respectively).

combinatorial control Strategy of gene regulation in which a relatively small number of time- and tissue-specific positive and negative regulatory elements are used in various combinations to control the expression of a much larger number of genes.

combinatorial joining The DNA splicing mechanism by which antibody variability is produced, resulting in many possible combinations of V and J regions in the formation of a light-chain gene and many possible combinations of V, D, and J regions in the formation of a heavy-chain gene.

common ancestor Any ancestor of both one's father and one's mother.

compartment In *Drosophila* development, a group of descendants of a small number of founder cells with a determined pattern of development.

compatible Blood or tissue that can be transfused or transplanted without rejection.

complementary DNA (cDNA) A DNA molecule made by copying RNA with reverse transcriptase; usually abbreviated cDNA.

complementation The phenomenon in which two recessive mutations with similar phenotypes result in a wildtype phenotype when both are heterozygous in the same genotype; complementation means that the mutations are in different genes.

complementation group A group of mutations that fail to complement one another.

complementation test A genetic test to determine whether two mutations are alleles (are present in the same functional gene).

complex A term used to refer to an ordered aggregate of molecules, as in an enzyme-substrate complex or a DNA-histone complex.

complex trait A trait whose expression is determined by multiple interacting genes and usually also by environmental factors.

compound-X chromosome *See* **attached-X chromosome.**

condensation The coiling process by which chromosomes become shorter and thicker during prophase of mitosis or meiosis.

conditional mutation A mutation that results in a mutant phenotype under certain (restrictive) environmental conditions but results in a wildtype phenotype under other (permissive) conditions.

congenital Present at birth.

conjugation A process of DNA transfer in sexual reproduction in certain bacteria; in *E. coli*, the transfer is unidirectional, from donor cell to recipient cell. Also, a mating between cells of *Paramecium*.

conjugative plasmid A plasmid encoding proteins and other factors that make possible its transmission between cells.

consanguineous mating A mating between relatives.

consensus sequence A generalized base sequence derived from closely related sequences found in many locations in a genome or in many organisms; each position in the consensus sequence consists of the base found in the majority of sequences at that position.

conserved sequence A base or amino acid sequence that changes very slowly in the course of evolution.

constant region The part of the heavy and light chains of an antibody molecule that has the same amino acid sequence among all antibodies derived from the same heavy-chain and light-chain genes.

constitutive mutation A mutation that causes synthesis of a particular mRNA molecule (and the protein that it encodes) to take place at a constant rate, independent of the presence or absence of any inducer or repressor molecule.

contact inhibition A phenomenon in normal mammalian cells in culture whereby cells cease to grow and divide when they are in close physical proximity.

contig A set of cloned DNA fragments overlapping in such a way as to provide unbroken coverage of a contiguous region of the genome; a contig contains no gaps.

continuous trait A trait in which the possible phenotypes have a continuous range from one extreme to the other rather than falling into discrete classes.

coordinate gene Any of a group of genes that establish the basic anterior-posterior and dorsal-ventral axes of the early embryo.

coordinate regulation Control of synthesis of several proteins by a single regulatory element; in prokaryotes, the proteins are usually translated from a single mRNA molecule.

core particle The aggregate of histones and DNA in a nucleosome, without the linking DNA.

co-repressor A small molecule that binds with an aporepressor to create a functional repressor molecule.

correlated response Change of the mean in one trait in a population accompanying selection for another trait.

correlation coefficient A measure of association between pairs of numbers, equaling the covariance divided by the product of the standard deviations.

cosuppression A phenomenon, originally discovered in transgenic plants, in which extra copies of a gene result in a shutdown of expression of all copies that are present in the genome.

Cot Mathematical product of the initial concentration of nucleic acid (C_0) and the time (t).

Cot curve A graph of percent nucleic acid renaturation as a function of the initial concentration (C_0) multiplied by time.

cotransduction Transduction of two or more linked genetic markers by one transducing particle.

cotransformation Transformation in bacteria of two genetic markers carried on a single DNA fragment.

counterselected marker A mutation used to prevent growth of a donor cell in an Hfr $\times$ F^- bacterial mating.

coupled transcription-translation In prokaryotes, the translation of an mRNA molecule before its transcription is completed.

coupling *See* ***cis*** **configuration.**

courtship song In *Drosophila*, the regular oscillation in the time interval between pulses in successive episodes of wing vibration that the male directs toward the female during mating behavior.

covalent bond A chemical bond in which electrons are shared.

covalent circle *See* **covalently closed circle.**

covalently closed circle A ring-shaped duplex DNA molecule whose ends are joined by covalent bonds.

covariance A measure of association between pairs of numbers that is defined as the average product of the deviations from the respective means.

CpG island A region of DNA with a high density of CpG dinucleotides; in mammalian cells it is often associated with the promoter regions of genes.

C region *See* **constant region.**

Crick strand The lower strand of duplex DNA when the strands are conventionally represented as parallel horizontal lines with the 5′ end of the top strand at the extreme left.

crossing-over A process of exchange between nonsister chromatids of a pair of homologous chromosomes that results in the recombination of linked genes.

crown gall tumor Plant tumor caused by infection with *Agrobacterium tumefaciens.*

CRP protein *See* **CAP protein.**

curie A unit of radiation equal to 3.7×10^{10} disintegrations/second, abbreviated Ci; equal to 3.7×10^{10} becquerels.

cut-and-paste elements Transposable elements that are not replicated in the process of transposition but are cleaved ("cut") from an existing site in the genome and inserted ("pasted") at a new target site.

C-value paradox The often massive, counterintuitive, and seemingly arbitrary differences in genome size observed among eukaryotic organisms.

cryptic splice site A potential splice site not normally used in RNA processing unless a normal splice site is blocked or mutated.

cyclic adenosine monophosphate (cAMP) *See* **cyclic AMP.**

cyclic AMP Cyclic adenosine monophosphate, usually abbreviated cAMP, used in the regulation of cellular processes; cAMP synthesis is regulated by glucose metabolism; its action is mediated by the CAP protein in prokaryotes and by protein kinases (enzymes that phosphorylate proteins) in eukaryotes.

cyclic AMP receptor protein (CRP) *See* **CAP protein.**

cyclin Any of a number of proteins that help regulate the cell cycle and whose abundance rises and falls rhythmically during the cell cycle.

cyclin–CDK complex Protein complex formed by the interaction between a cyclin and a cyclin-dependent protein kinase (CDK).

cyclin-dependent protein kinase (CDK) Any of a number of proteins that are activated by combining with a cyclin and that regulate the cell cycle by phosphorylation of other proteins.

cytogenetics The study of the genetic implications of chromosome structure and behavior.

cytokinesis Division of the cytoplasm.

cytological map Diagrammatic representation of a chromosome.

cytoplasmic inheritance Transmission of hereditary traits through self-replicating factors in the cytoplasm—for example, mitochondria and chloroplasts.

cytoplasmic male sterility Type of pollen abortion in maize with cytoplasmic inheritance through the mitochondria.

cytosine (C) A nitrogenous pyrimidine base found in DNA and RNA.

cytotoxic T cell Type of white blood cell that participates directly in attacking cells marked for destruction by the immune system.

D

daughter strand A newly synthesized DNA or chromosome strand.

deamination Removal of an amino group ($-NH_2$) from a molecule.

decaploid An organism with ten sets of chromosomes.

deficiency *See* **deletion.**

degeneracy The feature of the genetic code in which an amino acid corresponds to more than one codon; also called redundancy.

degenerate code *See* **degeneracy.**

degree of dominance The extent to which the phenotype of a heterozygous genotype resembles one of the homozygous genotypes.

degrees of freedom An integer that determines the significance level of a particular statistical test. In the goodness-of-fit type of chi-square test in which the expected numbers are not based on any quantities estimated from the data themselves, the number of degrees of freedom is one less than the number of classes of data.

deletion Loss of a segment of the genetic material from a chromosome; also called deficiency.

deletion mapping The use of overlapping deletions to locate a gene on a chromosome or a genetic map.

deme *See* **local population.**

denaturation Loss of the normal three-dimensional shape of a macromolecule without breaking covalent bonds, usually accompanied by loss of its biological activity; conversion of DNA from the double-stranded into the single-stranded form; unfolding of a polypeptide chain.

denaturation mapping An electron-microscopic technique for localizing regions of a double-stranded DNA molecule by noting the positions at which the individual strands separate in the early stages of denaturation.

denatured DNA Duplex DNA whose strands have become partially or completely disassociated.

deoxyribonuclease An enzyme that breaks sugar-phosphate bonds in DNA, forming either fragments or the component nucleotides; abbreviated DNase.

deoxyribonucleic acid *See* **DNA.**

deoxyribose The five-carbon sugar present in DNA.

depurination Removal of purine bases from DNA.

derepression Activation of a gene or set of genes by inactivation of a repressor.

deuteranomaly *See* **color blindness.**

deuteranopia *See* **color blindness.**

deviation In statistics, a difference from an expected value.

diakinesis The substage of meiotic prophase I, immediately preceding metaphase I, in which the bivalents attain maximum shortening and condensation.

dicentric chromosome A chromosome with two centromeres.

dideoxyribose A deoxyribose sugar lacking the 3' hydroxyl group; when incorporated into a polynucleotide chain, it blocks further chain elongation.

dideoxy sequencing method Procedure for DNA sequencing in which a template strand is replicated from a particular primer sequence and terminated by the incorporation of a nucleotide that contains dideoxyribose instead of deoxyribose; the resulting fragments are separated by size via electrophoresis.

differentiation The complex changes of progressive diversification in cellular structure and function that take place in the development of an organism; for a particular line of cells, this results in a continual restriction in the types of transcription and protein synthesis of which each cell is capable.

diffuse centromere A centromere that is dispersed throughout the chromosome, as indicated by the dispersed binding of spindle fibers, rather than being concentrated at a single point.

dihybrid Heterozygous at each of two loci; progeny of a cross between true-breeding or homozygous strains that differ genetically at two loci.

dimer A protein formed by the association of two polypeptide subunits.

diploid A cell or organism with two complete sets of homologous chromosomes.

diplotene The substage of meiotic prophase I, immediately following pachytene and preceding diakinesis, in which pairs of sister chromatids that make up a bivalent (tetrad) begin to separate from each other and chiasmata become visible.

direct repeat Copies of an identical or very similar DNA or RNA base sequence in the same molecule and in the same orientation.

discontinuous variation Variation in which the phenotypic differences for a trait fall into two or more discrete classes.

disease gene Any gene with one or more alleles that increase the risk of occurrence of a disease phenotype in an individual carrying such an allele.

disjunction Separation of homologous chromosomes to opposite poles of a division spindle during anaphase of a mitotic or meiotic nuclear division.

displacement loop Loop formed when part of a DNA strand is dislodged from a duplex molecule because of the partner strand's pairing with another molecule.

distribution In quantitative genetics, the mathematical relation that gives the proportion of members in a population that have each possible phenotype.

disulfide bond Two sulfur atoms covalently linked; found in proteins when the sulfur atoms in two different cysteines are joined.

dizygotic twins Twins that result from the fertilization of separate ova and that are genetically related as siblings; also called fraternal twins.

D loop *See* **displacement loop.**

DNA Deoxyribonucleic acid, the macromolecule, usually composed of two polynucleotide chains in a double helix, that is the carrier of the genetic information in all cells and many viruses.

DNA chip An array of tiny dots ("microarray") of DNA molecules immobilized on glass or on another solid support used for hybridization with a probe of fluorescently labeled nucleic acid.

DNA cloning *See* **cloned DNA.**

DNA damage checkpoint A mechanism that arrests the cell cycle while damaged DNA remains unrepaired.

DNA fingerprinting *See* **DNA typing.**

DNA gyrase One of a class of enzymes called topoisomerases, which function during DNA replication to relax positive supercoiling of the DNA molecule and which introduce negative supercoiling into nonsupercoiled molecules early in the life cycle of many phages.

DNA ligase An enzyme that catalyzes formation of a covalent bond between adjacent 5'-P and 3'-OH termini in a broken polynucleotide strand of double-stranded DNA.

DNA looping A mechanism by which enhancers that are distant from the immediate proximity of a promoter can still regulate transcription; the enhancer and promoter, both bound with suitable protein factors, come into indirect physical contact by means of the looping out of the DNA between them. The physical interaction stimulates transcription.

DNA marker Any feature of genomic DNA that differs among individuals and that can be used to disinguish homologous DNA molecules among the individuals in a population or in a pedigree.

DNA methylase An enzyme that adds methyl groups ($-CH_3$) to certain bases, particularly cytosine.

DNA polymerase Any enzyme that catalyzes the synthesis of DNA from deoxynucleoside 5'-triphosphates, using a template strand.

DNA polymerase I (Pol I) In *E. coli* DNA replication, the enzyme that fills in gaps in the lagging strand and replaces the RNA primers with DNA.

DNA polymerase III (Pol III) The major DNA replication enzyme in *E. coli*.

DNA polymorphism Any feature of genomic DNA that frequently differs among individuals; rare mutations are not usually considered polymorphisms. *See also* **DNA marker.**

DNA repair Any of several different processes for restoration of the correct base sequence of a DNA molecule into which incorrect bases have been incorporated or whose bases have been chemically modified.

DNA replication The copying of a DNA molecule.

DNase *See* **deoxyribonuclease.**

DNA typing Electrophoretic identification of individual persons by the use of DNA probes for highly polymorphic regions of the genome, such that the genome of virtually every person exhibits a unique pattern of bands; sometimes called DNA fingerprinting.

DNA uracyl glycosylase An enzyme that removes uracil bases when they occur in double-stranded DNA.

domain A folded region of a polypeptide chain that is spatially somewhat isolated from other folded regions.

dominance Condition in which a heterozygote expresses a trait in the same manner as the homozygote for one of the alleles. The allele or the corresponding phenotypic trait expressed in the heterozygote is said to be dominant.

dominant allele An allele associated with a phenotype that is expressed when the allele is either heterozygous or homozygous.

dominance variance The part of the genotypic variance that results from the dominance effects of alleles affecting the trait.

dominant-negative mutation A mutation whose presence knocks out the function of the homologous wildtype allele; the usual mechanism is one in which mutant protein subunits combine with wildtype subunits to produce inactive multimers.

dosage compensation A mechanism regulating X-linked genes such that their activities are equal in males and females; in mammals, random inactivation of one X chromosome in females results in equal amounts of the products of X-linked genes in males and females.

double-strand break-repair model A model of recombination in which the exchange process is initiated by a duplex DNA molecule containing a double-stranded break.

double-stranded DNA A DNA molecule consisting of two antiparallel strands that are complementary in nucleotide sequence.

double-Y syndrome The clinical features of the karyotype 47,XYY.

Down syndrome The clinical features of the karyotype 47,+21 (trisomy 21).

drug-resistance plasmid A plasmid that contains genes whose products inactivate certain antibiotics.

duplex DNA A double-stranded DNA molecule.

duplication A chromosome aberration in which a chromosome segment is present more than once in the haploid genome; if the two segments are adjacent, the duplication is a tandem duplication.

dynamic mutation Term applied to certain sequences composed of tandem repeats that show genetic instability owing to a relatively high frequency of change in the number of copies of the repeat from one generation to the next.

E

ectopic expression Gene expression that occurs in the wrong tissues or the wrong cell types.

ectopic rerecombination Recombination between homologous DNA sequences present at nonhomologous sites in the genome, such as between two copies of a transposable element.

editing function The activity of DNA polymerases that removes incorrectly incorporated nucleotides; also called the proofreading function.

effector A molecule that brings about a regulatory change in a cell, as by induction.

electrophoresis A technique used to separate molecules on the basis of their different rates of movement in response to an applied electric field, typically through a gel.

electroporation Introduction of DNA fragments into a cell by means of an electric field.

elongation Addition of amino acids to a growing polypeptide chain or of nucleotides to a growing nucleic acid chain.

embryo An organism in the early stages of development (the second through seventh weeks in human beings).

embryoid A small mass of dividing cells formed from haploid cells in anthers that can give rise to a mature haploid plant.

embryonic induction Developmental process in which the fate of a group of cells is determined by interactions with nearby cells in the embryo.

embryonic stem cells Cells in the blastocyst that give rise to the body of the embryo.

endonuclease An enzyme that breaks internal phosphodiester bonds in a single- or double-stranded nucleic acid molecule; usually specific for either DNA or RNA.

endoreduplication Doubling of the chromosome complement because of chromosome replication and centromere division without nuclear or cytoplasmic division.

endosperm Nutritive tissues formed adjacent to the embryo in most flowering plants; in most diploid plants, the endosperm is triploid.

endosymbiont theory Theory that mitochondria and chloroplasts were originally free-living organisms that invaded ancestral eukaryotes, first perhaps as parasites, later becoming symbionts.

enhancer A base sequence in eukaryotes and eukaryotic viruses that increases the rate of transcription of nearby genes; the defining characteristics are that it need not be adjacent to the transcribed gene and that the enhancing activity is independent of orientation with respect to the gene.

enhancer trap Mutagenesis strategy in which a reporter gene with a weak promoter is introduced at many random locations in the genome; tissue-specific expression of the reporter gene identifies an insertion near a tissue-specific enhancer.

environmental variance The part of the phenotypic variance that is attributable to differences in environment.

enzyme A protein that catalyzes a specific biochemical reaction and is not itself altered in the process.

epigenetic Inherited changes in gene expression resulting from altered chromatin structure or DNA modification (usually methylation) rather than changes in DNA sequence.

episome A DNA element that can exist in the cell either as an autonomously replicating entity or become incorporated into the genome.

epistasis A term referring to an interaction between nonallelic genes in their effects on a trait. Generally, *epistasis* means any type of interaction in which the genotype at one locus affects the phenotypic expression of the genotype at another locus. *See also* **hypostasis.**

equational division Term applied to the second meiotic division because the haploid chromosome complement is retained throughout.

equilibrium *See* **genetic equilibrium.**

equilibrium centrifugation in a density gradient A method for separating macromolecules of differing density by high-speed centrifugation in a solution of approximately the same density as the molecules to be separated; the centrifugation produces a gradient of density inside the centrifuge tube, and each macromolecule migrates to a position in the gradient that matches its own density.

equilibrium density-gradient centrifugation *See* **equilibrium centrifugation in a density gradient.**

erythroblastosis fetalis Hemolytic disease of the newborn; blood cell destruction that occurs when anti-Rh^+ antibodies in a mother cross the placenta and attack Rh^+ cells in a fetus.

E site *See* **exit site.**

estrogen A female sex hormone; of great interest in the study of cellular regulation in eukaryotes because the number of types of target cells is large.

ethidium bromide A fluorescent molecule that binds to DNA and changes its density; used to purify supercoiled DNA molecules and to localize DNA in gel electrophoresis.

euchromatin A region of a chromosome that has normal staining properties and undergoes the normal cycle of condensation; relatively uncoiled in the interphase nucleus (compared with condensed chromosomes), it apparently contains most of the genes.

Eukarya One of the major kingdoms of living organisms, in which the cells have a true nucleus and divide by mitosis or meiosis.

eukaryote A cell with a true nucleus (DNA enclosed in a membranous envelope) in which cell division takes place by mitosis or meiosis; an organism composed of eukaryotic cells.

euploid A cell or an organism having a chromosome number that is an exact multiple of the haploid number.

evolution Cumulative change in the genetic characteristics of a species through time, resulting in greater adaptation.

excision Removal of a DNA fragment from a molecule.

excisionase An enzyme that is needed for prophage excision; works together with an integrase.

excision repair Type of DNA repair in which segments of a DNA strand that are chemically damaged are removed enzymatically and then resynthesized, using the other strand as a template.

exit site The tRNA-binding site on the ribosome that binds each uncharged tRNA just prior to its release.

exon The sequences in a gene that are retained in the messenger RNA after the introns are removed from the primary transcript.

exon shuffle The theory that new genes can evolve by the assembly of separate exons from preexisting genes, each coding for a discrete functional domain in the new protein.

exonuclease An enzyme that removes a terminal nucleotide in a polynucleotide chain by cleavage of the terminal phosphodiester bond; nucleotides are removed successively, one by one; usually specific for either DNA or RNA and for either single-stranded or double-stranded nucleic acids. A 5'-to-3' exonuclease cleaves successive nucleotides from the 5' end of the molecule; a 3'-to-5' exonuclease cleaves successive nucleotides from the 3' end.

expressivity The degree of phenotypic expression of a gene.

extranuclear inheritance Inheritance mediated by self-replicating cellular factors located outside the nucleus—for example, in mitochondria or chloroplasts.

F

familial Tending to be present in multiple generations of a pedigree.

fate In development, the final product of differentiation of a cell or of a group of cells.

fate map A diagram of the insect blastoderm identifying the regions from which particular adult structures derive.

favism Destruction of red blood cells (anemia) triggered by eating raw fava beans or by coming into contact with certain other environmental agents; associated with a deficiency of the enzyme glucose-6-phosphate dehydrogenase.

feedback inhibition Inhibition of an enzyme by the product of the enzyme or, in a metabolic pathway, by a product of the pathway.

fertility The ability to mate and produce offspring.

fertility factor *See* **F plasmid.**

fertility restorer A suppressor of cytoplasmic male sterility.

F factor *See* **F plasmid.**

F_1 generation The first generation of descent from a given mating.

F_2 generation The second generation of descent, produced by intercrossing or self-fertilizing F_1 organisms.

first-division segregation Separation of a pair of alleles into different nuclei in the first meiotic division; happens when there is no crossing-over between the gene and the centromere in a particular cell.

first meiotic division The meiotic division that reduces the chromosome number; sometimes called the reduction division.

fitness A measure of the average ability of organisms with a given genotype to survive and reproduce.

5'-P group The end of a DNA or RNA strand that terminates in a free phosphate group not connected to a sugar farther along.

five prime end (5' end) The end of a polynucleotide chain that terminates with a 5' carbon unattached to another nucleotide.

fixed allele An allele whose allele frequency equals 1.0.

flagella The whip-like projections from the surface of bacterial cells that serve as the organs of locomotion.

fluctuation test A statistical test used to determine whether bacterial mutations occur at random or are produced in response to selective agents.

folded chromosome The form of DNA in a bacterial cell in which the circular DNA is folded to have a compact structure; contains protein.

folding domain A short region of a polypeptide chain within which interactions between amino acids result in a

three-dimensional conformation that is attained relatively independently of the folding of the rest of the molecule.

forward mutation A change from a wildtype allele to a mutant allele.

founder effect Random genetic drift that results when a group of founders of a population are not genetically representative of the population from which they were derived.

F plasmid A bacterial plasmid—often called the F factor, fertility factor, or sex plasmid—that is capable of transferring itself from a host (F^+) cell to a cell not carrying an F factor (F^- cell); when an F factor is integrated into the bacterial chromosome (in an Hfr cell), the chromosome becomes transferrable to an F^- cell during conjugation.

F' plasmid An F plasmid that contains genes obtained from the bacterial chromosome in addition to plasmid genes; formed by aberrant excision of an integrated F factor, taking along adjacent bacterial DNA.

fragile-X chromosome A type of X chromosome containing a site toward the end of the long arm that tends to break in cultured cells that are starved for DNA precursors; causes fragile-X syndrome.

fragile-X syndrome A common form of inherited mental retardation associated with a type of X chromosome prone to breakage (the fragile-X chromosome).

frameshift mutation A mutational event caused by the insertion or deletion of one or more nucleotide pairs in a gene, resulting in a shift in the reading frame of all codons following the mutational site.

fraternal twins Twins that result from the fertilization of separate ova and are genetically related as siblings; also called dizygotic twins.

free radical A highly reactive molecule produced when ionizing radiation interacts with water; free radicals are potent oxidizing agents.

frequency of recombination The proportion of gametes carrying combinations of alleles that are not present in either parental chromosome.

functional genomics Use of DNA microarrays and other methods to study the coordinated expression of many genes simultaneously.

G

gain-of-function mutation Mutation in which a gene is overexpressed or inappropriately expressed.

galactosidase *See* **β-galactosidase.**

gamete A mature reproductive cell, such as sperm or egg in animals.

gametophyte In plants, the haploid part of the life cycle that produces the gametes by mitosis.

gap gene Any of a group of genes that control the development of contiguous segments or parasegments in *Drosophila* such that mutations result in gaps in the pattern of segmentation.

gastrula Stage in early animal development marked by extensive cell migration.

G bands Bands in mammalian or other chromosomes resulting from treatment with the Giemsa staining reagent.

G-E association *See* **genotype–environment association.**

G-E interaction *See* **genotype–environment interaction.**

gel electrophoresis *See* **electrophoresis.**

gene The hereditary unit containing genetic information transcribed into an RNA molecule that is processed and either functions directly or is translated into a polypeptide chain; a gene can mutate to various forms called alleles.

gene amplification A process in which certain genes undergo differential replication either within the chromosome or extrachromosomally, increasing the number of copies of the gene.

gene cloning *See* **cloned DNA.**

gene conversion The phenomenon in which the products of a meiotic division in an *Aa* heterozygous genotype are in some ratio other than the expected $1A : 1a$—for example, $3A : 1a$, $1A : 3a$, $5A : 3a$, or $3A : 5a$.

gene dosage Number of gene copies.

gene expression The multistep process by which a gene is regulated and its product synthesized.

gene flow Exchange of genes among populations resulting from either dispersal of gametes or migration of individuals; also called migration.

gene fusion Juxtaposition of parts of two or more genes to create a new genetic unit capable of expression.

gene library A large collection of cloning vectors containing a complete (or nearly complete) set of fragments of the genome of an organism.

gene pool The totality of genetic information in a population of organisms.

gene product A term used for the polypeptide chain translated from an mRNA molecule transcribed from a gene; if the RNA is not translated (for example, ribosomal RNA), the RNA molecule is the gene product.

gene regulation Processes by which gene expression is controlled in response to external or internal signals.

gene silencing Any mechanism that results in the failure of a wildtype gene to be expressed.

gene targeting Procedure for genetic engineering of genes *in vivo*, usually through homologous recombination.

gene therapy Method for treatment of disease by means of DNA rather than proteins or other pharmaceuticals.

generalized transduction *See* **transducing phage.**

general transcription factor A protein molecule needed to bind with a promoter before transcription can proceed; transcription factors are necessary, but not sufficient, for transcription, and they are shared among many different promoters.

genetic code The set of 64 triplets of bases (codons) corresponding to the twenty amino acids in proteins and the signals for initiation and termination of polypeptide synthesis.

genetic differentiation Accumulation of differences in allele frequency between isolated or partially isolated populations.

genetic divergence *See* **genetic differentiation.**

genetic engineering The linking of two DNA molecules by *in vitro* manipulations for the purpose of generating a novel organism with desired characteristics.

genetic equilibrium In population genetics, a situation in which the allele frequencies remain constant from one generation to the next.

genetic linkage *See* **linkage.**

genetic map *See* **linkage map.**

genetic mapping Procedure in which mutant alleles or DNA markers are assigned relative positions along a chromosome on the basis of the frequencies of recombination between them.

genetic marker Any pair of alleles whose inheritance can be traced through a mating or through a pedigree.

genetics The study of biological heredity.

genome The total complement of genes contained in a cell or virus; in eukaryotes, commonly used to refer to all genes present in one complete haploid set of chromosomes.

genome equivalent The number of clones of a given size that contain the same aggregate amount of DNA as the haploid genome size of an organism.

genomic DNA The totality of DNA in a cell or organism.

genomic imprinting *See* **imprinting.**

genotype The genetic constitution of an organism or virus, typically with respect to one or a few genes of interest, as distinguished from its appearance, or phenotype.

genotype–environment association The condition in which genotypes and environments are not in random combinations.

genotype–environment interaction The condition in which genetic and environmental effects on a trait are not additive.

genotype frequency The proportion of members of a population that are of a prescribed genotype.

genotypic variance The part of the phenotypic variance that is attributable to differences in genotype.

germ cell A cell that gives rise to reproductive cells.

germinal mutation A mutation in a cell from which gametes are derived, as distinguished from a somatic mutation.

germ line Cell lineage consisting of germ cells.

germ-line mutation *See* **germinal mutation.**

glucose-6-phosphate dehydrogenase (G6PD) An enzyme used in one biochemical pathway for the metabolism of glucose. In human beings, a defective G6PD enzyme (G6PD deficiency) is associated with a form of anemia in which the red blood cells are prone to burst.

glyphosate A widely used herbicide.

goodness of fit The extent to which observed numbers agree with the numbers expected on the basis of some specified genetic hypothesis.

G protein One of a family of signaling proteins that is activated by binding to a molecule of guanosine triphosphate (GTP).

G6PD *See* **glucose-6-phosphate dehydrogenase.**

G6PD deficiency *See* **glucose-6-phosphate dehydrogenase.**

G_1 period *See* **cell cycle.**

G_1 restriction point The "start" point at which cells are arrested in the G_1 phase until they become committed to division.

G_1/S transition Transition between the first "growth" phase of the cell cycle to DNA synthesis.

G_2/M transition Transition between the second "growth" phase of the cell cycle to mitosis.

G_2 period *See* **cell cycle.**

gray A unit of absorbed radiation equal to 1 joule of radiation energy absorbed per kilogram of tissue, abbreviated Gy; equal to 100 rad.

guanine (G) A nitrogenous purine base found in DNA and RNA.

guide RNA The RNA template present in telomerase.

gynandromorph A sexual mosaic; an individual organism that exhibits both male and female sexual differentiation.

gyrase A type of topoisomerase II that cleaves and rejoins both strands of a DNA duplex.

H

Haldane's mapping function A mapping function based on the assumption of no chromosome interference.

haploid A cell or organism of a species containing the set of chromosomes normally found in gametes.

haplotype The allelic form of each of a set of linked genes present in a single chromosome, particularly genes of the major histocompatibility complex.

Hardy–Weinberg principle The genotype frequencies expected with random mating.

H chain *See* **heavy chain.**

heavy (H) chain One of the large polypeptide chains in an antibody molecule.

helicase A protein that separates the strands of double-stranded DNA.

helix *See* **α helix.**

helix–loop–helix A DNA-binding motif found in many regulatory proteins.

helper T cell A type of white blood cell involved in activating other cells in the immune system.

hemizygous gene A gene present in only one dose, such as the genes on the X chromosome in XY males.

hemoglobin The oxygen-carrying protein in red blood cells.

hemophilia A One of two X-linked forms of hemophilia; patients are deficient in blood-clotting factor VIII.

heritability A measure of the degree to which a phenotypic trait can be modified by selection. *See also* **broad-sense heritability** and **narrow-sense heritability.**

heterochromatin Chromatin that remains condensed and heavily stained during interphase; commonly present adjacent to the centromere and in the telomeres of chromosomes. Some chromosomes are composed primarily of heterochromatin.

heterochronic mutation A mutation in which an otherwise normal gene is expressed at the wrong time.

heteroduplex A duplex molecule of nucleic acid whose strands are derived from different sources, such as from different homologous chromosomes or even from different organisms.

heteroduplex region Region of a double-stranded nucleic acid molecule in which the two strands have different hereditary origins; produced either as an intermediate in recombination or by the *in vitro* annealing of single-stranded complementary molecules.

heterogametic Refers to the production of dissimilar gametes with respect to the sex chromosomes; in most animals, the male is the heterogametic sex, but in birds, moths, butterflies, and some reptiles, it is the female. *See also* **homogametic.**

heterogeneous nuclear RNA (hnRNA) The collection of primary RNA transcripts and incompletely processed products found in the nucleus of a eukaryotic cell.

heterokaryon A cell or individual having nuclei from genetically different sources, the result of cell fusion not accompanied by nuclear fusion.

heteroplasmy The presence of two or more genetically different types of the same organelle in a single organism.

heterosis The superiority of hybrids over either inbred parent with respect to one or more traits; also called hybrid vigor.

heterozygote superiority The condition in which a heterozygous genotype has greater fitness than either of the homozygotes.

heterozygous Carrying dissimilar alleles of one or more genes; not homozygous.

hexaploid A cell or organism with six complete sets of chromosomes.

H4 histone *See* **histone.**

Hfr An *E. coli* cell in which an F plasmid is integrated into the chromosome, making possible the transfer of part or all of the chromosome to an F^- cell.

high-copy-number plasmid A plasmid for which there are usually considerably more than 2 (and often more than 20) copies per cell.

highly repetitive sequence A DNA sequence present in thousands of copies in a genome.

highly significant *See* **statistically significant.**

histocompatability Acceptance by a recipient of transplanted tissue from a donor.

histocompatibility antigens Tissue antigens that determine transplant compatibility or incompatibility.

histocompatibility genes Genes coding for histocompatibility antigens.

histone Any of the small basic proteins bound to DNA in chromatin; the five major histones are designated H1, H2A, H2B, H3, and H4. Each nucleosome core particle contains two molecules each of H2A, H2B, H3, and H4. The H1 histone forms connecting links between nucleosome core particles.

Holliday junction A cross-shaped configuration of two DNA duplexes formed as an intermediate in recombination.

Holliday junction resolving enzyme An enzyme that catalyzes the breakage and rejoining of two DNA strands in a Holliday junction to generate two independent duplex molecules.

Holliday model A molecular model of genetic recombination in which the participating duplexes contain heteroduplex regions of the same length.

Holliday structure A cross-shaped configuration of two DNA duplexes formed as an intermediate in recombination.

holocentric chromosome A chromosome with a diffuse centromere. *See also* **diffuse centromere.**

holoenzyme *See* **RNA polymerase holoenzyme.**

homeobox A DNA sequence motif found in the coding region of many regulatory genes; the amino acid sequence corresponding to the homeobox has a helix-loop-helix structure.

homeotic gene Any of a group of genes that determine the fundamental patterns of development. *See also* **homeotic mutation.**

homeotic mutation A developmental mutation that results in the replacement of one body structure by another body structure.

homeotic selector gene *See* **homeotic gene.**

homogametic Producing only one kind of gamete with respect to the sex chromosomes. *See also* **heterogametic.**

homogentisic acid The molecule, formerly called alkapton, that is excreted in the urine of alkaptonurics and turns black upon oxidation.

homologous In reference to DNA, having the same or nearly the same nucleotide sequence.

homologous chromosomes Chromosomes that pair in meiosis and have the same genetic loci and structure; also called homologs.

homologous recombination Genetic exchange between identical or nearly identical DNA sequences.

homothallism The capacity of cells in certain fungi to undergo a conversion in mating type to make possible mating between cells produced by the same parental organism.

homozygous Having the same allele of a gene in homologous chromosomes.

H1 histone *See* **histone.**

Hoogstein base pairing An unusual type of base pairing in which four guanine bases form a hydrogen-bonded quartet.

hormone A small molecule in higher eukaryotes, synthesized in specialized tissue, that regulates the activity of other specialized cells. In animals, hormones are transported from their source to a target tissue in the blood.

hot spot A site in a DNA molecule at which the mutation rate is much higher than the rate for most other sites.

housekeeping gene A gene that is expressed at the same level in virtually all cells and whose product participates in basic metabolic processes.

H substance The carbohydrate precursor of the A and B red-blood-cell antigens.

H3 histone *See* **histone.**

H2A histone *See* **histone.**

H2B histone *See* **histone.**

human genome project A worldwide project to map genetically and sequence the human genome.

Huntington disease Dominantly inherited degeneration of the neuromuscular system, with onset in middle age.

hybrid An organism produced by the mating of genetically unlike parents; also, a duplex nucleic acid molecule produced of strands derived from different sources.

hybrid vigor *See* **heterosis.**

hydrogen bond A weak noncovalent linkage between two negatively charged atoms in which a hydrogen atom is shared.

hydrophobic interaction A noncovalent interaction between nonpolar molecules or nonpolar groups, causing the molecules or groups to cluster when water is present.

hypermorphic mutation A mutation in which the gene product is produced in greater abundance than wildtype.

hypomorphic mutation A mutation in which the gene product is produced in less abundance than wildtype.

hypostasis The condition in which the characteristic phenotype of a mutant allele is masked by the presence of a mutant allele of a different gene. For example, in *Drosophila*, the *white* (*w*) mutation causes white eyes, and its presence thereby masks the effects of any other genotype affecting eye color, such as homozygosity for *vermilion* (*v*), *brown* (*bw*), *garnet* (*g*), *scarlet* (*st*), and so forth. In classical terminology, *w* is said to be epistatic to *v, bw, g,* and *st;* and the latter are said to be hypostatic to *w.*

I

identical twins *See* **monozygotic twins.**

identity by descent The condition in which two alleles are both replicas of a single ancestral allele in the recent past.

imaginal disk Structures present in the body of insect larvae from which the adult structures develop during pupation.

immune response Tendency to develop resistance to a disease-causing agent after an initial infection.

immunity A general term for resistance of an organism to specific substances, particularly agents of disease.

immunoglobulin One of several classes of antibody proteins.

immunoglobulin class The category in which an immunoglobulin is placed on the basis of its chemical characteristics, including the identity of its heavy chain.

imprinting A process of DNA modification in gametogenesis that affects gene expression in the zygote; one probable mechanism is the methylation of certain bases in the DNA.

inborn error of metabolism A genetically determined biochemical disorder, usually in the form of an enzyme defect that produces a metabolic block.

inbreeding Mating between relatives.

inbreeding coefficient A measure of the genetic effects of inbreeding in terms of the proportionate reduction in heterozygosity in an inbred organism compared with the heterozygosity expected with random mating.

inbreeding depression The deterioration in a population's fitness or performance that accompanies inbreeding.

incompatible Refers to blood or tissue whose transfusion or transplantation results in rejection.

incomplete dominance A situation in which the heterozygous genotype has a phenotype that falls somewhere in the range between the two homozygous genotypes.

incomplete penetrance Condition in which a mutant phenotype is not expressed in all organisms with the mutant genotype.

independent assortment Random distribution of unlinked genes into gametes, as with genes in different (nonhomologous) chromosomes or genes that are so far apart on a single chromosome that the recombination frequency between them is 1/2.

individual selection Selection based on each organism's own phenotype.

induced mutation A mutation formed under the influence of a chemical mutagen or radiation.

inducer A small molecule that inactivates a repressor, usually by binding to it and thereby altering the ability of the repressor to bind to an operator.

inducible A gene that is expressed, or an enzyme that is synthesized, only in the presence of an inducer molecule. *See also* **constitutive.**

inducible transcription *See* **inducible.**

induction Activation of an inducible gene; prophage induction is the derepression of a prophage that initiates a lytic cycle of phage development.

initiation The beginning of protein synthesis.

inosine (I) One of a number of unusual bases found in transfer RNA.

insertional inactivation Inactivation of a gene by interruption of its coding sequence; used in genetic engineering as a means of detecting insertion of a foreign DNA sequence into the coding region of a gene.

insertion sequence A DNA sequence capable of transposition in a prokaryotic genome; such sequences usually code for their own transposase.

***in situ* hybridization** Renaturation of a radioactive probe nucleic acid to the DNA or RNA in a cell; used to localize particular DNA molecules to chromosomes or parts of chromosomes and to identify the time and tissue distribution of RNA transcripts.

integrase Any site-specific recombinase that can join two DNA molecules containing the target sequence.

integration The process by which one DNA molecule is inserted intact into another replicable DNA molecule, as in prophage integration and the integration of plasmid or tumor viral DNA into a chromosome.

integron In bacteria, a DNA structure consisting of a promoter, a flanking target site for a site-specific recombinase (integrase), a coding sequence for the integrase, and usually one or more protein-coding cassettes that have been "captured" by site-specific recombination.

intercalation Insertion of a planar molecule between the stacked bases in duplex DNA.

interference The tendency for crossing-over to inhibit the formation of another crossover nearby.

intergenic complementation Complementation between mutations in different genes. *See also* **complementation test.**

intergenic suppression Suppression of a mutant phenotype because of a mutation in a different gene; often refers specifically to a tRNA molecule able to recognize a stop codon or two different sense codons.

interphase The interval between nuclear divisions in the cell cycle, extending from the end of telophase of one division to the beginning of prophase of the next division.

interrupted-mating technique In an Hfr × F^- cross, a technique by which donor and recipient cells are broken apart at specific times, allowing only a particular length of DNA to be transferred.

intervening sequence *See* **intron.**

intragenic suppression Suppression of a mutant phenotype because of a compensatory mutation in the same gene.

intron A noncoding DNA sequence in a gene that is transcribed but is then excised from the primary transcript in forming a mature mRNA molecule; found primarily in eukaryotic cells. *See also* **exon.**

inversion A structural aberration in a chromosome in which the order of several genes is reversed from the normal order. A pericentric inversion includes the centromere within the inverted region, and a paracentric inversion does not include the centromere.

inversion loop Loop structure formed by synapsis of homologous genes in a pair of chromosomes, one of which contains an inversion.

inverted repeat Either of a pair of base sequences present in the same molecule that are identical or nearly identical but are oriented in opposite directions; often found at the ends of transposable elements.

***in vitro* experiment** An experiment carried out with components isolated from cells.

***in vivo* experiment** An experiment performed with intact cells.

ionizing radiation Electromagnetic or particulate radiation that produces ion pairs when dissipating its energy in matter.

IS element *See* **insertion sequence.**

isochromosome A chromosome with two identical arms containing homologous genes.

isotopes The forms of a chemical element that have the same number of electrons and the same number of protons but differ in the number of neutrons in the atomic nucleus; unstable isotopes undergo transitions to a more stable state and, in so doing, emit radioactivity.

J

J (joining) region Any of multiple DNA sequences that code for alternative amino acid sequences of part of the variable region of an antibody molecule; the J regions of heavy and light chains are different.

just-so story Made-up stories, not supported by hard evidence, about the presumed adaptive significance of a trait; named after the 1902 book *Just So Stories* by Rudyard Kipling.

K

kappa particle An intracellular parasite in *Paramecium* that releases a substance capable of killing sensitive cells.

karyotype The chromosome complement of a cell or organism; often represented by an arrangement of metaphase chromosomes according to their lengths and the positions of their centromeres.

kilobase pair (kb) Unit of length of a duplex DNA molecule; equal to 1000 base pairs.

kindred A group of relatives.

kinetochore The cellular structure, formed in association with the centromere, to which the spindle fibers become attached in cell division.

Klinefelter syndrome The clinical features of human males with the karyotype 47,XXY.

knockout mutation Any mutation that completely eliminates the function of the gene; synonyms include *null mutation*, *loss-of-function mutation*, and *amorphic mutation*.

Kosambi's mapping function A mapping function based on the assumption that the coincidence between crossovers is proportional to the distance between them.

kuru A neurological disease caused by a prion protein.

L

***lac* operon** The set of genes required to metabolize lactose in bacteria.

Lac repressor A protein coded by the *lacI* gene that, in the absence of lactose, binds with the operator and prevents transcription.

lactose Milk sugar; each molecule consists of a joined glucose and galactose.

lactose permease An enzyme responsible for transport of lactose from the environment into bacteria.

lagging strand The DNA strand whose complement is synthesized in short fragments that are ultimately joined together.

lariat structure Structure of an intron immediately after excision in which the 5' end loops back and forms a 5'-2' linkage with another nucleotide.

L chain *See* **light chain.**

leader *See* **leader polypeptide, leader sequence.**

leader polypeptide A short polypeptide encoded in the leader sequence of some operons coding for enzymes in amino acid biosynthesis; translation of the leader polypeptide participates in regulation of the operon through attenuation.

leader sequence The region of an mRNA molecule from the 5' end to the beginning of the coding sequence, sometimes containing regulatory sequences; in prokaryotic mRNA, it contains the ribosomal binding site.

leading strand The DNA strand whose complement is synthesized as a continuous unit.

leptotene The initial substage of meiotic prophase I during which the chromosomes become visible in the light microscope as unpaired thread-like structures.

lethal mutation A mutation that results in the death of an affected organism.

liability Risk, particularly toward a threshold type of quantitative trait.

library *See* **gene library.**

ligand The molecule that binds to a specific receptor.

light (L) chain One of the small polypeptide chains in an antibody molecule.

likelihood ratio The ratio of the probabilities of obtaining a set of observed data under two different hypotheses—for example, linkage at some specified value of recombination versus no linkage.

lineage The ancestral history of a cell in development or of a species in evolution.

lineage diagram A diagram of cell lineages and their developmental fates.

linkage The tendency of genes located in the same chromosome to be associated in inheritance more frequently than expected from their independent assortment in meiosis.

linkage equilibrium The condition in which the alleles of different genes in a population are present in gametes in proportion to the product of the frequencies of the alleles.

linkage group The set of genes present together in a chromosome.

linkage map A diagram of the order of genes in a chromosome in which the distance between adjacent genes is proportional to the rate of recombination between them; also called a genetic map.

linker In genetic engineering, synthetic DNA fragments that contain restriction-enzyme cleavage sites that are used to join two DNA molecules.

local population A group of organisms of the same species occupying an area within which most individual members find their mates; synonomous terms are *deme* and *Mendelian population.*

localized centromere The type of centromere organization found in most eukaryotic cells in which the centromere is located in one small region of the chromosome.

locomotor activity Movement of organisms from one place to another under their own power.

locus The site or position of a particular gene on a chromosome.

lod score The log-odds, or logarithm of the likelihood ratio; in genetic linkage studies, the logarithm is usually taken to the base 10.

loss-of-function mutation A mutation that eliminates gene function; also called a null mutation.

loss of heterozygosity Loss of the presence of the wild-type allele, or loss of its function, in a heterozygous cell, enabling the phenotype of a recessive mutant allele to be expressed; mechanisms for loss of heterozygosity include chromosome loss, gene conversion, and mutation.

lost allele An allele no longer present in a population; its frequency is 0.

LTR retrotransposon A type of transposable element that transposes via an RNA intermediate and that contains long terminal repeats (LTRs) in direct orientation at its ends.

lysis Breakage of a cell caused by rupture of its cell membrane and cell wall.

lysogen Clone of bacterial cells that have acquired a prophage.

lysogenic cycle In temperate bacteriophage, the phenomenon in which the DNA of an infecting phage becomes part of the genetic material of the cell.

lysozyme One of a class of enzymes that dissolves the cell wall of bacteria; found in chicken egg white, human tears, and coded by many phages.

lytic cycle The life cycle of a phage, in which progeny phage are produced and the host bacterial cell is lysed.

M

macrophage One of a class of white blood cells that processes antigens for presentation as a necessary step in stimulation of the immune response.

major groove In B-form DNA, the larger of two continuous indentations running along the outside of the double helix.

major histocompatibility complex (MHC) The group of closely linked genes coding for antigens that play a major role in tissue incompatibility and that function in regulation and other aspects of the immune response.

map-based cloning A strategy of gene cloning based on the position of a gene in the genetic map; also called positional cloning.

mapping function The mathematical relation between the genetic map distance across an interval and the observed percentage of recombination in the interval.

map unit A unit of distance in a linkage map that corresponds to a recombination frequency of 1 percent. Technically, the map distance across an interval in map units equals one-half the average number of crossovers in the interval expressed as a percentage. Map units are sometimes denoted centimorgans (cM).

masked mRNA Messenger RNA that cannot be translated until specific regulatory substances are available; present in eukaryotic cells, particularly eggs; storage mRNA.

maternal effect A phenomenon in which the genotype of a mother affects the phenotype of the offspring through substances present in the cytoplasm of the egg.

maternal effect gene A gene that influences early development through its expression in the mother and the presence of the gene product in the oocyte.

maternal inheritance Extranuclear inheritance of a trait through cytoplasmic factors or organelles contributed by the female gamete.

maternal sex ratio Aberrant sex ratio in certain *Drosophila* species caused by cytoplasmic transmission of a parasite that is lethal to male embryos.

mating system The norms by which members of a population choose their mates; important systems of mating include random mating, assortative mating, and inbreeding.

mating-type interconversion Phenomenon in homothallic yeast in which cells switch mating type as a result of the transposition of genetic information from an unexpressed cassette into the active mating-type locus.

Maxam–Gilbert method A technique for determining the nucleotide sequence of DNA by means of strand cleavage at the positions of particular nucleotides.

MCS Multiple cloning site. *See also* **polylinker.**

MCP *See* **methyl-accepting chemotaxis protein.**

mean The arithmetic average.

megabase pair Unit of length of a duplex nucleic acid molecule; equal to 1 million base pairs.

meiocyte A germ cell that undergoes meiosis to yield gametes in animals or spores in plants.

meiosis The process of nuclear division in gametogenesis or sporogenesis in which one replication of the chromosomes is followed by two successive divisions of the nucleus to produce four haploid nuclei.

melting curve A graph of the amount of denatured DNA present in a solution (measured by UV light adsorption) as a function of increasing temperature.

melting temperature The midpoint of the narrow temperature range at which the strands of duplex DNA denature.

Mendelian genetics The mechanism of inheritance in which the statistical relations between the distribution of traits in successive generations result from (1) particulate hereditary determinants (genes), (2) random union of gametes, and (3) segregation of unchanged hereditary determinants in the reproductive cells.

Mendelian population *See* **local population.**

meristem The mitotically active growing point of plant tissue.

meristic trait A trait in which the phenotype is determined by counting, such as number of ears on a stalk of corn or number of eggs laid by a hen.

merodiploid A bacterial cell carrying two copies of a region of the genome and therefore diploid for the genes in this region.

messenger RNA (mRNA) An RNA molecule transcribed from a DNA sequence and translated into the amino acid sequence of a polypeptide. In eukaryotes, the primary transcript undergoes elaborate processing to become the mRNA.

metabolic pathway A set of chemical reactions that take place in a definite order to convert a particular starting molecule into one or more specific products.

metabolite Any small molecule that participates in metabolism as a substrate, product, or cofactor.

metacentric chromosome A chromosome with its centromere about in the middle so that the arms are equal or almost equal in length.

metallothionein Any of a class of proteins that bind heavy metals.

metaphase In mitosis, meiosis I, or meiosis II, the stage of nuclear division in which the centromeres of the condensed chromosomes are arranged in a plane between the two poles of the spindle.

metaphase plate Imaginary plane, equidistant from the spindle poles in a metaphase cell, on which the centromeres of the chromosomes are aligned by the spindle fibers.

methyl-accepting chemotaxis protein Alternative designation of chemoreceptors in bacterial chemotaxis; abbreviated MCP.

methylation The modification of a protein or a DNA or RNA base by the addition of a methyl group ($-CH_3$).

methylation induced premeiotically (MIP) An epigenetic mechanism of gene silencing found in certain filamentous fungi in which genes present in multiple copies are heavily methylated prior to the onset of meiosis.

MHC *See* **major histocompatibility complex.**

micrococcal nuclease A nuclease that is isolated from a bacterium and that cleaves double-stranded DNA without regard to sequence.

microsatellite A type of DNA marker based on a short sequence (2–9 base pairs) that is present at one or more sites in the genome and is repeated in tandem at each site.

middle repetitive DNA sequence A DNA sequence present tens to hundreds of times per genome.

migration Movement of organisms among subpopulations; also, the movement of molecules in electrophoresis.

minimal medium A growth medium consisting of simple inorganic salts, a carbohydrate, vitamins, organic bases, essential amino acids, and other essential compounds; its composition is precisely known. Minimal medium contrasts with complex medium or broth, which is an extract of biological material (vegetables, milk, meat) that contains a large number of compounds and the precise composition of which is unknown.

minisatellite A type of DNA marker based on a sequence of 10–60 base pairs that is present at one or more sites in the genome and is repeated in tandem at each site.

minor groove In B-form DNA, the smaller of two continuous indentations running along the outside of the double helix.

mismatch An arrangement in which two nucleotides opposite one another in double-stranded DNA are unable to form hydrogen bonds.

mismatch repair Removal of one nucleotide from a pair that cannot properly hydrogen-bond, followed by replacement with a nucleotide that can hydrogen-bond.

missense mutation An alteration in a coding sequence of DNA that results in an amino acid replacement in the polypeptide.

mitosis The process of nuclear division in which the replicated chromosomes divide and the daughter nuclei have the same chromosome number and genetic composition as the parent nucleus.

mitotic spindle The set of fibers that arches through the cell during mitosis and to which the chromosomes are attached. *See also* **spindle.**

mobile DNA Any DNA sequence capable of moving from one position to another in the genome; usually refers to transposable elements.

monocistronic mRNA A mRNA molecule that codes for a single polypeptide chain.

monoclonal antibody Antibody directed against a single antigen and produced by a single clone of B cells or a single cell line of hybridoma cells.

monohybrid A genotype that is heterozygous for one pair of alleles; the offspring of a cross between genotypes that are homozygous for different alleles of a gene.

monomer Any of the individual polypeptide chains (subunits) that make up a protein composed of multiple polypeptide chains.

monomorphic gene A gene for which the most common allele is virtually fixed.

monoploid The basic chromosome set that is reduplicated to form the genomes of the species in a polyploid series; the smallest haploid chromosome number in a polyploid series.

monosomic Condition of an otherwise diploid organism in which one member of a pair of chromosomes is missing.

monozygotic twins Twins developed from a single fertilized egg that splits into two embryos at an early division; also called identical twins.

morphogen Substance that induces differentiation.

mosaic An organism composed of two or more genetically different types of cells.

M period *See* **cell cycle.**

mRNA *See* **messenger RNA.**

multifactorial trait A trait determined by the combined action of many factors, typically some genetic and some environmental.

multiple alleles The presence in a population of more than two alleles of a gene.

multiple cloning site *See* **polylinker.**

multiplication rule The principle that the probability that all of a set of independent events are realized simultaneously equals the product of the probabilities of the separate events.

multivalent An association of more than two homologous chromosomes resulting from synapsis during meiosis in a polysomic or polyploid individual.

mutagen An agent that is capable of increasing the rate of mutation.

mutagenesis The process by which a gene undergoes a heritable alteration; also, the treatment of a cell or organism with a known mutagen.

mutant Any heritable biological entity that differs from wildtype, such as a mutant DNA molecule, mutant allele, mutant gene, mutant chromosome, mutant cell, mutant organism, or mutant heritable phenotype; also, a cell or organism in which a mutant allele is expressed.

mutant screen A type of genetic experiment in which the geneticist seeks to isolate multiple new mutations that affect a particular trait.

mutation A heritable alteration in a gene or chromosome; also, the process by which such an alteration happens. Used incorrectly, but with increasing frequency, as a synonym for mutant, even in some excellent textbooks.

mutation pressure The generally very weak tendency of mutation to change allele frequency.

mutation rate The probability of a new mutation in a particular gene, either per gamete or per generation.

N

narrow-sense heritability The fraction of the phenotypic variance revealed as resemblance between parents and offspring; technically, the ratio of the additive genetic variance to the total phenotypic variance.

native conformation The three-dimensional shape that a molecule is thought to acquire inside the cell.

natural selection The process of evolutionary adaptation in which the genotypes genetically best suited to survive and reproduce in a particular environment give rise to a disproportionate share of the offspring and so gradually increase the overall ability of the population to survive and reproduce in that environment.

nature versus nurture Archaic term for genetics versus environment; usually a false dichotomy because most traits are influenced by both genetic and environmental factors.

negative assortative mating *See* **assortative mating.**

negative chromatid interference In meiosis, when two or more crossovers occur in a chromosome, a tendency for the nonsister chromatids that participate in one event to also participate in another event.

negative regulation Regulation of gene expression in which mRNA is not synthesized until a repressor is removed from the DNA of the gene.

negative supercoiling Supercoiling of DNA that counteracts underwinding of the DNA.

neomorph A mutation in which the mutant allele expresses either a novel gene product or a mutant gene product that has a novel phenotypic effect.

neutral allele An allele that has a negligible effect on the ability of the organism to survive and reproduce.

nick A single-strand break in a DNA molecule.

nicked circle A circular DNA molecule containing one or more single-strand breaks.

nitrous acid HNO_2, a chemical mutagen.

noncomplementation Failure of two mutant alleles to show complementation; noncomplementation implies that the alleles are mutant forms of the same gene.

nondisjunction Failure of chromosomes to separate (disjoin) and move to opposite poles of the division spindle; the result is loss or gain of a chromosome.

nonhistone chromosomal proteins A large class of proteins, not of the histone class, found in isolated chromosomes.

non-LTR retrotransposon A type of transposable element that transposes via an RNA intermediate and that does not contain long terminal repeats (LTRs) at its ends.

nonparental ditype An ascus containing two pairs of recombinant spores.

nonpermissive conditions Environmental conditions that result in expression of the phenotype of a conditional mutation.

nonselective medium A growth medium that allows growth of wildtype and of one or more mutant genotypes.

nonsense mutation A mutation that changes a codon specifying an amino acid into a stop codon, resulting in

premature polypeptide chain termination; also called a chain-termination mutation.

nonsense suppression Suppression of the phenotype of a polypeptide chain-termination mutation mediated by a mutant tRNA that inserts an amino acid at the site of the termination codon.

nonsense suppressor Any mutation that suppresses the phenotype of a chain-terminating ("nonsense") mutation; nonsense suppressors are usually mutations in tRNA genes.

nonsynonymous mutation A nucleotide substitution in a codon that changes the amino acid specified by the codon.

nontaster Individual with the inherited inability to taste the substance phenylthiocarbamide (PTC); phenotype of the homozygous recessive.

normal distribution A symmetrical bell-shaped distribution characterized by the mean and the variance; in a normal distribution, approximately 68 percent of the observations are within 1 standard deviation from the mean, and approximately 95 percent are within 2 standard deviations.

nuclease An enzyme that breaks phosphodiester bonds in nucleic acid molecules.

nucleic acid A polymer composed of repeating units of phosphate-linked five-carbon sugars to which nitrogenous bases are attached. *See also* **DNA** and **RNA.**

nucleic acid hybridization The formation of duplex nucleic acid from complementary single strands.

nucleoid A DNA mass, not bounded by a membrane, within the cytoplasm of a prokaryotic cell, chloroplast, or mitochondrion; often refers to the major DNA unit in a bacterium.

nucleolar organizer region A chromosome region containing the genes for ribosomal RNA; abbreviated NOR.

nucleolus (*pl.* nucleoli) Nuclear organelle in which ribosomal RNA is made and ribosomes are partially synthesized; usually associated with the nucleolar organizer region. A nucleus may contain several nucleoli.

nucleoside A purine or pyrimidine base covalently linked to a sugar.

nucleosome The basic repeating subunit of chromatin, consisting of a core particle composed of two molecules each of four different histones around which a length of DNA containing about 145 nucleotide pairs is wound, joined to an adjacent core particle by about 55 nucleotide pairs of linker DNA associated with a fifth type of histone.

nucleosome core particle A unit of chromatin structure composed of two molecules each of histones H2A, H2B, H3, and H4, around which is wound a segment of duplex DNA of about 145 nucleotide pairs.

nucleotide A nucleoside phosphate.

nucleotide analog A molecule that is structurally similar to a normal nucleotide and that is incorporated into DNA.

nucleotide substitution A mutation in which one nucleotide is replaced with a different nucleotide; in duplex DNA the partner nucleotide in the complementary strand is also changed to preserve the Watson–Crick base pairing.

nucleus The organelle, bounded by a membranous envelope, that contains the chromosomes in a eukaryotic cell.

nullisomic A cell or individual that contains no copies of a particular chromosome.

null mutation Any mutation that completely eliminates the function of the gene; synonyms include *knockout mutation, loss-of-function mutation,* and *amorphic mutation.*

nutritional mutation A mutation in a metabolic pathway that creates a need for a substance to be present in the growth medium or that eliminates the ability to utilize a substance present in the growth medium.

O

ochre codon Jargon for the UAA stop codon; an ochre mutation is a UAA codon formed from a sense codon.

octoploid A cell or organism with eight complete sets of chromosomes.

Okazaki fragment Any of the short strands of DNA produced during discontinuous replication of the lagging strand; also called a precursor fragment.

oligomerization domain A region of a polypeptide chain that binds to a region of another polypeptide chain to form a protein with two or more subunits.

oligonucleotide A short, single-stranded nucleic acid, usually synthesized for use in DNA sequencing , as a primer in the polymerase chain reaction, as a probe, or for oligonucleotide site-directed mutagenesis.

oligonucleotide primer A short, single-stranded nucleic acid synthesized for use in DNA sequencing or as a primer in the polymerase chain reaction.

oligonucleotide site-directed mutagenesis Replacement of a small region in a gene with an oligonucleotide that contains a specified mutation.

oncogene A gain-of-function mutation in a cellular gene, called a proto-oncogene, whose normal function is to promote cellular proliferation or inhibit apoptosis; oncogenes are often associated with tumor progression.

open reading frame In the coding strand of DNA or in mRNA, a region containing a series of codons uninterrupted by stop codons and therefore capable of coding for a polypeptide chain; abbreviated ORF.

operator A regulatory region in DNA that interacts with a specific repressor protein in controlling the transcription of adjacent structural genes.

operon A collection of adjacent structural genes regulated by an operator and a repressor.

ORF *See* **open reading frame.**

organelle A membrane-bounded cytoplasmic structure that has a specialized function, such as a chloroplast or mitochondrion.

origin of replication A DNA base sequence at which replication of a molecule is initiated.

origin recognition complex (ORC) A protein complex that initiates the formation of pre-replication complexes at origins of replication in eukaryotic cells.

outcrossing A mating system in which mating pairs are not genetically closely related; avoidance of close inbreeding.

overdominance A condition in which the fitness of a heterozygote is greater than the fitness of both homozygotes.

overlapping genes Genes that share part of their coding sequences.

ovule The structure in seed plants that contains the embryo sac (female gametophyte) and that develops into a seed after fertilization of the egg.

P

pachytene The middle substage of meiotic prophase I in which the homologous chromosomes are closely synapsed.

pair-rule gene Any of a group of genes active early in *Drosophila* development that specifies the fates of alternating segments or parasegments. Mutations in pair-rule genes result in loss of even-numbered or odd-numbered segments or parasegments.

palindrome In nucleic acids, a segment of DNA in which the sequence of bases on complementary strands reads the same from a central point of symmetry—for example 5'-GAATTC-3'; the sites of recognition and cleavage by restriction endonucleases are frequently palindromic.

paracentric inversion An inversion that does not include the centromere.

parasegment Developmental unit in *Drosophila* consisting of the posterior part of one segment and the anterior part of the next segment in line.

parental combination Alleles present in an offspring chromosome in the same combination as that found in one of the parental chromosomes.

parental ditype An ascus containing two pairs of nonrecombinant spores.

parental strand In DNA replication, the strand that served as the template in a newly formed duplex.

partial digestion Condition in which a restriction enzyme cleaves some, but not all, of the restriction sites present in a DNA molecule.

partial diploid A cell in which a segment of the genome is duplicated, usually in a plasmid.

partial dominance A condition in which the phenotype of the heterozygote is intermediate between the corresponding homozygotes but resembles one more closely than the other.

Pascal's triangle Triangular configuration of integers in which the nth row gives the binomial coefficients in the expansion of $(x + y)^{n-1}$. The first and last numbers in each row equal 1, and the others equal the sum of the adjacent numbers in the row immediately above.

paternal inheritance Extranuclear inheritance of a trait through cytoplasmic factors or organelles contributed by the male gamete.

path in a pedigree A closed loop in a pedigree, with only one change of direction, that passes through a common ancestor of the parents of an individual.

pathogen An organism that causes disease.

pattern formation The creation of a spatially ordered and differentiated embryo from a seemingly homogeneous egg cell.

PCR *See* **polymerase chain reaction.**

pedigree A diagram representing the familial relationships among relatives.

***P* element** A cut-and-paste type of transposable element in *Drosophila* used for germ-line transformation and other types of *in vivo* genetic engineering.

penetrance The proportion of organisms having a particular genotype that actually express the corresponding phenotype. If the phenotype is always expressed, penetrance is complete; otherwise, it is incomplete.

peptide bond A covalent bond between the amino group ($-NH_2$) of one amino acid and the carboxyl group ($-COOH$) of another.

peptidyl site The tRNA-binding site on the ribosome to which the tRNA bearing the nascent polypeptide becomes bound immediately after formation of the peptide bond.

peptidyl transferase The enzymatic activity of ribosomes responsible for forming a peptide bond.

percent G + C The proportion of base pairs in duplex DNA that consist of GC pairs.

pericentric inversion An inversion that includes the centromere.

permease A membrane-associated protein that allows a specific small molecule to enter the cell.

permissive condition An environmental condition in which the phenotype of a conditional mutation is not expressed; contrasts with nonpermissive or restrictive condition.

petite mutant A mutation in yeast resulting in slow growth and small colonies.

PEV *See* **position effect variegation.**

PFGE *See* **pulse-field gel electrophoresis.**

p53 transcription factor An important protein that helps regulate a mammalian cell's response to stress, especially to DNA damage.

P_1 generation The parents used in a cross, or the original parents in a series of generations; also called the P generation if there is no chance of confusion with the grandparents or more remote ancestors.

P group *See* **5'-P group.**

phage *See* **bacteriophage.**

phage-attachment site The base sequence in a bacterial chromosome at which bacteriophage DNA can integrate to form a prophage.

phage repressor Regulatory protein that prevents transcription of genes in a prophage.

phenotype The observable properties of a cell or an organism, which result from the interaction of the genotype and the environment.

phenotypic variance The total variance in a phenotypic trait in a population.

phenylalanine hydroxylase (PAH) The enzyme that converts phenylalanine to tyrosine and that is defective in phenylketonuria.

phenylketonuria (PKU) A hereditary human condition resulting from inability to convert phenylalanine into tyrosine; causes severe mental retardation unless treated in infancy and childhood by a low-phenylalanine diet; abbreviated PKU.

phenylthiocarbamide *See* **PTC tasting.**

Philadelphia chromosome Abnormal human chromosome 22, resulting from reciprocal translocation and often associated with a type of leukemia.

phosphodiester bond In nucleic acids, the covalent bond formed between the 5'-phosphate group (5'-P) of one nucleotide and the 3'-hydroxyl group (3'-OH) of the next nucleotide in line; these bonds form the backbone of a nucleic acid molecule.

photoreactivation The enzymatic splitting of pyrimidine dimers produced in DNA by ultraviolet light; requires visible light and the photoreactivation enzyme.

phylogenetic tree A diagram showing the genealogical relationships among a set of genes or species.

physical map A diagram showing the relative positions of physical landmarks in a DNA molecule; common landmarks include the positions of restriction sites and particular DNA sequences.

plaque A clear area in an otherwise turbid layer of bacteria growing on a solid medium, caused by the infection and killing of the cells by a phage; because each plaque results from the growth of one phage, plaque counting is a way of counting viable phage particles. The term is also used for animal viruses that cause clear areas in layers of animal cells grown in culture.

plasmid An extrachromosomal genetic element that replicates independently of the host chromosome; it may exist in one or many copies per cell and may segregate in cell division to daughter cells in either a controlled or a random fashion. Some plasmids, such as the F factor, may become integrated into the host chromosome.

pleiotropic effect Any phenotypic effect that is a secondary manifestation of a mutant gene.

pleiotropy The condition in which a single mutant gene affects two or more distinct and seemingly unrelated traits.

point mutation A mutation caused by the substitution, deletion, or addition of a single nucleotide pair.

polarity The 5'-to-3' orientation of a strand of nucleic acid.

pole cell Any of a group of cells, set off at the posterior end of the *Drosophila* embryo, from which the germ cells are derived.

Pol III holoenzyme A large protein complex containing the major DNA polymerase that elongates each growing DNA strand during replication.

Pol II holoenzyme A large protein complex containing the type of RNA polymerase used in transcribing most protein-coding genes.

poly-A tail The sequence of adenines added to the 3' end of many eukaryotic mRNA molecules in processing.

polycistronic mRNA An mRNA molecule from which two or more polypeptides are translated; found primarily in prokaryotes.

polygenic inheritance The determination of a trait by alleles of two or more genes.

polylinker A short DNA sequence that is present in a vector and that contains a number of unique restriction sites suitable for gene cloning.

polymer A regular, covalently bonded arrangement of basic subunits or monomers into a large molecule, such as a polynucleotide or polypeptide chain.

polymerase An enzyme that catalyzes the covalent joining of nucleotides—for example, DNA polymerase and RNA polymerase.

polymerase α The major DNA replication enzyme in eukaryotic cells.

polymerase chain reaction Repeated cycles of DNA denaturation, renaturation with primer oligonucleotide sequences, and replication, resulting in exponential growth in the number of copies of the DNA sequence located between the primers.

polymerase γ The enzyme responsible for replication of mitochondrial DNA.

polymorphic gene A gene for which there is more than one relatively common allele in a population.

polymorphism The presence in a population of two or more relatively common forms of a gene, chromosome, or genetically determined trait.

polynucleotide chain A polymer of covalently linked nucleotides.

polypeptide *See* **polypeptide chain.**

polypeptide chain A polymer of amino acids linked together by peptide bonds.

polyploidy The condition of a cell or organism with more than two complete sets of chromosomes.

polyprotein A protein molecule that can be cleaved to form two or more finished protein molecules.

polyribosome *See* **polysome.**

polysome A complex of two or more ribosomes associated with an mRNA molecule and actively engaged in polypeptide synthesis; a polyribosome.

polysomy The condition of a diploid cell or organism that has three or more copies of a particular chromosome.

polytene chromosome A giant chromosome consisting of many identical strands laterally apposed and in register, exhibiting a characteristic pattern of transverse banding.

P1 artificial chromosome A recombinant DNA plasmid in which a large fragment of DNA has been ligated into a vector derived from the *E. coli* bacteriophage P1.

P1 bacteriophage Temperate virus that infects *E. coli*; in cells lysogenic for P1, the phage DNA is not integrated into the chromosome and replicates as a plasmid.

population A group of organisms of the same species.

population genetics Application of Mendel's laws and other principles of genetics to entire populations of organisms.

population structure *See* **population substructure.**

population substructure Organization of a population into smaller breeding groups between which migration is restricted. Also called population subdivision.

positional cloning A strategy of gene cloning based on the position of a gene in the genetic map; also called map-based cloning.

positional information Developmental signals transmitted to a cell by virtue of its position in the embryo.

position effect A change in the expression of a gene depending on its position within the genome.

position-effect variegation Mosaic phenotype (variegation) due to variation in the level of expression of a gene in different cell lineages owing to its position in the genome.

positive assortative mating *See* **assortative mating.**

positive chromatid interference In meiosis, when two or more crossovers occur in a chromosome, a tendency for the nonsister chromatids that participate in one event not to participate in another event.

positive regulation Mechanism of gene regulation in which an element must be bound to DNA in an active form to allow transcription. Positive regulation contrasts with negative regulation, in which a regulatory element must be removed from DNA.

postmeiotic segregation Segregation of genetically different products in a mitotic division following meiosis, as in the formation in *Neurospora* of a pair of ascospores that have different genotypes.

postreplication repair DNA repair that takes place in nonreplicating DNA or after the replication fork is some distance beyond a damaged region.

post-transcriptional cosuppression A type of gene silencing in which transcripts of the silenced genes are produced but the gene products are not.

posttranslocation state The state of the ribosome immediately following the movement of the small subunit one codon farther along the mRNA; the aminoacyl (A) site is unoccupied.

precursor fragment *See* **Okazaki fragment.**

prediction equation In quantitative genetics, an equation used to predict the improvement in mean performance of a population brought about by artifical selection; it always includes heritability as one component.

pre-replication complex Protein complexes assembled at origins of replication that are activated by certain cyclin–CDKs, triggering the initiation of DNA synthesis.

pretranslocation state The state of the ribosome immediately prior to the movement of the small subunit one codon farther along the mRNA; the exit (E) site is unoccupied.

Pribnow box A base sequence in prokaryotic promoters to which RNA polymerase binds in an early step of initiating transcription.

primary transcript An RNA copy of a gene; in eukaryotes, the transcript must be processed to form a translatable mRNA molecule.

primase The enzyme responsible for synthesizing the RNA primer needed for initiating DNA synthesis.

primer In nucleic acids, a short RNA or single-stranded DNA segment that functions as a growing point in polymerization.

primer adapter A short, double-stranded oligonucleotide ligated to the end of a DNA duplex for use as a priming site for amplification by the polymerase chain reaction.

primer oligonucleotide A single-stranded DNA molecule, typically from 18 to 22 nucleotides in length, that can hybridize with a longer DNA strand and serve as a primer for replication. In the polymerase chain reaction, two primers anneal to opposite strands of a DNA duplex with their 3′-OH ends facing. *See also* **polymerase chain reaction.**

primosome The enzyme complex that forms the RNA primer for DNA replication in eukaryotic cells.

prion protein Protein infectious agent; a protein that can transmit disease.

probability A mathematical expression of the degree of confidence that certain events will or will not be realized.

probe A radioactive DNA or RNA molecule used in DNA-RNA or DNA-DNA hybridization assays.

processing A series of chemical reactions in which primary RNA transcripts are converted into mature mRNA, rRNA, or tRNA molecules or in which polypeptide chains are converted into finished proteins.

processivity Refers to the number of consecutive nucleotides in a template strand of nucleic acid that are traversed before a DNA polymerase or an RNA polymerase detaches from the template.

programmed cell death Cell death that happens as part of the normal developmental process. *See also* **apoptosis.**

prokaryote An organism that lacks a nucleus; prokaryotic cells divide by fission.

promoter A DNA sequence at which RNA polymerase binds and initiates transcription.

promoter fusion Joining of the promoter region of one gene with the protein-coding region of another.

promoter mutation A mutation that increases or decreases the ability of a promoter to initiate transcription.

promoter recognition The first step in transcription.

proofreading function *See* **editing function.**

prophage The form of phage DNA in a lysogenic bacterium; the phage DNA is repressed and usually integrated into the bacterial chromosome, but some prophages are in plasmid form.

prophage induction Activation of a prophage to undergo the lytic cycle.

prophase The initial stage of mitosis or meiosis, beginning after interphase and terminating with the alignment of the chromosomes at metaphase; often absent or abbreviated between meiosis I and meiosis II.

protanomaly *See* **color blindness.**

protanopia *See* **color blindness.**

protein A molecule composed of one or more polypeptide chains.

protein folding The process by which a polypeptide chain conforms itself in three dimensions to attain its native conformation.

protein subunit Any of the polypeptide chains in a protein molecule made up of multiple polypeptide chains.

proto-oncogene A eukaryotic gene that functions to promote cellular proliferation or inhibit apoptosis, in which gain-of-function mutations (oncogenes) are associated with tumor progression.

prototroph Microbial strain capable of growth in a defined minimal medium that ideally contains only a carbon

source and inorganic compounds. The wildtype genotype is usually regarded as a prototroph.

pseudoautosomal region In mammals, a small region of the X and Y chromosome containing homologous genes.

pseudogene A DNA sequence that is not expressed because of one or more mutations but that has a functional counterpart in the same organism; pseudogenes are regarded as mutated forms of ancient gene duplications.

P site *See* **peptidyl site.**

PTC tasting Genetic polymorphism in the ability to taste phenylthiocarbamide.

***P* transposable element** A *Drosophila* transposable element used for the induction of mutations, germ-line transformation, and other types of genetic engineering.

Punnett square A sort of cross-multiplication matrix used in the prediction of the outcome of a genetic cross, in which male and female gametes and their frequencies are arrayed along the edges.

pulse-field gel electrophoresis A type of electrophoresis in which the electric field is manipulated to make possible the separation of large DNA fragments.

purine An organic base found in nucleic acids; the predominant purines are adenine and guanine.

pyrimidine An organic base found in nucleic acids; the predominant pyrimidines are cytosine, uracil (in RNA only), and thymine (in DNA only).

pyrimidine dimer Two adjacent pyrimidine bases, typically a pair of thymines, in the same polynucleotide strand, between which chemical bonds have formed; the most common lesion formed in DNA by exposure to ultraviolet light.

Q

QTL *See* **quantitative trait locus.**

quantitative trait A trait—typically measured on a continuous scale, such as height or weight—that results from the combined action of several or many genes in conjunction with environmental factors.

quantitative trait locus A locus segregating for alleles that have different, measurable effects on the expression of a quantitative trait.

R

race A genetically or geographically distinct subgroup of a species.

rad Radiation absorbed dose; an amount of ionizing radiation resulting in the dissipation of 100 ergs of energy in 1 gram of matter; equal to 0.01 gray.

random amplified polymorphic DNA (RAPD) A type of DNA marker identified by a band in a gel after amplification of genomic DNA by the polymerase chain reaction using one or more short oligonucleotide primers of random sequence.

random genetic drift Fluctuation in allele frequency from generation to generation resulting from restricted population size.

random mating System of mating in which mating pairs are formed independently of genotype and phenotype.

random spore analysis In fungi, the genetic analysis of spores collected at random rather than from individual tetrads.

reading frame The phase in which successive triplets of nucleotides in mRNA form codons; depending on the reading frame, a particular nucleotide in an mRNA could be in the first, second, or third position of a codon. The reading frame actually used is defined by the AUG codon that is selected for chain initiation.

reannealing Reassociation of dissociated single strands of DNA to form a duplex molecule.

recessive Refers to an allele, or the corresponding phenotypic trait, expressed only in homozygotes.

reciprocal cross A cross in which the sexes of the parents are the reverse of another cross.

reciprocal translocation Interchange of parts between nonhomologous chromosomes.

recombinant A chromosome that results from crossing-over and that carries a combination of alleles differing from that of either chromosome participating in the crossover; the cell or organism containing a recombinant chromosome.

recombinant DNA A DNA molecule produced by the joining of two or more DNA fragments that are not normally in physical proximity and are often from different organisms.

recombinant DNA technology Procedures for creating DNA molecules composed of one or more segments from other DNA molecules.

recombination Exchange of parts between DNA molecules or chromosomes; recombination in eukaryotes usually entails a reciprocal exchange of parts, but in prokaryotes it is often nonreciprocal.

recombination repair Repair of damaged DNA by exchange of good for bad segments between two damaged molecules.

recruitment The process in which a transcriptional activator protein interacts with one or more components of the transcription complex and attracts it to the promoter.

red–green color blindness *See* **color blindness.**

reductional division Term applied to the first meiotic division because the chromosome number (counted as the number of centromeres) is reduced from diploid to haploid.

redundancy The feature of the genetic code in which an amino acid corresponds to more than one codon; also called degeneracy.

regulatory gene A gene with the primary function of controlling the expression of one or more other genes.

rejection An immune response against transfused blood or transplanted tissue.

relative fitness The fitness of a genotype expressed as a proportion of the fitness of another genotype.

relative risk The ratio of the probabilities of observing a particular phenotype, usually a disease, among individuals that differ in genotype.

relaxed DNA A DNA circle whose supercoiling has been removed either by introduction of a single-strand break or by the activity of a topoisomerase.

release factor A protein that functions in translation in the release of a completed polypeptide chain from the ribosome.

rem The quantity of any kind of ionizing radiation that has the same biological effect as one rad of high-energy gamma rays; rem stands for roentgen equivalent man.

renaturation Restoration of the normal three-dimensional structure of a macromolecule; in reference to nucleic acids, the term means the formation of a double-stranded molecule by complementary base pairing between two single-stranded molecules.

repair *See* **DNA repair.**

repair synthesis The enzymatic filling of a gap in a DNA molecule at the site of excision of a damaged DNA segment.

repellent Any substance that elicits chemotaxis of organisms in a direction away from itself.

repetitive sequence A DNA sequence present more than once per haploid genome.

replica plating Procedure in which a particular spatial pattern of colonies on an agar surface is reproduced on a series of agar surfaces by stamping them with a template that contains an image of the pattern; the template is often produced by pressing a piece of sterile velvet upon the original surface, which transfers cells from each colony to the cloth.

replication *See* **DNA replication; θ replication.**

replication fork In a replicating DNA molecule, the region in which nucleotides are added to growing strands.

replication origin The base sequence at which DNA synthesis begins.

replication slippage The process in which the number of copies of a small tandem repeat can increase or decrease during replication.

replicon A DNA molecule that has a replication origin.

reporter gene A gene whose expression can be monitored.

repression A regulatory process in which a gene is temporarily rendered unable to be expressed.

repressor A protein that binds specifically to a regulatory sequence adjacent to a gene and blocks transcription of the gene.

repulsion *See* ***trans*** **configuration.**

restriction endonuclease A nuclease that recognizes a short nucleotide sequence (restriction site) in a DNA molecule and cleaves the molecule at that site; also called a restriction enzyme.

restriction enzyme *See* **restriction endonuclease.**

restriction fragment A segment of duplex DNA produced by cleavage of a larger molecule by a restriction enzyme.

restriction fragment length polymorphism (RFLP) Genetic variation in a population associated with the size of restriction fragments that contain sequences homologous to a particular probe DNA; the polymorphism results from the positions of restriction sites flanking the probe, and each variant is essentially a different allele.

restriction map A diagram of a DNA molecule showing the positions of cleavage by one or more restriction endonucleases.

restriction site The base sequence at which a particular restriction endonuclease makes a cut.

restrictive condition A growth condition in which the phenotype of a conditional mutation is expressed.

retinoblastoma An inherited cancer caused by a mutation in the tumor-suppressor gene located in chromosome band *13q14*. Inheritance of one copy of the mutation results in multiple malignancies in retinal cells of the eyes in which the mutation becomes homozygous—for example, through a new mutation or mitotic recombination.

retinoblastoma protein Any of a family of proteins found in animal cells that functions to hold cells at the G_1/S restriction point ("start") by binding to and sequestering a transcription factor that initiates the cell cycle.

retrovirus One of a class of RNA animal viruses that cause the synthesis of DNA complementary to their RNA genomes on infection.

reverse genetics Procedure in which mutations are deliberately produced in cloned genes and introduced back into cells or the germ line of an organism.

reverse mutation A mutation that undoes the effect of a preceding mutation.

reverse transcriptase An enzyme that makes complementary DNA from a single-stranded RNA template.

reverse transcriptase PCR (RT-PCR) Amplification of a duplex DNA molecule originally produced by reverse trascriptase using an RNA template.

reversion Restoration of a mutant phenotype to the wild-type phenotype by a second mutation.

RFLP *See* **restriction fragment length polymorphism.**

R group *See* **side chain.**

Rh Rhesus blood-group system in human beings; maternal-fetal incompatibility of this system may result in hemolytic disease of the newborn.

Rh-negative Refers to the phenotype in which the red blood cells lack the D antigen of the Rh blood-group system.

rhodopsin One of a class of proteins that function as visual pigments.

Rh-positive Refers to the phenotype in which the red blood cells possess the D antigen of the Rh blood-group system.

ribonuclease Any enzyme that cleaves phosphodiester bonds in RNA; abbreviated RNase.

ribonucleic acid *See* **RNA.**

ribonucleoprotein particle Nuclear particle containing a short RNA molecule and several proteins; involved in intron excision and splicing and other aspects of RNA processing.

ribose The five-carbon sugar in RNA.

ribosomal RNA (rRNA) RNA molecules that are components of the ribosomal subunits; in eukaryotes, there are four rRNA molecules—5S, 5.8S, 18S, and 28S; in prokaryotes, there are three—5S, 16S, and 23S.

ribosome The cellular organelle on which the codons of mRNA are translated into amino acids in protein synthesis. Ribosomes consist of two subunits, each composed of RNA and proteins. In prokaryotes, the subunits are 30S

and 50S particles; in eukaryotes, they are 40S and 60S particles.

ribosome-binding site The base sequence in a prokaryotic mRNA molecule to which a ribosome can bind to initiate protein synthesis; also called the Shine–Dalgarno sequence.

ribozyme An RNA molecule able to catalyze one or more biochemical reactions.

ring chromosome A chromosome whose ends are joined; one that lacks telomeres.

risk factor Any genetic or environmental agent that increases the risk of a particular phenotype, usually a disease.

RIP (repeat-induced point mutation) An mutational mechanism of gene silencing found in certain filamentous fungi in which genes present in multiple copies undergo multiple nucleotide substitutions.

RNA Ribonucleic acid, a nucleic acid in which the sugar constituent is ribose; typically, RNA is single-stranded and contains the four bases adenine, cytosine, guanine, and uracil.

RNA editing Process in which certain nucleotides in RNA are chemically changed to other nucleotides after transcription.

RNA interference (RNAi) The ability of small fragments of double-stranded RNA to silence genes whose transcripts contain homologous sequences.

RNA polymerase An enzyme that makes RNA by copying the base sequence of a DNA strand.

RNA polymerase holoenzyme A large protein complex in eukaryotic nuclei consisting of Pol II in combination with at least 9 other subunits; Pol II itself includes 12 subunits.

RNA processing The conversion of a primary transcript into an mRNA, rRNA, or tRNA molecule; includes splicing, cleavage, modification of termini, and (in tRNA) modification of internal bases.

RNase Any enzyme that cleaves RNA.

RNA splicing Excision of introns and joining of exons.

Robertsonian translocation A chromosomal aberration in which the long arms of two acrocentric chromosomes become joined to a common centromere.

roentgen A unit of ionizing radiation, defined as the amount of radiation resulting in 2.083×10^9 ion pairs per cm^3 of dry air at 0°C and 1 atm pressure; abbreviated R; 1 R produces 1 electrostatic unit of charge per cm^3 of dry air at 0°C and 1 atm pressure.

rolling-circle replication A mode of replication in which a circular parent molecule produces a linear branch of newly formed DNA.

Rous sarcoma virus A type of retrovirus that infects the chicken.

R plasmid A bacterial plasmid that carries drug-resistance genes; commonly used in genetic engineering.

rRNA *See* **ribosomal RNA.**

RT-PCR *See* **reverse transcriptase PCR.**

S

S *See* **cell cycle.**

salvage pathway A minor source of deoxythymidine triphosphate (dTTP) for DNA synthesis that uses the enzyme thymidine kinase (TK).

satellite DNA Eukaryotic DNA that forms a minor band at a different density from that of most of the cellular DNA in equilibrium density gradient centrifugation; consists of short sequences repeated many times in the genome (highly repetitive DNA) or of mitochondrial or chloroplast DNA.

scaffold A protein-containing material in chromosomes, believed to be responsible in part for the compaction of chromatin.

second-division segregation Segregation of a pair of alleles into different nuclei in the second meiotic division, the result of crossing-over between the gene and the centromere of the pair of homologous chromosomes.

second meiotic division The meiotic division in which the centromeres split and the chromosome number is not reduced; also called the equational division.

segment Any of a series of repeating morphological units in a body plan.

segmentation gene Any of a group of genes that determines the spatial pattern of segments and parasegments in *Drosophila* development.

segment-polarity gene Any of a group of genes that determines the spatial pattern of development within the segments of *Drosophila* larvae.

segregation Separation of the members of a pair of alleles into different gametes in meiosis.

segregational petites Slow-growing yeast colonies deficient in respiration because of mutation in a chromosomal gene; inheritance is Mendelian.

segregation mutation Developmental mutation in which sister cells express the same fate when they should express different fates.

selected marker A genetic mutation that allows growth in selective medium.

selection In evolution, intrinsic differences in the ability of genotypes to survive and reproduce; in plant and animal breeding, the choosing of organisms with certain phenotypes to be parents of the next generation; in mutation studies, a procedure designed in such a way that only a desired type of cell can survive, as in selection for resistance to an antibiotic.

selection coefficient The amount by which relative fitness is reduced or increased.

selection differential In artificial selection, the difference between the mean of the selected organisms and the mean of the population from which they were chosen.

selection limit The condition in which a population no longer responds to artificial selection for a trait.

selection-mutation balance Equilibrium determined by the opposing effects of mutation tending to increase the frequency of a deleterious allele and selection tending to decrease its frequency.

selection pressure The tendency of natural or artificial selection to change allele frequency.

selectively neutral mutation A mutation that has no (or negligible) effects on fitness.

selective medium A medium that allows growth only of cells with particular genotypes.

self-fertilization The union of male and female gametes produced by the same organism.

selfish DNA DNA sequences that do not contribute to the fitness of an organism but are maintained in the genome through their ability to replicate and transpose.

semiconservative replication The usual mode of DNA replication, in which each strand of a duplex molecule serves as a template for the synthesis of a new complementary strand, and the daughter molecules are composed of one old (parental) and one newly synthesized strand.

semisterility A condition in which a significant proportion of the gametophytes produced by a plant or of the zygotes produced by an animal are inviable, as in the case of a translocation heterozygote.

sense strand Originally defined as the nontemplate strand in duplex DNA, which has the same nucleotide sequence as the transcript (although U replaces T in the RNA). Some later authors used the term in the opposite sense to mean the template DNA strand. Because both usages persist, the term is ambiguous and is best avoided. Use *template strand* or *transcribed strand* instead.

sequence-tagged site (STS) A DNA sequence, present once per haploid genome, that can be amplified by the use of suitable oligonucleotide primers in the polymerase chain reaction in order to identify clones that contain the sequence.

sex chromosome A chromosome, such as the human X or Y, that has a role in the determination of sex.

sex-influenced trait A trait whose expression depends on the sex of the individual.

sex-limited trait A trait expressed in one sex and not in the other.

sex-linked Refers to a trait determined by a gene on a sex chromosome, usually the X.

Shine–Dalgarno sequence *See* **ribosome-binding site.**

shotgun sequencing Method of genomic sequencing in which sequenced clones are chosen at random and the sequences are assembled by computer according to their overlaps.

shuttle vector A vector capable of replication in two or more organisms—for example, yeast and *E. coli*.

sib *See* **sibling.**

sibling A brother or sister, each having the same parents.

sibship A group of brothers and sisters.

sickle-cell anemia A severe anemia in human beings inherited as an autosomal recessive and caused by an amino acid replacement in the β-globin chain; heterozygotes tend to be more resistant to falciparum malaria than are normal homozygotes.

side chain In protein structure, the chemical group attached to the α-carbon atom of an amino acid by which the chemical properties of each amino acid are determined; also called R group.

sievert A unit of radiation exposure of biological tissue equal to the dose in gray multiplied by a quality factor that measures the relative biological effectiveness of the type of radiation; abbreviated Sv. Roughly speaking, 1 sievert is the amount of radiation equivalent in its biological effects to one gray of gamma rays; 1 Sv equals 100 rem.

sigma (σ) subunit The subunit of RNA polymerase needed for promoter recognition.

signal transduction The process by which a regulatory signal ultimately determines cell fate by activation of a series of regulatory proteins.

significant *See* **statistically significant.**

silencer A nucleotide sequence that binds with certain proteins whose presence prevents gene expression.

silent mutation A mutation that has no phenotypic effect.

simple sequence length polymorphism (SSLP) Tandem repeats of a short DNA sequence whose repeat number differs among homologous chromosomes.

simple tandem repeat polymorphism (STRP) A DNA polymorphism in a population in which the alleles differ in the number of copies of a short, tandemly repeated nucleotide sequence.

single-active-X principle In mammals, the genetic inactivation of all X chromosomes except one in each cell lineage, except in the very early embryo.

single-copy sequence A DNA sequence present only once in each haploid genome.

single nucleotide polymorphism (SNP) A DNA marker in which a single nucleotide pair differs in the DNA sequence of homologous chromosomes and in which each of the alternative sequences occurs relatively frequently.

single-stranded DNA binding protein (SSBP) A protein able to bind single-stranded DNA.

single-stranded DNA A DNA molecule that consists of a single polynucleotide chain.

sister chromatid exchange Recombination between duplex DNA molecules present in sister chromatids; it normally has no genetic consequence and does not occur in some organisms.

sister chromatids Chromatids produced by replication of a single chromosome.

site-directed mutagenesis Method of engineering mutations at specific nucleotide sites or in specific regions of DNA.

site-specific exchange *See* **site-specific recombination.**

site-specific recombinase An enzyme that catalyzes intermolecule recombination between two duplex DNA molecules at the site of a target sequence that they have in common.

site-specific recombination Recombination that is catalyzed by a particular enzyme and that always takes place at the same site between the same DNA sequences.

snRNP Abbreviation for *small nuclear ribonucleoprotein particle*.

small ribonucleoprotein particle (snRNP) Small nuclear particle that contains short RNA molecules and several proteins. They are involved in intron excision and splicing and other aspects of RNA processing.

somatic cell Any cell of a multicellular organism other than the gametes and the germ cells from which gametes develop.

somatic-cell genetics Study of somatic cells in culture.

somatic mosaic An organism that contains two genetically different types of cells.

somatic mutation A mutation arising in a somatic cell.

SOS repair An inducible, error-prone system for repair of DNA damage in *E. coli*.

Southern blot A nucleic acid hybridization method in which, after electrophoretic separation, denatured DNA is transferred from a gel to a membrane filter and then exposed to radioactive DNA or RNA under conditions of renaturation; the radioactive regions locate the homologous DNA fragments on the filter.

spacer sequence A noncoding base sequence between the coding segments of polycistronic mRNA or between genes in DNA.

specialized transduction *See* **transducing phage.**

species Genetically, a group of actually or potentially inbreeding organisms that is reproductively isolated from other such groups.

S period *See* **cell cycle.**

spindle A structure composed of fibrous proteins on which chromosomes align during metaphase and move during anaphase.

spindle checkpoint A mechanism that arrests the cell division cycle until the spindle is properly deployed.

splice acceptor The 5' end of an exon.

splice donor The 3' end of an exon.

spliceosome An RNA-protein particle in the nucleus in which introns are removed from RNA transcripts.

spontaneous mutation A mutation that happens in the absence of any known mutagenic agent.

sporadic An instance of a disease that is solitary, lacking other affected members in the same pedigree.

spore A unicellular reproductive entity that becomes detached from the parent and can develop into a new organism upon germination; in plants, spores are the haploid products of meiosis.

sporophyte The diploid, spore-forming generation in plants, which alternates with the haploid, gamete-producing generation (the gametophyte).

standard deviation The square root of the variance.

start codon An mRNA codon, usually AUG, at which polypeptide synthesis begins.

statistically significant Said of the result of an experiment or study that has only a small probability of happening by chance on the assumption that some hypothesis is true. Conventionally, if results as bad or worse would be expected less than 5 percent of the time, the result is said to be statistically significant; if less than 1 percent of the time, the result is called statistically highly significant; both outcomes cast the hypothesis into serious doubt.

sticky end A single-stranded end of a DNA fragment produced by certain restriction enzymes capable of reannealing with a complementary sequence in another such strand.

stop codon One of three mRNA codons—UAG, UAA, and UGA—at which polypeptide synthesis stops.

structural gene A gene that encodes the amino acid sequence of a polypeptide chain.

STRP *See* **simple tandem repeat polymorphism.**

STS *See* **sequence-tagged site.**

submetacentric chromosome A chromosome whose centromere divides it into arms of unequal length.

subpopulation Any of the breeding groups within a larger population between which migration is restricted.

subspecies A relatively isolated population or group of populations distinguishable from other populations in the same species by allele frequencies or chromosomal arrangements and sometimes exhibiting incipient reproductive isolation.

substrate A substance acted on by an enzyme.

subunit A component of an ordered aggegrate of macromolecules—for example, a single polypeptide chain in a protein containing several chains.

supercoiled DNA *See* **supercoiling.**

supercoiling Coiling of double-stranded DNA in which strain caused by overwinding or underwinding of the duplex makes the double helix itself twist; a supercoiled circle is also called a twisted circle or a superhelix.

suppressive petites A class of cytoplasmically inherited, small-colony mutants in yeast that are not corrected by the addition of normal mitochondria.

suppressor mutation A mutation that partially or completely restores the function impaired by another mutation at a different site in the same gene (intragenic suppression) or in a different gene (intergenic suppression).

suppressor tRNA Usually, a tRNA molecule capable of translating a stop codon by inserting an amino acid in its place, but a few suppressor tRNAs replace one amino acid with another.

synapsis The pairing of homologous chromosomes or chromosome regions in zygotene of the first meiotic prophase.

synaptonemal complex A complex protein structure that forms between synapsed homologous chromosomes in the pachytene substage of the first meiotic prophase.

syncytial blastoderm Stage in early *Drosophila* development formed by successive nuclear divisions without division of the cytoplasm.

syndrome A group of symptoms that appear together with sufficient regularity to warrant designation by a special name; also, a disorder, disease, or anomaly.

synonymous mutation A nucleotide substitution in a codon that does not alter the amino acid specified by the codon.

syntenic genes Genes present in the same chromosome; not equivalent to linked genes because syntenic genes can show independent assortment if they are far enough apart.

synteny The presence of two different genes on the same chromosome.

synteny group A group of genes present in a single chromosome in two or more species.

T

tandem duplication A pair of identical or closely related DNA sequences that are adjacent and in the same orientation.

taster One who has the inherited ability to taste the substance phenylthiocarbamide (PTC); the phenotype of the homozygous dominant and heterozygote.

TATA-box binding protein (TBP) A protein that binds to the TATA motif in the promoter region of a gene.

TATA box The base sequence 5'-TATA-3' in the DNA of a promoter.

tautomeric shift A reversible change in the location of a hydrogen atom in a molecule, altering the molecule from one isomeric form to another. In nucleic acids, the shift is typically between a keto group (keto form) and a hydroxyl group (enol form).

TBP TATA-box binding protein. *See also* **TATA-box binding protein.**

TBP-associated factor (TAF) Any protein found in close association with TATA binding protein.

T cells A class of white blood cells instrumental in various aspects of the immune reponse.

***T* DNA** Transposable element found in *Agrobacterium tumefaciens*, which produces crown gall tumors in a wide variety of dicotyledonous plants.

telomerase An enzyme that adds specific nucleotides to the tips of the chromosomes to form the telomeres.

telomere The tip of a chromosome, containing a DNA sequence required for stability of the chromosome end.

telophase The final stage of mitotic or meiotic nuclear division.

temperate phage A phage capable of both a lysogenic and a lytic cycle.

temperature-sensitive mutation A conditional mutation that causes a phenotypic change only at certain temperatures.

template strand A nucleic acid strand whose base sequence is copied in a polymerization reaction to produce either a complementary DNA or RNA strand.

terminal redundancy In bacteriophage T4, a short duplication found at opposite ends of the linear molecule; the duplicated region differs from one phage to the next.

termination Final stage in polypeptide synthesis when the ribosome encounters a stop codon and the finished polypeptide is released from the ribosome.

terminus of replication In a circular DNA molecule that undergoes bidirectional θ replication, the point at which the replication forks meet.

testcross A cross between a heterozygote and a recessive homozygote, resulting in progeny in which each phenotypic class represents a different genotype.

testis-determining factor (TDF) Genetic element on the mammalian Y chromosome that determines maleness.

tetrad The four chromatids that make up a pair of homologous chromosomes in meiotic prophase I and metaphase I; also, the four haploid products of a single meiosis.

tetrad analysis A method for the analysis of linkage and recombination using the four haploid products of single meiotic divisions.

tetramer A protein composed of four polypeptide subunits, which may or may not be identical.

tetraploid A cell or organism with four complete sets of chromosomes; in an autotetraploid, the chromosome sets are homologous; in an allotetraploid, the chromosome sets consist of a complete diploid complement from each of two distinct ancestral species.

tetratype An ascus containing spores of four different genotypes—one each of the four genotypes possible with two alleles of each of two genes.

thermophile An organism that normally lives at an unusually high temperature.

θ replication Bidirectional replication of a circular DNA molecule, starting from a single origin of replication.

30-nm fiber The level of compaction of eukaryotic chromatin resulting from coiling of the extended, nucleosome-bound DNA fiber.

three-point cross Cross in which three genes are segregating; used to obtain unambiguous evidence of gene order.

3'-OH group The end of a DNA or RNA strand that terminates in a sugar and so has a free hydroxyl group on the number-3' carbon.

threshold trait A trait with a continuously distributed liability or risk; organisms with a liability greater than a critical value (the threshold) exhibit the phenotype of interest, such as a disorder.

thymine (T) A nitrogenous pyrimidine base found in DNA.

thymine dimer *See* **pyrimidine dimer.**

time of entry In an Hfr $\times$ F^- bacterial mating, the earliest time that a particular gene in the Hfr parent is transferred to the F^- recipient.

***Ti* plasmid** A plasmid present in *Agrobacterium tumefaciens* and used in genetic engineering in plants.

topoisomerase An enzyme that introduces or eliminates either underwinding or overwinding of double-stranded DNA. It acts by introducing a single-strand break, changing the relative positions of the strands, and sealing the break.

topoisomerase I A nick-closing type of topoisomerase enzyme in which one supercoil is removed from DNA per reaction cycle.

topoisomerase II A topoisomerase enzyme able to remove both positive and negative supercoils from DNA. These enzymes can also cleave duplex DNA and pass another duplex DNA molecule through the gap.

total variance Summation of all sources of genetic and environmental variation.

totipotent cell A cell capable of differentiation into a complete organism; the zygote is totipotent.

trait Any aspect of the appearance, behavior, development, biochemistry, or other feature of an organism.

***trans* configuration** The arrangement in linked inheritance in which a genotype heterozygous for two mutant sites has received one of the mutant sites from each parent—that is, $a_1+/+a_2$.

transcript An RNA strand that is produced from, and is complementary in base sequence to, a DNA template strand.

transcription The process by which the information contained in a template strand of DNA is copied into a single-stranded RNA molecule of complementary base sequence.

transcriptional activator protein Positive control element that stimulates transcription by binding with particular sites in DNA.

transcriptional cosuppression A mechanism of gene silencing in which the silenced genes fail to be transcribed.

transcription complex An aggregate of RNA polymerase (consisting of its own subunits) along with other polypeptide subunits that makes transcription possible.

transducing phage A phage type capable of producing particles that contain bacterial DNA (transducing particles). A specialized transducing phage produces particles that carry only specific regions of chromosomal DNA; a generalized transducing phage produces particles that may carry any region of the genome.

transduction The carrying of genetic information from one bacterium to another by a phage.

transfer RNA (tRNA) A small RNA molecule that translates a codon into an amino acid in protein synthesis; it has a three-base sequence, called the anticodon, complementary to a specific codon in mRNA, and a site to which a specific amino acid is bound.

transformation Change in the genotype of a cell or organism resulting from exposure of the cell or organism to DNA isolated from a different genotype; also, the conversion of an animal cell, whose growth is limited in culture, into a tumor-like cell whose pattern of growth is different from that of a normal cell.

transformation mutation A mutation in which affected cells undergo a developmental fate that is characteristic of other cells.

transformation rescue The ability of a wildtype gene to complement a recessive mutation when introduced into a mutant organism by germ-line transformation.

transgenic Refers to animals or plants in which novel DNA has been incorporated into the germ line.

***trans* heterozygote** *See* ***trans* configuration.**

transition mutation A mutation resulting from the substitution of one purine for another purine or that of one pyrimidine for another pyrimidine.

translation The process by which the amino acid sequence of a polypeptide is synthesized on a ribosome according to the nucleotide sequence of an mRNA molecule.

translocation Interchange of parts between nonhomologous chromosomes; also, the movement of mRNA with respect to a ribosome during protein synthesis. *See also* **reciprocal translocation.**

transmembrane receptor A receptor protein containing amino acid sequences that span the cell membrane.

transmission genetics The study of the pattern of inheritance of genes and of other genetic elements from generation to generation.

transposable element A DNA sequence capable of moving (transposing) from one location to another in a genome.

transposase Protein necessary for transposition.

transposition The movement of a transposable element.

transposon A transposable element that contains bacterial genes—for example, for antibiotic resistance; also loosely used as a synonym for transposable element.

transposon tagging Insertion of a transposable element that contains a genetic marker into a gene of interest.

transversion mutation A mutation resulting from the substitution of a purine for a pyrimidine or that of a pyrimidine for a purine.

trinucleotide repeat A tandemly repeated sequence of three nucleotides; genetic instability in trinucleotide repeats is the cause of a number of human hereditary diseases.

triplet code A code in which each codon consists of three bases.

triplication The presence of three copies of a DNA sequence that is ordinarily present only once.

triploid A cell or organism with three complete sets of chromosomes.

trisomic A diploid organism with an extra copy of one of the chromosomes.

trisomy-X syndrome The clinical features of the karyotype 47,XXX.

trivalent Structure formed by three homologous chromosomes in meiosis I in a triploid or trisomic chromosome when each homolog is paired along part of its length with first one and then the other of the homologs.

tRNA *See* **transfer RNA.**

true-breeding Refers to a strain, breed, or variety of organism that yields progeny like itself; homozygous.

truncation point In artificial selection, the value of the phenotype that determines which organisms will be retained for breeding and which will be culled.

tumor-suppressor gene A gene that normally controls cell proliferation or that activates the apoptotic pathway, in which loss-of-function mutations are associated with cancer progression.

Turner syndrome The clinical features of human females with the karyotype 45,X.

twin spot A pair of adjacent, genetically different regions of tissue, each derived from one daughter cell of a mitosis in which mitotic recombination took place.

U

ultracentrifuge A centrifuge capable of speeds of rotation sufficiently high to separate molecules of different density in a suitable density gradient.

uncharged tRNA A tRNA molecule that lacks an amino acid.

uncovering The expression of a recessive allele present in a region of a structurally normal chromosome in which the homologous chromosome has a deletion.

underwound DNA A DNA molecule whose strands are untwisted somewhat; hence, some of its bases are unpaired.

unequal crossing-over Crossing-over between nonallelic copies of duplicated or other repetitive sequences—for example, in a tandem duplication, between the upstream copy in one chromosome and the downstream copy in the homologous chromosome.

uniparental inheritance Extranuclear inheritance of a trait through cytoplasmic factors or organelles contributed by only one parent. *See also* **maternal inheritance.**

unique sequence A DNA sequence that is present in only one copy per haploid genome, in contrast with repetitive sequences.

univalent Structure formed in meiosis I in a monoploid or a monosomic when a chromosome has no pairing partner.

unlinked genes Genes whose frequency of recombination is 50 percent (independent assortment); unlinked genes can be in the same chromosome if they are sufficiently far apart.

uracil (U) A nitrogenous pyrimidine base found in RNA.

V

variable expressivity Differences in the severity of expression of a particular genotype.

variable number of tandem repeats (VNTR) Any region of the genome that, among the members of a population, is highly poplymorphic in the number of tandem repeats of a short nucleotide sequence; the result is polymorphism in the size of the restriction fragment that contains the repeated sequence.

variable region The portion of an immunoglobulin molecule that varies greatly in amino acid sequence among antibodies in the same subclass. *See also* **V region.**

variance A measure of the spread of a statistical distribution; the mean of the squares of the deviations from the mean.

variegation Mottled or mosaic expression.

vector A DNA molecule, capable of replication, into which a gene or DNA segment is inserted by recombinant DNA techniques; a cloning vehicle.

viability The probability of survival to reproductive age.

viral oncogene A class of genes found in certain viruses that predispose to cancer. Viral oncogenes are the viral counterparts of cellular oncogenes. *See also* **cellular oncogene, oncogene.**

virulent phage A phage or virus capable only of a lytic cycle; contrasts with temperate phage.

virus An infectious intracellular parasite able to reproduce only inside living cells.

VNTR *See* **variable number of tandem repeats.**

Von Hippel–Lindau disease Hereditary disease marked by tumors in the retina, brain, other parts of the central nervous system, and various organs throughout the body. The disease is caused by an autosomal dominant mutation in chromosome 3.

V region One of multiple DNA sequences coding for alternative amino acid sequences of part of the variable region of an antibody molecule. *See also* **variable region.**

V-type position effect A type of position effect in *Drosophila* that is characterized by discontinuous expression of one or more genes during development; usually results from chromosome breakage and rejoining such that euchromatic genes are repositioned in or near centromeric heterochromatin.

W

Watson–Crick pairing Base pairing in DNA or RNA in which A pairs with T (or with U in RNA) and G pairs with C.

Watson strand The upper strand of duplex DNA when the strands are conventionally represented as parallel horizontal lines with the 5′ end of the top strand at the extreme left.

wildtype The most common phenotype or genotype in a natural population; also, a phenotype or genotype arbitrarily designated as a standard for comparison.

wobble The acceptable pairing of several possible bases in an anticodon with the base present in the third position of a codon.

X

χ^2 *See* **chi-square.**

X chromosome A chromosome that plays a role in sex determination and that is present in two copies in the homogametic sex and in one copy in the heterogametic sex.

xeroderma pigmentosum An inherited defect in the repair of ultraviolet-light damage to DNA, associated with extreme sensitivity to sunlight and multiple skin cancers.

X inactivation *See* **single-active-X principle.**

X-linked inheritance The pattern of hereditary transmission of genes located in the X chromosome; usually evident from the production of nonidentical classes of progeny from reciprocal crosses.

Y

YAC *See* **yeast artificial chromosome.**

Y chromosome The sex chromosome present only in the heterogametic sex; in mammals, the male-determining sex chromosome.

yeast artificial chromosome (YAC) In yeast, a cloning vector that can accept very large fragments of DNA; a chromosome introduced into yeast derived from such a vector and containing DNA from another organism.

Z

Z-form DNA One of the unusual three-dimensional structures of duplex DNA in which the helix is left-handed.

zinc finger A structural motif, found in many DNA-binding proteins, in which finger-like projections entrap a zinc ion.

zygote The product of the fusion of a female and a male gamete in sexual reproduction; a fertilized egg.

zygotene The substage of meiotic phrophase I in which homologous chromosomes synapse.

zygotic gene Any of a group of genes that control early development through their expression in the zygote.

Index

Note: A "C" following a page number indicates a Connection excerpt.

A

a1 α2 protein, 538
  in yeast mating-type regulation, 530–532
A antigens, in human blood, 113–115
ABC transporters, 562
aberrant chromosomes, 346, 357–375
ablation, 604
abnormalities, congenital, 771–772
ABO blood groups, 112–115
  allele frequencies of, 716–717
  locus of, 349
abortion, spontaneous, 357–358, 360
acceptor end, of introns, 458–460
ACE gene, 774–775
acentric chromosomes, 350
ACF (*ATP-dependent chromatin assembly and remodeling factor*), 514
acquired immune deficiency syndrome (AIDS)
  AZT treatment of, 287
  dideoxynucleoside analogs and, 247
  resistance, 708–709
  DNA amplification and, 64
  synthetic vaccines and, 567
acridine molecules, mutagenesis via, 289
acrocentric chromosomes, 350–351
acron, of *D. melanogaster*, 612, 614
acrylamide gel, in gel electrophoresis, 52
Activator (*Ac*) element
  methylation of, 519
  in transposition, 274–275, 278, 308
activator proteins, 492, 510, 514, 517
active repressor, in tryptophan synthesis, 501–502
acute leukemias, 666–667
adaptation, evolution and, 730
addition rule, in probability, 109, 158–159
additivity, of alleles, 761
adenine (A)
  cytosine methylation and, 285
  deamination of, 288
  in dideoxy DNA sequencing, 243–245
  as DNA base, 9–11, 18–19, 41–49, 224–225
  nucleotide substitution and, 268
  as RNA base, 19
  in sickle-cell anemia, 269–270
adenosine 5'-triphosphate (ATP), in RNA synthesis, 452
adenyl cyclase mutants, 497
adjacent-1 segregation, 370–372, 374
adjacent-2 segregation, 370–372, 374
Aegilops speltoides, endoreduplication in, 380
Aegilops tauschii, endoreduplication in, 380
Africa
  human origins in, 697–698
  sickle-cell anemia in, 735–737
agamous (*ag*) gene, 624–626
agar, growing bacteria on, 402
agarose gel, in gel electrophoresis, 52–53
agglutination, ABO blood groups and, 114–115
"agouti" mice, 117
agriculture
  complex traits in, 750
  cytoplasmic male sterility in, 688–691
  DNA typing in, 75, 79
  genetic engineering in, 556, 565–566, 585
  grass genomes in, 382–384
  of maize, 754–755
  polyploid plants in, 378
Agrobacterium tumefaciens
  genetic engineered rice and, 565
  genetic engineering of, 560–561
Ahmed, Shawn, 328C
alanine (Ala, A)
  DNA base sequence for, 16, 20, 476
  molecular structure of, 447
  in PAH, 21
albinism, in human pedigrees, 105–106
albino mice, 117
alcohol dehydrogenase genotypes, in *D. melanogaster*, 733
Alexandra, Tsarina
  hemophilia gene and, 154–155
  identification of remains of, 681, 702–703
Alexis, Tsarevich, 702
  as hemophiliac, 154–155
algae
  drug resistance in, 686–688
  as eukaryotes, 28
alkapton, 12–13. *See also* homogentisic acid
alkaptonuria, genetic basis of, 12–15
alkylating agents, as mutagens, 288–289
alkyl groups, of mutagens, 288
allele frequencies, 708–710
  DNA typing and, 716–720
  Hardy–Weinberg principle and, 712–713
  inbreeding and, 723–724
  irreversible mutation and, 729
  mating systems and, 712
  multiple alleles and, 715–716
  population differences in, 720–721
  in random genetic drift, 738–739
  random mating and, 713–714
  reversible mutation and, 729–730
  with X-linked genes, 721–722
alleles, 64–65. *See also* codominant alleles; dominant alleles; multiple alleles; recessive alleles; wildtype alleles
  with calico cats, 353–355
  of *CFTR* gene, 486
  chromosomes and, 134–135
  complementation tests and, 118–123
  for complex traits, 750, 755–757
  deletion of, 363–364
  epistasis and, 115–117
  fixed, 710
  gain-of-function, 604–607
  in gene recombination, 176–181
  in genetic mapping, 181–192
  hereditary diseases and, 74
  heritability of, 764–766
  in human pedigrees, 104–107
  with Huntington disease, 108–109
  in irreversible mutation, 728–729
  in lactose metabolism, 492–499
  loss-of-function, 604–607
  lost, 710
  of *lozenge* gene, 209
  in meiosis, 139, 147
  in migration, 728
  in mitotic recombination, 209–211
  nonchromosomal, 686–688
  of *Paramecium aurelia*, 692–693
  in pea plants, 88–93, 98–103, 115–116, 161–162
  population differences in, 720–721
  Principle of Assortment and, 99–101
  Principle of Complementation and, 120
  Principle of Segregation and, 94–95
  in random amplified polymorphic DNA analysis, 70
  restriction fragment length polymorphisms and, 67–68
  simple tandem repeat polymorphisms and, 72–73
  single-nucleotide polymorphisms and, 66
  of *SLC6A4* transporter gene, 775–776
  of T4 bacteriophage, 420–429
  in tetraploid plants, 378–379
  in transduction, 412–417
  typographical naming conventions for, 65
Allison, Anthony C., 736–737C
allohexaploidy
  in grasses, 382
  in *Triticum aestivum*, 380
allopolyploidy, 379–381
allotetraploidy, 379
α-amylase gene, 522–523
alpha carbon, in amino acids, 446
α-globin gene, introns in, 462
α helices (helix), in proteins, 448
α particles, 290
α polypeptide chains, of hemoglobin, 449
α satellite, in centromeres, 332
alternate segregation, 370–371, 374
alternative splicing, 522–523
Alu gene family, 278
alveolates, as eukaryotes, 28
amber stop codon, 479–480
amber suppressors, 479–481
Ames test, 302–303
amino acids, 13
  in alkaptonuria, 13–15
  codons for, 18–21, 472, 476

destabilizing replacements of, 25
in frameshift mutations, 271–272
in GAL4 protein, 510
in histones, 318
in LIN-12 protein, 605
molecular structures of (chart), 447
in mutation, 270
operons for biosynthesis of, 504–505
protein synthesis from, 16–21, 48–49, 446–449
synthetase enzymes for, 477–478
transfer RNA and, 17
in translation, 463–464
aminoacyl (A) site, in ribosomal subunit 60S, 464–465
aminoacyl-tRNA synthetases, 464, 477–478
amino group ($-NH_2$)
in amino acids, 446
in mutagenesis, 288
amino terminus, in polypeptide chains, 446–447
amniocentesis, 703
Down syndrome and, 360
amnion, 360
amorphic mutations, 267
amphibians
bivalent chromosomes of, 146
diploid chromosome numbers of, 134
genome sizes of, 72, 313
Amphiuma, genome size of, 72, 312–313
ampicillin resistance, plasmid for, 552–553
amplification
of DNA fragments, 49, 58–64, 549
of genes, 525
amplified fragment length polymorphisms (AFLPs), 71–72
amylase, production of, 522–523
amylopectin, in pea plants, 88
anaphase, 136–137, 637
in meiosis, 140
monocentric chromosomes in, 350–351
spindle checkpoint and, 654
anaphase I, 141, 143, 147–148
heterozygous reciprocal translocation during, 370–371
inversion during, 369
in *Lilium longiflorum*, 146
trisomic segregation during, 361–362
anaphase II, 142, 148
in *Lilium longiflorum*, 146
anaphase checkpoint, 649
anaphase-promoting complex (APC/C), in cell cycle, 648
anchor cells (ACs), in *Caenorhabditis elegans*, 604–605
anemia, sickle-cell, 269–271
aneuploid gametes, in segregation, 370–371, 374
aneuploidy, 357–358
Angelman syndrome, 521–522
angiosperms
genome sizes of, 313
polyploidy in, 378
angiotensin-converting enzyme (ACE), 774–775
animals
abnormal chromosomes in, 346
cloning of, 558
complex traits in, 750
diploid chromosome numbers of, 134
early embryonic development in, 594–598
as eukaryotes, 28
genome sizes of, 313
germ-line transformation in, 556–560, 562
life cycles of, 140
oocytes of, 598
RNA interference in, 524
sex chromosomes in, 149–150
telomeric DNA sequences of, 336
translocation in, 372
Animals (Scientific Procedure) Act of 1986, 558
annealing, in DNA amplification, 61–63
Anopheles gambiae, 736–737
Antennapedia gene complex (ANT-C), in *D. melanogaster*, 618–620
anterior genes, of *D. melanogaster*, 612–613
anther
of pea plants, 91
of primrose, 711
anthocyanin, 308
in pea plants, 115, 117
in snapdragons, 111
synthesis of, 520
anti-A antibodies, in human blood, 113–115
anti-B antibodies, in human blood, 113–115
antibiotic-resistant mutants, among bacteria, 402
antibiotics
algal resistance to, 686–688
bacterial resistance to, 394, 397–402, 407
antibodies, 113
from B cells, 525–528
with human blood groups, 113–115
numbers of possible, 527–528
antibody variability, programmed DNA sequence rearrangement for, 525–528
anticipation, 273
anticodons, in tRNA, 464, 477–478
antigens, 113
with human blood groups, 113–115
antimorphic mutations, 267–268
antiparallel strands, in DNA molecule, 47
Antirrhinum majus, incomplete dominance in, 111
APC gene, 660
AP endonuclease, 298–299
apetala (ap1, ap3) genes, 624–626
A polysaccharide, in human red blood cells, 112–113
apoptosis, 600, 602–604, 650, 652–653, 662
cancer cells and, 655–656
aporepressor protein
in negative regulation, 491
in tryptophan synthesis, 500–501
AP repair, 298–299
Apterous gene, in *D. melanogaster*, 618
apurinic nucleotides, AP repair of, 298
apyrimidinic nucleotides, AP repair of, 298
Arabidopsis thaliana
contig clones of, 555
cyclin B in, 622
diploid chromosome number of, 134
flower development in, 623–624, 630
genome size of, 312
genomic sequencing of, 574
Arber, Werner, 50
Archaea kingdom, 28
in eukaryote origins, 691
genomic sequencing of, 576–577
supercoiled DNA in organisms of, 315
archenteron, in embryonic development, 597
arginine (Arg, R)
DNA base sequence for, 20, 476
in histones, 318
mitochondrial codons for, 683
molecular structure of, 447
in mutant PAH, 22–23, 64
in plant genetic engineering, 561
arthritis, alkaptonuria and, 14
Arthrobacter luteus, restriction enzymes from, 51
arthropods, genome sizes of, 313
artificial chromosomes, 546, 569–571
artificial selection, 763–767
long-term, 766
AruI restriction enzyme, 51
Ascobolus immersus, MIP in, 519
ascospores, 201–203
of *Neurospora crassa*, 206
ascus (asci)
of *Neurospora crassa*, 206–208, 248–249
tetrads in, 201–203
Ashkenazi Jews, breast cancer among, 674
asparagine (Asn, N)
DNA base sequence for, 20, 476
molecular structure of, 447
aspartic acid (Asp, D)
DNA base sequence for, 20, 476
in Li–Fraumeni syndrome, 662
molecular structure of, 447
Aspergillus, mitotic recombination in, 209–211
Aspergillus ficuum, genetic engineered rice and, 565
assays, with DNA chips, 578–581
assortative mating, 711
aster, 637
Astyanax fasciatus, complex traits in, 760–762
atrial septal defect, heritability of, 771
attached-X chromosomes, 187–189
attC sequence, in integrons, 400–401
attenuation, in tryptophan synthesis, 502–505
attenuator region, in tryptophan synthesis, 502
attI sequence, in integrons, 400–401
AUG codon, 464–465, 476. *See also* start codon
autogamy, in *Paramecium aurelia*, 692–693
automated DNA sequencing, 244–247
autonomous development, 595–597
autonomously replicating sequences (ARSs), of DNA, 224
autopolyploidy, 379–380
autoradiograms, of DNA, 230
autoregulation
of genes, 492
in organism development, 593–594
autosomes, 150
auxotrophs, 402
Avery, Oswald T., 3–5, 8, 50, 544–545
axial flower position, in pea plants, 89, 93
azacytidine, methylation reversal via, 519
3'-azido-2',3'-dideoxythymidine (AZT), 247
AIDS treatment with, 287
molecular structure of, 287

B

BAC clones, 573–574
Bacillus amyloliquefaciens strain H, restriction enzymes from, 50–51
Bacillus subtilis, TRAP in, 503–504
Bacillus thuringiensis, insecticidal proteins from, 556

backcrosses
complex traits and, 761
cytoplasmic sterility and, 690
in transmission genetics, 98
bacteria
antibiotic resistance in, 394, 398–402
antigens versus, 113–114
bacteriophage and, 6–9, 394
cassettes in, 398–401
chromatin in, 342
chromosomes of, 316–317
in cloning, 40
colonies of, 3, 402
conjugation between, 395–396, 405–412
cotransduction in, 415–417
DNA of, 2–5
endosymbiont, 691–694
in evolution, 28
gene regulation in, 490–492
genetic sequencing of, 574–577
genetics of, 402–403
genome sizes of, 312–313
insertion sequences of, 396–397
integrons in, 398–401
in Meselson–Stahl experiment, 225–228
mismatch repair in, 296–298
mobile DNA in, 394
multiple-antibiotic-resistant, 401–402
plasmids in, 394–398
recombination in, 413
replicons of, 224
resistance in, 282
screening for recombinants in, 554
SOS repair in, 300
transduction between, 412–417, 427, 429–435
transposable elements of, 396–397
Bacteria kingdom, 28
bacterial artificial chromosomes (BACs), 569–570, 572
bacteriophage, 394, 440–441. *See also* viruses
DNA of, 6–9
fragmenting DNA from, 50
genetics of, 418–427
genome sizes of, 312–313
lysogeny via, 427, 429–433
lytic cycle of, 6, 413, 418–422
as organisms, 417
and plasmid origin of replication, 552–553
repressor proteins of, 433
specialized transduction via, 433–435
therapy via, 441
transduction via, 412–417
virulent, 413
BAC vectors, 569–570, 572
balanced translocations, 357
*Bal*I restriction enzyme, 544
*Bam*HI restriction enzyme
in DNA amplification, 59
DNA cleavage with, 50–52, 54
in DNA hybridization, 54–55
in site-directed mutagenesis, 563
bands. *See also* phenotypes
in automated DNA sequencing, 244–246
in *D. melanogaster* chromosome, 364
in gel electrophoresis, 52–54, 88
in human karyotype, 348–350
in polytene chromosomes, 323–324
B antigens, in human blood, 113–115
Bar duplication, in *D. melanogaster*, 364–365
Barnett, Leslie, 472–473C
Bar phenotype, of *D. melanogaster*, 282–283
Barr bodies, 352–353
barred-feathers mutation, in chickens, 155
basal transcription factors, in eukaryotes, 511–513
base-analog mutations, 286–287
base analogs, 286
base composition, of DNA, 43–44
base pairing
complementary pairing with RNA, 18
in DNA, 44–49, 224
in DNA hybridization, 56–57
Hoogstein base pairing, 337
Watson–Crick pairing, 10–11
bases, in DNA, 9–11, 41–49
base stacking, in DNA, 45–47
Bax gene, 652–653
Bax protein, 650–653, 659–660, 662
BCD activator protein, 514
B cells, 525–528
Bcl2 protein, 652–653
Beadle, George W., 14, 413, 449–451C, 754–755C
becquerel (Bq), 290
Beijersbergen, Roderick L., 652C
Benzer, Seymour, 427–429C, 472
β-carotene, in transgenic rice, 565–566
β-galactosidase. *See also lacZ* gene
in lactose metabolism, 492–494, 497
pBlueprint II plasmid and, 552–553
β-globin gene
in embryonic and adult chickens, 518
exons in, 462
frameshift mutations in, 271–272
sickle-cell anemia and, 476
beta-glucuronidase genes, 523–524
β particles, 290
β polypeptide chains, of hemoglobin, 449
β sheets, in proteins, 448
B form of DNA, 44–45
bicoid (*bcd*) gene, in *D. melanogaster*, 514, 612–613, 615
bicoid protein, in *D. melanogaster*, 612–613
bidirectional replication
of DNA, 230–231
of eukaryote DNA, 232–233
binomial distribution, probability and, 158–161
binomial expansion, probability and, 159–161
biochemical pathways, 13
of protein folding, 24–25
bio mutants, of *E. coli*, 415–417, 432–435
bio-transducing phage, 434–435
birth defects, 391, 784
bithorax gene complex (BX-C), in *D. melanogaster*, 618–620
bivalent chromosomes, during meiosis, 143–144, 146–147, 361–362
Blackburn, Elizabeth H., 334C
black urine disease, 12. *See also* alkaptonuria
blastoderm, in *D. melanogaster* development, 608
blastula, in embryonic development, 594
bleomycin, genes for resistance to, 397
blood-clotting factors, genetic engineering of, 566
blood groups (types), 112–115
blunt ends, of cleaved DNA, 52, 544
BOB' site, in lysogeny, 432–433
bone marrow-derived (B) cells, programmed DNA sequence rearrangement in, 525–528
BOP' site, in specialized transduction, 435
Borrelia burgdorferi, genome size of, 312–313
Botstein, David, 68C
boundary enhancer, in *D. melanogaster*, 618–619
B polysaccharide, in human red blood cells, 112–113
Brachystola magna, sex chromosomes of, 145
branch migration
in Holliday recombination, 250
in single-strand break and repair model, 253–254
BRCA1 gene
breast cancer and, 74, 462, 664, 674
introns in, 462
BRCA2 gene, breast cancer and, 664, 674
breast cancer, 674
BRCA1 gene and, 74, 462
familial, 661
risk factor for, 201
Brenner, Sydney, 457C, 472–473C
Bridges, Calvin B., 156–157, 173
British royal family, hemophilia gene in, 154–155
broad-sense heritability, 762–763
bromodeoxyuridine (BUdR), as DNA label, 228–229
5-bromouracil (5-Bu)
as DNA label, 228–229
as mutagen, 286–287
Brooks, Mary W., 652C
brown-eyes mutation, in *D. melanogaster*, 180
B strain, of *E. coli*, 428–429
Bub protein, 654
Bub1 protein, 660
budding yeast, 638
cell cycle of, 637–639

C

Caenorhabditis elegans
autonomous development of, 596
contig clones of, 555
development of, 598–607
diploid chromosome number of, 134
genome of, 39, 312
genome studies of, 38
genomic sequencing of, 574, 576
germ-line transformation of, 556
proteome of, 27, 578
RNA interference in, 524
segregation mutations in, 601
telomere replication in, 328
calico cats, X inactivation in, 353–355
Campbell, Allan, 432
Campbell, Keith H. S., 558C
cAMP–CRP complex, in lactose metabolism, 497–498
Campostoma anomalum, random amplified polymorphic DNA analysis of, 70
cancer, 636, 655–667. *See also* breast cancer
acute leukemias as, 666–667
Ames test and, 302–303
from defective DNA repair, 664–666
hereditary syndromes of, 660–666
from ionizing radiation, 292–293
telomeres and, 652
cancer cells, properties of, 655–656
candidate genes, for complex traits, 774–776
capping, of mRNA, 458, 466
carbon-source mutants, among bacteria, 403
carboxyl group (–COOH), in amino acids, 446
carboxyl terminus, in polypeptide chains, 446–447
cardiomyopathy, 679

carnation-colored eyes mutant, in *D. melanogaster*, 186–187
Carothers, E. Eleanor, 145C
carriers, in human pedigrees, 105–106, 154–155
cascade effect, in *D. melanogaster* development, 614
caspases, 652
cassettes, 398–401, 529–530
cats
  pleiotropy in, 26
  X inactivation in calico, 353–355
caudal gene, in *D. melanogaster*, 615
cave fish, complex traits in, 760–762
CCA1 gene, in yeast, 486
CCR5 gene, 709
  allele frequency of, 708–709, 713
CCR5 polypeptide, 708
Cdc2 cyclin, 642–643, 645, 651
cdc13 mutant, cell cycles and, 640–642
Cdc20 protein, 654
Cdc28 cyclin, 642–643
CDE1 through CDE4 regions, in yeast centromeres, 331–332
Cdk2 cyclin, 642–643, 645–647
Cdk4 cyclin, 645–646, 660
  cancer and, 656–658
Cdk6 cyclin, 645
ced-3 (*cell death abnormal*) gene, 602
celiac disease, heritability of, 771
cell cycle, 136, 636–639. *See also* mitosis
  checkpoints in, 648–655
  genetic analysis of, 640–642
  mutations affecting regulation of, 640–642
  regulation of, 642–648
  temperature-sensitive mutants and, 640–642
  transcriptional program of, 638–639
cell differentiation, 558
cell division, chromosomes and, 134–135. *See also* meiosis; mitosis
cell division cycle (*cdc*) mutants, 640
cell fates
  in embryonic development, 595
  of nematode cells, 604–607
cell lineages, 599–600
  of *Caenorhabditis elegans*, 602–603
  of *D. melanogaster*, 607–609
cell membranes, transmembrane receptors in, 593–594
cell migrations, in *D. melanogaster* development, 607–608
cell nucleus
  in cell cycle, 636–639
  chromosomes in, 134–135, 312
  cloning animals from, 558
  discovery of, 134
  DNA in, 2, 40
  in meiosis, 139–148
  in mitosis, 135–139
  RNA processing in, 470
cell senescence, 656
centimorgans, map distance and, 182
central dogma, of molecular genetics, 16–21
Centre d'Etude du Polymorphisme Humain (CEPH), 68, 197, 201
  human genome database at, 198–199
centrioles, in cell cycle, 636
centromere cores, 331–332
centromeres, 331, 342
  in anaphase, 350–351
  in attached-X chromosomes, 187–188
  in cell cycle, 636
  chromosome stability and, 350–351
  in human karyotype, 348–349
  in inversion, 368–369
  localized, 331
  during meiosis, 139, 142–143, 146, 148
  during mitosis, 137–138
  in mitotic recombination, 210
  molecular structure of, 330–332
  in *Neurospora crassa*, 207–208
  in polytene chromosomes, 323
centromeric core particles, 332
centrosome, in cell cycle, 636–637
centrosome duplication checkpoint, 649, 653–654
cereals, synteny groups in, 383
cesium-137, from Chernobyl, 294
cesium chloride (CsCl), in Meselson–Stahl experiment, 225–228
CFTR gene, 486
CG dinucleotides, 518
chain elongation, in transcription, 455
chain initiation, in transcription, 455
chain termination, in transcription, 455–456
chalcone synthetase (CHS), 520
Chapalote corn, 755
chaperone proteins, in protein folding, 23, 446
characteristics, of species, 2
Chargaff, Erwin, 43–44, 46
Chargaff's rules, 44, 46
charged tRNA, 464
Chase, Martha, 6, 8C
checkpoint genes, 328
checkpoints
  during interphase, 136, 648–653
  during metaphase, 138
  during mitosis, 653–655
chemokine, 708
chemokine receptor genes, 708–709
Chernobyl nuclear accident, 293–295, 308
chiasma (chiasmata)
  with chromosomes, 143, 145–146, 148, 181
  reciprocal translocation and, 370
chickens
  β-globin gene in, 518
  diploid chromosome numbers of, 134
  epistasis in, 117
  ovalbumin mRNA from, 550
  sex-linked mutations in, 155
children, with Down syndrome, 358–360
chimeric genes, in color blindness, 366–367
chimpanzees, chromosome complement of, 351
chi-square (χ^2) method
  testing goodness of fit with, 161–164
  testing linkage with, 179–180
  testing mating randomness with, 713–714
Chissoe, Stephanie L., 572C
Chlamydomonas, drug resistance in, 686–688
chlorofluorocarbons (CFCs), ozone hole and, 289
chlorophyll, plants lacking, 684–685
chloroplasts
  of *Chlamydomonas*, 686–688
  DNA in, 312, 678, 682–683
  evolution of, 691, 694
  of four-o'clock plants, 684–685
  genetic code for, 683–684
chorion, 360
chorionic villus sampling (CVS), 360, 703
CHRAC (*chromatin accessibility complex*), 514
chromatid interference
  in multiple crossing-over, 192
  in *Neurospora crassa*, 208–209
chromatids, 203, 648
  and chromosome interference, 194
  crossing-over and, 181—183
  inversion and, 367, 369
  during meiosis, 139, 142, 146–148
  during mitosis, 137–138
  in mitotic recombination, 210
  in multiple crossing-over, 191–192
  in *Neurospora crassa*, 208–209
  structure of, 321–322
chromatin, 318–323, 342. *See also* euchromatin; heterochromatin
  mutable gene loci and, 274–275
  remodeling of, 514–515
  in *Saccharomyces cerevisiae* centromeres, 331–332
chromatin-remodeling complexes (CRCs), 514–515
chromocenters, in polytene chromosomes, 323
chromomeres, 144
chromosome 2, evolution of human, 351
chromosome 21, Down syndrome and, 349, 358–361, 373
chromosome 22, sequencing of, 572
chromosome complements, 135, 346. *See also* karyotypes
  of chimpanzee, 351
  of *D. melanogaster*, 152
  of human male, 135
  stability of, 134–135
  table of, 134
chromosome interference, 194–195
chromosome maps, 181. *See also* genetic mapping
chromosome numbers, table of, 134, 312
chromosome painting, 346–347
  of *Crocus* species, 380–381
chromosomes, 342. *See also* bivalent chromosomes; diploid chromosomes; haploid chromosomes; homologous chromosomes; human karyotype; karyotypes; sex chromosomes; X chromosomes; Y chromosomes
  with aberrant structures, 349–375
  abnormalities in, 346
  acute leukemias and, 666–667
  artificial, 546, 569–571
  attached-X, 187–189
  bacterial, 316–317
  bridges, 350–351
  in cell cycle, 636–639
  compound-X, 187–189
  condensation of, 320–323
  discovery of, 134
  DNA in, 2–3, 18, 320–323
  DNA markers in, 39–41
  electrophoretic DNA fragment separation and, 567–568
  of *E. coli*, 410–411
  of eukaryotes, 318–323
  gene recombination in, 176–181
  genes and, 134
  in genetic analysis, 64
  in hereditary diseases, 74–75
  histones in, 312
  holocentric, 331
  ionizing radiation damage to, 292
  linkage groups in, 184–186
  linkage maps of human, 220
  lod scores and, 198–201
  in meiosis, 139–148
  missing, 358
  in mitosis, 135–139

chromosomes (cont'd)
molecular organization of, 312–337
mutable loci in, 274–275
polytene, 323–324
semiconservative DNA replication in, 228–230
sets of, 377–378
stability of, 134–135
telomeres and, 332–337
in tetrad analysis, 201–205
in three-point testcrosses, 192–197
chromosome theory of heredity, 157
Chrysanthemum, polyploidy in, 377–379
Chs gene, 520
cinnabar eye color (*cn*) gene, in *D. melanogaster*, 180, 197
cI protein, in λ bacteriophage, 507
circular DNA
in lysogeny, 431–432
in prokaryote genomes, 312–313
from restriction enzymes, 542–544
rolling-circle replication of, 231–232
supercoiling of, 314
T4 bacteriophage and, 420–422
θ replication of, 230–231
circular permutation, in T4 bacteriophage DNA, 422
cis configuration, of alleles, 178–179
cis-dominant mutations, 495
cisplatin, 659
cistrons
as bacteriophage genes, 424–425, 427
polycistronic mRNA and, 471
Class 1 integrons, 399–401
classical genetics, 2
Clb method, measuring mutation rate via, 282–284
cleavage, of DNA, 40, 50–52, 54, 57–58
cleavage divisions, in embryonic development, 594
cleft lip, heritability of, 771
Cln cyclins, 642–643
cloning, 40
of DNA, 54, 58, 542, 545–551
of "Dolly," 558
ethical issues in, 558
in genetic engineering, 546–551
positional, 554–555
closed loops, in pedigrees, 724–726
Clostridium perfringens, DNA base composition of, 44
club foot, heritability of, 771
coding sequence, in mRNA, 458
codominant alleles
in pea-plant genetics, 90
restriction fragment length polymorphisms and, 67–68
codon 408, in R408W mutant, 22–23
codons
in amino acid operons, 504–505
for amino acids, 18–21, 464, 472, 476
in frameshift mutations, 271–272
in Li–Fraumeni syndrome, 662
sickle-cell anemia and, 269
table of, 20
triplet nature of, 473–475
for tryptophan in tryptophan synthesis, 502–504
coefficient of coincidence, in chromosome interference, 194–195
cohesive ends
in cosmids, 546
in lysogeny, 431
cointegrates, plasmids as, 398–399
colchicine, 382
Coleoptera order, delta-endotoxins versus insects in, 556
colinearity, 450
of genes and peptides, 449–451
colon cancer, 674
colony hybridization, in genetic engineering, 554
color blindness, 391, 744
genetic basis of red–green, 365–367
colored (*C*) gene, 308
common ancestry
in inbreeding, 712
unity of life and, 28–29
compartments, in *D. melanogaster* metamorphosis, 617
complementary DNA (cDNA), 550–551
classification of sequences of, 575
in genetic engineering, 566, 574
complementary pairing
of DNA and RNA bases, 18
of DNA bases, 9–11, 41, 44
DNA hybridization and, 56–57
complementation, 119
principle of, 120
groups, 121
test, 117–123
complete penetrance, of a genetic disorder, 104–105
complex inheritance, 750
complex pedigrees, inbreeding coefficient and, 726
complex traits, 750–754
artificial selection for, 763–767
candidate genes for, 774–776
correlation between relatives and, 767–770
genes affecting, 761–762, 772–776
genetic analysis of, 760–763
genotype versus environment in, 754–760
normal distributions of, 751–753
threshold, 770–772
compound-X chromosomes, 187–189
concatemeric molecules, 422
concatemers, in T4 bacteriophage DNA, 422
condensation, of chromosomes, 320–323
cone pigments, 365
confocal microscope, 347
congenital abnormalities, 771–772
narrow-sense heritability of (table), 771
conidia, of *Neurospora crassa*, 206
conjugation
between bacteria, 395–396, 403, 405–412
in *Paramecium aurelia*, 692–693
conjugative plasmids, 395–396
Hfr cells and, 405–407
in multiple-antibiotic-resistant bacteria, 401–402
nonconjugative plasmids and, 398
transposons and, 397
consanguineous individuals, in human pedigrees, 106
consanguineous mating, 712
consensus sequences
of enhancers, 511
in transcription, 454, 458–460
conservative structure, of histones, 318
constant (C) regions, in antibodies, 527
constant antibody regions, 526
constitutive synthesis, in lactose metabolism, 493–494
constricted pods, in pea plants, 89, 93
contact inhibition, 655
cancer cells and, 655–656
contig clones, 555
continuous traits, 751
coordinate genes, of *D. melanogaster*, 611–614, 616
coordinate regulation, 490, 500
in translation, 471
copia element, in *D. melanogaster*, 277–278
Coprinus cinereus, MIP in, 519
core particles, in nucleosomes, 320
co-repressors
in negative regulation, 491
in tryptophan synthesis, 500–501
core proteomes, of organisms, 576
corn. *See* maize
corn borer, genetic engineering against, 556
correlated response, to selection, 766
correlation, 767–769
correlation coefficient, 767–768
cosmids, 546
cloning vectors from, 545–546
Cot curves, 326–327
Cot expression, 326
cotransduction, 415–417
cotransformation, in bacteria, 404–405
cotransformation frequencies, 405
Counter, Christopher M., 652C
counterselected markers, 407–408
coupling, of transcription and translation, 470
coupling configuration, of alleles, 178–179
covariance, 767–768
CpG dinucleotides, 519
CpG islands, 518
Cre recombinase, 398–399
cress, diploid chromosome number of, 134
Crick, Francis H. C., 9–10, 16, 32, 43–44, 48C, 49, 224, 472–473C, 475, 478
criminal cases, DNA typing in, 64
Crocus angustifolius, chromosome painting of, 380–381
Crocus flavus, chromosome painting of, 380–381
cro protein, in λ bacteriophage, 507
crossing-over, 186–187
in bacteriophage, 417
double, 183–184
duplication and, 365
frequency of, 182
genetic mapping and, 181–192
in human color blindness, 365–367
in inversion, 368–369
in male and female *D. melanogaster*, 180–181
during meiosis, 144, 146–147, 152–153, 156
multiple, 189–192
in *Neurospora crassa*, 206–209
in tetrad analysis, 202–205
timing of, 187–189
unequal, 365–367
cryotherapy, 674
cryptic splice site, in mRNA, 462
Culotti, Joseph, 644
curie (Ci), 290
cut-and-paste elements, 276–277
C-value paradox, 312–314
cyclic adenosine monophosphate (cAMP), 497
in lactose metabolism, 497–498
cyclic AMP receptor protein (CRP), 499
in lactose metabolism, 497–498
cyclin A, 642–643, 645–647
molecular structure of, 643

cyclin B, 643, 645, 647–648, 651, 654
in plants, 622
cyclin–CDK complexes, 642, 646–647, 651
in cell cycle, 642–648
molecular structure of, 643
cyclin D, 643, 645–646, 660
cancer and, 656–658
cyclin-dependent protein kinases (CDKs), 642
in mitosis, 642–648
cyclin E, 642–643, 645–647
cyclins, 642, 644
in mitosis, 136, 642–648
cysteine (Cys, C)
DNA base sequence for, 20, 476
in *E. coli* intragenic suppression, 301–302
molecular structure of, 447
cystic fibrosis, 486, 559
allele frequencies of, 714–715
cystic fibrosis transmembrane conductance regulator (CFTR), 486
cytidine 5'-triphosphate (CTP), in RNA synthesis, 452
cytochrome *b* gene, post-Chernobyl mutations in, 294–295
cytogenetic map, 573
cytokinesis, 135, 636. *See also* meiosis; mitosis
cytological maps, of polytene chromosomes, 324
cytoplasmic exchange, in *Paramecium aurelia*, 692–693
cytoplasmic inheritance, 678
in *Chlamydomonas*, 686–688
cytoplasmic male sterility, 688–691
cytoplasmic transmission, of symbionts, 691–695
cytosine (C)
in dideoxy DNA sequencing, 243–245
as DNA base, 9–11, 19, 41–49, 224–225
hypoxanthine and, 288
molecular structure of, 518
nucleotide substitution and, 268
as RNA base, 19
cytosine methylase, 284
cytosine methylation, 518–519. *See also* 5-methylcytosine
mutation hot spots and, 284–285
RIP mutations and, 280
in X inactivation, 352

D

damage rate, of human DNA, 295
Dam methylase, in mismatch repair, 296
Darwin, Charles, 29, 730
data, from Mendel's experiments, 165–167
daughter strands, in DNA, 224–225
Davis, Ronald W., 68C
DDT, insect resistance to, 282
deadenylation-dependent pathway, 522–523
deadenylation-independent pathway, 523
deamination
of adenine, 288
cytosine methylation and, 285
decaploid chromosome set, 377
deficiencies, 363. *See also* deletion
degeneracy, of genetic code, 475
degradation
of protein, 647–648
in single-strand break and repair model, 253–254
degree of dominance, 734
degrees of freedom, in statistics, 164
Delbrück, Max, 8
deletion, 349, 362–364
in bacteriophage, 423–425
genetic code and, 473–475
mutation via, 271–272
in pericentric inversion, 368
reciprocal translocations and, 280
deletion mutations, in site-directed mutagenesis, 563
"Delicious" apple, as somatic mutation, 266
Δ*32* mutant, in HIV, 708–709
delta-endotoxins, 556
denaturation, 55
of DNA, 55–58, 60–61, 63
denominator genes, 157–158
denticles, in *D. melanogaster* larvae, 610
deoxyadenosine 5'-diphosphate (dADP), 42
deoxyadenosine 5'-monophosphate (dAMP), 42
deoxyadenosine 5'-triphosphate (dATP)
in DNA synthesis, 59
as nucleotide, 42
deoxycytidine 5'-diphosphate (dCDP), 42
deoxycytidine 5'-monophosphate (dCMP), 42
deoxycytidine 5'-triphosphate (dCTP)
5-bromouracil and, 287
in DNA synthesis, 59
as nucleotide, 42
deoxyguanosine, 42
deoxyguanosine 5'-diphosphate (dGDP), 42
deoxyguanosine 5'-monophosphate (dGMP), 42
deoxyguanosine 5'-triphosphate (dGTP)
in DNA synthesis, 59
as nucleotide, 42
deoxyribonuclease (DNase), 316
in DNA experiments, 5
in *E. coli*, 317
deoxyribonucleotides, in DNA synthesis, 224–225
deoxyribose
dideoxyribose versus, 242–243
in DNA, 17, 41–45, 47, 235–236, 452
deoxythymidine 5'-diphosphate (dTDP), 42
deoxythymidine 5'-monophosphate (dTMP), 42
deoxythymidine 5'-triphosphate (dTTP)
in DNA synthesis, 59
nucleoside triphosphate regulation by, 286–287
as nucleotide, 42
depression, 775
depurination, 298
deuteranomaly, 366–367
deuteranopia, 366–367, 391
developmental plans
genetic regulation of, 592–594
of species, 2
deviation, in normal distribution, 752
dextrally coiled shells, 695–696
DHU loop, in tRNA, 477–478
diakinesis, 141, 143, 146
in *Lilium longiflorum*, 144
dicentric chromosomes, 350–351
Dicotyledoneae, genetic engineering of, 560–561
2',3'-didehydro-2',3'-dideoxythymidine (D4T), 247
2',3'-dideoxycytidine (ddC), 247
2',3'-dideoxyinosine (ddI), 247
dideoxynucleoside analogs, clinical uses of, 247
dideoxy nucleotide analogs, as mutagens, 287
dideoxyribose, in DNA sequencing, 242–247
dideoxy sequencing method, 242–247
dihybrid crosses
of chickens, 117
of pea plants, 92, 99, 115–116
dipeptides, 447
diphosphates, of nucleosides, 42
diploid chromosome numbers, table of, 134
diploid chromosomes, 135
in *E. coli* lactose metabolism, 494
from monoploid chromosomes, 382
in *Paramecium aurelia*, 692–693
selection in organisms with, 732
sets of, 377–378
sex chromosomes and, 148–149
tetraploid chromosomes from, 379
in yeast galactose metabolism, 509
diplotene, 141, 143, 145–146
in *Lilium longiflorum*, 144
Diptera order
delta-endotoxins versus insects in, 556
evolution of wings in, 620–621
direct selection, in genetic engineering, 549
"disease genes," 73–74
disjunctional segregation, 370–371
dislocation of hip, heritability of, 771
displacement (D) loop
in double-strand break and repair model, 254–255
in single-strand break and repair model, 253–254
Dissociation (*Ds*) element, in transposition, 274–278, 308
distal alleles, in recombination, 181
diversity, in evolution, 28–29
DNA (deoxyribonucleic acid), 2–9
in alleles, 64
amplification of, 58–62
for antibodies, 525–528
automated sequencing of, 244–247
bacterial exchange of, 403
in bacteriophage, 6–9
bacteriophage transduction of, 412–417, 433–435
B form of, 44–45
bidirectional replication of, 232–233
chemical fragmentation of, 50–52
in chromosomes, 64, 320–323
classification of human, 575
cloning of genomic, 569–571
condensation of, 320–323
denaturation of, 55–56
dideoxy sequencing of, 242–247
double helix of, 9–10, 41–49
ectopic recombination in, 279–280
electrophoresis of, 52–54, 567–569
enhancers in, 455, 510–511
enzymes encoded in, 15
in evolution, 27–28
in genes, 64
as genetic material, 2–9, 48–49
genomes and, 39–41
histones and, 318–320
Holliday recombination in, 250–252, 254
homologous recombination in, 254–255
hybridization, 57
of λ bacteriophage, 427–435
large-scale sequencing of, 38–39
in lysogeny, 427, 429–433
manipulation of, 40
mismatch repair of, 248–249
in mitosis, 135–136
mobile, 394–402
molecular structure of, 9–10, 41–49

DNA (cont'd)
monomorphic, 70
multiple alleles and, 111–112
mutation hot spots in, 284–285
mutation of, 16–25, 266–295, 300–303
nonconjugative plasmids and, 397–398
in nucleosomes, 318–319
in organelles, 678
programmed rearrangements in, 524–532
random amplified polymorphic, 69–70
recombinant, 542–544
recombination of, 248–255
renaturation of, 55–56
repair of, 295–300
repetitive sequences in, 324–330
replication of, 10–11, 224–247
replication proteins for, 234–235
reverse transcriptase and, 278
rolling circle replication of, 231–232
satellite, 324
selective joining of, 547–549
selective replication of, 49, 58–64
semiconservative replication of, 225–233
separation and identification of, 49–58
spontaneous damage rate of, 295
strands, 9–11, 41–49, 55–56, 59–63, 244–255, 278, 320–323
supercoiling of, 314–316
of T4 bacteriophage, 420–422
θ replication of, 230–231
telomeres and replication of, 332–337
topoisomerases and, 315–316
transcription of, 16–18
in transduction, 412–417, 427, 429–435
in transformation, 403–405
variant forms of, 47
Z form of, 47
dnaB gene, in *E. coli*, 234
DnaB helicase, 234
DNA chips, 578–581
DNA cleavage, 40
restriction maps of, 54
site-specific, 50–52
in Southern blot, 57–58
DNA cloning, 54, 58, 542, 545–546
DNA damage checkpoint, 655
during cell cycle, 649–653
DNA fingerprinting. *See also* DNA typing
for individual identification, 74–75
simple tandem repeat polymorphisms and, 72
dnaG gene, in *E. coli*, 235
DnaG primase, 235
DNA ligase, 234
amplified fragment length polymorphisms and, 71
in AP repair, 298–299
in excision repair, 299
genes encoding, 578
joining defined DNA fragments with, 542–544, 547–548
joining DNA precursor fragments via, 240
in mismatch repair, 296
DNA looping, in enhancers, 511–512
DNA markers (polymorphisms), 39–41, 65–73
applications of, 73–76
with candidate genes, 774–776
in DNA typing, 716–719
in genomic DNA, 65–73
in Holliday recombination, 250–251
in mapping QTLs, 772–773
positional cloning and, 554–555
DNA methylase, 518
DNA polymerase I (Pol I), 234, 238, 240–241
DNA polymerase III (Pol III), 234, 237–238, 240
DNA polymerases, 332. *See also* polymerase chain reaction (PCR)
in AP repair, 298–299
chain reactions of, 60–64
in DNA replication, 235
in DNA synthesis, 59–60, 236–238
exonuclease activity of, 295–296
heat-stable, 62
in postreplication repair, 299
processivity of, 651
replication slippage in, 297
telomeres and, 335
DNA rearrangement, gene regulation via, 490, 524–532
DNA repair, 295–300, 308. *See also* mismatch repair
defects in, 664–666
mutation and, 266
DNA–RNA hybridization, 461
DNA sequencing, 54, 242–247
of human genome, 68, 574–576
of organisms (table), 577
uses of, 576–581
DNA synthesis, 59–60
initiation of, 235–236
during interphase, 136
DNA typing, 716–720, 744
in criminal cases, 64, 718–720
in domesticated animals, 75
in domesticated plants, 75
for individual identification, 74–75
simple tandem repeat polymorphisms and, 72
of Tsar Nicholas II and family, 702
DNA uracil glycosylase, 285, 295
Dobzhansky, Theodosius, 28
dogma, 16–17
"Dolly," cloning of, 558
dominant alleles
incomplete dominance, 111–115, 127
in human pedigrees, 104–106
in natural selection, 732
in pea plants, 88–90, 93
probability and, 161
in random amplified polymorphic DNA analysis, 70
dominant-negative mutations, 267–268
donor end, of introns, 458–460
dorsal gene, 614
dosage, of genes, 346
dosage compensation
in calico cats, 353–355
in mice, 354
of X-linked genes, 351–353
double crossing-over
chromosome interference in, 194–195
recombination and, 183–184
in tetrad analysis, 203–205
in three-point testcrosses, 192–194
double helix
DNA molecule as, 9–10, 41–49
"unzipping" of, 55
double-strand break and repair model, of DNA repair, 254–255
double-Y karyotype, 362
Down syndrome, 272, 349, 362, 391
chromosome abnormality in, 358–361
frequency of, 360
Robertsonian translocation in, 373, 375
symptoms of, 358–359
D-raf gene, 614
dragonfly, evolution of wings in, 620–621
Drosophila melanogaster
apoptosis in, 604
artificial chromosomes for, 570–571
attached-X chromosomes in, 187–189
bidirectional DNA replication in, 232–233
chromosome deletion in, 363–364
chromosome interference in, 194
chromosome size in, 567–569
complex traits in, 750
contig clones of, 555
core proteome of, 576
development of, 607–621
diploid chromosome number of, 134
DNA base composition of, 44
ectopic expression in, 267
enhancers in, 513–514
fitness experiments with, 732–733
genetic maps of, 184
genome of, 38–39, 312
genomic sequencing of, 574
germ-line transformation in, 556–557
heterochromatin in, 197
highly repetitive DNA sequences in, 329–330
homeotic mutations in, 600–601
instars of, 680
interactive fly Web site, 630
intron mutations in, 462
larvae of, 607–609
lethal mutations in, 282–284
life cycle of, 607
map distances in, 196–197
maternal sex ratio in, 694–695
middle repetitive DNA sequences in, 330
mitochondrial genome of, 682
mitotic recombination in, 209
mutation rates in, 282–284
nondisjunction in, 155–157
oocyte division pattern in, 599
polytene chromosomes of, 323–324
positional cloning in, 555
position-effect variegation in, 375–377
promoters in, 516–517
proteome of, 27, 578
random genetic drift in, 739
recombination experiments with, 186–187, 190
recombination in, 180–181
recombination within genes of, 209–210
RNA interference in, 524
sex determination in, 157–158
silencers in, 511, 513–514
telomeres in, 332
transcriptional activation in, 513
transcriptional cosuppression in, 520
transformation rescue in, 562
translational control in, 524
transposable elements in genome of, 278
transposition in, 279
triploid females of, 357
X-linked inheritance in, 150–153, 173, 177–181, 190
x-ray-induced mutations in, 286, 291–292
Drosophila virilis, highly repetitive DNA sequences in, 329
Duchenne muscular dystrophy (DMD), 173
Dunham, Ian, 572C
Duroc-Jersey pigs, epistasis in, 117
dwarf stem length, in pea plants, 89, 93
dynamic mutations, 272–273
spontaneous mutations, 280–285

E

E2F transcription factor, 646–647, 660
easter gene, 614
ecdysone, of *D. melanogaster*, 617
ecology, DNA typing in, 75
*Eco*RI restriction enzyme, 51, 54
  amplified fragment length polymorphisms and, 71–72
  in DNA recombination, 543–544, 547–548
  in pea-plant genetics, 90
  restriction fragment length polymorphisms and, 66–67
  restriction sites on, 542
  in site-directed mutagenesis, 563
ectopic expression, 267
ectopic recombination, 279–280
eggs. *See* oocytes
EF (elongation factors), 466–468
eIF (elongation factors), 464–465
eight-cutter enzymes, cleaving DNA with, 568–569
electrophoresis, 52. *See also* gel electrophoresis
  of large DNA fragments, 567–569
electrophoretic mobility, of DNA fragments, 53
electroporation, 546
elongation
  of DNA strands, 59–62, 71–72, 224–225, 236–241
  of RNA strands, 455
  of telomeres, 333–337
  in translation, 464, 466–468
elongation factors, 464–468
embryoids, of monoploid plants, 381–382
embryonic induction, 596–597
embryonic stem cells, germ-line transformation via, 557–559
embryos
  of *D. melanogaster*, 606–621
  genetic regulation of development of, 592, 594–598, 600–604
Emerson, Rollins A., 755
encephalopathy, 679
endoreduplication, tetraploidy via, 378–379
endosperm, in meiosis, 141
endosymbiont theory, of organelle evolution, 691, 694
endotoxins, 556
engrailed gene, in *D. melanogaster*, 616, 618
enhancers, in transcription, 455, 510–511, 516–518
environment
  complex traits and, 750, 754–755, 757–760
  genes and, 25–26
  genetic variance and, 758–759
environmental variance, 757–758
enzymes. *See also* DNA polymerases; exonuclease activity; restriction enzymes; topoisomerases
  in bacterial antibiotic resistance, 398–400
  DNA encoding of, 15–16
  in DNA replication, 234–238
  in DNA transcription, 452–454
  in Holliday recombination, 252
  in metabolic pathways, 13–15
  in mismatch repair, 296
  plasmids and, 395
  in protein degradation, 24–26
  as proteins, 12
  in RNA splicing, 461
  with supercoiled DNA, 315–316
  with telomeres, 332–335
  with tRNA, 477–478
  uracil-removing, 285
epidemiology, DNA typing and, 75
epidermal growth factor (EGF), 605
epidermal growth factor receptor (EGFR), 658
epigenetic mechanisms, of gene regulation, 518–522
episomes, F plasmids as, 405
epistasis, 111, 115–117
equational division, 148
equilibrium
  of allele frequency, 729–730
  between mutation and selection, 733–735
equilibrium density-gradient centrifugation, in Meselson–Stahl experiment, 225–228
error-prone repair, 300
Erwinia uredovora, genetic engineered rice and, 565
Escherichia coli
  amplification of DNA of, 62
  artificial chromosomes in, 569–571
  bacteriophage of, 6–9, 50, 54
  bacteriophage recombination in, 418–422
  as *Caenorhabditis elegans* food, 598
  chromatin of, 342
  chromosome map of, 411
  chromosome of, 317
  circular genetic map of, 410
  cloning of, 54, 543
  cloning vectors from, 545–546
  conjugation in, 405
  conjugative plasmid with, 395–396
  contig clones of, 555
  cotransduction in, 415–417
  DNA base composition of, 44
  DNA initiation in, 235–236
  DNA replication rate in, 232
  generalized transduction in, 414–415
  gene regulation in, 492–505
  genes of, 410–411
  genetic code and, 473–476
  genome size of, 312–313
  genomic sequencing of, 574
  helicase in, 234
  Hfr cells of, 405–412, 440
  Holliday recombination in, 252
  intragenic suppression in, 301
  joining DNA precursor fragments in, 240
  laboratory experimental selection of, 730–731
  lac operon of, 495–496, 509
  λ bacteriophage in, 427–435, 505–507
  lysogeny of, 429–431
  in Meselson–Stahl experiment, 225–226, 228
  mismatch repair in, 296
  mRNA caps in, 466
  mutants of, 402–403
  mutation rate in, 284
  nonconjugative plasmids with, 398
  nonsense suppression in, 479–481
  nucleoid of, 317
  pBluescript II plasmid from, 552–553
  phage-resistant, 282
  plasmids in, 394–396
  polycistronic mRNA of, 471
  polysomes of, 470
  promoters and terminators in, 456
  as recombinant DNA host, 542
  recombination in, 413
  replicon of, 224
  restriction enzymes from, 51, 542
  restriction fragments of DNA of, 52
  ribosomal subunits in, 468
  rolling-circle DNA replication in, 232
  SOS repair in, 300
  supercoiled DNA of, 314–315, 317
  T4 bacteriophage in, 423, 428–429
  θ DNA replication in, 230–231
  transcription in, 454–456, 470
  translation in, 470
  trpA gene of, 450–451
  tryptophan synthesis in, 499–505
  wobble in, 479
ethical issues
  in cloning, 558
  in genetic testing, 108–109, 584–585
ethidium bromide, in gel electrophoresis, 52–53
ethyl methanesulfonate (EMS), as mutagen, 288
eubacteria, genomic sequencing of, 577
euchromatin
  in chromosomes, 196–197
  in highly repetitive DNA sequences, 329–330
  position-effect variegation in, 375–377
Eukarya kingdom, 28
eukaryotes
  centromeres of, 331–332
  chemical mutagens of, 288–289
  chromatin of, 318–323
  chromatin remodeling in, 514–515
  chromosomes of, 318–323
  colinearity in, 450
  core proteomes of, 578
  cyclin–CDK complexes in, 642–648
  DNA replication rate in, 232
  enhancers in, 511
  in evolution, 28
  excision repair in, 299
  gene regulation in, 490
  genome sizes of, 312–314
  genomic sequencing of, 576–577
  housekeeping genes in, 507
  mismatch repair in, 297
  monocistronic mRNA of, 471
  origin of, 691, 694
  promoters in, 516–518
  recombination in, 418
  release factors of, 468
  repetitive nucleotide sequences of, 324–330
  RNA polymerase processivity in, 452–454
  RNA processing in, 458–463, 470
  semiconservative DNA replication in, 228–230
  supercoiled DNA in, 315
  telomeres of, 332–337
  transcription complex in, 511–514
  translation in, 464
  translation initiation in, 466
euploid gametes
  alternate segregation in, 370–371
  in segregation, 374
euploidy, 357
European corn borer, genetic engineering against, 556
evening primroses, translocation complexes in, 372–373
even-skipped gene, 615
evolution, 27–29, 727
  cell development and, 602–603
  chromosome reduplication in, 346
  of DNA polymerase α, 262

evolution (cont'd)
DNA typing to trace, 76
genetics and, 727–728
of grass genomes, 382–384
of *Homo sapiens*, 696–698
HOX genes in, 620–621
of imprinting, 522
of introns, 463
natural selection and, 730–736
of organelles, 691
of plant polyploidy, 377–382
random genetic drift and, 736–739
of RNA and DNA, 17
evolutionary genetics, 29
DNA typing in, 75
excision
of F plasmid, 411–412
in RNA splicing, 458–460
excisionase, 433
excision repair, 299
execution mutations, 602, 603
exit (E) site, in ribosomal subunit 60S, 464–465
exons, in mRNA, 458–462
exon-shuffle model, 462–463
exonuclease activity, 332
of DNA polymerases, 238, 295–296
in joining DNA precursor fragments, 240–241
expressed sequence tags (ESTs), 574
classification of, 575
extra chromosomes, in human karyotype, 349–350
extranuclear inheritance, 678–698
described, 678
evolution of, 691
maternal, 695–696
from "mitochondrial Eve," 696–698
patterns of, 678–682
through cytoplasm, 691–695
through organelles, 682–691
eye-antenna disks, of *D. melanogaster*, 617
eye color, 784
eyeless gene, in *D. melanogaster*, 267

F

F (fertility) factor, with *E. coli*, 395–396
F_1 generation
in *D. melanogaster* studies, 150–153, 177
epistasis and, 115–117
of pea plants, 92–93, 98, 119–120
Principle of Assortment and, 100
Principle of Segregation and, 95
F_2 generation
complex traits in, 755–757
in *D. melanogaster* studies, 150–153, 177
epistasis and, 115–117
of pea plants, 93, 95–100, 119
Principle of Assortment and, 99–101
Principle of Segregation and, 95–98
in trihybrid crosses, 102–103
F_3 generation, of pea plants, 96–97
factorials, 160–161
table of, 160
factor VIII, in hemophilia, 154–155
familial cancer, 655, 660–666
adenomatous polyposis of the colon, 661
breast cancer, 74, 661
melanoma, 661
retinoblastoma, 661–665, 674
fatal infantile multisystem disorder, 679
fate maps, of *D. melanogaster*, 608, 609
fava beans, diploid chromosome number of, 134
FC0 allele, genetic code and, 473–475
female germ line, genomic imprinting in, 520–522
females
distribution of heights in, 752
fragile-X syndrome and, 272
mutation rates in *D. melanogaster*, 282–283
recombination in, 180–181, 186–187
with sex-chromosome abnormalities, 362
sex chromosomes and, 148–158
and white-eye *D. melanogaster* mutation, 150–153, 155–157, 177–181
X inactivation in, 352–353
ferns, diploid chromosome number of, 134
fertility, in fitness, 733
fertility restorers, in cytoplasmic male sterility, 691
fertilization
independent, 110
in pea plants, 96–97
fetus, 360
fibroblast growth factor receptor (FGFR), 658
fibroin, translational control of, 524
59-base element, in integrons, 400–401
first-division segregation, in *Neurospora crassa*, 207–209
first meiotic division, 141–148
first nuclear division, in meiosis, 140
Fischer, Emil, 12
fish
cave fish, 760–762
genetically engineered salmon, 564–565
random amplified polymorphic DNA analysis of, 70
sex chromosomes in, 155
Fisher, Ronald Aylmer, 164, 166–167C
fitness
components of, 732–733
relative, 731
5'-phosphate (5'-P) group, in polynucleotide chains, 43
5' cap, of mRNA, 458
fixed alleles, 710
floral mutations, of *Arabidopsis thaliana*, 624
floral organs, of plants, 622–626, 711
"Florida" navel orange, as somatic mutation, 266
flower development, 622–626, 630
in *Arabidopsis thaliana*, 623–624, 630
in four-o'clock plant, 684–685
flowers
of pea plants, 89, 91, 93, 115–123, 161–162
of *Primula officalis*, 711
of snapdragons, 111, 127
fluorescence
in automated DNA sequencing, 244–246
in chromosome painting, 347
fluoxetine, 784
foldbacks, in DNA molecule, 279
folding domains, 462–463
Fölling, Asbjörn, 32
Fölling, Ivar, 32
food safety science, DNA typing in, 75
forward mutations, 300, 729
four-o'clock plant, leaf variegation in, 684–685
four-strand double crossing-over, 191–192
in tetrad analysis, 204–205
four-strand stage of meiosis, crossovers during, 187–189
14-3-3σ protein, 650–651, 653
foxtail millet, synteny groups in, 383–384
F plasmids, 395–396
in artificial chromosomes, 569–570
in conjugation, 405–412
R plasmids and, 402
F' plasmids, 411–412
in *E. coli* lactose metabolism, 492–494
fragile site mental retardation-1 (*FMR1*) gene, 273
fragile-X chromosome, trinucleotide repeat and, 272–273
fragile-X syndrome, 272
frameshift mutations, 271–272, 303
genetic code and, 473–476
Franklin, Rosalind, 48
fraternal twins, 762–763
free radicals, ionizing radiation and, 290–291
frequency of cotransduction, 417, 433
frequency of recombination, 177–178
in bacteriophage genetic mapping, 420–422
cotransformation frequencies and, 405
in coupling versus repulsion configurations, 178–179
with Hfr cells, 405
in human pedigrees, 198–201
of linked genes, 180
map distance and, 182–186
mapping functions and, 195–196
minimum and maximum, 190–192
Friedrich ataxia, dynamic mutation and, 273
fringe gene, in *D. melanogaster*, 618
fs(1)K10 mutation, in *D. melanogaster*, 555
F transfer, Hfr transfer versus, 407
fucose, human ABO blood groups and, 113
functional genomics, 578–581
fungi. *See also Aspergillus; Neurospora crassa; Saccharomyces cerevisiae*
as eukaryotes, 28
gene conversion in, 248–249, 254
genetic mapping of, 201–209, 254
genetic studies of, 450–451
genome sizes of, 313
methylation in, 519
mitotic recombination in, 209–211
RIP mutations in, 280
telomeric DNA sequences of, 336

G

G_0 period, 648–649
G_1 (gap 1) period, of interphase, 136, 636–639
G_1 restriction point, 646
G_1/S checkpoint, 649–650, 652–653
cancer cells and, 656
G_1/S transition, 646–647, 649–652
cancer cells and, 656
G_2 (gap 2) period, of interphase, 136, 636–639
G_2/M checkpoint, 649–652
G_2/M transition, 646–647, 649–652
GADD45 protein, 650–651, 653, 659
gain-of-function alleles, 604–607
gain-of-function mutations, 267
GAL genes, in yeast, 508–509
gal4 mutation in yeast, 508–509
GAL4 protein
chemical structure of, 510
three-dimensional structure of, 509
as transcriptional activator, 510
in yeast, 508–509

galactose
human ABO blood groups and, 113
metabolism in yeast of, 508–509
gal mutants, of *E. coli*, 415–417, 432–435
gal-transducing phage, 434–435
gametes
chromosomes in, 135
of males and females, 149–150
meiosis and, 139–148
Principle of Assortment and, 99–101
Principle of Segregation and, 94–95
in trisomic segregation, 361
gametophytes, in meiosis, 141
γ-globin gene, in humans, 518
γ rays, 290
gap genes, of *D. melanogaster*, 611–612, 614, 616
Garrod, Archibald E., 12–15C, 449–450
gastrula
of *D. melanogaster*, 607–608
in embryonic development, 594
gastrulation-defective gene, 614
Gautier, Marthe, 358–359C
G-bands, in human karyotype, 347–348
gel electrophoresis
with dideoxy sequencing method, 243–244
in DNA purification, 549
of genomic DNA, 52–54
pulsed-field, 568–569
in Southern blot, 57–58
gene amplification, 525
gene conversion
Holliday recombination and, 250–252, 254
mismatch repair and, 248–249
gene expression, 446
in organism development, 593–594
gene fusion, acute leukemias from, 667
gene inactivation, in cloning vectors, 551–553
gene patents, 566
gene pools
evolution and, 727
populations as, 708
gene products, 15–16
generalized transducing phages, 413–415
generalized transduction, between bacteria, 413–415
gene regulation, 490–532, 538
described, 490
epigenetic mechanisms of, 518–522
in eukaryotes, 507–518
in λ bacteriophage, 505–507
operon system in, 492–500
of organism development, 592–594
programmed DNA rearrangements in, 524–532
in prokaryotes, 490–492
RNA processing in, 522–524
translational control in, 524
tryptophan operon in, 499–505
genes
affecting complex traits, 761–762, 772–776
analysis of complex, 567–578
antibodies and, 525–528
in bacterial resistance to antibiotics, 394, 398–402
characteristic recombination frequency of, 180
chromosomes and, 134
cistrons as, 426–427
cloning of, 542
complementation tests and, 118–123
in development and maturation, 25
DNA chip assays of, 578–581
dosages of, 346
effects of multiple, 25–26
effects of mutations of, 267–268
environment and, 25–26
enzymes encoded in, 15–16
in evolution, 27–29
expression of, 16–21
in genetic analysis, 64
in genomes, 27, 39–41
for germ-line extinction, 328
homeotic, 613
housekeeping, 507
introns in, 462
linkage groups of, 184–186
linkage of, 98, 101, 108–109, 176–181, 184
of maize, 754–755
maternal-effect, 609, 612–614
during meiosis, 147
in mitochondrial DNA, 680–681
multiple alleles of, 111–112
mutation of, 21–25
mutation rates of, 282–284
nonchromosomal, 686–688
in organism development, 592–594
origin of new, 462–463
polypeptide chains and, 450–451
proteins and, 11–16, 49
recombination of, 176–192
recombination within, 209
segmentation, 611–612
segregation of multiple, 98–103
segregation of single, 90–98
in sex chromosomes, 355–357
sex-specific imprinting of, 520–522
syntenic, 178
target, 560
three distinct meanings of, 426–427
in trait inheritance, 2
traits and, 25–26
transcriptional cosuppression of, 520
transposition of, 274–280
unlinked, 101
zygotic, 609, 614
gene targeting, 560
gene therapy
human, 525, 584
with retroviruses, 567
genetically engineered crops, 585
genetic analysis
complementation tests in, 117–123
of complex traits, 760–763
probability in, 107–111, 158–161
terminology of, 64–65
genetic code, 15–16, 19–21, 49, 472–481
of organelles, 683–684
redundancy of, 478–479
table of codons in, 20, 476
triplet nature of, 472–475
wobble in, 478–479
genetic crosses
backcross, 98, 690, 761
dihybrid, 92, 99, 115–117
hybrid, 89–98
monohybrid, 89,92, 93
outcrossing, 91
probability in analysis of, 107–111
reciprocal, 92
with T4 bacteriophage, 420–422, 424–425
testcross, 98, 101, 119, 176–178, 192–197, 363
trihybrid, 92, 102–103
genetic drift, 728, 736–739
genetic engineering, 54, 542
applications of, 564–567
cloning strategies for, 546–551
of crops, 585
patents for, 566
genetic exchange, among bacteria, 403
genetic instability, cancer cells and, 656
genetic linkage. *See* linkage.
genetic linkage map, 573–574
genetic mapping, 73–74, 176, 181–192
of bacteriophage, 418–427
of complex genomes, 572–578
cotransduction in, 416–417
cotransformation and, 404–405
deletions in, 363–364
in human pedigrees, 197–201
mitotic recombination and, 210–211
positional cloning and, 554–555
of quantitative trait loci, 772–773
by tetrad analysis, 201–209
in three-point testcrosses, 192–197
time-of-entry, 408–411
of *trpA* gene, 451
genetic mapping functions, 195–196
genetic markers, 39–41, 73–74, 181
cotransformation of, 404–405
linkage groups and, 184–186
mapping functions and, 195–196
in mitotic recombination, 210
recombination and, 182–183
genetics
evolution and, 28–29, 727–728
fundamental concept of, 2
types of, 2
genetic switch, λ bacteriophage proteins as, 507
genetic transfer, among bacteria, 403
genital disk, of *D. melanogaster*, 617
genomes, 27, 312
of grasses, 382–384
individual differences and, 39–41
large-scale sequencing of, 38–39, 262
positional cloning of, 554–555
repetitive DNA sequences in, 329–330
sizes of, 72, 312–314, 327–329
of T4 bacteriophage, 420–427
genomic DNA, 40
chemical isolation of, 50–52
cloning of large, 569–571
electrophoretic separation of large fragments of, 567–569
gel electrophoresis of, 52–54
in nucleosomes, 318–319
in random amplified polymorphic DNA, 69–70
for recombinant DNA, 542–544
selective joining of, 547–549
selective replication of, 49, 58–64
separation and identification of, 49–58
in transduction, 412–417, 427, 429–435
in transformation, 403–405
genomic imprinting, 520–522
evolution of, 522
genomics, 27
functional, 578–581
genotype, complex traits and, 750
genotype–environment (G–E) interaction, 759–760
genotype frequencies, 708–710
Hardy–Weinberg principle and, 712–715
mating systems and, 712
multiple alleles and, 715–716
X-linked genes and, 721–722

genotypes, 65, 96. *See also* alleles; heterozygous genotypes; homozygous genotypes
of bacteriophage, 418–422
causes of variation in, 754–760
complementation tests and, 118–123
designating bacterial, 403
epistasis and, 115–117
heritability of, 764–766
with human blood groups, 112–113
in human pedigrees, 104–107
multiple alleles and, 111–112
Principle of Assortment and, 99–101
Principle of Complementation and, 120
Principle of Segregation and, 94–95
probability and, 107–111, 158–161
in random amplified polymorphic DNA analysis, 70
relative fitness of, 731
semisterility of, 372
in transmission genetics, 88, 90–93
genotypic variation, 755–762
environmental variance versus, 758–759
in twins, 762–763
germ cells
cancers from, 660–666
chromosomes in, 135
in familial retinoblastoma, 662, 665
germ-line extinction, in *Caenorhabditis elegans*, 328
germ-line mutations, 266
germ-line transformation
in animals, 556–560, 562
cancer and, 664
giant chromosomes. *See* polytene chromosomes
giant mutation, in *D. melanogaster*, 555, 615
giant salmon, from genetic engineering, 564–565
gibbon, diploid chromosome number of, 134
Giemsa reagent, 347
gigabases (Gb), 313
glossy leaf (*gl*) allele, in maize, 192–194
glutamic acid (Glu, E)
DNA base sequence for, 20, 476
in *E. coli* intragenic suppression, 301–302
molecular structure of, 447
in PAH, 21
sickle-cell anemia and, 269–270, 476
glutamine (Gln, Q)
DNA base sequence for, 20, 476
molecular structure of, 447
glycine (Gly, G)
DNA base sequence for, 20, 476
in *E. coli* intragenic suppression, 301–302
genetic code and, 475
in Li–Fraumeni syndrome, 662
molecular structure of, 447
glycosyl transferase, with human blood groups, 112–113
goodness of fit, 161–167
G protein, 658
G-quartet structure, 337
Gramineae, genome evolution in, 382–384
grasses, genome evolution in, 382–384
gray (Gy), 290
"green-blindness" (deuteranopia), 366–367
green-pigment genes, 391
in color blindness, 365–366
green–red chimeric genes, in color blindness, 366–367
Greider, Carol W., 334C
Griffith, Frederick, 3–4
growth-factor receptors, 656–658
GTPase, 658
GTP hydrolysis, in protein release, 468
guanine (G)
in dideoxy DNA sequencing, 243–245
as DNA base, 9–11, 19, 41–49, 224–225
in G-quartet, 337
nucleotide substitution and, 268
as RNA base, 19
guanine methylation, 299
guanosine, in capping mRNA, 458–459
guanosine 5'-triphosphate (GTP)
in elongation, 466
in RNA synthesis, 452
guide RNA, in telomeres, 334–335
Guthrie, Woody, 108
gyrase, in DNA replication, 234–235

H

*Hae*II restriction enzyme, 51
mitochondria and, 678–679
Haemanthus
chromosomes of, 134
mitosis of, 138–139
Haemophilus, core proteome of, 576
Haemophilus aegyptus, restriction enzymes from, 51
Haemophilus influenzae, restriction enzymes from, 51
Hahn, William C., 652C
hairpins, in DNA molecule, 279, 455
hairy gene, 615
Haldane, John Burdon Sanderson, 195–196
Haldane's mapping function, 195–196
halteres
of *D. melanogaster*, 619
evolution of, 620
haploid chromosomes, 135
of bacteria and bacteriophage, 402
in meiosis, 140–143, 148
numbers, table of, 312
RIP mutations in, 280
haploid chromosome set, monoploid chromosome set versus, 378
Hardy, Godfrey, 712–713
Hardy–Weinberg principle
demonstration of, 716–717
DNA typing and, 716–720
heterozygous genotype frequencies and, 714–715
inbreeding and, 723–724
multiple alleles and, 715–716
random mating and, 712–714
Hartwell, Leland H., 644C
HB activator protein, 514
HD allele, human pedigrees of, 104–105
heavy (H) chain, in immunoglobulin, 525–527, 666–667
hedgehog gene, in *D. melanogaster*, 618
height, distribution in British women, 752
helicase proteins, in DNA replication, 234–235
hemoglobin, 268
in high-F disease, 518
protein subunits of, 449
in sickle-cell anemia, 269–270
hemophilia, in human pedigrees, 154–155
hepatitis virus, synthetic vaccines and, 567
hereditary diseases
cancer as, 660–666
causes of, 12–15
DNA polymorphisms and, 73–74
mitochondrial, 678–680
hereditary nonpolyposis colorectal cancer (HNPCC), 298, 661
heredity
chromosomes and, 134, 157
DNA in, 2
heritability
broad-sense, 762, 763
distribution of traits, 751–753
narrow-sense, 764–766, 769–770
hermaphrodites, of *Caenorhabditis elegans*, 598–599
herpes simplex virus, genome size of, 312
Hershey, Alfred D., 6, 8C, 417C
Herskowitz, Ira, 530C
heterochromatin
in chromosomes, 196–197
in highly repetitive DNA sequences, 329–330, 332
in polytene chromosomes, 323
position-effect variegation in, 375–377
heteroduplex regions
in DNA repair, 249
in Holliday recombination, 250–252
in single-strand break and repair model, 253–254
heterogametic sex, 150
in chickens, 155
heterogeneous populations, in genetic analysis, 760–761
heteroplasmy, 680–681
heteropyknotic X chromosome, 354
heterosis, 766–767
heterozygote superiority, 735–736
heterozygous genotypes, 65
allele configurations in, 178–179
allele frequencies of, 708–710
in artificial selection, 763–764
with calico cats, 353–355
complementation tests and, 118–123
with Δ*32* mutant, 708–709
in *D. melanogaster*, 152
effects of deletions from, 363
epistasis and, 115–117
in fitness experiments, 733
gene recombination and, 176–178
with human blood groups, 112–113
in human pedigrees, 104–107
inbreeding and, 722–727
incomplete dominance in, 111
inversion in, 367
in mitotic recombination, 210
multiple alleles and, 715–716
in natural selection, 732
of *Paramecium aurelia*, 693
in pea-plant genetics, 90–93, 95–103, 166–167
polyploid, 377–379
Principle of Assortment and, 99–101
Principle of Complementation and, 120
Principle of Segregation and, 94–95
probability and, 107–111
in random amplified polymorphic DNA analysis, 70
rare alleles in, 714–715
reciprocal translocation in, 370–372
restriction fragment length polymorphisms and, 67–68
in selection–mutation balance, 733–735
with sickle-cell trait, 270, 736–737
simple tandem repeat polymorphisms and, 72
single-nucleotide polymorphisms and, 66
testing, 161–162
in three-point testcrosses, 192–197

translocation complexes with, 372–373
translocation in, 370
hexaploid chromosome set, 377
hexaploidy, in *Triticum aestivum*, 380
Hfr (high frequency of recombination) cells, 405–412, 440
Hfr transfer, F transfer versus, 407
Hfr × F^- mating, 405–408
genetic mapping with, 408–411
HGD gene, 16
Hicks, James B., 530C
high-copy-number plasmids, 394–395
high-F disease, 518
highly repetitive sequences, of DNA, 329–330
high significance, in statistics, 163
high-throughput genomic sequencing, 574–576
*Hin*dIII restriction enzyme, 51
hip dislocation, heritability of, 771
histidine (His, H)
in Ames test, 303
DNA base sequence for, 20, 476
molecular structure of, 447
histidine operon, 504–505
histones, 312, 318–323
types of, 318–320
HML genes, 529–530, 538
HMR genes, 529–530, 538
Hodgkin, Jonathan, 328C
HO endonuclease, 529, 532
HO gene, 529–530
Holley, Robert W., 472
Holliday, Robin, 250, 254C
Holliday junction-resolving enzyme (RuvC), 252
Holliday junctions, in Holliday recombination, 251–252, 263
Holliday model of recombination, 250–252, 254, 263
Holliday structures, 251–252
holocentric chromosomes, centromeres of, 331
holoenzymes
in DNA replication, 237–238
for RNA polymerases, 452
for transaction activation, 512–515
homeobox, in *HOX* genes, 620
homeotic genes, 613
of *D. melanogaster*, 616, 618–620
homeotic mutations, 600–601
homogametic sex, 150
in chickens, 155
homogentisic acid, alkaptonuria and, 12–15
homogentisic acid 1,2-dioxygenase (HGD), in tyrosinemia, 15–16
homologous chromosomes, 64–65
in genetic mapping, 181
in meiosis, 139–148, 176
in mitotic recombination, 210
homologous recombination, 254–255
homoplasmy, 681
Homo sapiens
amplification of DNA of, 62
centromeres of, 332
characteristics of, 2
chromosomes of male of, 135
cloning of, 558
complex traits in, 750
contig clones of, 555
diploid chromosome number of, 134
DNA base composition of, 44
DNA typing of, 716–720
γ-globin gene in, 518
genetically engineered genes for, 566
genome size of, 312–313
Hardy–Weinberg principle and, 716–717
inbreeding in, 724–727
mitochondrial DNA in, 696–698
mitochondrial family tree of, 697
mitochondrial mutations in, 678–680
sex chromosomes in, 149–150
spontaneous DNA damage rate for, 295
twin studies of, 762–763
homothallism, in yeast, 528–530
homozygous genotypes, 65
allele frequencies of, 708–710
in artificial selection, 763–764
complementation tests and, 118–123
with Δ*32* mutant, 708–709
effects of deletions from, 363
environmental variance in, 757–758
epistasis and, 115–117
in fitness experiments, 733
gene recombination and, 176–178
with human blood groups, 112–113
in human pedigrees, 104–107
inbreeding and, 722–727
in mitotic recombination, 210
multiple alleles and, 715–716
in natural selection, 732
of *Paramecium aurelia*, 693
in pea-plant genetics, 90–93, 95–103, 166–167
Principle of Assortment and, 99–101
Principle of Complementation and, 120
Principle of Segregation and, 94–95
probability and, 107–111
in random amplified polymorphic DNA analysis, 70
rare alleles in, 714–715
restriction fragment length polymorphisms and, 67–68
in selection–mutation balance, 733–735
with sickle-cell trait, 736–737
simple tandem repeat polymorphisms and, 72
single-nucleotide polymorphisms and, 66
testing, 161–162
translocation complexes with, 373
translocation in, 370
Hoogstein base pairing, 337
horsetails, diploid chromosome number of, 134
host organisms, for recombinant DNA, 542
host range (h^+) allele, 418–420
hot spots, of mutation, 284–285
housekeeping genes, 507
HOX genes
of *D. melanogaster*, 618–620
in evolution, 620–621
*Hpa*II enzyme, 518–519
HPD gene, 15
H polysaccharide, in human blood cells, 113
hTERT gene, 652
huckbein gene, in *D. melanogaster*, 615
human ABO blood groups, 111–115
allele frequencies of, 716–717
human beings. *See Homo sapiens*
human color blindness, 365–367
human genes
classification of, 574–575
genetic engineering of, 566
inbreeding coefficient of, 724–725
human genome
linkage maps of, 220
mapping risk factors in, 197–201
non-LTR retrotransposons in, 278
sequencing, 68, 572
Human Genome Project, 38–39, 572, 574–576
human growth hormone, genetic engineering of, 566
human immunodeficiency virus (HIV)
AZT inhibition of, 287
detecting with DNA amplification, 64
dideoxynucleoside analogs and, 247
resistance to, 708–709
synthetic vaccines and, 567
human karyotype, 346–353, 355–357, 391
human pedigrees, 104–107
human polymorphisms, 68
human population history, DNA typing in, 75
human pregnancies, chromosome abnormalities in (table), 357
hunchback (*hb*) gene, in *D. melanogaster*, 514, 613–615
Huntington disease (HD), 108–109
dynamic mutation and, 273
human pedigrees of, 104–105
Huntington's Disease Collaborative Research Group, The, 108–109C
hybrid crosses, of pea plants, 89, 90–98
hybridization
in chromosome painting, 347
DNA–RNA, 461
hybridization methods. *See* nucleic acid hybridization
hybrid vigor, 767
hydrogen bonds, in DNA, 46
hydrophobic amino acids, in protein folding, 23
hydrophobic surfaces, of DNA molecule, 46–47
4-hydroxyphenylpyruvic acid, 13
in tyrosinemia, 15
4-hydroxyphenylpyruvic acid dioxygenase (HPD), in tyrosinemia, 15
hypermorphic mutations, 267
hyphae, of *Neurospora crassa*, 206
hypomorphic mutations, 267
hypospadias, heritability of, 771
hypoxanthine, 288

I

I^A allele, with human blood groups, 112–113
I^B allele, with human blood groups, 112–113
identical twins, 762–763
correlation between, 770
identification, via DNA typing, 74–75
identity by descent, 724–725
Ig antibodies, 525–526
imaginal disks, in *D. melanogaster* metamorphosis, 616–617
imino acid, proline as, 447
immune system. *See also* acquired immune deficiency syndrome (AIDS); human immunodeficiency virus (HIV)
diseases of, 584
immunity, to bacteriophage, 433
immunoglobulins, 525–527
imprinting, in embryonic development, 597–598
inborn errors of metabolism
alleles and, 64
environment and, 25
genetic basis of, 12–15, 38, 449–451
mutations and, 21
Inborn Errors of Metabolism (Garrod), 14, 450

inbreeding, 712, 722–727
deleterious effects of, 726–727
Hardy–Weinberg principle and, 723–724
in human pedigrees, 106
inbreeding coefficient, 722–724, 766–767
inbreeding depression, 766–767
incomplete dominance, 111–115, 127
incorporation errors, in DNA strands, 224–225
independent assortment, 110
chromosomes and, 134
gene recombination and, 176–178
principle of, 99–101
in trihybrid crosses, 102–103
independent fertilization, 110
independent possibilities, in probability, 110–111, 158–159
independent segregation, 99–101, 110
in trihybrid crosses, 102–103
individual selection, 764
inducers
in lactose metabolism, 495–496
in negative regulation, 491
inducible systems
in lactose metabolism, 493–494
in negative regulation, 491
infertility, of triploid plants, 378
inherited cancer syndromes, table of, 661
inherited traits
extranuclear, 678–698
mutations as, 266
of species, 2
initiation
of DNA strands, 224, 235–236
of RNA strands, 455
in translation, 464–466
initiation complex 48S
in translation, 464–465
insecticidal proteins, 556
insects
DDT-resistant, 282
delta-endotoxins versus, 556
diploid chromosome numbers of, 134
evolution of *HOX* genes in, 620–621
evolution of wings in, 620–621
genome sizes of, 72
germ-line transformations in, 562
polytene chromosomes of, 323–324
proteomes of, 27
sex chromosomes in, 149–150, 155
telomeric DNA sequences of, 336
insertion
genetic code and, 473–475
mutation via, 271–272, 276
insertion sequences (IS elements), 396–397
in situ hybridization, polytene chromosomes and, 324
instars, of *D. melanogaster*, 608
insulin, genetic engineering of, 566
Int1 integrase, 399–401
integrases
for cassettes, 399–401
in lysogeny, 431
integration, of phage DNA into bacterial genomes, 431
integrons, 398–401
interbands, in human karyotype, 348
intercalation, mutagenesis via, 289
intercellular signaling, 606
interference, chromatid versus chromosome, 194
intergenic suppression, 302
interphase, 135–137
interrupted-mating technique, genetic mapping with, 408–411
intersex genes, 158
intragenic suppression, 301–302
introns
evolutionary origin of, 463
mutations in, 462–463
numbers of, 462
in reverse transcription, 551
in RNA, 458–462
invasion
in double-strand break and repair model, 254–255
in single-strand break and repair model, 253
inversion loops, 367–369
inversions, 349, 362, 367–369
in DNA molecule, 279
in sex-chromosome evolution, 355–357
invertebrates, telomeric DNA sequences of, 336
I^O allele, with human blood groups, 112–113
iodine-131, from Chernobyl, 294
ionizing radiation
annual exposure to sources of (table), 292
genetic effects of (table), 293
as mutagen, 290–293
in tumor therapy, 292
iron deficiency, genetic engineered rice and, 565
irreversible mutation, 728–729
IS50 insertion sequence, 397
isochromosomes, 349
isoleucine (Ile, I)
DNA base sequence for, 20, 476
mitochondrial codons for, 683
molecular structure of, 447
isoleucine operon, 504
isopropylthiogalactoside (IPTG), in lactose metabolism, 493–494

J

J (joining) regions, in antibodies, 527–528
Jacob, François, 457C, 495, 500C, 505
Jordan, Michael, 774

K

K12(λ) strain, of *E. coli*, 423, 428, 431, 473–474
kanamycin resistance, 570
kappa particles, 693–694
of *Caedobacter taeniospiralis*, 691
karyotypes, 347
of humans, 346–353, 355–357, 391
with sex-chromosome abnormalities, 362
karyotype symbols, table of, 349
Kennedy disease, dynamic mutation and, 273
K gene, in *Paramecium aurelia*, 691
Khorana, Har Gobind, 472
killer strains, of *Paramecium aurelia*, 691–694
kilobases (kb), 50, 313
Kind, Alex J., 558C
kinetochores, 654
in cell cycle, 636–637
during mitosis, 138
kinetosomes, origin of, 694
kitchen blender experiment, with bacteriophage, 6–9
Klinefelter syndrome, 362
knirps gene, in *D. melanogaster*, 615
knockout mutations, 267, 276, 562–564, 662
Knudson, Alfred G., 662, 665C
Kosambi's mapping function, 195–196
Krüppel gene, in *D. melanogaster*, 614–615
K strain, of *E. coli*, 428–429

L

laboratory selection, effects of, 730–731
lac genes, in lactose metabolism, 493–498, 500
lac mutants, 403
of *E. coli*, 410, 412, 492–499
lac operon, 495–496
positive regulation of, 497–499
repression loop of, 499
lactose, bacteria unable to utilize, 403
lactose metabolism
in *E. coli*, 492–499
lactose permease in, 492–493
regulatory elements in, 495–496
regulatory genes in, 500
repressible systems in, 493–494
lacZ gene. *See also* β-galactosidase
in lactose metabolism, 493–499
in pBluescript II plasmid, 552–553
promoter mutations of, 456
lagging strand, in DNA replication, 225, 238–241
λ bacteriophage, 427–435, 440
cloning vectors from, 545–546
gene regulation in, 505–507
genetic map of, 427–429, 506
genome size of, 312
plaques from, 419
large-scale genomic sequencing, 38–39, 574–576
large-T antigen, 652
lariat, in RNA splicing, 459–460
lateral inhibition, 594
lawns, of *E. coli*, 419
lazy growth habit (*lz*) allele, in maize, 192–194
leader polypeptide
in histidine synthesis, 505
in phenylalanine synthesis, 505
in tryptophan synthesis, 502–503
leading strand, in DNA replication, 225, 239
leaky mutations, 267
Lederberg, Esther, 281
Lederberg, Joshua, 281, 405, 413C
Lejeune, Jerome, 358–359C
Lepidoptera order
delta-endotoxins versus insects of, 556
evolution of wings in, 620–621
leptotene, 141–142
in *Lilium longiflorum*, 144
Les6 allele, in maize, 185
let-23 gene, 607
LET-23 protein, 606–607
lethal mutations
deletion as, 363
in *D. melanogaster*, 282–284
translocation as, 372
leucine (Leu, L), 549
DNA base sequence for, 16, 20, 476
mitochondrial codons for, 683
molecular structure of, 447
in PAH, 21
in selective medium, 407–408
leucine auxotroph, 403
leucine operon, 504
leukemias
acute, 666–667
from ionizing radiation, 293

leu mutants
cloning, 549
of *E. coli*, 407–408, 410, 414–416
Lewis, Edward B., 613
LHON (Leber's hereditary optic neuropathy), 679
liability, of threshold traits, 751, 770–772
libraries, of clones, 553, 572
Li–Fraumeni syndrome, 661–662
pedigree of, 663
ligands, 607
ligation, in double-strand break and repair model, 254–255
light (L) chain, in immunoglobulin, 525–527
likelihood ratios, 200
Lilium longiflorum
later meiotic stages in, 146
prophase I substages in, 144
lily
diploid chromosome number of, 134
prophase I in, 144
Limnaea peregra, shell coiling in, 695–696
lin-3 gene, 606–607
LIN proteins, 605–607
lin mutants, 601, 603–605
lineage diagrams, 599–600
of *Caenorhabditis elegans*, 630
linkage
chi-square testing of, 179–180
frequency of recombination and, 180
gene recombination and, 176–181
of genes, 98, 101, 108–109, 176, 184
in hereditary diseases, 74–75
human pedigrees and, 197–201
in tetrad analysis, 202–205
linkage groups, 184–186
linkage maps, 181. *See also* genetic mapping
of human chromosomes, 220
linker DNA, in nucleosomes, 320
Liskay, R. Michael, 297C
localized centromeres, 331
local populations, 708
migration and, 728
locus (loci)
complementation tests and, 121–122
crossing-over and, 186–187
of genes in chromosomes, 64–65
of genetic risk factors, 74–75
in human genome, 68
mutable, 274–275
quantitative-trait, 772–773
lod (log-odds) scores, 198–201
long interspersed elements (LINEs), 278
long-term artificial selection, 766
long terminal repeats (LTRs), in transposition, 277
loss-of-function alleles, 604–607
loss-of-function mutations, 267
loss of heterozygosity, cancer and, 664
lost alleles, 710
low-copy-number plasmids, 394–395
loxP sequence, 398–399
lozenge (*lz*) gene, in *D. melanogaster*, 209
LTR retrotransposons, in transposition, 277–278
Lundberg, Ante S., 652C
Luria, Salvador, 8
Lwolf, André, 500
Lycopersicon chmielewskii, quantitative trait loci in, 773
Lycopersicon esculentum, quantitative trait loci in, 773
Lyon, Mary F., 353–354C
lysine (Lys, K)
DNA base sequence for, 20, 476
in histones, 318
molecular structure of, 447
lysis, of bacteria, 6, 418
lysogen, 431
lysogeny, 429–430
of λ bacteriophage, 427–433, 505–507
lytic cycle
of bacteriophage, 6, 413, 418–422
of λ bacteriophage, 427, 431

M

M (mitosis) period, 636–639
interphase and, 136
M1V mutant, of *PAH* gene, 22
M13 bacteriophage
cloning vectors from, 545–546
genome size of, 312
MacLeod, Colin M., 3–5, 8, 50, 544–545
macronucleus, of *Paramecium aurelia*, 692–693
Mad protein, 654
MADS box gene family, 624
maize (corn). *See also Zea mays*
agriculture of, 754–755
complex traits in, 750
cytoplasmic male sterility in, 690
diploid chromosome number of, 134
DNA base composition of, 44
genetic engineering of, 556
genetic maps of, 184–186
genome size of, 313
genotype–environment interaction in, 759–760
methylation in, 519
synteny groups in, 383–384
telomeres in, 332
three-point testcrosses with, 192–194
translocation complexes in, 373
transposable elements in genome of, 278
transposition in, 274–276, 308
major groove, in DNA molecule, 45, 47
malaria
geographical distribution of, 735
sickle-cell anemia and, 270, 735–737
synthetic vaccines and, 567
male germ line, genomic imprinting in, 520–522
males
of *Caenorhabditis elegans*, 598–599
cytoplasmic sterility in, 688–691
with fragile-X syndrome, 272
mutation rates in *D. melanogaster*, 282–284
recombination in, 180–181, 186–187
red–green color blindness in, 365
with sex-chromosome abnormalities, 362
sex chromosomes and, 148–158
and white-eye *D. melanogaster* mutation, 150–153, 155–157, 177–181
4-maleylacetoacetic acid, 13
in tyrosinemia, 15
Malthus, Thomas, 730
maltose gene cluster, of *E. coli*, 410
mammals
apoptosis in, 604
cloning, 558
cyclin–CDK complexes in, 645
genomic imprinting in, 520–522
germ-line transformation in, 557–559
X chromosome in, 262
mannose-binding protein, 448
manual dideoxy sequencing, of DNA, 243–244
map-based cloning, 554–555
map distance
attached-X chromosomes and, 186–189
in genetic mapping, 182–186, 196–197
mapping functions and, 195–196
multiple crossovers and, 189–192
in tetrad analysis, 204–205
map units, 182
Marchantia polymorpha, chloroplast genome of, 683
Margulis [Sagan], Lynn, 694C
maternal-effect genes, of *D. melanogaster*, 609, 612–614
maternal effects, 681–682
in shell coiling, 695–696
maternal inheritance, 678
effects of, 681–682
of mitochondrial diseases, 678–680
maternal PKU, 33
maternal sex ratio, in *D. melanogaster*, 694–695
MAT genes, 529–531, 538
mating systems, 711–722
mating type, transcriptional control of, 529–532
mating-type interconversion, 528–529
maturation-promoting factor, 647
maximum likelihood, lod scores and, 198–201
McCarty, Maclyn, 3–5, 8, 50, 544–545
McClintock, Barbara, 274–275C, 308, 332
McDonald, Margaret, 6, 8
McWhir, Jim, 558C
Mdm2 protein, 650, 656–659
mean, in normal distribution, 751–753
media, for bacteria, 402–403
megabases (Mb), 313
megaspores, in meiosis, 141
meiocytes, in meiosis, 140–141
meiosis, 139–148, 176
of *Chlamydomonas*, 687
chromosome scaffold during, 322–323
in fungi, 201–202
inversion in, 367
in monoploid organisms, 381
in *Neurospora crassa*, 206–207
in *Paramecium aurelia*, 692–693
polyploidy and, 377–378
recombination during, 176–192
Robertsonian translocation during, 373–374
sex chromosomes and, 149
translocation during, 370–372
trisomic segregation in, 361–362
melanin, in Siamese cats, 266–267
MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis, stroke-like episodes), 679
melatonin, 784
melting temperature, of DNA, 55
Mendel, Gregor, 2, 12, 88–89, 96–97C, 98–99, 102, 107, 164
correctness of data of, 164–167
Principle of Assortment and, 99–101
Principle of Segregation and, 90–98
Mendelian (transmission) genetics, 88–123
and congenital abnormalities, 771–772
described, 88
Down syndrome and, 358
Hardy–Weinberg principle and, 716–717
human pedigrees in, 104–107
mating in, 711–722

Mendelian genetics (cont'd)
of pea plants, 88–103
in population genetics, 708–710
Principle of Assortment in, 99–101
Principle of Complementation in, 120
Principle of Segregation in, 94–95
probabilities in, 107–111, 158–161
Mendel's first law (Principle of Segregation), 94–95
Mendel's second law (Principle of Assortment), 99–101
mental retardation
from fragile-X syndrome, 272
from PKU, 15
meristems, 622
meristic traits, 751
merodiploids, F' plasmid and, 412
MERRF (myoclonic epilepsy associated with ragged-red muscle fibers), 679–681
Meselson, Matthew, 225, 228C, 457C
Meselson–Stahl experiment, 225–228
messenger RNA (mRNA), 16, 456–458, 464
abundance of, 551
of chloroplasts, 683
cloning from, 549–551
DNA chips and, 579, 581
identification of, 457
of mitochondria, 683
polycistronic, 470–471
in protein synthesis, 18–19, 21, 446
reading frame of, 473
stability, gene regulation via, 490, 522–533
translation of, 463–469
metabolic pathways, 13
blocks in, 117
for yeast galactose metabolism, 508
metabolic processes, genetic control of, 449–451
metabolites, in metabolic pathways, 14–15
metacentric chromosomes, 350
metallothionein, in genetic engineering, 564–565
metamorphosis, of *D. melanogaster*, 616–618
metaphase, 136–137, 637
in meiosis, 140
structure of chromosomes during, 321–323
metaphase I, 141, 143, 147
heterozygous reciprocal translocation during, 371
in *Lilium longiflorum*, 146
Robertsonian translocation during, 373–374
trisomic segregation during, 361
metaphase II, 143, 148
in *Lilium longiflorum*, 146
metaphase plate
during meiosis, 143
during mitosis, 138
metaphase spread, 346–348
Methanococcus jannaschii, genome size of, 312–313
methionine (Met, M)
DNA base sequence for, 16, 19–20, 476
molecular structure of, 447
in PAH, 21–22
in polypeptide chains, 20
methionine auxotroph, 402
methylation
of cytosine, 518–519
in genomic imprinting, 520–522
in transcriptional cosuppression, 520
transcriptional inactivation and, 519
methylation induced premeiotically (MIP) process, 519
methylation reversal, 519
5-methylcytosine
in DNA, 44, 284–285
molecular structure of, 518
methyl group ($-CH_3$), 518
7-methylguanosine (MeG), in capping mRNA, 458–459
methyl transferase, 299
microarrays, 578–581
micrococcal nuclease, 318–319
micromeres, in embryonic development, 597
micronuclei, of *Paramecium aurelia*, 692–693
microsatellite markers, 744
microsatellite STRPs, 72
microscope, confocal, 347
microspores, in meiosis, 141
middle-repetitive sequences, of DNA, 329–330
Miescher, Friedrich, 2
migration, evolution and, 728
millet, synteny groups in, 383–384
miniature wing (*m*) gene, in *D. melanogaster*, 178–179, 190
minimal medium, for auxotrophs, 402
minisatellite STRPs, 72
minor groove, in DNA molecule, 45, 47
Mirabilis jalapa, leaf variegation in, 684–685
misalignment, mutagenesis via, 289
mismatch repair, 295–298
in DNA replication, 248–249
dynamic mutation and, 273
mispairing, in duplication, 365
missense mutations, 269–271
mitochondria, 702
in *Chlamydomonas*, 688
DNA in, 312, 678–680
evolution of, 691, 694
genetic code of, 683–684
genetic sequencing of, 574–576
heteroplasmy in, 680–681
introns in, 461
of *Saccharomyces cerevisiae*, 688–689
mitochondrial DNA (mtDNA), 678–680, 683, 702
human, 696–698
maternal inheritance via, 681–682
from rat liver, 682
mitochondrial DNA typing, 702–703
"mitochondrial Eve," 696–698
mitochondrial genetic diseases, 678–680
table of, 679
mitogens, 645
mitosis, 135–139. *See also* cell cycle
in cell cycle, 636–639
of *Chlamydomonas*, 687
chromosome scaffold during, 322–323
in *D. melanogaster* development, 607–609
inversion in, 367
in *Neurospora crassa*, 206–207
in *Paramecium aurelia*, 692–693
recombination during, 209–211
triploidy in, 378
mitotic recombination, 209–211
mitotic spindles, 654–655
in cell cycle, 636–637
centrosome duplication checkpoint and, 653–654
during meiosis, 146
during mitosis, 137–139
mlh1 mutant, mismatch repair and, 297
mob1-77 mutant, cell cycle and, 642
mobile DNA, 394–402
moderately abundant mRNA, 551
"molecular fossil," RNA as, 17
molecular genetics, 2
central dogma of, 16–21
complementation tests and, 122–123
molecular markers, 79
in human pedigrees, 106–107
molecular phenotypes, 88–90, 92–93
complementation tests and, 118
monocentric chromosomes, 350
monocistronic mRNA, of eukaryotes, 471, 490
Monod, Jacques, 457, 495, 500, 505
monohybrid crosses, of pea plants, 89, 92–93
monomers
nucleotide, 224
in PAH molecule, 24–25
monomorphic DNA, 70
monophosphates, nucleotides as, 42
monoploid chromosome set, 377, 381
monoploidy, 381–382
monosomy, 358
of X chromosome in females, 362
Montgomery, Hugh E., 774–775C
Morgan, Thomas Hunt, 150–152C, 156, 173, 176–178, 182, 190
morphogens, in embryonic development, 595
morphological phenotypes, 88–90, 92–93
Morton, W. J., 290
mosaic (mos) individuals, 349–350
mosaicism
in calico cats, 353–355
of gene loci, 274
in mice, 354
position-effect variegation as, 376
of X inactivation in females, 353
mouse (mice). *See Mus musculus*.
Mps 1 protein, 654
mrt mutants, in *Caenorhabditis elegans*, 328
MS2 bacteriophage, genome size of, 312
msh2 mutant, mismatch repair and, 297
MspI enzyme, 518–519
mt cells, of *Chlamydomonas*, 686–688
Muller, Hermann J., 173, 267, 286, 291C, 332
Mullis, Kary B., 60, 78–79
multicellular organisms, evolution of, 28–29
multifactorial traits, 750
multiple alleles, 111–112
allele frequencies of, 715–716
complementation tests and, 118
with human blood groups, 112–115
simple tandem repeat polymorphisms and, 72–73
multiple cloning site (MCS), in pBluescript II plasmid, 552–553
multiple crossing-over, 189–192
in tetrad analysis, 203–205
multiple genes, traits influenced by, 750
multiple origins of replication, for eukaryote DNA, 232–233
multiplication rule, in probability, 110, 158–160
Murry, Jeffrey C., 198–199C
Mus musculus (mouse)
chromosomes of, 322
coat color of, 117
contig clones of, 555
diploid chromosome number of, 134
epistasis in, 117
genomic sequencing of, 574, 576
germ-line transformation in, 559
knockout mutations in, 662

life span of, 662
X chromosome dosage compensation in, 354
mutable genes, 274
mutable loci, 274–275
mutagenic primers, 563
mutagens, 286–295
base-analog, 286–287
chemical agents as, 288–289, 303
chemically detecting, 303
from Chernobyl nuclear accident, 293–295
defined, 286
ionizing radiation as, 290–293
misalignment as, 289
site-directed, 562–564
ultraviolet radiation as, 289–290
mutants
bacterial, 402–403
of bacteriophage, 417
inborn errors of metabolism and, 38
in maize genetics, 184–186
from mutations, 16
in pea-plant genetics, 88–90
mutant screens, 118
complementation tests and, 117–123
mutation, 16–25, 266–295, 300–303
affecting cell lineages, 600
affecting DNA mismatch repair, 297
cancer and, 636
complementation tests and, 118–123
conditional, 266–267
defined, 266
DNA polymorphisms and, 73–74
as DNA sequence change, 10, 49
in *D. melanogaster*, 150–153
evolution and, 727–730
in four-o'clock plants, 684–685
genetic code and, 472–475
germ-line, 266
hemophilia gene as, 154–155
of introns, 462–463
knockout, 562–564
in λ bacteriophage, 505–507
in mitochondria, 678–680
molecular basis of, 268–273
in monoploid plants, 382
multiple alleles and, 112
mutagens in, 286–295
natural selection versus, 733–735
nonadaptiveness of, 281–282
in *PAH* gene, 21–23
Principle of Complementation and, 120
of promoters, 456
reverse, 300–303
somatic, 266
spontaneous, 280–285
suppressor, 300–303
in T4 bacteriophage, 422–427
transposable elements as agents of, 278–280
transposition as, 274–280
types of, 266–268
types of cell-lineage, 600–604
x rays and, 211
mutation hot spots, 284–285
mutation rates, of genes, 282–284
Mut proteins, in mismatch repair, 296–297
mut genes, 297–298
mutually exclusive possibilities, in probability, 109, 158
Mycoplasma genitalium, genes of, 449
myelin, in PKU, 14–15
myopathy, 679
myotonic dystrophy, dynamic mutation and, 273
myxothiozol, algal resistance to, 688

N

^{14}N (normal nitrogen) isotope, in Meselson–Stahl experiment, 225–228
^{15}N (heavy nitrogen) isotope, in Meselson–Stahl experiment, 225–228
N-acetylgalactosamine, human ABO blood groups and, 113
N-acetylglucosamine, human ABO blood groups and, 113
"Nagel anomaloscope," 391
nail–patella syndrome (NPS), 220
nanos mutant, 613–614
Narcissus pseudonarcissus, genetic engineered rice and, 565
NARP (neurogenic muscle weakness, ataxia, retinitis pigmentosa), 679
narrow-sense heritability, 764–766
of congenital abnormalities, 771–772
estimating, 769–770
National Academy of Sciences, 293
National Research Council, 293
native conformation, in protein folding, 23
natural selection
evolution and, 28–29, 728, 730–736
harmful recessive alleles in, 732
necdin gene, in sex-specific imprinting, 521–522
negative assortative mating, 711
negative autoregulation, in organism development, 593
negative chromatid interference, in multiple crossing-over, 192
negatively supercoiled DNA, 314–315
negative regulation
of genes, 490–491
of yeast mating type, 531
nematodes
cell lineages of, 599–600
development of, 598–607
diploid chromosome numbers of, 134
proteomes of, 27
neomorphic mutations, 267
neomycin, genes for resistance to, 397
net wing veins (*net*) gene, in *D. melanogaster*, 197
Neurospora, 754
gene conversion in, 248–249
meiosis in, 202
methylation in, 519
Neurospora crassa, 14
analysis of ordered tetrads in, 206–209
DNA base composition of, 44
functions of genes of, 449–451
life cycle of, 206
RIP mutations in, 280
neuroticism, 775–776
neutral petites, 688–689
Nicholas II, Tsar
hemophilia gene and, 154–155
identification of remains of, 681, 702–703
Nicotiana longiflora, narrow-sense heritability in, 764–765
Nirenberg, Marshall W., 472
nitrocellulose, in Southern blot, 55, 57–58
nitrogen
in DNA, 41
in Meselson–Stahl experiment, 225–228
nitrogen isotopes, in Meselson–Stahl experiment, 225–228
nitrous acid (HNO_2), as mutagen, 288
Nomarski microscopy, 602
nonadaptiveness, of mutation, 281–282
nonchromosomal (NC) genes, 686–688
noncomplementation, 118
nonconjugative plasmids, 397–398
nondisjunction
in Down syndrome, 360
in *D. melanogaster*, 155–157
nondisjunctional segregation, 370–371
non-LTR retrotransposons, 278
non-Mendelian inheritance, 678
nonparental-ditype (NPD) tetrads, 202–205
ordered, 206
nonreciprocal translocations, 373
nonrecombination
crossing-over and, 182–183
in Holliday recombination, 250–252
in human pedigrees, 198–201
in single-strand break and repair model, 253–254
three-point testcrosses and, 192
nonselective medium, for bacteria, 403
nonsense mutations, 270–271, 479–481
nonsense suppressors, 302, 479–481
nonsynonymous mutations, 269
nopaline, in plant genetic engineering, 561
normal distribution, 751–753
normal wing, in *D. melanogaster*, 178–179
Notch gene, in *D. melanogaster*, 363, 605–606
*Not*I fragment, 569
N protein, in λ bacteriophage, 507
nuclear division, cell division and, 134
nuclear fusion, in *Paramecium aurelia*, 692
nucleases, in single-strand break and repair model, 253–254
nucleic acid hybridization
DNA chips and, 579–580
polymerase chain reaction versus, 64
polytene chromosomes and, 324
restriction fragment length polymorphisms and, 67–68
separation and identification of DNA fragments via, 49, 54–58
nucleoids, 317
nucleolus (nucleoli), in mitosis, 137–139
nucleoside analogs, clinical uses of, 247
nucleosides, in DNA, 42, 247
nucleosomes, 318–320
nucleotide analogs, as mutagens, 287
nucleotides
in DNA, 9–11, 16–17, 41–49, 224–225
in genes, 64
in multiple alleles, 111
repetitive sequences of, 324–330
in RNA, 452–453
nucleotide substitution, mutation via, 268–269
nucleus
in cell cycle, 636–639
chromosomes in, 134–135, 312
cloning animals from, 558
discovery of, 134
DNA in, 2, 40
in meiosis, 139–148
in mitosis, 135–139
RNA processing in, 470
null mutations, 267
numerator genes, 157–158
NURF (*nucleosome remodeling factor*) complex, 514
Nüsslein-Volhard, Christiane, 613C
nutrients, auxotrophs for, 402
nutritional mutants, among bacteria, 402

O

oat seeds, hull color and epistasis in, 117
ochre stop codon, 479–480
ochre suppressors, 479–481
octopine, in plant genetic engineering, 561
octoploid chromosome set, 377, 379
odds, 200
odd-skipped gene, 615
Oedipina poelzi, bivalent chromosomes of, 146
Oenothera, translocation complexes in, 372–373
Ohno's law, 262
oil yellow-1 (*oy-1*) gene, in maize, 185
Okazaki fragments
  in DNA replication, 234, 239
  joining of, 240–241
oligomerization domains, in PAH molecule, 24–25
oligonucleotides, 59
  amplified fragment length polymorphisms and, 71–72
  in DNA amplification, 60–63
  in polymerase chain reaction, 60–61
  primers as, 59–60
Oliver, C. P., 209
oncogenes, 656
  in acute leukemias, 666–667
  cancer and, 652–653
  table of, 657
oncogenesis. *See* cancer
Oncopeltus fasciatus, chromosome of, 322
On the Origin of Species (Darwin), 29
oocytes
  composition of, 598
  developmental regulation of, 593–594
  development of fertilized, 592–598
  gene amplification in *Xenopus*, 525
  in meiosis, 140–141
  organization of, 598
  sex chromosomes and, 149
open reading frame (ORF), in mRNA, 458
operator region, for lactose metabolism, 495
operators, in lactose metabolism, 495–496
operon model
  of gene regulation, 495–496, 500
  in lactose metabolism, 497–499
  in tryptophan synthesis, 499–505
orange pericarp-2 (*orp-2*) gene, in maize, 185–186
ordered tetrads, analysis of, 206–209
organelles, 678
  evolution of, 691
  genetic codes of, 683–684
  heredity in, 682–691
organisms
  bacteriophage as, 417
  core proteomes of, 576
  development of, 592
  DNA of, 39–41
  DNA sequencing of (table), 577
  evolution of, 28–29
  genetics of, 2
  populations of, 708
  transgenic, 542, 556–562, 564–567
origin of replication
  for DNA, 224, 230–231
  multiple, 232–233
  in plasmids, 552–553
origin recognition complex (ORC), 647
oriS gene, 570
orthologs, 576–578
Oryza sativa, synteny groups in, 384. *See also* rice
outcrossing, of pea plants, 91
ovalbumin mRNA, 550
overdominance, of heterozygous genotypes, 735–736
ovule, of pea plants, 91
Oxytricha, G-quartet base pairing in, 337
ozone layer, UV exposure and, 288–289

P

p ("petite") arm of chromosomes, 348–349
P1 artificial chromosomes (PACs), 569–570, 572
P1 bacteriophage
  artificial chromosomes and, 569–570
  cotransduction via, 415–417
  Cre recombinase gene in, 398–399
  generalized transduction via, 414, 415
  in positional cloning, 555
P_1 generation
  in *D. melanogaster* studies, 153
  of pea plants, 92–93, 96, 119
P1 promoter, 516–518
P2 promoter, 517–518
p16 gene, 661
p16 protein, 659
p19ARF protein, 659
p21 protein, 650–651, 653
  loss of function of, 659
p34 kinase subunit, 136
p53 gene, 661–663
p53 pathway, cancer and, 657
p53 transcription factor, 650–651
  loss of function of, 659
PAC/BAC contig, 574
PAC clones, 573–574
pachytene, 141–142, 144
  in *Lilium longiflorum*, 144
PAC vectors, 569–570, 572
PAH gene, 15, 17–18, 64
  mutations in, 21–23
PAH molecules, abnormal folding in, 24–25
pair-rule genes, of *D. melanogaster*, 611–612, 614–616
palindrome sequences, in DNA, 52, 542
panmixis, 716
P_{am} promoter, in integrons, 400–401
paracentric inversion, 368–369
parA gene, 570
paralogs, in protein families, 576
Paramecium aurelia
  killer strains of, 691–694
  nucleus of, 692–693
parasegments, of *D. melanogaster*, 610–612
parasitemia, 736–737
parB gene, 570
parental-ditype (PD) tetrads, 202–205
  ordered, 206
parental strands, in DNA, 224–225
parental types, alleles of, 177
PARp region, 355–356
PARq region, 355–356
partial digestion, 549
partial diploids, F^1 plasmid and, 412
Pascal's triangle, 160
patent ductus arteriosus, heritability of, 771
patents, for genetic engineering, 566
paternal inheritance, 678, 682
paths, in pedigrees, 724–725
pattern formation, in organism development, 592
pBluescript II plasmid, 552–553
PcG (Polycomb group) proteins, 511
Pds1p protein, 648
peach eye color, in *D. melanogaster*, 279
pea plants. *See Pisum sativum*.
pecanex mutation, in *D. melanogaster*, 555
pedigrees, 38, 104
  complex traits in, 750
  conventional charting symbols for human, 104
  human, 104–107, 154–155, 197–201, 272, 663, 724–726
  inbreeding and, 724–725
  inbreeding coefficient from, 725–726
  probability and, 107–111
P element, in germ-line transformation, 556–557
penetrance, of a genetic disorder, 104–105
PEO (progressive external ophthalmoplegia), 679
peptide bonds, in polypeptide chains, 446–447
peptide groups, in polypeptide chains, 447
peptidyl (P) site, in ribosomal subunit 60S, 464–465
peptidyl transferase, in elongation, 466
percent G + C, of DNA, 43–44, 324
percent recombination, 182
pericentric inversion, 368–369
periodicity, in *D. melanogaster* development, 616
period mutation, in *D. melanogaster*, 555
permissive conditions, in conditional mutations, 266
permissive temperature, in cell division cycle mutants, 640–641
Perrin, David, 500C
Petes, Thomas D., 297–298C
petite mutants, of *Saccharomyces cerevisiae*, 688–689
Petunia hybrida, transcriptional cosuppression in, 520
phage. *See* bacteriophage
Phaseolus vulgaris, genetic engineered rice and, 565
phenotypes, 65, 96. *See also* bands
  of bacteriophage, 418–420
  complementation tests and, 118–123
  complex traits in, 750–751
  designating bacterial, 403
  distributions of complex traits in, 751–753
  effects of deletions on, 363–364
  effects of mutations on, 267–268
  epistasis and ratios of, 115–117
  of four-o'clock plants, 684–685
  heritability of, 764–766
  with human blood groups, 112–113
  in human pedigrees, 104–107
  incomplete dominance in, 111
  from mitochondrial mutations, 679 (table), 680
  morphological versus molecular, 88–90
  mutant bacterial, 402–403
  Principle of Assortment and, 99–101
  Principle of Complementation and, 120
  Principle of Segregation and, 94–95
  in random amplified polymorphic DNA analysis, 70
  somatic mosaics and, 266
  in three-point testcrosses, 192–197
  transformation of bacterial, 403–405
  in transmission genetics, 88, 90–93
phenotypic ratios
  in epistasis, 115–117
  in Mendel's experiments, 164–167
  probability and, 158–161
  in transmission genetics, 93

phenotypic variance, 759
in genetic analysis, 760–761
phenylalanine (Phe, F)
alkaptonuria and, 13–15
DNA base sequence for, 20, 476
genetic code and, 475–476
molecular structure of, 447
in PKU, 14–15
phenylalanine hydroxylase (PAH), 18, 21, 64, 267–268
folding of, 23–25
nucleotide sequence for, 16
in PKU, 14–15, 21–25
three-dimensional structure of, 24–25
in tyrosinemia, 15
phenylalanine operon, 504–505
phenylketonuria (PKU), 64
dietary treatment of, 22, 25–26, 32–33
discovery of, 32–33
genetic basis of, 14–15
mutations with, 21–23
pleiotropic effects with, 26
ϕX174 bacteriophage, genome size of, 312
Phlox cuspidata
self-fertilization and inbreeding in, 724
testing mating randomness in, 714
phosphates, in DNA, 42, 44
phosphodiester bonds, in DNA, 42–43, 60, 235–238
phosphoglucomutase (*pgm*) genes, of *Phlox cuspidata*, 714
phosphoric acid, in DNA, 41–44
phosphorus, in DNA, 6–8
phosphorylation, RNA polymerase and, 514
photoreactivation, reversal of UV DNA damage via, 299
photosynthesis, evolution of, 694
phylogenetic trees, 697–698
physical distance, map distance and, 196–197
physical map, of complex genomes, 572–574
physical performance gene, 774–775
phytates, genetic engineered rice and, 565–566
pigments, in human vision, 365
pigs
DNA base composition of, 44
epistasis in, 117
pilus (pili), 395–396
pin flowers, of *Primula officalis*, 711
Pingelap disease, 744
pink flowers, of snapdragons, 111
pistillata (*pi*) gene, 624–626
Pisum sativum (pea plant)
diploid chromosome number of, 134
epistasis in, 115–116
fertilization in, 96–97
flowers of , 89, 91, 93, 115–123, 161–162
hybrid crosses of, 89–98
Mendel's experiments on, 88, 96–97
monohybrid crosses, 89, 92–93
mutants in, 88–90
outcrossing of, 91
ovule of, 91
P_1 generation of, 92–93, 96, 119
pod types in, 89, 93
recessive alleles in, 89, 93
reciprocal crosses of, 92
RFLPs in 66–68
seed colors in, 89,93, 98–103, 160–161
seed types in, 88–103, 164
SBEI genes in , 88–90, 278
stem length in, 89, 93
terminal flower position of, 89, 93
trihybrid crosses of, 92, 102–103
traits of, 88–89
transmission genetics of, 88–103, 107, 160–161, 164–167
transposable elements in genome of, 278
true-breeding traits of, 91–92
wildtype alleles in, 88–90
placenta, 360
plants
abnormal chromosomes in, 346
chloroplast genomes of, 682
complex traits in, 750
cytoplasmic male sterility in, 688–691
diploid chromosome numbers of, 134
as eukaryotes, 28
genetic developmental control in, 622–626
genetic engineering in, 560–561
genome sizes of, 313
life cycles of, 141
monoploid, 381–382
paternal inheritance in, 682
plastic growth patterns of, 622–623
polyploidy in, 377–382
posttranscriptional cosuppression in, 523–524
self-fertilization and inbreeding of, 722–724
sex chromosomes in, 149–150
somatic mutation propagation in, 266
telomeric DNA sequences of, 336
translocation in, 372–373
plaque formation, 418
plasmids, 224, 312, 394–395
as cloning vectors, 543–546, 549–553
conjugative, 395–396
constructing recombinant, 547
for *E. coli* cloning, 543
in lactose metabolism, 492–493
nonconjugative, 397–398
pBluescript II, 552–553
in plant genetic engineering, 560–561
Plasmodium berghei, synthetic vaccines and, 567
Plasmodium falciparum, 737
geographical distribution of, 735
sickle-cell anemia and, 270
synthetic vaccines and, 567
pleiotropic effects, 26
pleiotropy, 26
PM2 bacteriophage, supercoiled DNA of, 314
pms1 mutant, mismatch repair and, 297
pneumonia, bacterial, 3–5
POB' site, in specialized transduction, 435
Podisma pedestris, genome size of, 72
point centromeres, 331
polar granules, in embryonic development, 596
polarity
of DNA strands, 10, 18
in *D. melanogaster* development, 616
of polynucleotide chains, 43
of water molecule, 46–47
pole cells, in *D. melanogaster* development, 608, 613–614
pole hole mutation, in *D. melanogaster*, 555
poles
during meiosis, 147
during mitosis, 137
Pol II holoenzyme, 512–514
Pol III holoenzyme, 237–238
pollen
of pea plants, 91
of primroses, 711
self-sterility genes and, 112
poly(GT) tracts, stability of, 297
poly-A tail, of mRNA, 458, 461
polycistronic mRNA
in complex translation, 470–471
in gene regulation, 490
polylinker, in pBluescript II plasmid, 552
polylysine, genetic code and, 476
polymerase α (DNA POLA), 238, 262
polymerase chain reaction, selective DNA replication via, 49
polymerase chain reaction (PCR), 78–79, 88, 198–199
in DNA amplification, 60–64, 69, 549
random amplified polymorphic DNA versus, 69
with reverse transcriptase, 550, 551
sequence-tagged sites and, 574
in site-directed mutagenesis, 563
polymerase γ, 238
polymerase δ, 238
polymers
DNA as, 41–44
dominant-negative mutations in, 267–268
polynucleotide chains, in DNA, 41–44
polypeptide backbones, in proteins, 448
polypeptide chains
dominant-negative mutations and, 267–268
folding of, 23–25
genes and, 450–451
mutation of, 21–23
in proteins, 16, 446–449
synthesis of, 17–21
polypeptide synthesis, translation in, 463–469
polyphenylalanine, genetic code and, 475–476
polyploidy, 377–382
spindle checkpoint and, 654–655
polyproline, genetic code and, 476
polysaccharide coat, of *Streptococcus pneumoniae*, 3–5
polysaccharides, in human red blood cells, 112
polysomes, in complex translation, 468–470
polytene chromosomes, 323–324
polyuridylic acid (poly-U), genetic code and, 475
POP' site, in lysogeny, 432–433
population genetics, 29, 708
migration and, 728
random genetic drift in, 736–739
populations, 708
genetic variations among, 720–721
trait distribution throughout, 751–753
population studies, DNA typing in, 76
positional cloning, 554–555
positional information, in embryonic development, 595–596
position effect, 375
position-effect variegation (PEV), 375–377
positive assortative mating, 711
positive autoregulation, in organism development, 593–594
positive chromatid interference, in multiple crossing-over, 192
positive regulation
of genes, 490–492
of *lac* operon, 497–499
of yeast mating type, 531

posterior genes, of *D. melanogaster*, 612–614
postreplication repair, 295, 299–300
posttranscriptional cosuppression, 523–524
posttranslational control, gene regulation via, 490
posttranslocation states, of ribosomes, 466–467
Prader–Willi syndrome, 521–522
precursor (Okazaki) fragments
  in DNA replication, 234, 239
  joining of, 240–241
precursor RNA, in RNA splicing, 461
premutation, 272–273
pre-replication complexes, 647
pretranslocation states, of ribosomes, 466–467
primary transcript, mRNA as, 456–458
primases, in DNA replication, 234–236
primates, characteristics of, 2
primer adapters, amplified fragment length polymorphisms and, 71–72
primers
  in DNA amplification, 60–63
  in DNA replication, 235–236
  in DNA synthesis, 59–60, 224
  mutagenic, 563
  removal of, 240–241
  for reverse transcriptase, 550–551
  in RNA synthesis, 452
primosomes, in DNA replication, 234–236
primroses, assortative mating in, 711
Primula officalis, assortative mating in, 711
Principle of Complementation, 120
Principle of Independent Assortment, 99–101
  in trihybrid crosses, 102–103
Principle of Segregation, 94–95
Pringle, John R., 644C
probability
  inbreeding coefficient and, 725
  lod scores and, 198–201
  pedigrees and, 107–111
  in predicting progeny distributions, 158–161
  in random genetic drift, 737–738
probe DNA
  in DNA fragment identification, 56–58
  in recombinant screening, 554
probe RNA, in DNA fragment identification, 58
product molecules, in metabolic pathways, 13
proflavin, mutagenesis via, 289, 472–473
programmed cell death. *See* apoptosis
prokaryotes
  chemical mutagens of, 288
  colinearity in, 450
  DNA replication rate in, 232
  in eukaryote origins, 691, 694
  in evolution, 28
  excision repair in, 299
  genome sizes of, 312–313
  intergenic suppression in, 302
  introns in, 463
  mismatch repair in, 297
  mRNA caps in, 466
  polycistronic mRNA of, 470–471
  RNA polymerase processivity in, 452–454
  transcriptional regulation in, 490–492
  translation in, 464
  translation initiation in, 466
proliferating cell nuclear antigen (PCNA), 650–651
proline (Pro, P)
  DNA base sequence for, 20, 476
  molecular structure of, 447
Prolla, Thomas A., 297C
promoter fusion, acute leukemias from, 666–667
promoter mutations, 456
promoter recognition, 454–455
promoter region, for lactose metabolism, 495
promoters, 452, 516–518
  genetic evidence for, 456
  in integrons, 400–401
  in lactose metabolism, 495–496
  in transcription, 454–455, 458
proofreading
  in DNA replication, 225, 236–238, 297
  in mismatch repair, 295
  SOS repair and, 300
prophage, 431–433
  in specialized transduction, 433–434
prophage induction, 433
prophase, 136–138
prophase I, 141–146
  in *Lilium longiflorum*, 144
prophase II, 143, 148
protanomaly, 365–367
protanopia, 365–367, 391
proteases
  in apoptosis, 652
  in DNA experiments, 5
  in enzymatic protein breakdown, 25
proteasomes, 648, 650
  in enzymatic protein breakdown, 24–25
protein kinases, during mitosis, 136
protein oncogenes, table of, 657
proteins
  in cells, 11–12
  chemical structure of, 446–449
  in chromosomes, 312
  complementation tests and, 122–123
  degradation of, in cell cycle, 647–648
  degradation of misfolded, 24–25
  destabilizing amino acid replacements in, 25
  DNA and, 6, 48–49
  in DNA replication, 234
  domains, 462
  in evolution, 27–28
  families of, 27, 576
  folding in, 23–25, 446–449, 462
  as gene products, 15–16
  genes and, 11–21
  as genetic material, 3
  histones as, 318
  insecticidal, 556
  in λ bacteriophage lysogeny, 505–507
  orthologs of, 576–578
  paralogs of, 576
  production of useful, 566
  in proteomes, 27
  subunits of, 449
  synthesis of, 16–21
  three-dimensional structure of, 23–25
proteolysis, in cell cycle, 647–648
proteomes, 27, 576
  of *Caenorhabditis elegans*, 630
proteomics, 27
Proteus vulgaris, restriction enzymes from, 51
protists, genome sizes of, 313
protomitochondria, 694
proto-oncogenes, 656
  acute leukemias from, 666
protoperithecium, of *Neurospora crassa*, 206
protoplastids, 694
prototrophs, 402
protozoa
  as eukaryotes, 28
  genome sizes of, 313
  killer symbionts in, 691–694
  telomeres in, 332–337
  telomeric DNA sequences of, 336
Providencia stuartii, restriction enzymes from, 51
pseudoautosomal inheritance, 355
pseudoautosomal regions, 355
pseudogenes, 39
*Pst*I restriction enzyme, 51
pulsed-field gel electrophoresis (PFGE), 568–569
Punnett, Reginald C., 95
Punnett square, 94–95, 99–100, 106–107, 115, 152–153
  Hardy–Weinberg principle and, 713
  multiple alleles in, 715–716
  with X-linked genes, 721–722
pupa, in *D. melanogaster* development, 607–608, 617
purification, of DNA, 549
purines
  in DNA, 41–44, 48
  nucleotide substitution and, 268
  in RNA splicing, 460
purple eye color (*pr*) gene, in *D. melanogaster*, 197
*Pvu*II restriction enzyme, 51
pyloric stenosis, heritability of, 771
pyrimidine dimers, from UV irradiation, 288
pyrimidines
  in DNA, 41–44, 48
  nucleotide substitution and, 268
  in RNA splicing, 460
pyrophosphate groups, from DNA synthesis, 60

Q

q ("not-p") arm of chromosomes, 348–349
Q protein, in λ bacteriophage, 507
quadrant enhancer, in *D. melanogaster*, 618–619
quantitative-trait loci (QTLs), 772, 774
  genetic mapping of, 772–773
quartets, of cells, 640

R

R1-mb allele, in maize, 185–186
R408W mutant, of *PAH* gene, 22–23
rad, 290
radiation units, 290–291
  table of, 290
radioactivity, as mutagen, 290, 293–295
radium, as mutagen, 290
radium dances, 290
radium roulette, 290
Ramanis, Zenta, 686C
random amplified polymorphic DNA (RAPD), 68–70, 88
random genetic drift, evolution and, 728, 736–739
random mating, 711–722
  DNA typing and, 716–720
  Hardy–Weinberg principle and, 712–713
  heterozygous genotype frequency in, 714–715
  multiple alleles and, 715–716
  population differences and, 720–721

test for, 713–714
X-linked genes and, 721–722
randomness, of mutations, 280–281
random sampling, in population genetics, 737
random-spore analysis, 205
RAP1 gene, 335
Rap 1 protein, telomeres and, 335
rapid lysis (r^-) allele, 418–420
rare alleles
allele frequency of, 714–715
in natural selection, 732
rare mRNA, 551
Ras gene, 652, 660–661
Ras–GTP complex, 658
Ras protein, 656–658
ratios. *See* likelihood ratios; maternal sex ratio; phenotypic ratios; recombination ratios; sex ratios
RB gene, 662
RB1 gene, 661, 664–665
reading frame, of mRNA, 473
Rec8 protein, in centromeres, 148
RecA protein, in Holliday recombination, 252
RecBCD nuclease, in Holliday recombination, 252
recessive alleles
complementation tests and, 118–123
in human pedigrees, 104–106
in natural selection, 732
in pea plants, 89, 93
probability and, 161
in sickle-cell anemia, 269–270
uncovering of, 363, 664
for X-linked genes, 722
reciprocal crosses, of pea plants, 92
reciprocal recombination, of F plasmid, 405–407
reciprocal translocations, 349, 370–372
in DNA molecule, 280
recombinant DNA, 542
complex genome analysis via, 567–578
in genetic engineering applications, 564–567
host organisms for, 542
molecules of, 544–545, 551–555
restriction enzymes for, 542–544
vectors for, 542, 545–546
recombinants
alleles of, 177
in *D. melanogaster*, 190
screening for, 554
recombination
between bacteriophage, 417–422
in *Chlamydomonas*, 687–688
cotransformation and, 404
of DNA, 248–255
ectopic, 279–280
in *E. coli*, 413
frequency of, 177–180, 182–186, 190–192, 195–196
of genes, 176–192, 209, 250–252, 254, 413
Holliday model of, 250–252, 254
in human pedigrees, 198–201
mitotic, 209–211
ratios, in chi-square testing, 179–180
in single-strand break and repair model, 253–254
in three-point testcrosses, 192–197
recruitment
of CRCs, 514–515
transcriptional activation by, 511–513
"red-blindness" (protanopia), 365–367, 391
red blood cells, in sickle-cell anemia, 269–270
red clover, self-sterility genes in, 112
red eyes, in *D. melanogaster*, 150–153, 177–181, 562
red–green color blindness, 391
genetic basis of, 365–367
red-pigment genes, 391
in color blindness, 365–366
reductional division, 141
redundancy, of genetic code, 478–479
regional centromeres, 331
regionalization, in *D. melanogaster* development, 616
Reid, Brian J., 644C
relationships, DNA typing to trace, 76
relative fitness, of genotypes, 731
relative risk, of hereditary diseases, 74
relatives, correlation between, 767–770
relaxed DNA, 314
release factors (RFs), in translation, 464, 468–469
rem, 290
renaturation, of DNA, 55–58, 61, 324–329
renaturation rate, of DNA, 324–329
repeat-induced point (RIP) mutations, 280, 519
repE gene, 570
repetitive nucleotide sequences, 324–330
replica plating, 281–282
replication
of DNA, 10–11, 48, 140, 224–247, 332–337
by plasmids, 395
selective, 49, 58–64
replication bubbles, 232–233
replication forks
lagging and leading DNA strands and, 238–241
in θ replication, 230–231
replication rate, of DNA, 232
replication slippage, 273, 297
replicons, 224
in artificial chromosomes, 569–570
repressible systems
in lactose metabolism, 493–494
of negative regulation, 491
repressible transcription, in negative regulation, 491
repressor protein
in lactose metabolism, 494–496, 498
in negative regulation, 491
in tryptophan synthesis, 500–502
reproduction, evolution and, 730
reptiles, sex chromosomes in, 155
repulsion configuration, of alleles, 178–179
respiration-defective mutants, of *Saccharomyces cerevisiae*, 688–689
restriction endonucleases, DNA cleavage with, 50–52
restriction enzymes, 54–55, 59, 90
DNA cleavage with, 50–52, 54, 542, 568–569
production of defined DNA fragments with, 542–544, 547–549
table of, 51
restriction fragment length polymorphisms (RFLPs), 66–68
gene segregation and, 95
of pea plants, 92–93
restriction fragments, of DNA, 52
restriction maps, 54, 572
restriction sites
in enzymes, 50–52
on restriction enzymes, 542
restrictive conditions, in conditional mutations, 266
restrictive temperature, in cell division cycle mutants, 640–641
retina, 365
retinoblastoma, 661–665, 674
cancer and, 657, 659–660
and cell cycle, 646
retrotransposons, 277–278
retroviruses, genetic engineering with, 566–567
reverse genetics, 563
reverse mutations, 300–301, 729
reverse transcriptase, 278
cloning from mRNA with, 550–551
reverse transcriptase PCR (RT-PCR), 550–551
reversible mutation, 729–730
reversions, 300–301
R groups, in amino acids, 446
Rhesus (Rh) blood group locus, 349
Rhodopseudomonas sphaeroides, restriction enzymes from, 51
rhodopsin, 365
"ribbon" diagrams, of proteins, 448
ribonucleoproteins (RNPs), telomeres and, 334
ribose, in RNA, 17, 235–236, 452
ribosomal RNA (rRNA), 17
ribosomal subunits
in elongation, 466–467
in protein release, 468–469
in translation, 464–466
ribosome-binding site, in mRNA, 466
ribosomes, 17, 19, 457, 464
in complex translation, 468–471
in elongation, 466–468
in lactose metabolism, 497–499
protein release from, 468–469
structure of, 468
in tryptophan synthesis, 502–503
ribozymes, 466–467
in RNA splicing, 461
rice
genetic engineering of, 565–566
genome of, 382–383
synteny groups in, 383–384
rII cistrons, of T4 bacteriophage, 424–425, 427, 472–473
rII gene, of T4 bacteriophage, 422–429, 472–475
ring chromosomes, 349
in *Oenothera*, 373
risk factors
alleles as, 74, 75
in exposure to ionizing radiation, 293
in human genome, 197–201
RME1 gene, 529–530, 532
RNA (ribonucleic acid)
Barr bodies and, 353
in DNA initiation, 235–236
DNA versus, 42, 224
in evolution, 27–28
evolution of, 17
as genetic material, 4, 9, 452–456
in joining DNA precursor fragments, 240–241
messenger, 16–17, 456–458
in organelles, 678
in protein synthesis, 16–21
reverse transcriptase and, 278
ribosomal, 17

RNA (cont'd)
splicing of, 458–462
in telomeres, 334–335
transfer, 17
translation of, 16–19, 463–471
turnover of, 523–524
types of, 17
RNA decay, gene regulation via, 522–524
RNA editing, 683
RNA interference (RNAi), 524
RNA polymerase holoenzyme, 452, 515
RNA polymerases, 235, 446, 470, 511–514
processivity of, 452–454
in transcription, 454–456
RNA primase, 234, 332
RNA primase complex, in DNA replication, 234
RNA processing, 446, 458
in eukaryotes, 458–463
gene regulation via, 490, 522–524
RNAse, in DNA experiments, 5
RNA splicing, 458–462, 490, 516
RNA synthesis, 235, 452–453
RNA tumor viruses, reverse transcriptase in, 550
RNA viruses, germ-line transformation via, 557–559
Robertsonian translocations, 349, 373–375
rod cells, 365
Roe, Bruce A., 572C
roentgen (R), 290
rolling-circle replication
in conjugative plasmids, 396, 406–407
of DNA, 231–232
Romanov family, identification of remains of, 681, 702–703
Roman, Raquel, 417C
R plasmids, 402
*Rsa*I restriction enzyme, 51
RSC (*remodels the structure of chromatin*) complex, 514
R strain, of *Streptococcus pneumoniae*, 3–5
ruby eye color (*rb*) gene, in *D. melanogaster*, 184
rudimentary wing (*r*) gene, in *D. melanogaster*, 190

S

S (DNA synthesis) period, of interphase, 136, 138, 636–639
sacB gene, 569–570
Saccharomyces cerevisiae, 220
cell cycle of, 637, 644
centromere structure of, 331–332, 342
contig clones of, 555
diploid chromosome number of, 134
DNA base composition of, 44
electrophoretic DNA fragment separation in, 567–568
galactose metabolism in, 508–509
genome size of, 312–313
genome studies of, 38
genomic sequencing of, 576
germ-line extinction gene in, 328
knockout mutations in, 564
life cycle of, 202, 528
mating-type interconversion in, 528–530, 538
proteome of, 27, 578
as recombinant DNA host, 542
respiration-defective mutants of, 688–689
topoisomerase II in, 315–316
wobble in, 479
Saccharum officinarum, synteny groups in, 384
S-adenosylmethionine, in cytosine methylation, 284
Sager, Ruth, 686C
salmon
DNA base composition of, 44
genetically engineered, 564–565
Salmonella typhi, 744
Salmonella typhimurium, in Ames test, 303
Sanchez, Carmen, 500C
Sanger, Fred, 262–263
Sanger Centre, 262–263
satellite DNA, 324
SBEI genes, in pea plants, 88–90
scaffold, in chromosomes, 321–323
scatter diagrams, 769
Scc1p protein, 648
Schierenberg, E., 602–603C
schizophrenia, 784
Schizosaccharomyces pombe
cell cycle of, 637
germ-line extinction gene in, 328
Schnicke, Angelika E., 558C
scorpion, diploid chromosome number of, 134
sea urchins, embryonic induction in, 596–597
second-division segregation, in *Neurospora crassa*, 207–209
second meiotic division, 141, 148
second nuclear division, in meiosis, 140
segmentation genes, of *D. melanogaster*, 611–616
segment-polarity genes, of *D. melanogaster*, 611–613, 616
segments, of *D. melanogaster*, 610–612
segregation
in bacteriophage, 417
in human pedigrees, 106–107
independent, 99–103
of multiple genes, 98–103
in *Neurospora crassa*, 207–209
of nonchromosomal genes, 686
principle of, 94–95
and Principle of Independent Assortment, 99–101
and Principle of Segregation, 94–95
in reciprocal translocation, 370–372
in Robertsonian translocations, 373–374
of single genes, 90–98
in snapdragons, 111
of three independent genes, 755–757
trisomic, 361–362
segregational petites, 688–689
segregation mutations, 601
selected markers, 407–408
selection coefficient, 731
selection limit, 766
selection–mutation balance, 733–735
selective breeding, effects of, 762
selectively neutral mutations, 728
selective medium, for bacteria, 403
selective replication, 49
of genomic DNA fragments, 58–64
self-fertilization, inbreeding via, 722–724
self-splicing RNA precursor, 461
self-sterility genes, 112
semiconservative replication
bidirectional replication as, 232–233
of DNA, 225–230
rolling-circle replication as, 231–232
θ replication as, 230–231
semisterility
in *Oenothera*, 372–373
reciprocal translocation and, 370, 372
septum, in yeast cell cycle, 638
sequence logos, 486
sequences. *See also* DNA sequencing
of DNA bases, 10–11, 88–90
programmed rearrangements in DNA, 524–532
repetitive nucleotide, 324–330
sequence-tagged sites (STSs), 573–574
sequoia, diploid chromosome number of, 134
Sequoia sempervirens, paternal inheritance in, 682
serine (Ser, S)
DNA base sequence for, 16, 19–20, 476
molecular structure of, 447
in PAH, 21
serine proteases, in *D. melanogaster*, 614
serotonin, 775, 784
serotonin transporter, 775–776
Serrate gene, in *D. melanogaster*, 618
seryl-tRNA-synthetase, 477–478
structure of, 478
Setaria itallica, synteny groups in, 384
sex-chromatin body, in mice, 354
sex chromosomes, 148. *See also* X chromosomes; Y chromosomes
abnormalities of, 362
of *Caenorhabditis elegans*, 598–599
in determining sex, 148–158
genetic crosses and, 92
in human karyotype, 347–348
in polyploid plants, 378
pseudoautosomal regions in, 355
Sex-lethal (Sxl) protein, 157–158
sex-linked inheritance
in *D. melanogaster*, 150–153
of hemophilia mutation, 154–155
of mutations in chickens, 155
sex ratios
in *D. melanogaster*, 157–158, 694–695
probability and, 158–159
in sibships, 110
sex-specific imprinting, 520–522
sexual reproduction, in *E. coli*, 413
*Sfi*I fragment, 569
sheep, cloning of, 558
shell coiling, maternal effect in, 695–696
Shimizu, Nobuyoshi, 572C
Shine-Dalgarno sequence, 466
shortened wing crossvein (*cv*) gene, in *D. melanogaster*, 184
short interspersed elements (SINEs), 278
shotgun dideoxy sequencing, of DNA, 244–247
shrunken (*sh*) gene, in maize, 275–276, 278
shuttle vectors, 570
Siamese cats, temperature sensitivity in, 266–267
siblings (sibs)
correlation between, 770
in human pedigrees, 104
sibships
in human pedigrees, 104
probabilities of traits in, 161
sex ratios in, 110, 158–159
sickle-cell anemia
geographical distribution of, 735
malaria and, 270, 735–737
as result of missense mutation(s), 269–271, 476
sievert (Sv), 290

signaling, in *D. melanogaster* development, 619
significance
high, 163
statistical, 163–164
silencers, in transcription, 511
"silencing"
of sex-chromosome abnormalities, 362
of telomeres, 335
of transposable elements, 280
silent mutations, 268
silk fibroin mRNA, translational control of, 524
silkworms, translational control in, 524
simple sequence length polymorphisms (SSLPs), 72
simple tandem repeat polymorphisms (STRPs), 72–73, 198–199
from Chernobyl nuclear accident, 294
in human pedigrees, 106–107, 111
single-active-X principle, 353, 362
single-nucleotide polymorphisms (SNPs), 66
in human pedigrees, 106
restriction fragment length polymorphisms and, 66–67
single-strand break and repair model, of DNA repair, 253–254
single-stranded DNA, 10, 239
in prokaryote genomes, 312–313
from reverse transcriptase, 550–551
in rolling-circle replication, 231–232
single-stranded DNA-binding protein (SSB), 234
sinistrally coiled shells, 695–696
sister-cell segregation, in *Caenorhabditis elegans*, 601
sister-chromatid exchange, in DNA replication, 229–230
sister chromatids
during meiosis, 139, 142, 148
during mitosis, 138
site-directed mutagenesis, 562–564
site-specific DNA cleavage, 50–52
site-specific recombinases, in bacterial antibiotic resistance, 398–400
site-specific recombination, lysogeny and, 429–433
Skolnick, Mark, 68C
SLC6A4 protein, 775
SLC6A4 serotonin transporter gene, 775–776
slime molds, telomeric DNA sequences of, 336
smallpox, synthetic vaccines and, 567
smallpox virus, genome size of, 312
snake gene, 614
snapdragons
diploid chromosome number of, 134
incomplete dominance in, 111
red flowers in, 111, 127
snf mutant, 514
snRNPs (small nuclear ribonucleoprotein particles), 460–461
SNRPN gene, in sex-specific imprinting, 521–522
social implications
of cloning, 558
of genetic engineering, 584–585
of genetic testing, 584–585
sodium hydroxide (NaOH), as DNA denaturant, 554
solute carrier (SLC), 775
somatic cell hybridization map, 573–574
somatic cells
cancer cells from, 656
cancer in, 636
chromosomes in, 135
development of nematode, 599–600
in familial retinoblastoma, 662, 665
programmed DNA sequence rearrangements in, 524–532
somatic mosaics, 211, 266
somatic mutations, 266
Sorghum vulgare, synteny groups in, 383–384
SOS repair, 300
Southern, Edward, 57
Southern blot, 57–58
random amplified polymorphic DNA and, 68–69
restriction fragment length polymorphisms and, 66–67
simple tandem repeat polymorphisms and, 72
species, 2
specification, in *D. melanogaster* development, 616
speck of wing pigment (*sp*) gene, in *D. melanogaster*, 197
spermatocytes
in meiosis, 140–141, 146
sex chromosomes in, 149
spina bifida, heritability of, 771
spindle checkpoint, 649, 654–655
spindle pole bodies, in cell cycle, 637–638
spindles. *See* mitotic spindles
spinocerebellar ataxia type I, dynamic mutation and, 273
spliceosomes, 460–461
splicing
alternative, 522–523
in antibodies, 528
intron mutations and, 462–463
of mRNA, 458–462, 490, 516
spontaneous abortion
chromosome abnormalities in, 357–358
Down syndrome and, 360
spontaneous mutation, 280–285
dynamic mutations, 272–273
sporadic cancer, 655, 660
spores, in meiosis, 141
sporophytes, in meiosis, 141
SRY gene, 355
S strain, of *Streptococcus pneumoniae*, 3–5
stabilization, in DNA replication, 234–235
Stahl, Franklin W., 225, 228C
stamen
of pea plants, 93
of primrose, 711
standard deviation, in normal distribution, 752–753
Staphylococcus aureus, nucleosomes of, 318–319
starch-branching enzyme I (SBEI), in pea plants, 88, 278
start codon, 20, 476. *See also* AUG codon
statistical significance, 163–164
sterility, of monoploid organisms, 381
Stern, Curt, 186, 209–210
sticky ends, of cleaved DNA, 52, 543–544, 547–548
stigma
of pea plants, 91
of primrose, 711
stop codons, 20, 476
for mitochondrial DNA, 683
nonsense mutations and, 270–271, 479–481
in translation, 468–469
tRNA and, 477
storage, of cloned DNA, 54
stramenopiles, as eukaryotes, 28
Strand, Micheline, 297
strand exchange, in postreplication repair, 299–300
Strathern, Jeffrey N., 530C
Streisinger, George, 420
Streptococcus pneumoniae
DNA base composition of, 44
DNA function in, 3–5
genomic sequencing of, 576
transformation in, 403
streptomycin
algal resistance to, 687
genes for resistance to, 397, 407
mutant bacterial resistance to, 402
stress relief, in DNA replication, 234–235
Strongylocentrotus purpuratus, embryonic induction in, 596–597
str-r gene, of *Chlamydomonas*, 687
structural genes, in lactose metabolism, 495–496, 500
STS markers, 574
Sturtevant, Alfred H., 173, 181, 190C
submetacentric chromosomes, 350
subpopulations, 708
migration and, 728
random genetic drift in, 738–739
substrate molecules, in metabolic pathways, 13
subunits, of proteins, 136, 449
sugar cane, synteny groups in, 383–384
sugary endosperm (*su*) allele, in maize, 192–194
sul1 coding region, 399–401
sulfur, DNA and, 6–8
Sulston, John E., 602–603C
supercoiling, of DNA, 314–316
super-repressor, in lactose metabolism, 494
suppressor mutations, 300–303
surface exclusion, with F plasmid, 396
survival, evolution and, 730
SV40 virus, genome size of, 312–313
swi mutant, 514
SWI–SNF complex, 514
Sxl gene, 158
symbionts, cytoplasmic transmission of, 691–695
symbiosis, in organelle evolution, 691
synapsis, 144, 181
inversion and, 367, 369
in *Lilium longiflorum*, 144
during meiosis, 142
synchronized cell cultures, 638–639
syncytial blastoderm, in *D. melanogaster* development, 608
synonymous codons, 475–476
synonymous mutations, 268
syntenic genes, 178
coupling versus repulsion of, 178–179
in grass genomes, 383–384
synteny groups, in grass genomes, 383–384
synthetic vaccines, 567

T

T1 bacteriophage, in replica plating, 281–282
T2 bacteriophage, 6–9
genome size of, 312
history of, 33
in identification of mRNA, 457
lytic cycle of, 6

T2 bacteriophage (cont'd)
recombination between, 417
structure of, 6
T3 bacteriophage, 8
T4 bacteriophage, 6, 8
DNA renaturation time in, 325
genetic code and, 472–473
genetic maps of, 420–421, 424–426
genome size of, 312
history of, 33
ligase from, 544
linear DNA of, 422
nonsense mutations in, 479
recombination in, 418–422
rII gene of, 422–427
T6 bacteriophage
genome size of, 312
history of, 33
T7 bacteriophage, DNA renaturation time in, 325–326
tailless gene, in *D. melanogaster*, 615
tandem duplication, 364–365
Taq polymerase, 62
*Taq*I restriction enzyme, 51
target genes, 560
TATA box, 486, 511–512
in transcription, 454
TATA-box-binding protein (TBP), 511–512, 514–515
TAT gene, 15
Tatum, Edward L., 14, 413C, 449–451
Tay-Sachs disease, 391
TBP-associated factors (TAFs), 511–512
T cells, 525–528
T DNA, in plant genetic engineering, 560–561
technical knockout mutation, in *D. melanogaster*, 555
telocentric chromosomes, Down syndrome and, 358–359
telomerase, 332–335
telomerase activity, cell senescence and, 656
telomeres, 332, 342
of *Caenorhabditis elegans*, 328
cancer and, 652, 656
molecular structure of, 330, 332–337
in telocentric chromosomes, 358
telomere terminal transferase, structure of, 334
telomeric DNA, 333–337
sequences of (table), 336
telomeric silencing, 335
telophase, 136–137, 637
telophase I, 141, 143, 148
telophase II, 142, 148
in *Lilium longiflorum*, 146
telson, of *D. melanogaster*, 612–614
temperate bacteriophage, 429
temperature-sensitive mutations, 266–267
cell cycles and, 640–642, 644
templates, DNA strands as, 10–11, 18–19, 41, 455–457
template strand, 455
teocentli, 755
teopod-2 (*tp-2*) gene, in maize, 185
teosinte, maize and, 754–755
teosinte branched gene, 754
terminal (ter) parts of chromosome arms, 349
terminal flower position, in pea plants, 89, 93
terminal genes, of *D. melanogaster*, 612, 614
terminal inverted repeats, 277
terminal redundancy, of T4 bacteriophage DNA, 422
termination
of protein synthesis, 468–469
of RNA strands, 455–456
termination protein, in transcription, 455
terminators
genetic evidence for, 456
in transcription, 455–456, 458
terminus of replication, in θ replication, 230–231
testcrosses
in complementation tests, 119
for gene recombination, 176–178
genetic mapping in three-point, 192–197
mapping deletions with, 363
in transmission genetics, 98, 101
with unlinked genes, 101
testicular embryonal carcinoma, 659
testis-determining factor (TDF), 355
tetracycline, mutant bacterial resistance to, 402
tetracycline resistance, in bacteria, 395
tetrad analysis, genetic mapping via, 201–209, 254
tetrads, 220
during meiosis, 144
types of, 202–204
Tetrahymena
artificial chromosomes in, 569–571
G-quartet base pairing in, 337
RNA splicing in, 461
telomeres in, 332–337
tetramers
dominant-negative mutations in, 268
PAH molecules as, 24–25
tetrapeptides, 447
tetraploid chromosomes, 357
tetraploid chromosome set, 377–379
tetratype (TT) tetrads, 202–205
ordered, 206
"T-even phages," history of, 33
TFII factors, 511–513, 515
thermophiles, heat-stable DNA polymerases from, 62
Thermus aquaticus
heat-stable polymerase from, 62
restriction enzymes from, 51
RNA polymerase of, 453
θ replication, of circular DNA, 230–231
lagging and leading strands in, 239
30-nm fiber, in chromosomes, 321–322
Thomson, J. N., 602–603C
thorax, of *D. melanogaster*, 612, 620–621
3'-hydroxyl (3'-OH) group, in polynucleotide chains, 43
three-point crosses, 192
genetic mapping in, 192–197
translocation complexes as, 373
3' trimming, of mRNA, 458
three-strand double crossing-over, 191–192
in tetrad analysis, 204–205
threonine (Thr, T)
DNA base sequence for, 16, 19–20, 476
mitochondrial codons for, 683
molecular structure of, 447
in PAH, 21
threshold traits, 751
heritabilities of, 770–772
thrum flowers, of *Primula officalis*, 711
thymine (T)
5-bromouracil and, 286
cytosine methylation and, 285
in dideoxy DNA sequencing, 243–245
as DNA base, 9–11, 17–19, 41–49, 224–225, 235–236, 452
nucleotide substitution and, 268
as radioactive tracer, 230
in sickle-cell anemia, 269–270
thymine dimer, from UV irradiation, 288
thymus-derived (T) cells, programmed DNA sequence rearrangement in, 525–528
TIIF factor, 511, 513
time of entry, in time-of-entry mapping, 408–409
time-of-entry mapping, 408–411
Tn5 transposon, 397
toad, diploid chromosome number of, 134
tolerance, antibodies and, 114
tomato, quantitative trait loci in, 773
topoisomerases
in DNA replication, 234–235, 315–316
three-dimensional structure of, 315
torso gene, 614
TψC loop, in tRNA, 477–478
traits
complex, 750
continuous, 751
genes and, 25–26
in human pedigrees, 104–106
meristic, 751
morphological versus molecular, 88–90
normal distribution of, 751–753
probabilities of, 161
quantitative, 751
of species, 2
threshold, 751
in transmission genetics, 88
trans-activation, in transposition, 277
trans configuration, of alleles, 178–179
transcript, RNA as, 18, 454–458
transcription, 16–18, 21, 446, 452–456
polycistronic mRNA in, 470–471
polysomes in, 468–470
transcriptional activation, in eukaryotes, 510–514
transcriptional activator proteins (TAPs), 510, 514–515
in positive regulation, 492
transcriptional cosuppression, 520
through RNA turnover, 523–524
transcriptional enhancers, 455, 510–511
transcriptional program, of cell cycle, 638–639
transcriptional regulation, 492
of mating type, 529–532
in prokaryotes, 492–494
in yeast, 508–509
transcriptional silencers, 510–511
transcription factors (TFs), 538, 646–647, 650
in eukaryotes, 511–513
transcription start sites, 452
transducing phages, 412
transduction, between bacteria, 403, 412–417
transfer RNA (tRNA), 17, 486
in elongation, 466–468
from mitochondrial DNA, 679–681
reverse transcriptase and, 278
in translation, 18–21, 464–469, 477–478
transformation
among bacteria, 403
DNA-mediated, 403–405
placing recombinant DNA into cells via, 544–545
in *Streptococcus pneumoniae*, 3–5
transformation mutations, 600–601

transformation rescue, 561–562
transgenes, 520
transgenic organisms, 542, 556–562
  genetically engineered, 564–567
transition mutations, 268
translation, 16–21, 446, 463–469
  complex units in, 468–471
  polycistronic mRNA in, 470–471
  polysomes in, 468–470
translational control
  in gene regulation, 524
  gene regulation via, 490
translocation complexes, 372–373
translocations, 349, 357–358, 362, 370–375
  acute leukemias from, 666–667
  balanced translocations, 357
  Down syndrome from, 360–361
  in human pregnancies, 357
  reciprocal translocations, 280, 349, 370–372
  unbalanced translocations, 357–358
transmembrane receptors, in organism development, 593–594
transmission (Mendelian) genetics, 88–123
  and congenital abnormalities, 771–772
  described, 88
  Down syndrome and, 358
  Hardy–Weinberg principle and, 716–717
  human pedigrees in, 104–107
  mating in, 711–722
  of pea plants, 88–103
  in population genetics, 708–710
  Principle of Assortment in, 99–101
  Principle of Complementation in, 120
  Principle of Segregation in, 94–95
  probabilities in, 107–111, 158–161
transposable elements, 396–397
  in chromosomes, 88, 274–280
transposase, 277
  in germ-line transformation, 556–557
transposition, 308
  of genes, 88, 274–280
transposons, 396–397
transversion mutations, 268
TRAP (*trp* RNA-binding attenuation protein)
  molecular structure of, 504
  in tryptophan synthesis, 503–504
tree diagrams
  of *Homo sapiens* phylogeny, 697–698
  in tetrad analysis, 205
Tribolium, 762
trihybrid crosses, of pea plants, 92, 102–103
trimming, of mRNA, 458, 522–523
trinucleotide expansion, 273
trinucleotide repeats
  dynamic mutation via, 272–273
  with Huntington disease, 108
triphosphates, of nucleosides, 42, 60, 242–243, 452
triploid chromosomes, 357
triploid chromosome set, 378
trisomic segregation, 361–362
trisomies, 357
  in Down syndrome, 358–361, 375
trisomy-X, 362
Triticum aestivum
  endoreduplication in, 380
  synteny groups in, 384
Triticum turgidum, endoreduplication in, 380
Triticum urartu, endoreduplication in, 380
Triturus, transcription in, 456
trivalent chromosomes, during meiosis, 361–362
tRNA suppressors, 480–481
trpa (attenuator) gene, in tryptophan synthesis, 500
trpA gene, 301
  of *E. coli*, 450–451
  in tryptophan synthesis, 499
TrpA polypeptide, 301
TrpA protein, 451
trp genes, in tryptophan synthesis, 499–500
trp operon, 499–505
true-breeding traits, of pea plants, 91–92
truncation point, in individual selection, 764–765
Trypanosoma brucei
  mitochondrial DNA in, 683
  RNA editing in, 683
tryptophan (Trp, W)
  DNA base sequence for, 20, 476
  E. coli genes for synthesizing, 455
  genetic code and, 476
  mitochondrial codons for, 683
  molecular structure of, 447
  in mutant PAH, 22–23, 64
  operon model in synthesis of, 499–505
tryptophan synthetase, 450–451
  in *E. coli* intragenic suppression, 301
tumorigenic cells, 655
tumor-suppressor genes, 656, 659–660
tumor therapy, with ionizing radiation, 292
Turner syndrome, 362
Turpin, Raymond, 358–359C
twin spots, 210
twin studies, 762–763
two-mutation hypothesis, for familial retinoblastoma, 662, 665
two-strand double crossing-over, 191–192
  in tetrad analysis, 204–205
two-strand stage of meiosis, crossovers during, 188–189
type II restriction endonucleases, DNA cleavage with, 50–52
type II tyrosinemia, 15
type III tyrosinemia, 15
tyrosine (Tyr, Y)
  alkaptonuria and, 13–15
  base changes in codon for, 270–271
  DNA base sequence for, 20, 476
  in *E. coli* intragenic suppression, 301–302
  molecular structure of, 447
tyrosine aminotransferase (TAT), in tyrosinemia, 15
tyrosine kinase, 606
tyrosinemia, 15

U

U2–U6 interaction, in RNA splicing, 460
UBE3A gene, in sex-specific imprinting, 521–522
ubiquitin, in enzymatic protein breakdown, 24–25
ubiquitin-conjugating enzyme, in enzymatic protein breakdown, 24–25
ubiquitin ligase, 648
Ultrabithorax (*Ubx*) gene, in *D. melanogaster*, 620–621
ultracentrifuges, 225–226
ultraviolet (UV) light
  in gel electrophoresis, 52–53
  in Meselson–Stahl experiment, 227
  as mutagen, 289–290
  photoreactivation reversal of damage from, 299
  in Southern blot, 58
umber stop codon, 479–480
unbalanced translocations, 357–358
unc-86 (*uncoordinated*) mutant, 600–601
uncovering of the recessive allele, 363
underwound DNA, 314–315
unequal crossing-over, 365
  in red–green color blindness, 365–367
unidirectional replication, of DNA, 231
uniform populations, in genetic analysis, 760–761
uniparental inheritance, 678
units of radiation, 290–291
  table of, 290
"unity of life," 27–29
univalent chromosomes, during meiosis, 361–362
universal donors, ABO blood groups and, 115
universal recipients, ABO blood groups and, 115
unlinked genes, 101
  chromosomes and, 134
  in tetrad analysis, 202
unordered tetrads, analysis of, 202–205
unstable genes, 274
unstable repeats, 297
unwinding, in DNA replication, 234–235
uracil (U)
  in cytosine methylation, 284–285
  as RNA base, 17–19, 224, 235–236, 452
uridine 5'-triphosphate (UTP), in RNA synthesis, 452
urine
  alkaptonuria and, 12
  PKU and, 32–33

V

V (variable) regions, in antibodies, 527–528
vaccines, synthetic, 567
vaccinia virus, synthetic vaccines and, 567
valine (Val, V)
  DNA base sequence for, 20, 476
  molecular structure of, 447
  in PAH, 21, 22
  sickle-cell anemia and, 269–270, 476
variable antibody regions, 526–528
variable number of tandem repeats (VNTR) polymorphism, 72
  in DNA typing, 718–720
  in human pedigrees, 106–107
  population differences and, 720–721
variance, 754
  in normal distribution, 751–753
variation, causes of, 754–760
variegation, 375
  of four-o'clock leaves, 684–685
  of gene loci, 274
vectors
  artificial chromosomes as, 569–571
  gene targeting with, 560
  for recombinant DNA, 542, 545–546, 550–555
vegetative organs, of plants, 622
vegetative propagation, of somatic mosaics, 266
ventral uterine (VU) precursor cell, of *Caenorhabditis elegans*, 604–605

vermilion-eyes (*v*) gene, in *D. melanogaster*, 190
vertebrates
cell types in, 490
evolution of sex chromosomes in, 355–357
vertebrates (cont'd)
introns in, 463
telomeric DNA sequences of, 336
vesicular stomatitis virus, synthetic vaccines and, 567
vestigial gene, in *D. melanogaster*, 618–619
viability, in fitness, 732–733
Victoria, Queen, as hemophilia carrier, 154–155
virulent bacteriophage, 413
viruses. *See also* bacteriophage.
antigens versus, 113–114
in bacterial transduction, 403
bacteriophage as, 6
endosymbiont, 691
genetic material of, 9
genome sizes of, 312–313
germ-line transformation via, 557–559
posttranscriptional cosuppression and, 523–524
retroviruses, 566–567
reverse transcriptase in, 550
synthetic vaccines and, 567
V–J joining, in antibodies, 527–528
voles, post-Chernobyl mutations in, 294–295
v-rel gene, 614
vulva, of *Caenorhabditis elegans*, 604–607

W

wall cress, diploid chromosome number of, 134
water, DNA and, 46–47
Watson, James D., 9–10, 16, 32, 43–44, 48–49C, 224
Watson–Crick pairing. *See* base pairing
Watson–Crick structure, of DNA, 9–10, 44, 228
Watts-Tobin, R. J., 472–473C
W chromosome, 155
Weinberg, Robert A., 652C
Weinberg, Wilhelm, 712–713, 716–717C
wells, in gel electrophoresis, 52–53
wheat. *See also* emmer wheat
diploid chromosome number of, 134
DNA base composition of, 44
genome of, 382
hexaploidy in, 380
synteny groups in, 383–384
White, J. G., 602–603C
White, Raymond L., 68C
white-eosin gene, in *D. melanogaster*, 190
white eyes (*w*) gene, in *D. melanogaster*, 150–153, 155–157, 173, 177–181, 190, 562
white flowers
of pea plants, 89, 93, 115–123, 161–162
of snapdragons, 111
White Leghorn chickens, 117
white plants, as four-o'clock phenotype, 684–685
white-spotting gene *S*
in cats, 354–355
White Wyandotte chickens, 117
whorls, in plant flowers, 623–626
Wieschaus, Eric, 613C
wildtype alleles
complementation tests and, 118–123
in *D. melanogaster*, 150–153, 155–157, 177–181
multiple, 111–112
in pea plants, 88–90
Wilkins, Maurice H. F., 48
Wilmut, Ian, 558C
wing disks, of *D. melanogaster*, 617
wings clipped mutant, in *D. melanogaster*, 557
W-linked inheritance, in chickens, 155
wobble, in genetic code, 478–479
Wright, Sewall, 165
wrinkled seeds, in pea plants, 88–103, 164

X

X chromosomes, 148–158
abnormalities of, 362
cancer cells and, 656
dosage compensation of genes in, 351–353
in *D. melanogaster*, 150–153, 176–181, 184, 186–187, 209, 282–284, 323, 364
evolution of, 355–357
genotype frequencies and, 721–722
in grasshoppers, 145
in human karyotype, 347–348
in mammals, 262
methylation of, 519
in mice, 354
pseudoautosomal regions in, 355
telomeres of, 358
Xenopus, oocytes of, 598
Xenopus laevis, gene amplification in, 525
xeroderma pigmentosum, 661
defective DNA repair and, 664–666
UV sensitivity in, 288
X inactivation, 351–353
in calico cats, 353–355
X-inactivation center (*XIC*), 352–354
X-inactivation–specific transcript (*Xist*), 352–353
X-linked inheritance, 262
dosage compensation of genes in, 351–353
in *D. melanogaster*, 150–153, 176–181, 187–189
of Duchenne muscular dystrophy, 173
fragile-X syndrome as, 272–273
genotype frequencies and, 721–722
X-linked severe combined immunodeficiency disease, 584
x-ray film, in Southern blot, 58
"x-ray martyrs," 290
x rays
Hermann Muller and, 291C
mutations via, 211, 286, 290–291
XXX trisomy, 362
XXY trisomy, 362
XYY trisomy, 362

Y

YAC clones, 573–574
YAC vectors, 569–571
Y chromosomes, 148–158
abnormalities of, 362
in *D. melanogaster*, 323, 330
evolution of, 355–357
genotype frequencies and, 721–722
in grasshoppers, 145
in human karyotype, 347–348
pseudoautosomal regions in, 355
yeast. *See also Saccharomyces cerevisiae*
autonomously replicating sequences in, 224
CCA1 gene in, 486
cell cycle of, 637–639, 644
centromeres of, 331–332
chromosome interference in, 194
core proteome of, 576
cyclin–CDK complexes in, 642–643
diploid chromosome number of, 134
DNA chip assays of, 580–581
galactose metabolism in, 508–509
genome size of, 313
mating type in, 530–532
mating-type interconversion in, 528–529
meiosis in, 202
mismatch repair in, 297
petite mutants of, 688–689
proteomes of, 27
silencers in, 511
topoisomerase II in, 315–316
wobble in, 479
yeast artificial chromosomes (YACs), 569–571
yellow body (*y*) gene
in *D. melanogaster*, 184, 190
mutation rate of, 284
yellow pods, in pea plants, 89, 93
yellow seeds, in pea plants, 89, 93, 98–103, 160–161
Y-linked inheritance, dosage compensation of genes in, 351

Z

Z1.ppp cells, in nematodes, 604–605
Z4.aaa cells, in nematodes, 604–605
Z chromosome, 155
Zea mays. *See also* maize
diploid chromosome number of, 134
genome size of, 312
life cycle of, 141
linkage groups in, 184–186
synteny groups in, 384
zebra necrotic-1 (*zn*-1) gene, in maize, 185
zeste (*z*) mutation, in *D. melanogaster*, 364, 555
Z form of DNA, 47
Ziegfeld, Florenz, 290
zinc finger, 510, 614
zinc ion, in transcriptional activators, 510
zw2 gene, in *D. melanogaster*, 364
zygotene, 141–142, 144
in *Lilium longiflorum*, 144
zygotes. *See also* gametes; heterozygous genotypes; homozygous genotypes
of *Caenorhabditis elegans*, 596
in Down syndrome, 360
of *Neurospora crassa*, 206
nonchromosomal genes from, 686–688
zygotic genes, of *D. melanogaster*, 609, 614
zygotic genome, activation of, 597–598

GENETICS

ANALYSIS OF GENES AND GENOMES

Daniel L. Hartl
Elizabeth W. Jones

Fifth Edition

GENETICS: Analysis of Genes and Genomes, 5/e, presents an integrated view of the modern world of genetics: treating classical, molecular, and population genetics as unified subdisciplines in a field that, even in our post-genomic era, still goes by the name "genetics." This approach to teaching genetics is a logical progression in an era when the various subdisciplines of genetics are so closely interwoven. GENETICS, 5/e, delivers a clear, comprehensive, and balanced approach to genetics and genomics at the college level. The text provides students with an understanding of the basic processes of gene transmission, mutation, expression, and regulation. Equally important, it familiarizes students with modern and historical experimental methods used by geneticists and gives insight into the historical and social context of genetics and genomics. Once again, Hartl and Jones break new ground in the field of genetics education.

FEATURES

- Connecting past and present, **Connections** are unique excerpts from the original literature of genetics, each introduced with a brief explanatory passage. The papers range from Mendel's early writing to the latest research and serve to enrich the material in the text.
- The text's author-written **Problems** are plentiful and diverse in level and format. Each chapter contains approximately 40 problems for solution. They are designed to reinforce the concepts introduced within the text and to challenge students to achieve a full understanding of the material.
- A list of **Principles** and an **Outline** introduce the main focus of each chapter, while each chapter is concluded with a complete, concise **Summary** and a list of **Key Terms.** At the end of the book, there is a **Concise Dictionary of Genetics,** where students can check their knowledge of terms encountered within the text or on the Internet.
- **GeNETics on the Web** is the text-specific site for GENETICS, 5/e. It serves as a guide to the vast amount of information on the Internet. Questions, exercises, and informative **Online Resources** are provided for each chapter. Frequently updated, and tuned in to breaking news in the field of genetics, it is a valuable tool for both students and instructors. Please visit us at

  http://www.jbpub.com/genetics

- **Supplements** include an **Instructor's Toolkit CD-ROM,** the **Student Solutions Manual and Supplemental Problems,** and **An Electronic Companion to Genetics™, ©Cogito Learning Media.**



Jones and Bartlett Publishers
40 Tall Pine Drive
Sudbury, MA 01776
978-443-5000
info@jbpub.com



