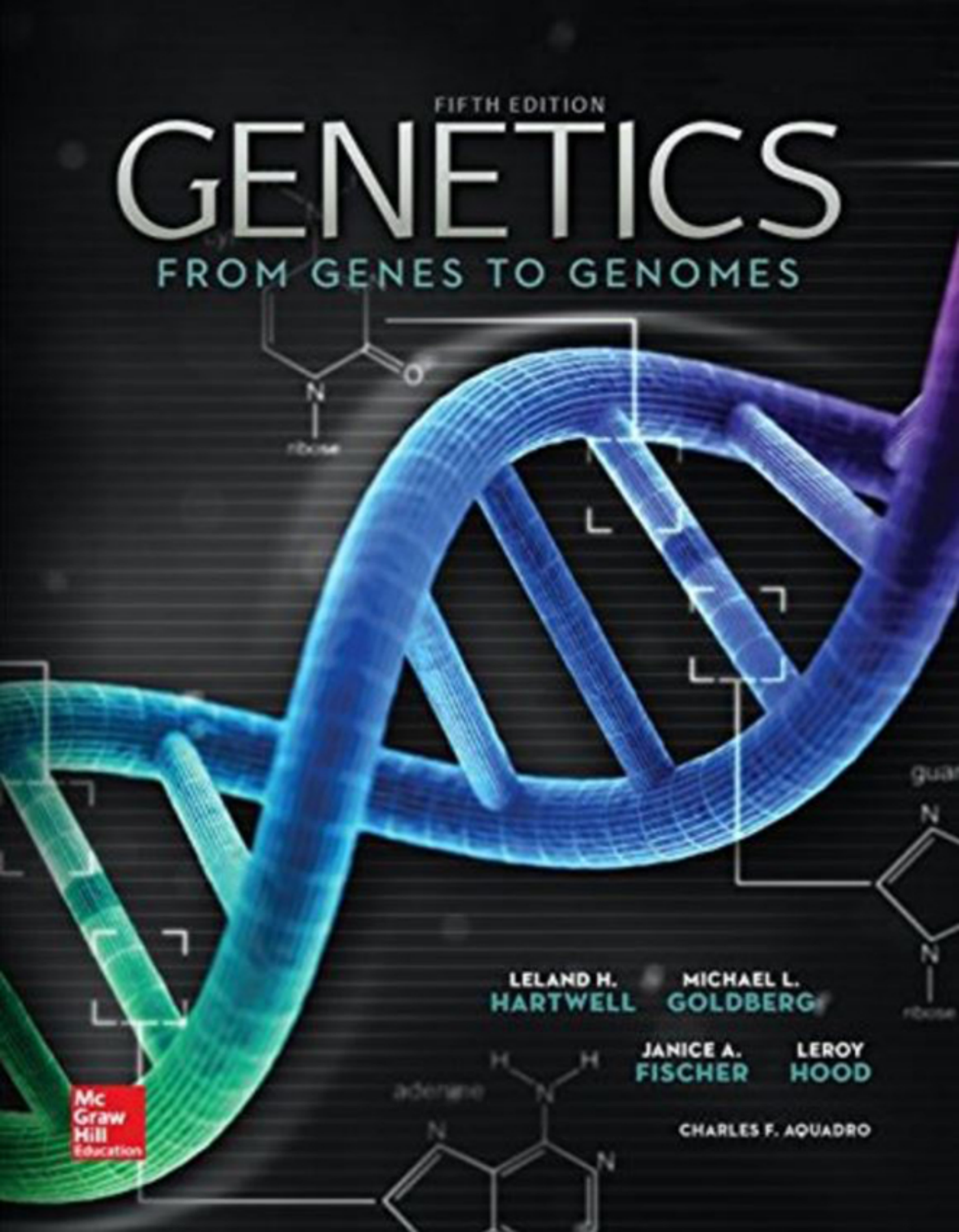
FIFTH EDITION

GENETICS

FROM GENES TO GENOMES



LELAND H. HARTWELL

MICHAEL L. GOLDBERG

JANICE A. FISCHER

LEROY HOOD

CHARLES F. AQUADRO

Mc Graw Hill Education

About the Authors

Dr. Leland Hartwell is President and Director of Seattle's Fred Hutchinson Cancer Research Center and Professor of Genome Sciences at the University of Washington.

Dr. Hartwell's primary research contributions were in identifying genes that control cell division in yeast, including those necessary for the division process as well as those necessary for the fidelity of genome reproduction. Subsequently, many of these same genes have been found to control cell division in humans and often to be the site of alteration in cancer cells.

Dr. Hartwell is a member of the National Academy of Sciences and has received the Albert Lasker Basic Medical Research Award, the Gairdner Foundation International Award, the Genetics Society Medal, and the 2001 Nobel Prize in Physiology or Medicine.

Dr. Michael Goldberg is a professor at Cornell University, where he teaches introductory genetics and human genetics. He was an undergraduate at Yale University and received his Ph.D. in biochemistry from Stanford University. Dr. Goldberg performed postdoctoral research at the Biozentrum of the University of Basel (Switzerland) and at Harvard University, and he received an NIH Fogarty Senior International Fellowship for study at Imperial College (England) and fellowships from the Fondazione Cenci Bolognetti for sabbatical work at the University of Rome (Italy). His current research uses the tools of *Drosophila* genetics and the biochemical analysis of frog egg cell extracts to investigate the mechanisms that ensure proper cell cycle progression and chromosome segregation during mitosis and meiosis.

Dr. Janice Fischer is a Professor at The University of Texas at Austin, where she is an award-winning teacher of genetics and Director of the Biology Instructional Office. She received her Ph.D. in biochemistry and molecular biology from Harvard University, and did postdoctoral research at The University of California at Berkeley and The Whitehead Institute at MIT. In her current research, Dr. Fischer uses *Drosophila* to examine the roles of ubiquitin and endocytosis in cell signaling during development.

Dr. Lee Hood received an M.D. from the Johns Hopkins Medical School and a Ph.D. in biochemistry from the California Institute of Technology. His research interests include immunology, cancer biology, development, and the development of biological instrumentation (for example, the protein sequencer and the automated fluorescent DNA sequencer). His early research played a key role in unraveling the mysteries of antibody diversity. More recently he has pioneered systems approaches to biology and medicine.

Dr. Hood has taught molecular evolution, immunology, molecular biology, genomics and biochemistry and has co-authored textbooks in biochemistry, molecular biology, and immunology, as well as *The Code of Codes*—a monograph about the Human Genome Project. He was one of the first advocates for the Human Genome Project and directed one of the federal genome centers that sequenced the human genome. Dr. Hood is currently the president (and co-founder) of the cross-disciplinary Institute for Systems Biology in Seattle, Washington.

Dr. Hood has received a variety of awards, including the Albert Lasker Award for Medical Research (1987), the Distinguished Service Award from the National Association of Teachers (1998) and the Lemelson/MIT Award for Invention (2003). He is the 2002 recipient of the Kyoto Prize in Advanced Biotechnology—an award recognizing his pioneering work in developing the protein and DNA synthesizers and sequencers that provide the technical foundation of modern biology. He is deeply involved in K–12 science education. His hobbies include running, mountain climbing, and reading.

Dr. Charles Aquadro (Chip) is Professor of Population Genetics, the Charles A. Alexander Professor of Biological Sciences, and Director of the Center for Comparative and Population Genomics at Cornell University. He obtained his Ph.D. in genetics from the University of Georgia, was a postdoc at the National Institute for Environmental Health Sciences/NIH, and joined the faculty at Cornell University in 1985 where he is now a professor. He has served as President of the Society of Molecular Biology and Evolution, is an elected Fellow of the AAAS, is a member of the Scientific Advisory Board for National Geographic Society's Genographic Project, was a member of the Scientific Advisory Board for the WGBH/NOVA TV series "Evolution," and has been a visiting scholar at Cambridge University (England, 1993) and Harvard University (2007). His research and teaching focuses on molecular population genetics, molecular evolution, and comparative genomics. While *Drosophila* is his primary research system, recent work has also involved yeast, humans, and plants. At Cornell, he teaches a university-wide course to nonmajors on personal genomics and medicine, and a major's course in population genetics.

Digital Author

In today's world of learning through technology, it is highly important to have the content of the text be mirrored and delivered in a digital manner and format which leads to classroom success for students while simultaneously delivering individual and/or classroom progress information to instructors.

Enter the digital author. With this fifth edition we are pleased to add professor **Bruce Bejcek** from Western Michigan University to the Hartwell Genetics team.

Dr. Bruce Bejcek received his Ph.D. from St. Louis University. After postdoctoral fellowships at the Jewish Hospital of St. Louis and University of Minnesota he joined the faculty at Western Michigan University in the Department of Biological Sciences. Currently a professor, his research interests have focused on the establishment and maintenance of tumors, particularly those that involve the expression of platelet-derived growth factor and, more recently, herpes simplex virus. His research also includes the discovery of anti-tumor compounds from plants native to Michigan. He has taught a variety of courses including cancer biology, cell biology, and general genetics.

Contributors

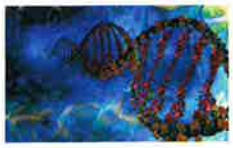
Genetics research tends to proceed down highly specialized paths. A number of experts in specific areas generously provided information in their areas of expertise. We thank them for their contributions to this edition of our text.

Jody Larson, *Instructional Designer, Textbook Development*

Brief Contents

Contents

PART II



What Genes Are and What They Do 173

chapter 6

DNA Structure, Replication, and Recombination 173

chapter 7

Anatomy and Function of a Gene: Dissection Through Mutation 206

chapter 8

Gene Expression: The Flow of Information from DNA to RNA to Protein 254

PART III

Analysis of Genetic Information 298

chapter 9

Digital Analysis of Genomes 298

chapter 10

Analyzing Genomic Information 342

PART IV

How Genes Travel on Chromosomes 381

chapter 11

The Eukaryotic Chromosome 381

Preface

A Note from the Authors

The science of genetics is less than 150 years old, but its accomplishments within that short time have been astonishing. Gregor Mendel first described genes as abstract units of inheritance in 1865; his work was ignored and then "rediscovered" in 1900. Thomas Hunt Morgan and his students provided experimental verification of the idea that genes reside within chromosomes during the years 1910–1920. By 1944, Oswald Avery and his coworkers had established that genes are made of DNA. James Watson and Francis Crick published their pathbreaking structure of DNA in 1953. Remarkably, less than 50 years later (in 2001), an international consortium of investigators deciphered the sequence of the 3 billion nucleotides in the human genome. Twentieth century genetics made it possible to identify individual genes and to understand a great deal about their functions.

Today, scientists are able to access the enormous amounts of genetic data generated by the sequencing of many organisms' genomes. Analysis of these data will result in a deeper understanding of the complex molecular interactions within and among vast networks of genes, proteins, and other molecules that help bring organisms to life. Finding new methods and tools for analyzing these data will be a significant part of genetics in the twenty-first century.

Our fifth edition of *Genetics: From Genes to Genomes* emphasizes both the core concepts of genetics and the cutting-edge discoveries, modern tools, and analytic methods that will keep the science of genetics moving forward.

The authors of the fifth edition have worked together in revising each and every chapter in an effort not only to provide the most up-to-date information, but also to provide continuity and the clearest possible explanations of difficult concepts in one voice.

Our Focus—An Integrated Approach

Genetics: From Genes to Genomes represents a new approach to an undergraduate course in genetics. It reflects the way we, the authors, currently view the molecular basis of life.

We integrate:

- **Formal genetics:** the rules by which genes are transmitted.
- **Molecular genetics:** the structure of DNA and how it directs the structure of proteins.
- **Digital analysis and genomics:** recent technologies that allow a comprehensive analysis of the entire gene set and its expression in an organism.
- **Human genetics:** how genes contribute to health and diseases, including cancer.
- **The unity of life-forms:** the synthesis of information from many different organisms into coherent models.
- **Molecular evolution:** the molecular mechanisms by which biological systems, whole organisms, and populations have evolved and diverged.

The strength of this integrated approach is that students who complete the book will have a strong command of genetics as it is practiced today by both academic and corporate researchers. These scientists are rapidly changing our understanding of living organisms, including ourselves. Ultimately, this vital research may create the ability to replace or correct detrimental genes—those "inborn errors of metabolism," as researcher Archibald Garrod called them in 1923, as well as the later genetic alterations that lead to the many forms of cancer.

The Genetic Way of Thinking

Modern genetics is a molecular-level science, but an understanding of its origins and the discovery of its principles is a necessary context. To encourage a genetic way of thinking, we begin the book by reviewing Mendel's principles and the chromosomal basis of inheritance. From the outset, however, we aim to integrate organism-level genetics with fundamental molecular mechanisms.

Chapter 1 presents the foundation of this integration by summarizing the main biological themes we explore. In Chapter 2, we tie Mendel's studies of pea trait inheritance to the actions of enzymes that determine whether a pea is round or wrinkled, yellow or green, etc. In the same chapter, we point to the relatedness of the patterns of heredity in all organisms. Chapters 3–5 cover extensions to Mendel, the chromosome theory of inheritance, and the fundamentals of gene linkage and mapping. Starting in Chapter 6, we focus on the physical characteristics of DNA, on mutations, and on how DNA encodes, copies, and transmits biological information.

Beginning in Chapter 9, we move into the digital revolution in DNA analysis with a look at modern genetics techniques, including gene cloning, PCR, microarrays, and high-throughput genome sequencing. We explore how bioinformatics, an emergent analytical tool, can aid in discovery of genome features.

The understanding of molecular and computer-based techniques carries into our discussion of chromosome specifics in Chapters 11–14, and also informs our analysis of gene regulation in Chapters 15 and 16. Chapter 17 describes the technology that scientists can use to manipulate genomes at will – for research and practical purposes including gene therapy. Chapter 18 describes the use of genetic tools at the molecular level to uncover the complex interactions of eukaryotic development. In Chapter 19, we explain how our understanding of genetics and the development of molecular genetic technologies is enabling us to comprehend cancer and in some cases to cure it.

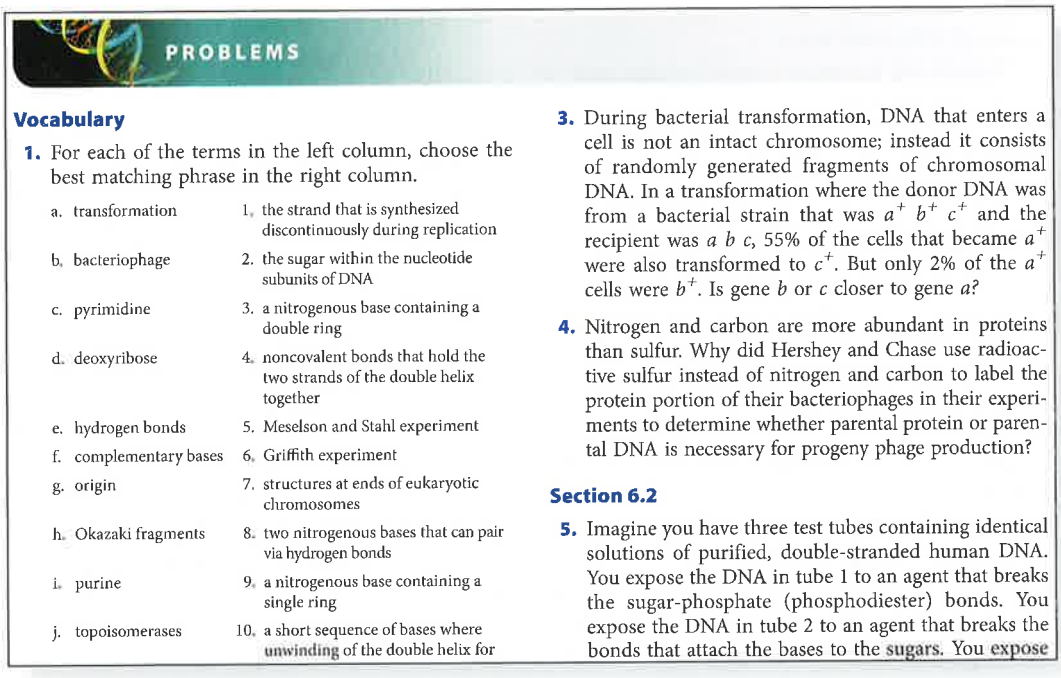
Chapters 20 and 21 cover population genetics, with a view of how molecular tools have provided information on species relatedness and on genome changes at the molecular level over time. In addition, we explain how bioinformatics can be combined with population genetics to understand inheritance of complex traits and to trace human ancestry.

Throughout our book, we present the scientific reasoning of some of the ingenious researchers of the field—from Mendel, to Watson and Crick, to the collaborators on the Human Genome Project. We hope student readers will see that genetics is not simply a set of data and facts, but also a human endeavor that relies on contributions from exceptional individuals.

Student-Friendly Features

We have taken great pains to help the student make the leap to a deeper understanding of genetics. Numerous features of this book were developed with that goal in mind.

- ***One Voice Genetics:*** *Genes to Genomes* has a friendly, engaging reading style that helps students master the concepts throughout this book. The writing style provides the student with the focus and continuity required to make the book successful in the classroom.
- ***Visualizing Genetics*** The highly specialized art program developed for this book integrates photos and line art in a manner that provides the most engaging visual presentation of genetics available. Our Feature Figure illustrations break down complex processes into step-by-step illustrations that lead to greater student understanding. All illustrations are rendered with a consistent color theme—for example, all presentations of phosphate groups are the same color, as are all presentations of mRNA.
- ***Accessibility*** Our intention is to bring cutting-edge content to the student level. A number of more complex illustrations are revised and segmented to help the student follow the process. Legends have been streamlined to highlight only the most important ideas, and throughout the book, topics and examples have been chosen to focus on the most critical information.
- ***Problem Solving*** Developing strong problem-solving skills is vital for every genetics student. The authors have carefully created problem sets at the end of each chapter that allow students to improve upon their problem-solving ability.
- ***Solved Problems*** These cover topical material with complete answers provide insight into the step-by-step process of problem solving.
- ***Review Problems*** More than 600 questions involving a variety of levels of difficulty that develop excellent problem-solving skills. The problems are organized by chapter section and in order of increasing difficulty within each section for ease of use by instructors and students. Answers to selected problems are in the back of the book. The companion online Study Guide and Solutions Manual, completely revised for the 5th edition by Michael Goldberg and Janice Fischer, provides detailed analysis of strategies to solve all of the end-of-chapter problems.



PROBLEMS

Vocabulary

1. For each of the terms in the left column, choose the best matching phrase in the right column.

a. transformation	1. the strand that is synthesized discontinuously during replication
b. bacteriophage	2. the sugar within the nucleotide subunits of DNA
c. pyrimidine	3. a nitrogenous base containing a double ring
d. deoxyribose	4. noncovalent bonds that hold the two strands of the double helix together
e. hydrogen bonds	5. Meselson and Stahl experiment
f. complementary bases	6. Griffith experiment
g. origin	7. structures at ends of eukaryotic chromosomes
h. Okazaki fragments	8. two nitrogenous bases that can pair via hydrogen bonds
i. purine	9. a nitrogenous base containing a single ring
j. topoisomerases	10. a short sequence of bases where unwinding of the double helix for

3. During bacterial transformation, DNA that enters a cell is not an intact chromosome; instead it consists of randomly generated fragments of chromosomal DNA. In a transformation where the donor DNA was from a bacterial strain that was a^+ b^+ c^+ and the recipient was a b c, 55% of the cells that became a^+ were also transformed to c^+. But only 2% of the a^+ cells were b^+. Is gene b or c closer to gene a?

4. Nitrogen and carbon are more abundant in proteins than sulfur. Why did Hershey and Chase use radioactive sulfur instead of nitrogen and carbon to label the protein portion of their bacteriophages in their experiments to determine whether parental protein or parental DNA is necessary for progeny phage production?

Section 6.2

5. Imagine you have three test tubes containing identical solutions of purified, double-stranded human DNA. You expose the DNA in tube 1 to an agent that breaks the sugar-phosphate (phosphodiester) bonds. You expose the DNA in tube 2 to an agent that breaks the bonds that attach the bases to the sugars. You expose

Changes in the 5th Edition: A Chapter-by-Chapter Summary

The fifth edition represents a major revision of the fourth edition. Coverage of every topic was clarified and updated. In total, over 200 Figures and Tables are completely new, many more figures and tables were revised, and greater than 200 new end-of-chapter problems were created. Two new Chapters—Chapter 17 (Manipulating the Genomes of Eukaryotes) and Chapter 21 (Genetics of Complex Traits)—are included.

Chapter 1 Genetics: The Study of Biological Information
- Modernized to reflect new human genomics technologies
- All new set of end-of-chapter problems

Chapter 2 Mendel's Principles of Heredity
- New emphasis on biochemical explanations for dominant and recessive alleles

Chapter 3 Extensions to Mendel's Laws
- Major revisions emphasize biochemical basis for gene interactions
- New Comprehensive Example: gene interactions in dog coat color and pattern

Chapter 4 The Chromosome Theory of Inheritance
- New material emphasizes human genetics: human sex chromosomes; human sex determination; X inactivation
- New Fast Forward Box: Transgenic Mice Prove that *SRY* Is the Maleness Factor

Chapter 5 Linkage, Recombination, and the Mapping of Genes on Chromosomes
- Significant revisions clarify linkage analysis, especially: the relationship between crossing over and physical distance between genes; tetrad analysis

Chapter 6 DNA Structure, Replication, and Recombination
- Major update to molecular mechanism of recombination: the double-strand break repair and anticrossover helicase model; completely revised Feature Figure illustrating recombination
- More extensive coverage of gene conversion
- Better coverage of how tetrad analysis supports the recombination model

Chapter 7 Anatomy and Function of Gene: Dissection Through Mutation
- Sharper focus on point mutations in the human germ line and human disease
- New Fast Forward Box: Trinucleotide repeat disease genes
- Updated description of DNA repair mechanisms
- Improvements to description of Benzer's experiments

Chapter 8 Gene Expression: The Flow of Information from DNA to RNA to Protein
- Better explanation of Crick and Brenner triplet codon experiment
- Inclusion of Brenner's stop codon experiments
- Clarified coverage of tRNA synthetases
- Updated Wobble rules
- Clearer explanation of the five classes (morphs) of mutant alleles
- Amino acids 21 and 22 included

Chapter 9 Digital Analysis of Genomes
- Complete revision includes some material from Chapters 9, 10, and 20 of the fourth edition
- Students progress rapidly from cloning and Sanger sequencing to genome assembly, to a description of what genomes look like and how to examine genomes online.
- Outmoded techniques (for example, Southern blotting) replaced by new technologies (automated DNA sequencing; bioinformatics)

Chapter 10 Analyzing Genomic Information
- Complete revision of former Chapter 11 in the fourth edition
- Bookended by a case history showing how whole exome/genome sequencing was used to identify a disease mutation and to cure the patient
- Updated description of polymorphism types (SNPs, InDels, SSRs)
- Updated coverage of SNP analysis by PCR
- Updated coverage of forensic DNA fingerprinting
- Modernized coverage of DNA microarrays for whole genome analysis
- In-depth coverage of positional cloning
- New Box on Lod scores
- New section on high-throughput DNA sequencing of whole genomes to identify disease genes

Chapter 11 The Eukaryotic Chromosome
- Updated coverage of FISH techniques for karyotyping
- Modernized and expanded coverage of histone tail modifications and their effects on gene expression
- Updated coverage of X chromosome inactivation mechanism
- New material about nucleosome re-assembly after DNA replication
- Updated coverage of telomeres and shelterin
- Modernized coverage of kinetochores and cohesion complexes

Chapter 12 Chromosomal Rearrangements and Changes in Chromosome Number
- Updated modern methods (FISH, PCR, and genome sequencing) for detecting and analyzing chromosomal rearrangements and ploidy
- Reorganized coverage of chromosomal rearrangement origins, phenotypic consequences in humans, and their uses in research
- Updated coverage of transposons in human genome
- New material about Barbara McClintock's discovery of transposons

Chapter 13 Bacterial Genetics
- Bacterial genetics and organellar genetics in separate chapters
- Updated coverage includes pangenomes and human microbiome
- Clearer explanations of classical experiments and gene mapping methods
- Modernized coverage of genetic methods: gene targeting; transposon mutagenesis; gene identification by plasmid transformation
- New Comprehensive Example: molecular mechanisms of drug resistance

Chapter 14 Organellar Genetics
- Entire chapter now devoted to organellar genetics
- New coverage of Correns's experiment showing chloroplast maternal inheritance
- Expanded explanation of human mitochondrial disease inheritance
- New material about mtDNA "gene therapy"

Chapter 15 Gene Regulation in Prokaryotes
- Better explanation of the classic *lac* operon experiments
- New material on *cis*- and *trans*-acting RNAs (for example, riboswitches)
- Modernized coverage of transcriptome analysis (including RNA-Seq)
- New comprehensive example: quorum sensing in *V. fischeri*

Chapter 16 Gene Regulation in Eukaryotes
- Revised heavily to clarify the basic elements of eukaryotic gene regulation, and to update the coverage of mechanisms and technology used to discover them
- Clarified roles of *cis*- and *trans*-acting regulators of eukaryotic gene expression (activators/coactivators, repressors/corepressors, indirect repressors, enhancers, and insulators)
- Modernized coverage of the role of DNA methylation in imprinting (epigenetics)
- Updated coverage of small RNAs
- Updated technology for discovery of transcription factors
- Updated Comprehensive Example: *Drosophila* sex determination

Chapter 17 Manipulating the Genomes of Eukaryotes
- A completely new chapter highlighting techniques for genome manipulation and the rationale for altering genomes
- Technology described includes: creation and uses of transgenic organisms; cloning by somatic nuclear transfer; targeted mutagenesis (knockouts, conditional knockouts, knockins, TALENs); human gene therapy

Chapter 18 The Genetic Analysis of Development
- Revised significantly to sharpen focus on the use of genetic analysis as a tool to study development
- Detailed descriptions of primary and modifier mutant screens
- Coverage of methods for identifying genes corresponding to mutant phenotypes
- Reorganized coverage of how mutations and cloned genes are used to determine where and when genes act
- Expanded coverage of ordering genes in a pathway using epistasis
- Comprehensive Example of *Drosophila* body patterning includes more detail about mutant screens for maternal effect and zygotically-acting genes

Chapter 19 Understanding Cancer
- Updated significantly to emphasize: identifying "driver" mutations from tumor genome sequences; the genetic landscape of cancers; comparisons of the whole genome sequences of tumors from many patients to find patterns that can suggest treatments

Chapter 20 Variation and Selection in Populations
- Heavy revision to simplify and modernize coverage of population genetics
- Hardy-Weinberg analysis related to forensics/CODIS
- Coverage of MRCA concept
- New material on Y chromosome and mtDNA analysis for tracing human ancestry

Chapter 21 Genetics of Complex Traits
- Largely new chapter focuses on the use of modern methods for whole genome analysis
- In-depth presentation of mapping QTLs by crosses in plants
- GWAS in humans
- Analysis of DNA variation to trace human ancestry

Online Teaching and Learning Resources

Help Your Students Prepare for Class

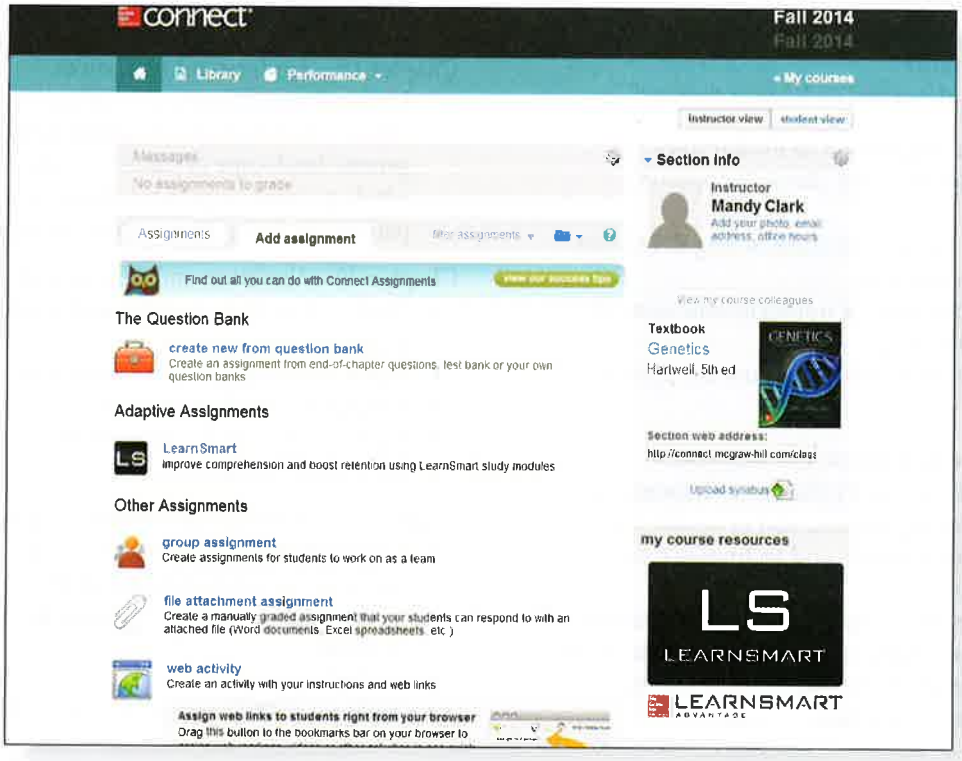
Digital resources can help you achieve your instructional goals—making your students more responsible for learning outside of class by meeting your students where they live: on the go and online. Use the text and digital tools to empower students to come to class more prepared and ready to engage!

McGraw-Hill Connect Genetics provides online presentation, assignment, and assessment solutions. It connects your students with the tools and resources they'll need to achieve success. With **Connect Genetics** you can deliver assignments, quizzes, and tests online. A robust set of questions and activities is presented in the Question Bank and a separate set of questions to use for exams is presented in the Test Bank. As an instructor, you can edit existing questions and author entirely new problems. Track individual student performance—by question, by assignment, or in relation to the class overall—with detailed grade reports. Integrate grade reports easily with Learning Management Systems such as Blackboard and Canvas—and much more. **ConnectPlus Genetics** provides students with all the advantages of Connect Biology plus 24/7 online access to an eBook. This media-rich version of the book is available through the McGraw-Hill ConnectPlus platform and allows seamless integration of text, media, and assessments.

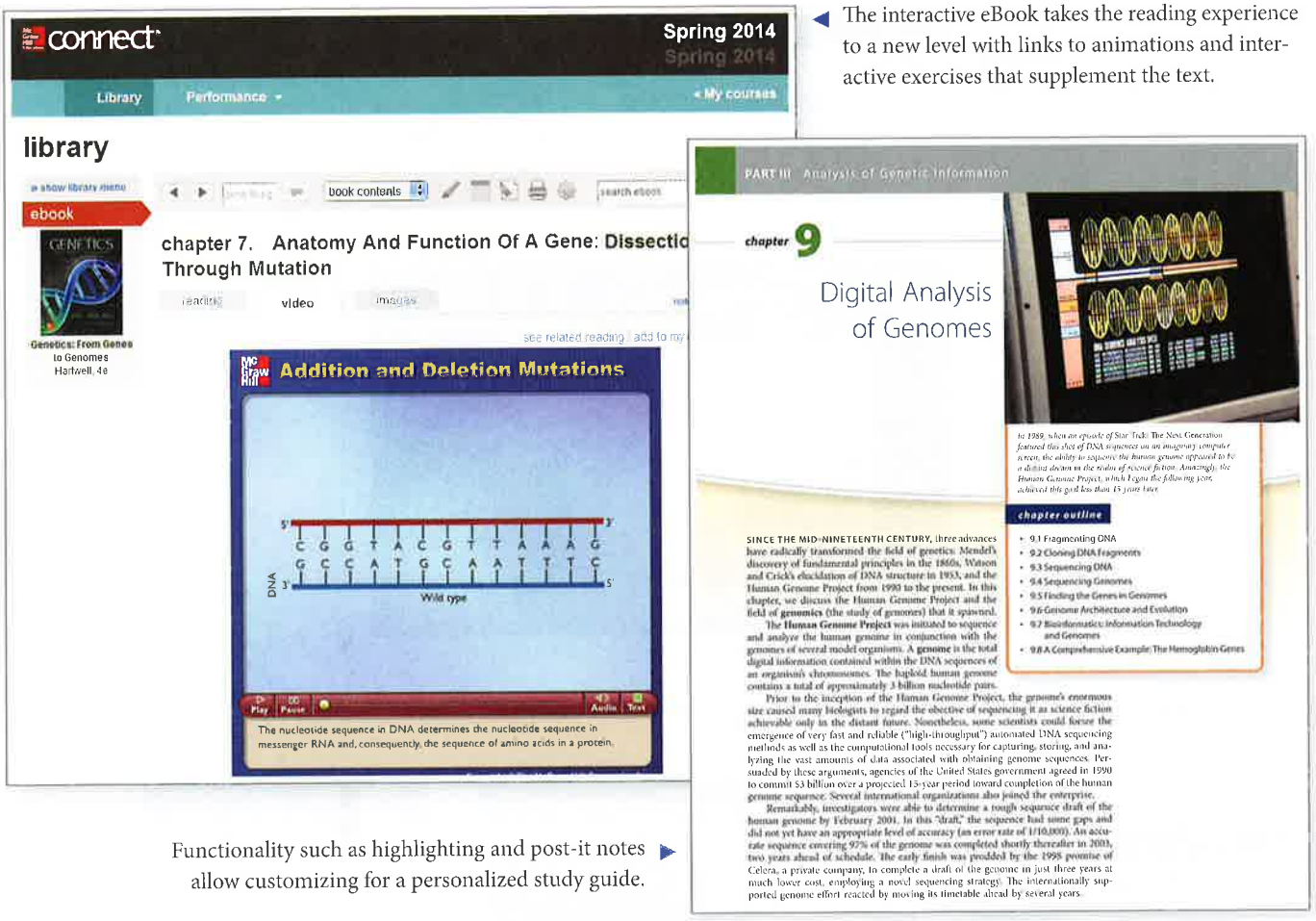
To learn more, **visit www.mcgrawhillconnect.com.**

Help Your Students by Making Assignments—Reading, Homework, and LearnSmart

Connect content can be assigned as homework for before class to help students with basic concepts so they can better understand classroom presentations and projects. Quizzes taken after class can also evaluate their comprehension. These assignments support the rich assessment presented in the text so that student and professors can gauge the level of understanding of concepts and the mastery of skills.



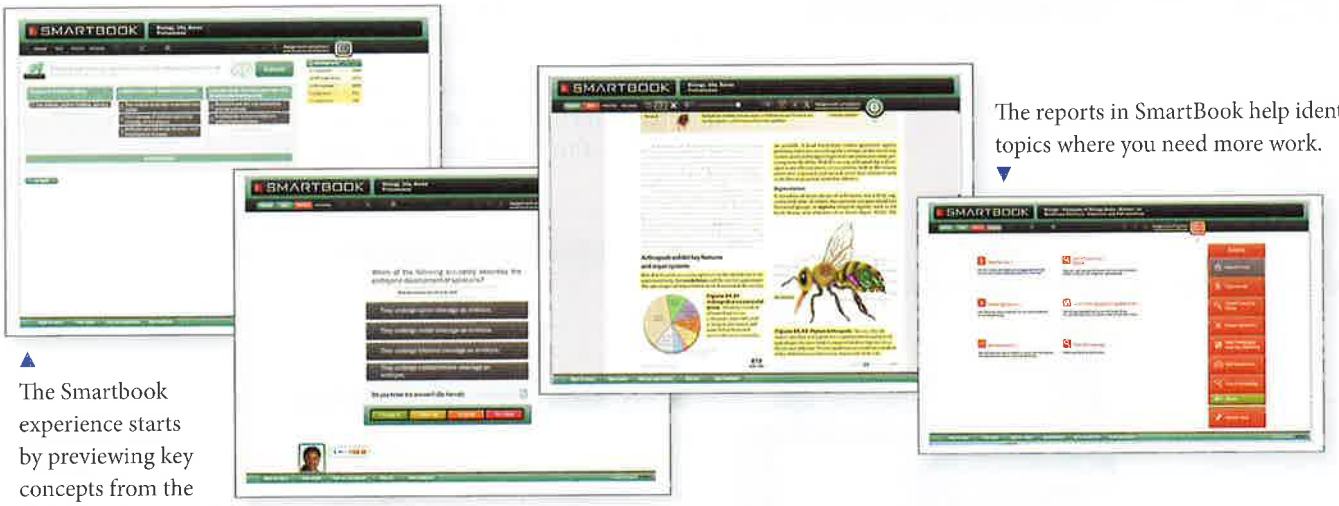
Assignments can include reading assignments from the ConnectPlus eBook or SmartBook, homework or quizzes, LearnSmart, your own web or short answer activities, and more.



The interactive eBook takes the reading experience to a new level with links to animations and interactive exercises that supplement the text.

Functionality such as highlighting and post-it notes allow customizing for a personalized study guide.

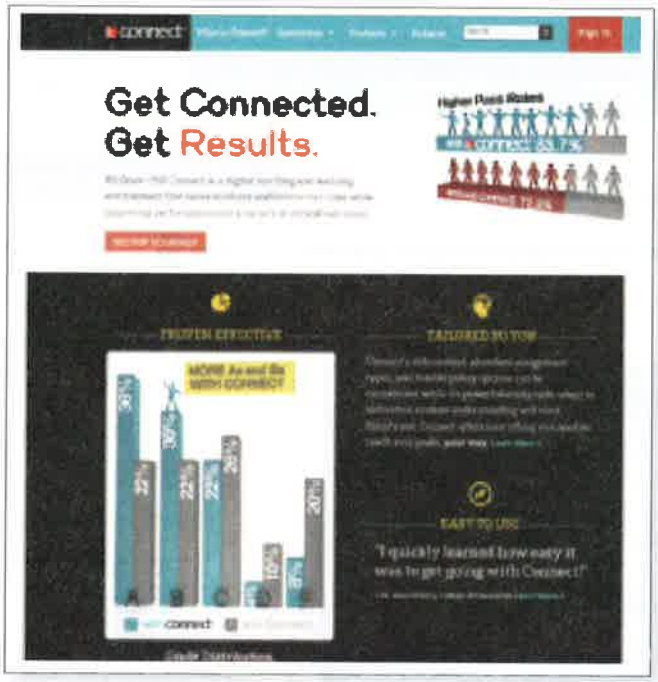
SMARTBOOK™ Powered by an intelligent diagnostic and adaptive engine, **SmartBook** facilitates the reading process by identifying what content a student knows and doesn't know through adaptive assessments.



The reports in SmartBook help identify topics where you need more work.

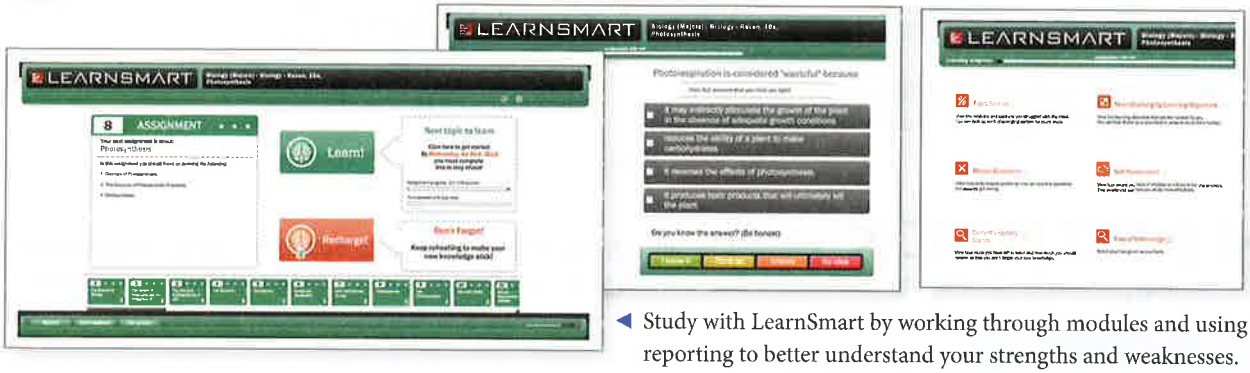
The Smartbook experience starts by previewing key concepts from the chapter and ensuring that you understand the big ideas.

SmartBook asks you questions that identify gaps in your knowledge. The reading experience then continuously adapts in response to the assessments, highlighting the material you need to review based on what you don't know.



◀ Help your students prepare for class and revisit content by assigning homework in Connect and LearnSmart.

LEARNSMART®



◀ Study with LearnSmart by working through modules and using LearnSmart's reporting to better understand your strengths and weaknesses.

Gauge your students' progress using reports in LearnSmart and Connect. Students can run these same reports in LearnSmart to track their own progress.

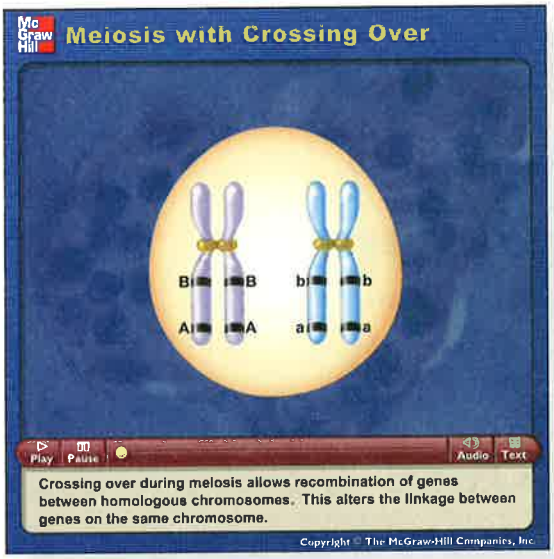
▲ The Tree of Knowledge tracks your progress, reporting on short-term successes and long-term retention.

▲ Download the LearnSmart app from iTunes or Google Play and work on LearnSmart from anywhere!

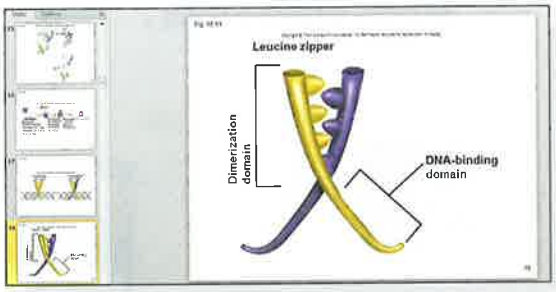
◀ Reports in Connect and LearnSmart help you monitor student assignments and performance, allowing for "just-in-time" teaching to clarify concepts that are more difficult for your students to understand.

Presentation Tools

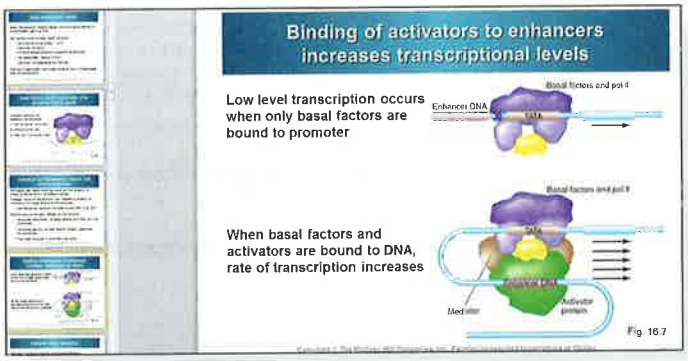
Within Connect, you will find presentation materials to enhance your class all in one place.



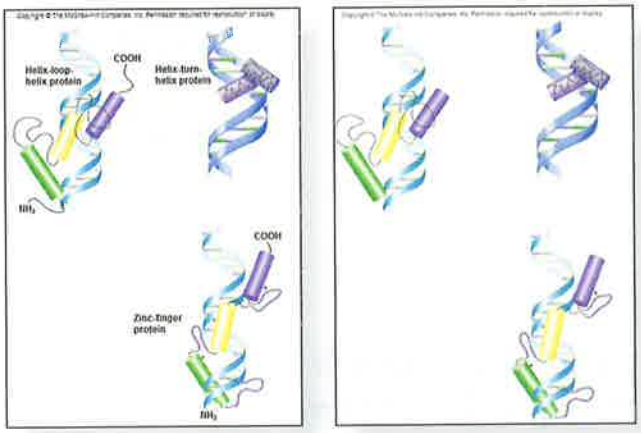
◀ Animation PowerPoints contain full-color animations illustrating important processes, which are fully embedded in PowerPoint slides for easy use in your presentations.



▲ FlexArt PowerPoints contain editable art from the text. For all figures, labels and leader lines are editable and some figures also have editable or stepped out art allowing you to customize your PowerPoint presentations.



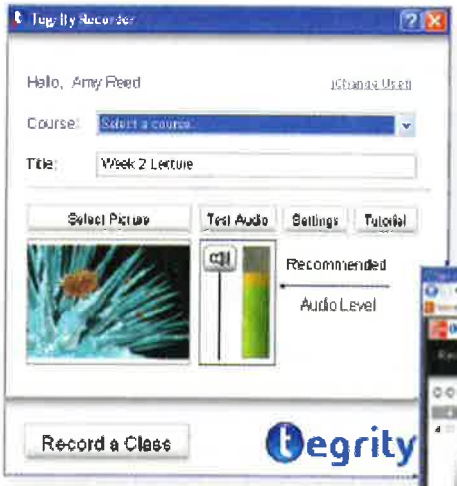
▲ Lecture PowerPoints with animations fully embedded.



▲ Labeled and unlabeled JPEG files of all art and photos in the text to be readily incorporated into presentations, exams, or custommade classroom materials.

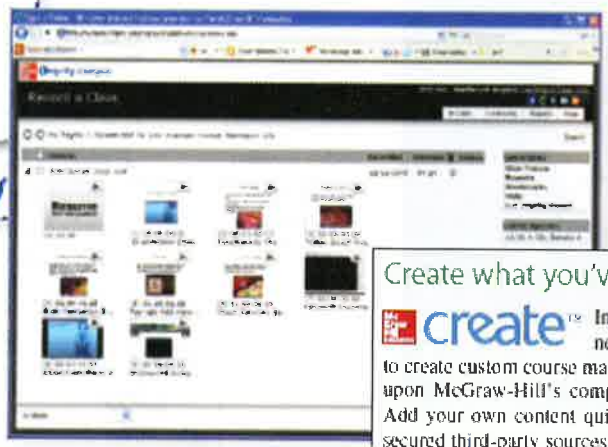
Fully Developed Test Bank

All questions have been updated to fully align with the learning objectives and content of the text. Provided within a computerized test bank powered by McGraw-Hill's flexible electronic testing program EZ Test Online, instructors can create paper and online tests or quizzes in this easy-to-use program! A new tagging scheme allows you to sort questions by difficulty level, topic, and section. Imagine being able to create and access your test or quiz anywhere, at any time, without installing the testing software. Now, with EZ Test Online, instructors can select questions from multiple McGraw-Hill test banks or author their own, and then either print the test for paper distribution or give it online.



Take your course online—*easily*—with one-click Digital Lecture Capture

McGraw-Hill Tegrity Campus™ records and distributes your lecture with just a click of a button. Students can view them anytime/anywhere via computer, iPod, or mobile device. Tegrity Campus indexes as it records your slideshow presentations and anything shown on your computer so **students can use keywords to find exactly what they want to study.**



Create what you've only imagined

create™ Introducing **McGraw-Hill Create™**—a new, self-service website that allows you to create custom course materials—print and eBooks—by drawing upon McGraw-Hill's comprehensive, cross-disciplinary content. Add your own content quickly and easily. Tap into other rights-secured third-party sources as well. Then, arrange the content in a way that makes the most sense for your course. Even personalize your book with your course name and information! Choose the best format for your course: color print, black-and-white print, or eBook. The eBook is now even viewable on an iPad! And, when you are done you will receive a free PDF review copy in just minutes!

Visit McGraw-Hill Create—**www.mcgrawhillcreate.com**—today and begin building your perfect book.

Finally, a way to quickly and easily create the course materials you've always wanted.

Imagine that.

Solutions Manual/ Study Guide Online

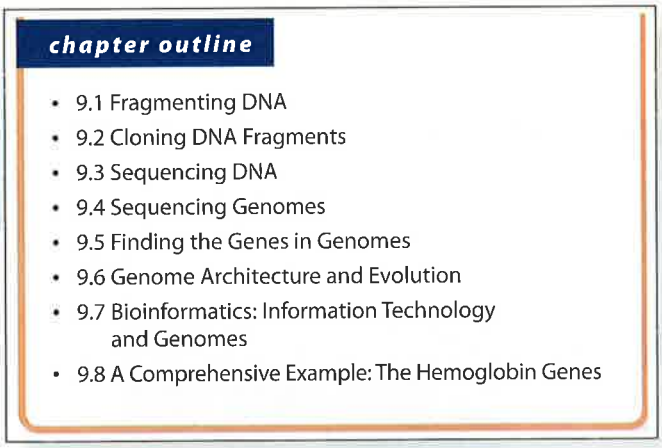
Extensively revised by the authors of the fifth edition, this manual presents the solutions to the end-of-chapter problems and questions along with the step-by-step logic of each solution delivered in a digital format for the first time! The manual also includes a synopsis, the objectives, and problem-solving tips for each chapter. Key figures and tables from the textbook are referenced throughout to guide student study.

Integrating Genetic Concepts

Genetics: From Genes to Genomes takes an integrated approach in its presentation of genetics, thereby giving students a strong command of genetics as it is practiced today by academic and corporate researchers. Principles are related throughout the text in examples, essays, case histories, and Connections sections to make sure students fully understand the relationships between topics.

Chapter Outline

Every chapter opens with a brief outline of the chapter contents.

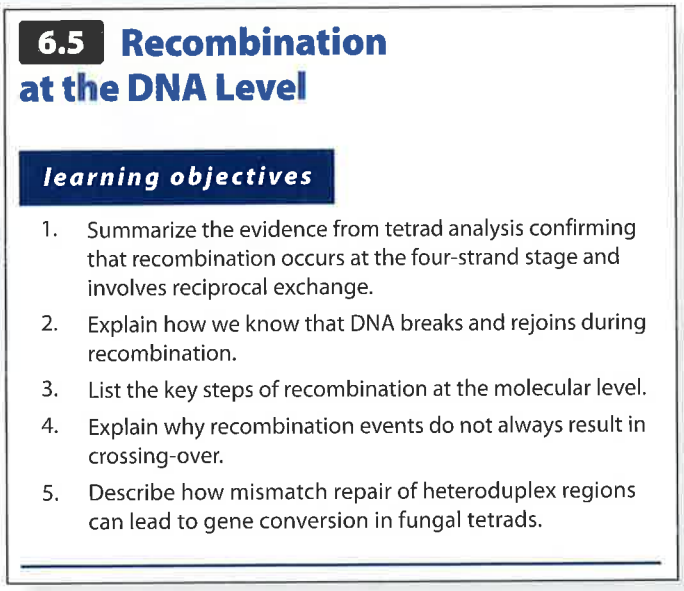


chapter outline

- 9.1 Fragmenting DNA
- 9.2 Cloning DNA Fragments
- 9.3 Sequencing DNA
- 9.4 Sequencing Genomes
- 9.5 Finding the Genes in Genomes
- 9.6 Genome Architecture and Evolution
- 9.7 Bioinformatics: Information Technology and Genomes
- 9.8 A Comprehensive Example: The Hemoglobin Genes

New! Learning Objectives

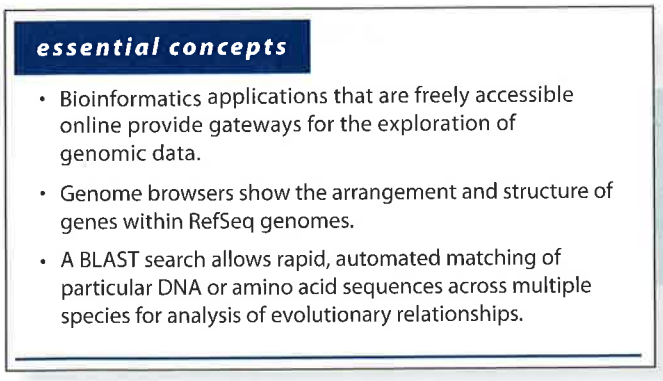
Learning Objectives appear before each section, and are carefully written to clearly outline expectations.



6.5 Recombination at the DNA Level

learning objectives

1. Summarize the evidence from tetrad analysis confirming that recombination occurs at the four-strand stage and involves reciprocal exchange.
2. Explain how we know that DNA breaks and rejoins during recombination.
3. List the key steps of recombination at the molecular level.
4. Explain why recombination events do not always result in crossing-over.
5. Describe how mismatch repair of heteroduplex regions can lead to gene conversion in fungal tetrads.



essential concepts

- Bioinformatics applications that are freely accessible online provide gateways for the exploration of genomic data.
- Genome browsers show the arrangement and structure of genes within RefSeq genomes.
- A BLAST search allows rapid, automated matching of particular DNA or amino acid sequences across multiple species for analysis of evolutionary relationships.

Essential Concepts

After each section, the most relative points of content are now provided in concise, bulleted statements to reinforce critical concepts and learning objectives for students.

WHAT'S NEXT

Most of the methods that bacteria use to regulate their genes are available to eukaryotes as well. For example, both types of organisms can use diffusible regulatory proteins to increase or decrease transcription initiation. In both prokaryotes and eukaryotes, transcription and translation are also regulated after the intiation step by sRNAs.

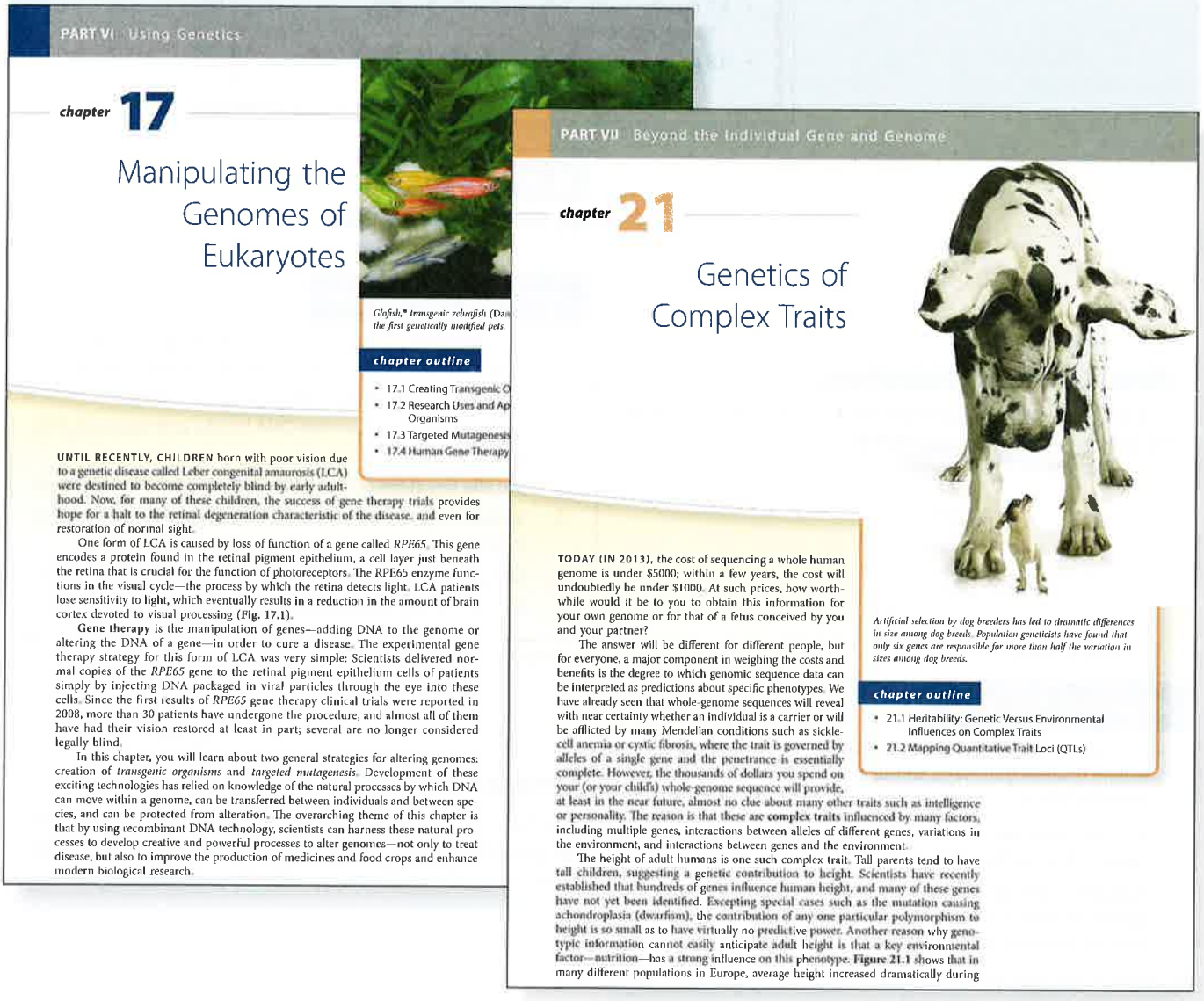
Several features unique to eukaryotes nonetheless dictate that the mechanisms these organisms utilize to regulate gene expression cannot all be the same as those used in prokaryotes. In eukaryotes, transcription in the nucleus is physically separated by the nuclear membrane from the sites of translation on the ribosomes in the cytoplasm. Thus, eukaryotes cannot employ mechanisms such as attenuation that depend on the coupling between transcription and translation. However, mRNAs of eukaryotic genes must be spliced, modified at their 5′ and 3′ ends, and transported from the nucleus into the cytoplasm where they are translated. In addition, eukaryotic chromosomes are wound up in chromatin. You will see in Chapter 16 that all of these processes, as well as multicellularity, provide and necessitate additional avenues for the regulation of gene expression in eukaryotic organisms.

New! What's Next

Each chapter closes with a What's Next section that serves as a bridge between the topics in the chapter just completed to those in the upcoming chapter or chapters. This spirals the learning and builds connections for students.

New! Exciting Revised Content and Two New Chapters

The fifth edition represents a major revision of the fourth edition. Coverage of every topic was clarified and updated. In total, over 200 Figures and Tables are completely new, many more figures and tables were revised, and greater than 200 new end-of-chapter problems were created. Two new Chapters—Chapter 17 (Manipulating the Genomes of Eukaryotes) and Chapter 21 (Genetics of Complex Traits)—are included.


PART VI Using Genetics

chapter 17

Manipulating the Genomes of Eukaryotes

Glofish,® transgenic zebrafish (Da... the first genetically modified pets.

chapter outline

- 17.1 Creating Transgenic O...
- 17.2 Research Uses and Ap... Organisms
- 17.3 Targeted Mutagenesis
- 17.4 Human Gene Therapy

UNTIL RECENTLY, CHILDREN born with poor vision due to a genetic disease called Leber congenital amaurosis (LCA) were destined to become completely blind by early adulthood. Now, for many of these children, the success of gene therapy trials provides hope for a halt to the retinal degeneration characteristic of the disease, and even for restoration of normal sight.

One form of LCA is caused by loss of function of a gene called *RPE65*. This gene encodes a protein found in the retinal pigment epithelium, a cell layer just beneath the retina that is crucial for the function of photoreceptors. The RPE65 enzyme functions in the visual cycle—the process by which the retina detects light. LCA patients lose sensitivity to light, which eventually results in a reduction in the amount of brain cortex devoted to visual processing (**Fig. 17.1**).

Gene therapy is the manipulation of genes—adding DNA to the genome or altering the DNA of a gene—in order to cure a disease. The experimental gene therapy strategy for this form of LCA was very simple: Scientists delivered normal copies of the *RPE65* gene to the retinal pigment epithelium cells of patients simply by injecting DNA packaged in viral particles through the eye into these cells. Since the first results of *RPE65* gene therapy clinical trials were reported in 2008, more than 30 patients have undergone the procedure, and almost all of them have had their vision restored at least in part; several are no longer considered legally blind.

In this chapter, you will learn about two general strategies for altering genomes: creation of *transgenic organisms* and *targeted mutagenesis*. Development of these exciting technologies has relied on knowledge of the natural processes by which DNA can move within a genome, can be transferred between individuals and between species, and can be protected from alteration. The overarching theme of this chapter is that by using recombinant DNA technology, scientists can harness these natural processes to develop creative and powerful processes to alter genomes—not only to treat disease, but also to improve the production of medicines and food crops and enhance modern biological research.

PART VII Beyond the Individual Gene and Genome

chapter 21

Genetics of Complex Traits

Artificial selection by dog breeders has led to dramatic differences in size among dog breeds. Population geneticists have found that only six genes are responsible for more than half the variation in sizes among dog breeds.

chapter outline

- 21.1 Heritability: Genetic Versus Environmental Influences on Complex Traits
- 21.2 Mapping Quantitative Trait Loci (QTLs)

TODAY (IN 2013), the cost of sequencing a whole human genome is under $5000; within a few years, the cost will undoubtedly be under $1000. At such prices, how worthwhile would it be to you to obtain this information for your own genome or for that of a fetus conceived by you and your partner?

The answer will be different for different people, but for everyone, a major component in weighing the costs and benefits is the degree to which genomic sequence data can be interpreted as predictions about specific phenotypes. We have already seen that whole-genome sequences will reveal with near certainty whether an individual is a carrier or will be afflicted by many Mendelian conditions such as sickle-cell anemia or cystic fibrosis, where the trait is governed by alleles of a single gene and the penetrance is essentially complete. However, the thousands of dollars you spend on your (or your child's) whole-genome sequence will provide, at least in the near future, almost no clue about many other traits such as intelligence or personality. The reason is that these are **complex traits** influenced by many factors, including multiple genes, interactions between alleles of different genes, variations in the environment, and interactions between genes and the environment.

The height of adult humans is one such complex trait. Tall parents tend to have tall children, suggesting a genetic contribution to height. Scientists have recently established that hundreds of genes influence human height, and many of these genes have not yet been identified. Excepting special cases such as the mutation causing achondroplasia (dwarfism), the contribution of any one particular polymorphism to height is so small as to have virtually no predictive power. Another reason why genotypic information cannot easily anticipate adult height is that a key environmental factor—nutrition—has a strong influence on this phenotype. **Figure 21.1** shows that in many different populations in Europe, average height increased dramatically during

FAST FORWARD

Gene Mapping May Lead to a Cure for Cystic Fibrosis

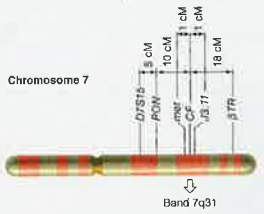
For 40 years after the symptoms of cystic fibrosis were first described in 1938, no molecular clue—no visible chromosomal abnormality transmitted with the disease, no identifiable protein defect carried by affected individuals—suggested the genetic cause of the disorder. As a result, there was no effective treatment for the 1 in 2000 Caucasian Americans born with the disease, most of whom died before they were 30. In the 1980s, however, geneticists were able to combine recently invented techniques for looking directly at DNA with maps constructed by linkage analysis to pinpoint a precise chromosomal position, or *locus*, for the cystic fibrosis gene.

The mappers of the cystic fibrosis gene faced an overwhelming task. They were searching for a gene that encoded an unknown protein, a gene that had not yet even been assigned to a chromosome. It could lie anywhere among the 23 pairs of chromosomes in a human cell.

- A review of many family pedigrees confirmed that cystic fibrosis is most likely determined by a single gene (*CF*). Investigators collected white blood cells from 47 families with two or more affected children, obtaining genetic data from 106 patients, 94 parents, and 44 unaffected siblings.
- They next tried to discover if any other trait is reliably transmitted with cystic fibrosis. Analyses of the easily obtainable serum enzyme paraoxonase showed that its gene (*PON*) is indeed linked to *CF*. At first, this knowledge was not that helpful because *PON* had not yet been assigned to a chromosome.
- Then, in the early 1980s, geneticists developed a large s of DNA markers, based on new techniques that enable them to recognize variations in the genetic material. A **marker** is a piece of DNA representing a specific locus, comes in identifiable variations. These allelic variations gate according to Mendel's laws, which means it is poss to follow their transmission as you would any gene's. Chapter 10 explains the discovery and use of DNA mar greater detail; for now, it is only important to know that exist and can be identified.

By 1986, linkage analyses of hundreds of DNA marker shown that one marker, known as *D7S15*, is linked with bot and *CF*. Researchers computed recombination frequencies and that the distance from the DNA marker to *CF* was 15 cM: fro DNA marker to *PON*, 5 cM; and from *PON* to *CF*, 10 cM. They conc

Figure A How molecular markers helped locate the gene for cystic fibrosis (*CF*).



that the order of the three loci was *D7S15-PON-CF* (**Fig. A**). Because *CF* could lie 15 cM in either of two directions from the DNA marker, the area under investigation was approximately 30 cM. And because the human genome consists of roughly 3000 cM, this step of linkage analysis narrowed the search to 1% of the human genome.

Fast Forward

This feature is one of the methods used to integrate the Mendelian principles introduced early in the content with the molecular content that will follow.

TOOLS OF GENETICS

The Lod Score Statistic

The Lod score is a mathematical answer to the question: How much more likely is it that the particular allele transmission pattern observed in a pedigree would have been seen if the loci were linked at any given recombination frequency (RF) less than 50% than if they were not linked? The Lod score, as its name implies (log of the odds) is the logarithm of the ratio between these two probabilities:

$$\text{Lod} = \log \frac{\text{P(obtaining observed results if loci are linked at a given RF)}}{\text{P(obtaining observed results if loci are unlinked)}}$$

Here, we illustrate the Lod score calculation for the pedigree in Fig. 10.21a. The pedigree suggests that the *NF* gene is linked to a particular SNP on chromosome 17. The calculation will allow us to determine our degree of confidence in this preliminary conclusion.

1. **Tabulate which progeny are parental and which are recombinant.** In Fig. 10.20a, you can see that the first 7 children in generation III have the parental (P) configuration of alleles and that only child III-8 has the recombinant (R) configuration. We'll abbreviate these data as **PPPPPPPR**.
2. **Calculate the Lod score denominator.** If two loci are unlinked, it is equally likely that any one child would be P or R (that is, the RF = 50%). The probability of P is thus ½, and the probability of R is also ½. The probability of obtaining children in the particular birth order **PPPPPPPR** if the *NF* gene and the SNP locus are unlinked is:

$$\text{P(RF50\%)} = \left(\frac{1}{2}\right)^8 = \frac{1}{256}$$

You can see that a generalized formula for this part of th calculation is simply:

$(½)^n$, where n is the total number of tabulated individ

3. **Calculate the Lod score numerator.** Loci could be link the RF is any value less than 50%, but the calculation re us to assume an RF value. The pedigree in Fig. 10.20a in cates an RF of 1/8 = 12.5%, so we will use this as our be current estimate. With RF = 1/8, the expected frequenc of P progeny is 7/8, and R progeny is 1/8. The probabilit

7 parentals and 1 recombinant in the particular birth order **PPPPPPPR** is:

$$\text{P(RF12.5\%)} = \left(\frac{7}{8}\right)^7 \left(\frac{1}{8}\right)^1 \approx \frac{1}{20}$$

A generalized formula for calculating the Lod score numerator is:

$$(1 - \text{RF}_{obs})^{\#P} \times (\text{RF}_{obs})^{\#R}$$

where RF_{obs} is the RF indicated by the data, #P is the number of parentals, and #R is the number of recombinants.

4. **Calculate the likelihood ratio.** This is simply the ratio of the values you found in steps 2 and 3. For this example,

$$\text{P(RF12.5\%)/P(RF50\%)} = \left(\frac{1}{20}\right) \Big/ \left(\frac{1}{256}\right) = 12.8$$

This likelihood ratio means that it is 12.8 times more likely that the *NF* gene and the SNP are linked with RF = 12.5% than that they are not linked (RF = 50%).

5. **Calculate the Lod score.** The Lod score is simply the base 10 logarithm of the likelihood ratio. For the example in Fig. 10.20a:

$$\text{Lod score} = \log (12.8) = 1.1$$

6. **Interpret the Lod score.** The convention among human geneticists is that a Lod score ≥3 (that is, a likelihood ratio ≥1000) is required to be confident of linkage. The Lod score of 1.1 indicates that the data in Fig. 10.20a are insufficient to

Tools of Genetics Essays

Current readings explain various techniques and tools used by geneticists, including examples of applications in biology and medicine.

GENETICS AND SOCIETY

Mitochondrial DNA Tests as Evidence of Kinship in Argentine Courts

Between 1976 and 1983, the military dictatorship of Argentina kidnapped, incarcerated, and killed more than 10,000 university students, teachers, union members, and others who did not support the regime. Many very young children disappeared along with the young adults, and close to 120 babies were born to women in detention centers. In 1977, the grandmothers of some of these infants and toddlers held vigils in the main square of Buenos Aires to bear witness and inform others about the disappearance of their children and grandchildren (**Fig. A**). They soon formed a human rights group—the "Grandmothers of the Plaza de Mayo."

The grandmothers' goal was to locate the more than 200 grandchildren they suspected were still alive, and to reunite them with their biological families. To this end, they gathered information from eyewitnesses, such as midwives and former jailers, and set up a network to monitor the papers of children entering kindergarten. They also contacted organizations outside the country, including the American Association for the Advancement of Science (AAAS).

What the grandmothers asked of AAAS was help with genetic analyses that would stand up in court. By the time a democracy had replaced the military regime and the grandmothers could argue their legal cases before an impartial court, children abducted at age 2 or 3 or born in 1976 were 7–10 years old. Although the external features of the children had changed, their genes—relating them unequivocally to their biological families—had not. The grandmothers, who had educated themselves about the potential of genetic tests, sought help with the details of obtaining and analyzing such tests. Starting in 1983, the courts agreed to accept their test results as proof of kinship.

In 1983, the best way to confirm or exclude the relatedness of two or more individuals was to compare proteins called human lymphocyte antigens (HLAs). People carry a unique set of HLA markers on their white blood cells, or lymphocytes, and these markers are diverse enough to form a kind of molecular fingerprint. HLA analyses can be carried out even if a child's parents are no longer alive, because for each HLA marker, a child inherits one allele from the maternal grandparents and one from the paternal grandparents. Statistical analyses can establish the probability that a child shares genes with a set of grandparents. In some cases, HLA analysis was sufficient to make a very strong case that a tested child belonged to the family claiming him or her. But in other cases, a reliable match could not be accomplished through HLA typing.

The AAAS put the grandmothers in touch with Mary Claire King, then at the University of California. King and two colleagues—

Figure A Grandmothers of the Plaza de Mayo (1977).

C. Orrego and A. C. Wilson—developed an mtDNA test based on the then new techniques of PCR amplification and direct sequencing of a highly variable noncoding region of the mitochondrial genome. Maternal inheritance and lack of recombination of mtDNAs mean that as long as a single maternal relative is available for matching, the approach can resolve cases of disputed relatedness. The extremely polymorphic noncoding region makes it possible to identify grandchildren through a direct match with the mtDNA of only one person—their maternal grandmother, or their mother's sister or brother—rather than through statistical calculations.

To validate their approach King and colleagues amplified sequences from three children and their three maternal grandmothers without knowing who was related to whom. The mtDNA test unambiguously matched the children with their grandmothers. Thus, after 1989, the grandmothers included mtDNA data in their archives.

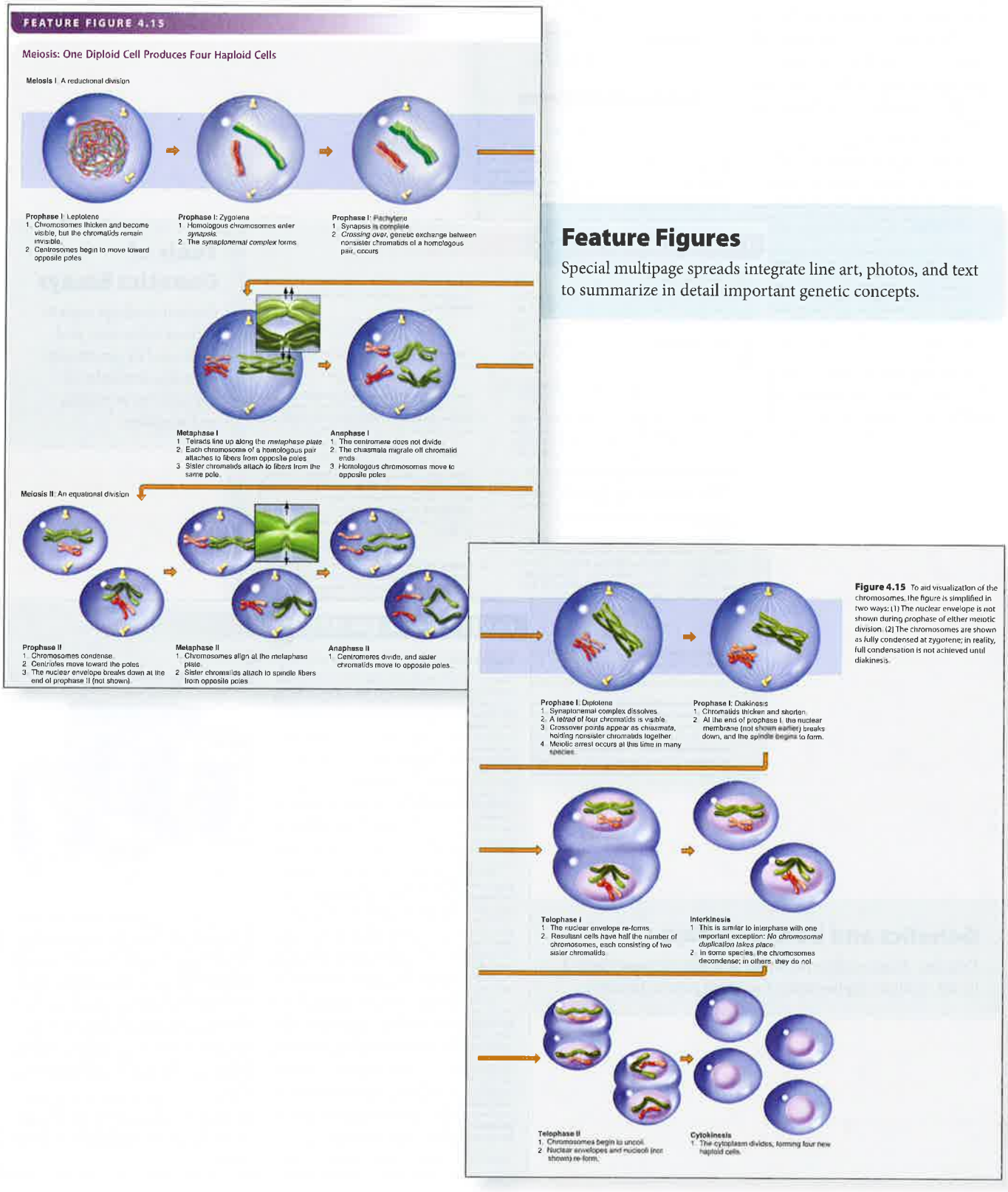
Today, the grandchildren—the children of "the disappeared"—have reached adulthood and attained legal independence. Although most of their grandmothers have died, the grandchildren may still discover their biological identity and what happened to their families through the mtDNA data the grandmothers left behind.

Genetics and Society Essays

Dramatic essays explore the social and ethical issues created by the multiple applications of modern genetic research.

Visualizing Genetics

Full-color illustrations and photographs bring the printed word to life. These visual reinforcements support and further clarify the topics discussed throughout the text.



Feature Figures

Special multipage spreads integrate line art, photos, and text to summarize in detail important genetic concepts.

Process Figures

Step-by-step descriptions allow the student to walk through a compact summary of important details.

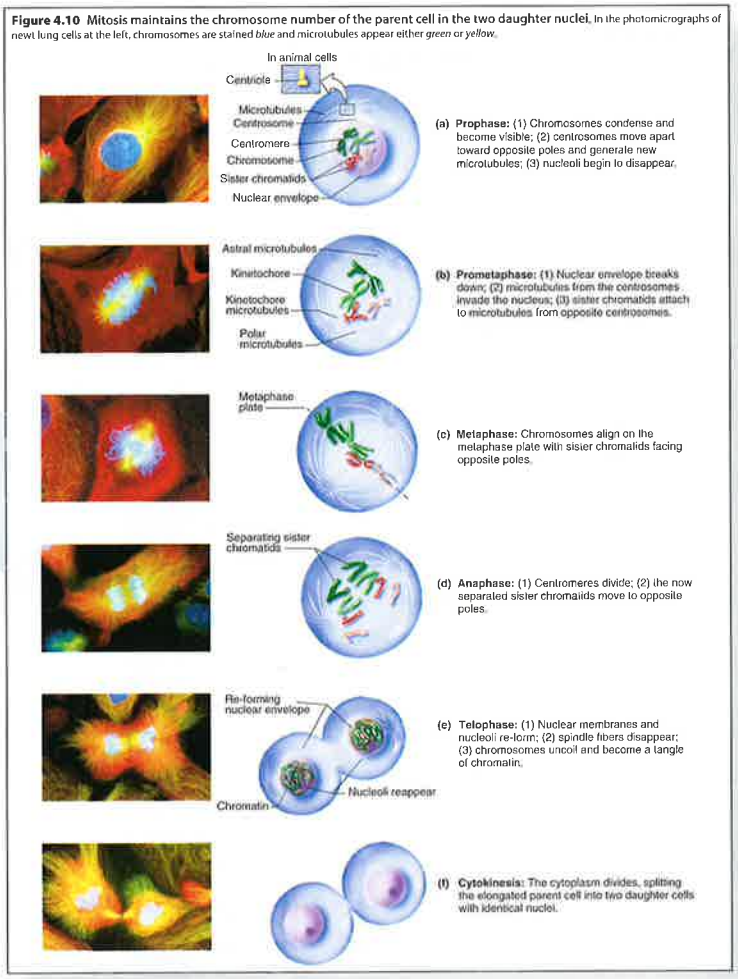
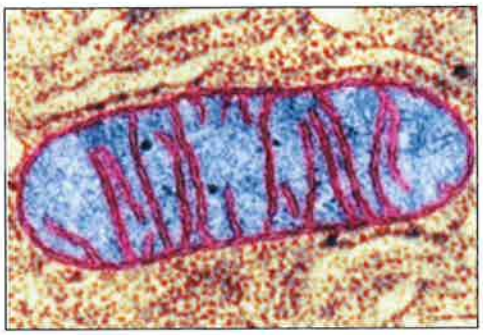
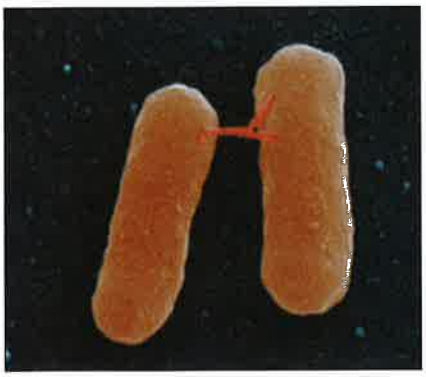


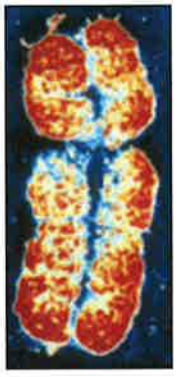
Figure 4.10 **Mitosis maintains the chromosome number of the parent cell in the two daughter nuclei.** In the photomicrographs of newt lung cells at the left, chromosomes are stained *blue* and microtubules appear either *green* or *yellow*.

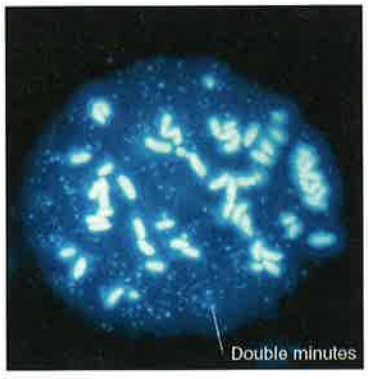
Micrographs

Stunning micrographs bring the genetics world to life.









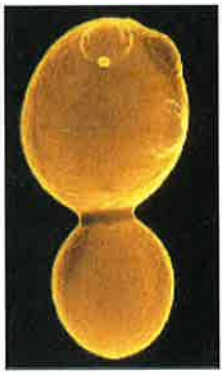
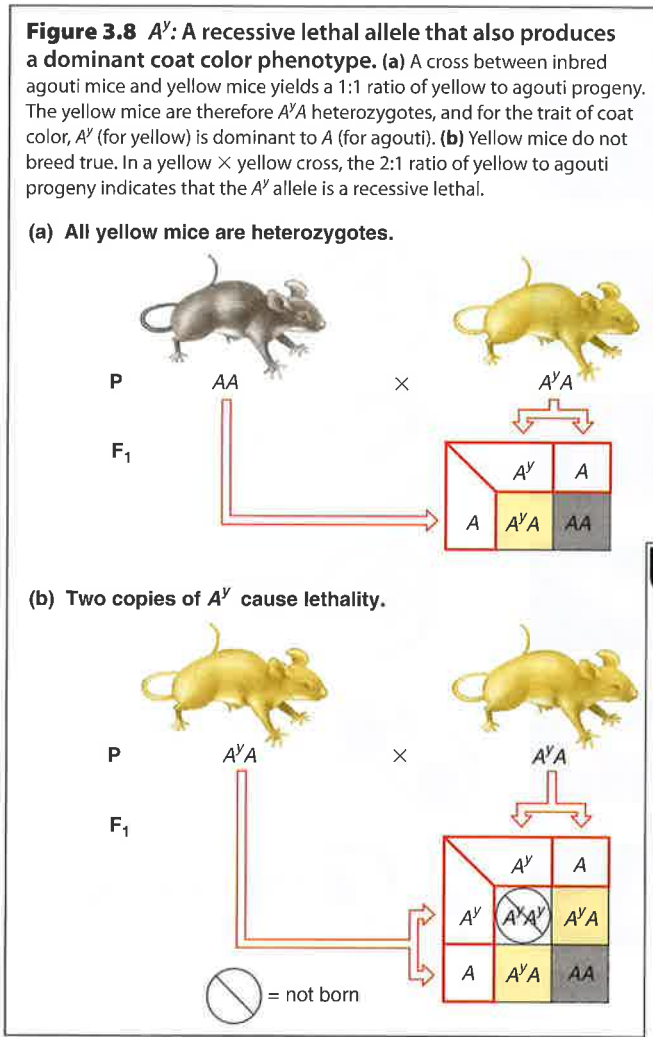
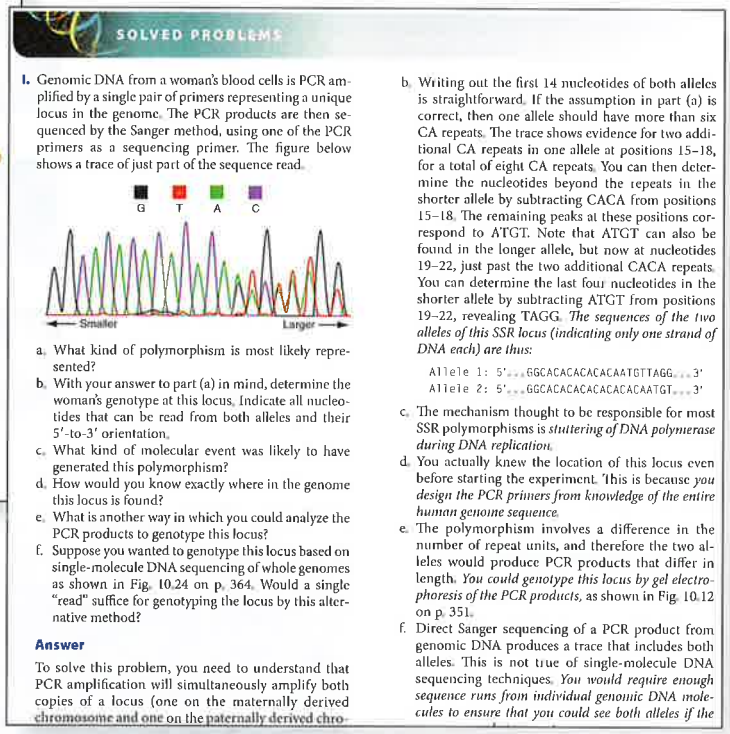


Figure 3.8 A^y: A recessive lethal allele that also produces a dominant coat color phenotype. **(a)** A cross between inbred agouti mice and yellow mice yields a 1:1 ratio of yellow to agouti progeny. The yellow mice are therefore A^yA heterozygotes, and for the trait of coat color, A^y (for yellow) is dominant to A (for agouti). **(b)** Yellow mice do not breed true. In a yellow × yellow cross, the 2:1 ratio of yellow to agouti progeny indicates that the A^y allele is a recessive lethal.



Comparative Figures

Comparison illustrations lay out the basic differences of often confusing principles.



SOLVED PROBLEMS

1. Genomic DNA from a woman's blood cells is PCR amplified by a single pair of primers representing a unique locus in the genome. The PCR products are then sequenced by the Sanger method, using one of the PCR primers as a sequencing primer. The figure below shows a trace of just part of the sequence read.

G T A C

Smaller Larger

a. What kind of polymorphism is most likely represented?
b. With your answer to part (a) in mind, determine the woman's genotype at this locus. Indicate all nucleotides that can be read from both alleles and their 5′-to-3′ orientation.
c. What kind of molecular event was likely to have generated this polymorphism?
d. How would you know exactly where in the genome this locus is found?
e. What is another way in which you could analyze the PCR products to genotype this locus?
f. Suppose you wanted to genotype this locus based on single-molecule DNA sequencing of whole genomes as shown in Fig. 10.24 on p. 364. Would a single "read" suffice for genotyping the locus by this alternative method?

Answer

To solve this problem, you need to understand that PCR amplification will simultaneously amplify both copies of a locus (one on the maternally derived chromosome and one on the paternally derived chro-

b. Writing out the first 14 nucleotides of both alleles is straightforward. If the assumption in part (a) is correct, then one allele should have more than six CA repeats. The trace shows evidence for two additional CA repeats in one allele at positions 15–18, for a total of eight CA repeats. You can then determine the nucleotides beyond the repeats in the shorter allele by subtracting CACA from positions 15–18. The remaining peaks at these positions correspond to ATGT. Note that ATGT can also be found in the longer allele, but now at nucleotides 19–22, just past the two additional CACA repeats. You can determine the last four nucleotides in the shorter allele by subtracting ATGT from positions 19–22, revealing TAGG. *The sequences of the two alleles of this SSR locus (indicating only one strand of DNA each) are thus:*

```
Allele 1: 5'...GGCACACACACACAATGTTAGG...3'
Allele 2: 5'...GGCACACACACACACAATGT...3'
```

c. The mechanism thought to be responsible for most SSR polymorphisms is *stuttering of DNA polymerase during DNA replication.*
d. You actually knew the location of this locus even before starting the experiment. This is because *you design the PCR primers from knowledge of the entire human genome sequence.*
e. The polymorphism involves a difference in the number of repeat units, and therefore the two alleles would produce PCR products that differ in length. *You could genotype this locus by gel electrophoresis of the PCR products,* as shown in Fig. 10.12 on p. 351.
f. Direct Sanger sequencing of a PCR product from genomic DNA produces a trace that includes both alleles. This is not true of single-molecule DNA sequencing techniques. *You would require enough sequence runs from individual genomic DNA molecules to ensure that you could see both alleles if the*

Solving Genetics Problems

The best way for students to assess and increase their understanding of genetics is to practice through problems. Found at the end of each chapter, problem sets assist students in evaluating their grasp of key concepts and allow them to apply what they have learned to real-life issues.

Review Problems

Problems are organized by chapter section and in order of increasing difficulty to help students develop strong problem-solving skills. The answers to select problems can be found in the back of this text.

Solved Problems

Solved problems offer step-by-step guidance needed to understand the problem-solving process.

Acknowledgements

The creation of a project of this scope is never solely the work of the authors. We are grateful to our colleagues around the world who took the time to review the previous edition and make suggestions for improvement. Their willingness to share their expectations and expertise was a tremendous help to us.

Andrew Andres, *University of Nevada, Las Vegas*
Daniel A Barbash, *Cornell University*
Miriam Barlow, *University of California Merced*
Christine Barton, *Centre College*
Mary A. Bedell, *University of Georgia*
Kerry Bloom, *University of North Carolina at Chapel Hill*
Robb Brumfield, *Louisiana State University*
Jeff Chen, *University of Texas, Austin*
Brian Condie, *University of Georgia*
Aboubaker Elkharroubi, *Johns Hopkins University*
Tom Fox, *Cornell University*
Josh Gamse, *Vanderbilt University*
Brinda Govindan, *San Francisco State University*
Nancy Hollingsworth, *Stony Brook University*
Mike Ibba, *Ohio State University*
Bob Ivarie, *University of Georgia*
Rohinton T. Kamakaka, *University of California Santa Cruz*
Thomas R. King, Ph.D., *Central Connecticut State University*
Alan Lloyd, *University of Texas, Austin*
Paul Macdonald, *University of Texas, Austin*
Cindy S. Malone, *California State University Northridge*
Dmitri A. Maslov, *University of California–Riverside*
Philip Meneely, *Haverford College*
Kyle Miller, *University of Texas, Austin*
Debra Nero, *Cornell University*
Michael Reagan, *College of St. Benedict/St. John's University*
Ajna Rivera, *University of the Pacific*
Malcolm Schug, *University of North Carolina Greensboro*
Elaine Sia, *University of Rochester*
Ron Siu, *University of California Los Angeles*
Randall Small, *University of Tennessee*
Tara Stoulig, *Southeastern Louisiana University*
Steve Vokes, *University of Texas, Austin*
James Walker, *University of Texas, Austin*
Yunqiu Wang, *University of Miami*
Scott Weitze, *San Francisco State University*
Marvin Whiteley, *University of Texas, Austin*
Patricia Wittkopp, *University of Michigan*
Millie Worku, *North Carolina A&T State University*

A special thank-you to Jody Larson for her extensive feedback on this fifth edition. We would also like to thank the highly skilled publishing professionals at McGraw-Hill who guided the development and production of the fifth edition of ***Genetics: From Genes to Genomes:*** Rebecca Olson for her support; Mandy Clark for her organizational skills and tireless work to tie up all loose ends; and Sheila Frank and the entire production team for their careful attention to detail and ability to move the schedule along.

chapter 1

Genetics: The Study of Biological Information

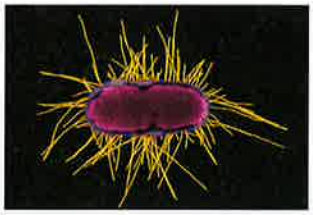
Information can be stored in many ways, including the patterns of letters and words in books and the sequence of nucleotides in DNA molecules.

chapter outline

GENETICS, THE SCIENCE of heredity, is at its core the study of biological information. All living organisms—from single-celled bacteria and protozoa to multicellular plants and animals—must store, replicate, transmit to the next generation, and use vast quantities of information to develop, reproduce, and survive in their environments (**Fig. 1.1**). Geneticists examine how organisms pass biological information on to their progeny and how they use it during their lifetimes.

This book introduces you to the field of genetics as currently practiced in the early twenty-first century. Several broad themes recur throughout this presentation. First, we know that biological information is encoded in DNA, and that the proteins responsible for an organism's many functions are built from this code. We also have found that all living forms are closely related at the molecular level, and recent technology has revealed that genomes have a modular construction that has allowed rapid evolution of complexity. With the aid of high-speed computers and other technologies, we can now study genomes at the level of DNA sequence. Finally, our focus in this book is on human genetics and the application of genetic discoveries to human problems.

Figure 1.1 The biological information in DNA generates an enormous diversity of living organisms.


(a) Bacteria


(b) Clown fish


(c) Lion


(d) Oak tree


(e) Poppies


(f) Hummingbird


(g) Red-eyed tree frog


(h) Humans

1.1 DNA: The Fundamental Information Molecule of Life

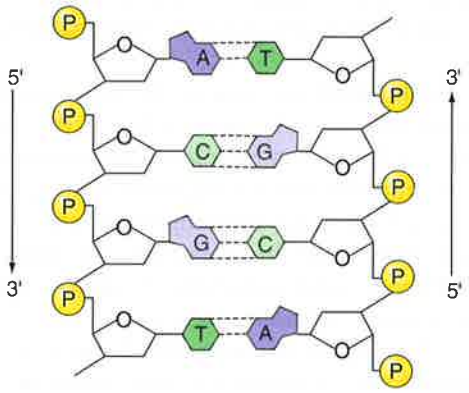
learning objectives

1. Relate the structure of DNA to its function.
2. Differentiate between a chromosome, DNA, a gene, a base pair, and a protein.

The process of evolution has taken close to 4 billion years to generate the amazing mechanisms for storing, replicating, expressing, and diversifying biological information seen in organisms now inhabiting the earth. The linear DNA molecule stores biological information in units known as nucleotides. Within each DNA molecule, the sequence of the four letters of the DNA alphabet—G, C, A, and T—specify which proteins an organism will make as well as when and where protein synthesis will occur. The letters refer to the bases—guanine, cytosine, adenine, and thymine—that are components of the nucleotide building blocks of DNA. The DNA molecule itself is a double strand of nucleotides carrying complementary G–C or A–T base pairs (**Fig. 1.2**). These complementary base pairs bind together through hydrogen bonds. The molecular complementarity of double-stranded DNA is its most important property and the key to understanding how DNA functions.

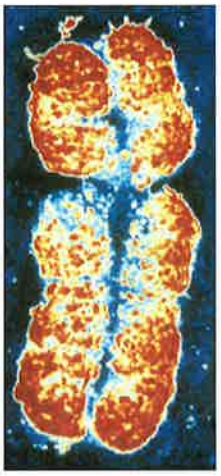
Although the DNA molecule is three-dimensional, most of its information is one-dimensional and digital.

Figure 1.2 Complementary base pairs are a key feature of the DNA molecule. A single strand of DNA is composed of nucleotide subunits each consisting of a deoxyribose sugar (*white pentagons*), a phosphate (*yellow circles*), and one of four nitrogenous bases—adenine, thymine, cytosine, or guanine (designated as *lavender* or *green* As, Ts, Cs, or Gs). Hydrogen bonds enable A to associate tightly with T, and C to associate tightly with G. Thus the two strands are complementary to each other. The arrows labeled 5′ to 3′ show that the strands have opposite orientation.



The information is one-dimensional because it is encoded as a specific sequence of letters along the length of the molecule. It is digital because each unit of information—one of the four letters of the DNA alphabet—is discrete. Because genetic information is digital, it can be stored as readily in a computer memory as in a DNA molecule. Indeed, the combined power of DNA sequencers, computers, and DNA synthesizers makes it possible to interpret, store, replicate, and transmit genetic information

Figure 1.3 A human chromosome. Each chromosome contains hundreds to thousands of genes.



electronically from one place to another anywhere on the planet.

The DNA regions that encode proteins are called *genes.* Just as the limited number of letters in a written alphabet places no restrictions on the stories one can tell, so too the limited number of letters in the genetic code alphabet places no restrictions on the kinds of proteins and thus the kinds of organisms genetic information can define.

Within the cells of an organism, DNA molecules carrying the genes are assembled into *chromosomes:* organized structures containing DNA and proteins that package and manage the storage, duplication, expression, and evolution of DNA (**Fig. 1.3**). The DNA within the entire collection of chromosomes in each cell of an organism is its *genome.* Human cells, for example, contain 24 distinct kinds of chromosomes carrying approximately 3×10^9 base pairs and roughly 25,000 genes. The amount of information that can be encoded in this size genome is equivalent to 6 million pages of text containing 250 words per page, with each letter corresponding to one *base pair,* or pair of nucleotides.

To appreciate the long journey from a finite amount of genetic information easily storable on a computer disk to the production of a human being, we must examine proteins, the primary molecules that determine how complex systems of cells, tissues, and organisms function.

essential concepts

- DNA, a double-stranded macromolecule composed of four nucleotides, is the repository of genetic information.
- DNA is organized into chromosomes (of 24 different types in humans) that collectively constitute an organism's genome.
- The human genome contains about 25,000 genes, most of which encode proteins.

1.2 Proteins: The Functional Molecules of Life Processes

learning objectives

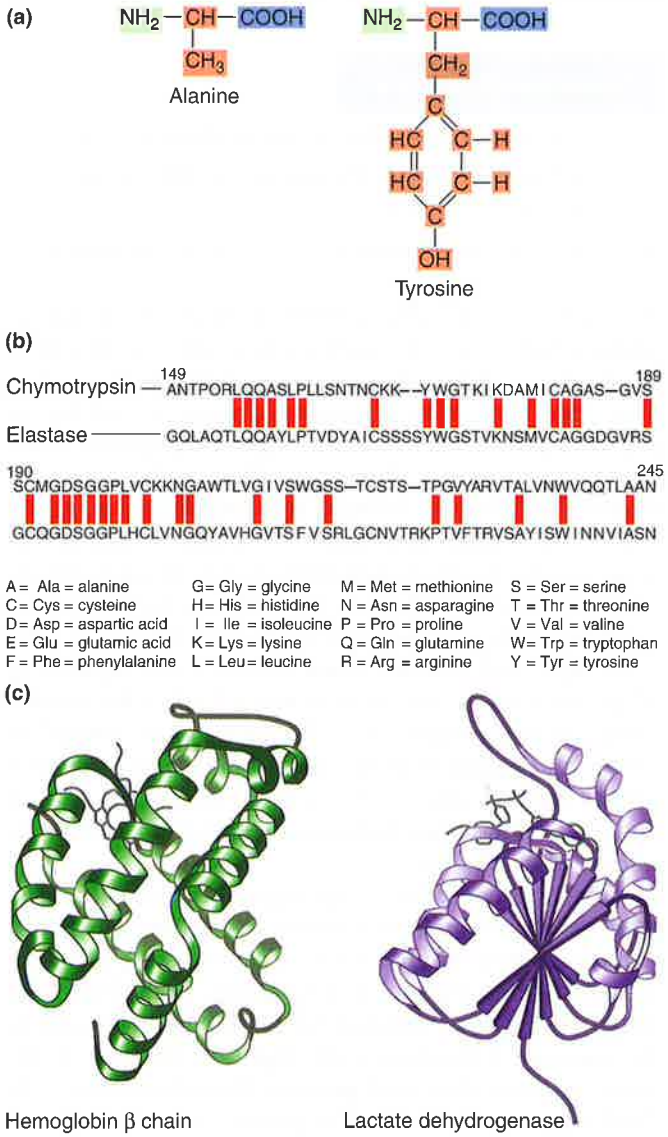
1. Compare the chemical structures of DNA and proteins.
2. Differentiate between the functions of DNA and the functions of proteins.

Although no single characteristic distinguishes living organisms from inanimate matter, you would have little trouble deciding which entities in a group of 20 objects are alive. Over time, these living organisms, governed by the laws of physics and chemistry as well as a genetic program, would be able to reproduce themselves. Most of the organisms would also have an elaborate and complicated structure that would change over time—sometimes drastically, as when an insect larva metamorphoses into an adult. Yet another characteristic of life is the ability to move. Animals swim, fly, walk, or run, while plants grow toward or away from light. Still another characteristic is the capacity to adapt selectively to the environment. Finally, a key characteristic of living organisms is the ability to use sources of energy and matter to grow—that is, the ability to convert foreign material into their own body parts. The chemical and physical reactions that carry out these conversions are known as *metabolism.*

Most properties of living organisms ultimately arise from the class of molecules known as *proteins*—large polymers composed of hundreds to thousands of amino acid subunits strung together in long chains. Each chain folds into a specific three-dimensional conformation dictated by the sequence of its amino acids (**Fig. 1.4**). There are 20 different amino acids in most proteins. The information in the DNA of genes dictates, via a genetic code, the order of amino acids in a protein molecule.

You can think of proteins as constructed from a set of 20 different kinds of snap beads distinguished by color and shape; if you were to arrange the beads in any order, make strings of a thousand beads each, and then fold or twist the chains into shapes dictated by the order of their beads, you would be able to make a nearly infinite number of different three-dimensional shapes. The astonishing diversity

Figure 1.4 Proteins are polymers of amino acids that fold in three dimensions. The specific sequence of amino acids in a chain determines the precise three-dimensional shape of the protein. **(a)** Structural formulas for two amino acids: alanine and tyrosine. All amino acids have a basic amino group (–NH) at one end and an acidic carboxyl group (–COOH) at the other. The specific side chain determines the amino acid's chemical properties. **(b)** A comparison of equivalent segments in the chains of two digestive proteins, chymotrypsin and elastase. The *red lines* connect sites in the two sequences that carry identical amino acids; the two chains differ at all the other sites shown. **(c)** Schematic drawings of the hemoglobin β chain (*green*) and lactate dehydrogenase (*purple*) show the different three-dimensional shapes determined by different amino acid sequences.



of three-dimensional protein structure generates the extraordinary diversity of protein function that is the basis of each organism's complex and adaptive behavior. The structure and shape of the hemoglobin protein, for example, allow it to transport oxygen in the bloodstream and release it to the tissues. The proteins myosin and actin can slide together to allow muscle contraction. Chymotrypsin and elastase are enzymes that help break down other proteins. Most of the properties associated with life emerge from the constellation of protein molecules that an organism synthesizes according to instructions contained in its DNA.

essential concepts

- Proteins are responsible for most biological functions of cells and organisms.
- A protein is a macromolecule consisting of amino acids linked in a linear sequence.
- The sequences of amino acids in proteins are encoded by genes within the DNA.

1.3 Molecular Similarities of All Life-Forms

learning objective

1. Summarize the molecular evidence for the common origin of living organisms.

The evolution of biological information is a fascinating story spanning the 4 billion years of earth's history. Many biologists think that RNA was the first information-processing molecule to appear. Very similar to DNA, RNA molecules are also composed of four subunits: the bases G, C, A, and U (for uracil, which replaces the T of DNA). Like DNA, RNA has the capacity to store, replicate, mutate, and express information; like proteins, RNA can fold in three dimensions to produce molecules capable of catalyzing the chemistry of life. In fact, you will learn that the ultimate purpose of some genes is to encode RNA molecules instead of proteins. RNA molecules, however, are intrinsically unstable. Thus, it is probable that the more stable DNA took over the linear information storage and replication functions of RNA, while proteins, with their far greater capacity for diversity, preempted in large part the functions derived from RNA's three-dimensional folding. With this division of labor, RNA became primarily an intermediary in converting the information in DNA into the sequence of amino acids in protein (**Fig. 1.5a**). The separation that