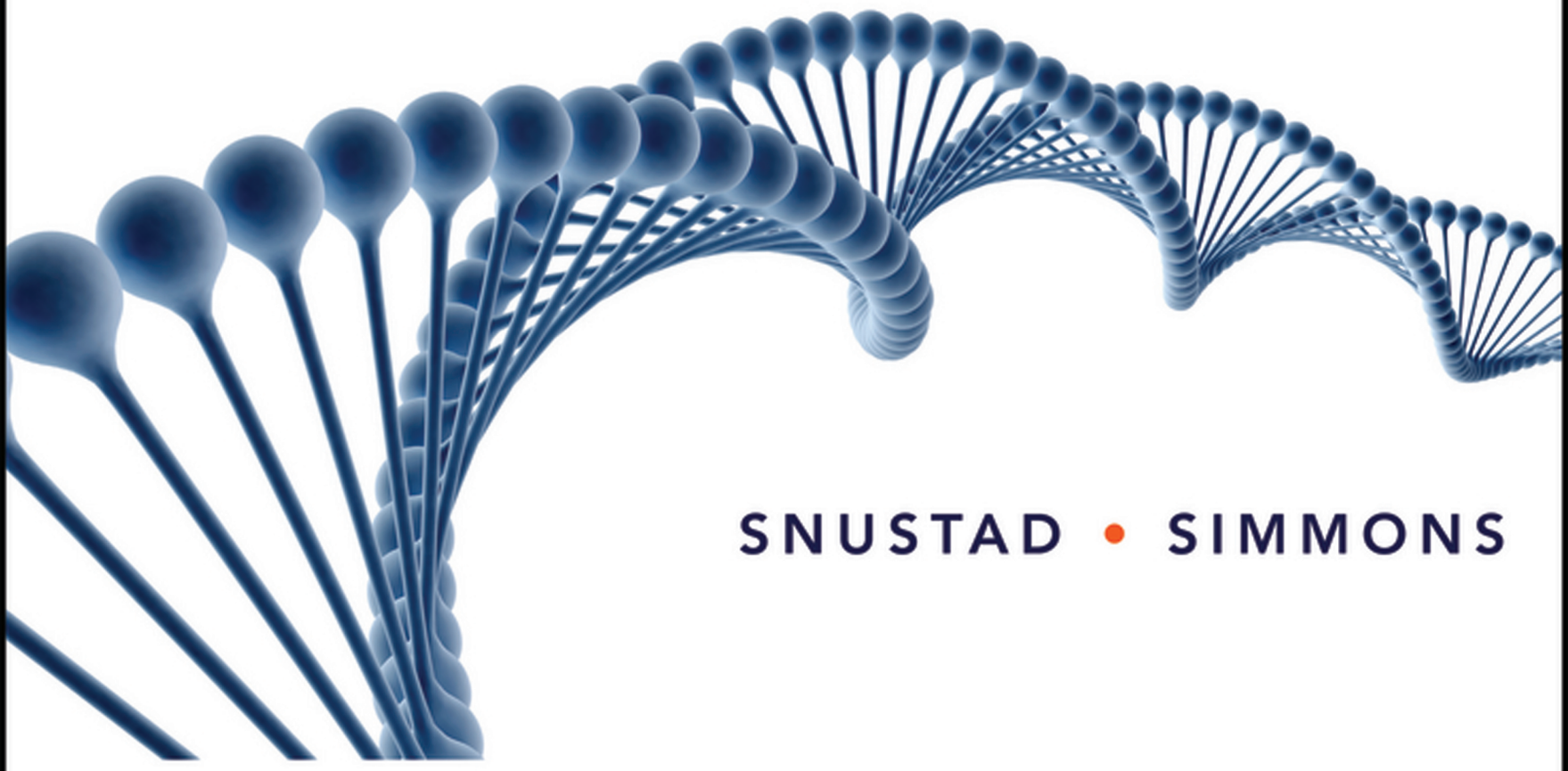
PRINCIPLES OF
GENETICS

SEVENTH EDITION



SNUSTAD • SIMMONS

Wiley Binder Version

WILEY

Principles *of* GENETICS

SEVENTH EDITION

D. Peter Snustad
University of Minnesota

Michael J. Simmons
University of Minnesota

WILEY

VICE PRESIDENT & DIRECTOR — Petra Recter
DIRECTOR — Kevin Witt
SENIOR ACQUISITIONS EDITOR — Bonnie Roth
EXECUTIVE MARKETING MANAGER — Clay Stone
PRODUCT DESIGNER — Melissa Edwards
PROJECT MANAGER — Gladys Soto
PROJECT SPECIALIST — Marcus Van Harpen
MARKETING SOLUTIONS ASSISTANT — Carolyn Thompson
SENIOR CONTENT MANAGER — Ellinor Wagner
PRODUCTION EDITOR — Swathi Chandrasekar
PHOTO EDITOR — Mary Ann Price
COVER PHOTO CREDIT — © Jezper/Shutterstock

This book was set in 10/12 JansonText by SPi Global and printed and bound by Quad Graphics Versailles.

This book is printed on acid free paper. ∞

Founded in 1807, John Wiley & Sons, Inc. has been a valued source of knowledge and understanding for more than 200 years, helping people around the world meet their needs and fulfill their aspirations. Our company is built on a foundation of principles that include responsibility to the communities we serve and where we live and work. In 2008, we launched a Corporate Citizenship Initiative, a global effort to address the environmental, social, economic, and ethical challenges we face in our business. Among the issues we are addressing are carbon impact, paper specifications and procurement, ethical conduct within our business and among our vendors, and community and charitable support. For more information, please visit our website: *www.wiley.com/go/citizenship*.

ISBN: 9781119142287 (BRV)
ISBN: 9781119232605 (EVALC)

Library of Congress Cataloging-in-Publication Data

Snustad, D. Peter, author.
Principles of genetics / D. Peter Snustad, Michael J. Simmons. —Seventh edition.
p. ; cm.
Includes index.
ISBN 978-1-119-14228-7 (looseleaf)
I. Simmons, Michael J., author. II. Title.
[DNLM: 1. Genetics. 2. Genetic Phenomena. QU 450]
QH430
576.5—dc23
2015025356

The inside back cover will contain printing identification and country of origin if omitted from this page. In addition, if the ISBN on the back cover differs from the ISBN on this page, the one on the back cover is correct.

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Dedication

To the memory of D. Peter Snustad, who skillfully guided this book through so many editions.

About the Authors

D. Peter Snustad received his B.S. degree from the University of Minnesota and his M.S. and Ph.D. degrees from the University of California, Davis. He began his faculty career in the Department of Agronomy and Plant Genetics at Minnesota in 1965, became a charter member of the new Department of Genetics in 1966, and moved to the Department of Plant Biology in 2000. During his 43 years at Minnesota, he taught courses ranging from general biology to biochemical genetics. His initial research focused on the interactions between bacteriophage T4 and its host, *E. coli*. In the 1980s, his research switched to the cytoskeleton of *Arabidopsis* and the glutamine synthetase genes of corn. His honors include the Morse-Amoco and Dagley Memorial teaching awards and election to Fellow of the American Association for the Advancement of Science.

Michael J. Simmons received his B.A. degree in biology from St. Vincent College in Latrobe, Pennsylvania, and his M.S. and Ph.D. degrees in genetics from the University of Wisconsin, Madison. As a member of the Department of Genetics, Cell Biology and Development at the University of Minnesota, Dr. Simmons taught a variety of courses, including genetics and population genetics. Early in his career he received the Morse-Amoco teaching award from the University of Minnesota in recognition of his contributions to undergraduate education. Dr. Simmons's research focuses on the genetic significance of transposable elements in the genome of *Drosophila melanogaster*. He has served on advisory committees at the National Institutes of Health and was a member of the Editorial Board of the journal *Genetics* for 21 years.

Preface

The science of genetics has been evolving rapidly. The DNA of genomes, even large ones, can now be analyzed in great detail; the functions of individual genes can be studied with an impressive array of techniques; and organisms can be changed genetically by introducing alien or altered genes into their genomes. The ways of teaching and learning genetics have also been changing. Electronic devices to access and transmit information are ubiquitous; engaging new media are being developed; and in many colleges and universities, classrooms are being redesigned to incorporate "active learning" strategies. This edition of *Principles of Genetics* has been created to recognize these scientific and educational advances.

Goals

Principles of Genetics balances new information with foundational material. In preparing this edition, we have been guided by four main goals:

- **To focus on the basic principles of genetics** by presenting the important concepts of classical, molecular, and population genetics carefully and thoroughly. We believe that an understanding of current advances in genetics and an appreciation for their practical significance must be based on a strong foundation. Furthermore, we believe that the breadth and depth of coverage in the different areas of genetics—classical, molecular, and population—must be balanced, and that the ever-growing mass of information in genetics must be organized by a sturdy—but flexible—framework of key concepts.
- **To focus on the scientific process** by showing how scientific concepts develop from observation and experimentation. Our book provides numerous examples to show how genetic principles have emerged from the work of different scientists. We emphasize that science is an ongoing process of observation, experimentation, and discovery.
- **To focus on human genetics** by incorporating human examples and showing the relevance of genetics to societal issues. Experience has shown us that students are keenly interested in the genetics of their own species. Because of this interest, they find it easier to comprehend complex concepts when these concepts are illustrated with human examples. Consequently, we have used human examples to illustrate genetic principles wherever possible. We have also included discussions of the Human Genome Project, human gene mapping, genetic disorders, gene therapy, and genetic counseling throughout the text. Issues such as genetic screening, DNA profiling, genetic engineering, cloning, stem cell research, and gene therapy have sparked vigorous debates about the social, legal, and ethical ramifications of genetics. We believe that it is important to involve students in discussions about these issues, and we hope that this textbook will provide students with the background to engage in such discussions thoughtfully.
- **To focus on developing critical thinking skills** by emphasizing the analysis of experimental data and problems. Genetics has always been a bit different from other disciplines in biology because of its heavy emphasis on problem solving. In this text, we have fleshed out the analytical nature of genetics in many ways—in the development of principles in classical genetics, in the discussion of experiments in molecular genetics, and in the presentation of calculations in population genetics. Throughout the book we have emphasized the integration of observational and experimental evidence with logical analysis in the development of key

concepts. Each chapter has two sets of worked-out problems—the *Basic Exercises* section, which contains simple problems that illustrate basic genetic analysis, and the *Testing Your Knowledge* section, which contains more complex problems that integrate different concepts and techniques. A set of *Questions and Problems* follows the worked-out problems so that students can enhance their understanding of the concepts in the chapter and develop their analytical skills. Another section, *Genomics on the Web*, poses issues that can be investigated by going to the National Center for Biotechnology Information web site. In this section, students can learn how to use the vast repository of genetic information that is accessible via that web site, and they can apply that information to specific problems. Each chapter also has a *Problem-Solving Skills* feature, which poses a problem, lists the pertinent facts and concepts, and then analyzes the problem and presents a solution. Each chapter also has two examples of another feature, *Solve It*, to provide students with opportunities to test their understanding of concepts as they encounter them in the text. Step-by-step explanations of the answers to the Solve It problems are presented on the book's web site, some in video format.

Content and Organization of the Seventh Edition

The organization of this edition of *Principles of Genetics* is similar to that of the previous edition. However, the content has been sifted and winnowed to allow thoughtful updating. In selecting material to be included in this edition of *Principles of Genetics*, we have tried to be comprehensive but not encyclopedic.

The printed text comprises 20 chapters. Four more chapters can be found on the companion website and within WileyPLUS; we have moved these chapters online to create a slimmer, more compact book that is suitable for most courses in genetics. Chapters 1–2 introduce the science of genetics, basic features of cellular reproduction, and some of the model genetic organisms; Chapters 3–8 present the concepts of classical genetics and the basic procedures for the genetic analysis of microorganisms; Chapters 9–13 present the topics of molecular genetics, including DNA replication, transcription, translation, and mutation; Chapters 14–16 cover more advanced topics in molecular genetics and genomics; Chapters 17 and 18 deal with the regulation of gene expression, and Chapters 19 and 20 present the concepts of quantitative and population genetics. Chapters 21–24, which are on the companion website and within WileyPLUS, deal with the genetics of transposable elements, animal development, cancer, and evolution.

As in previous editions, we have tried to create a text that can be adapted to different course formats. Many instructors prefer to present the topics in much the same way as we have, starting with classical genetics, progressing into molecular genetics, and finishing with quantitative and population genetics. However this text is constructed so that teachers can present topics in different orders. They may, for example, begin with basic molecular genetics (Chapters 9–13), then present classical genetics (Chapters 3–8), progress to more advanced topics in molecular genetics (Chapters 14–18), and finish the course with quantitative and population genetics (Chapters 19 and 20). Alternatively, they may wish to insert quantitative and population genetics between classical and molecular genetics.

Pedagogy of the Seventh Edition

The text includes special features designed to emphasize the relevance of the topics discussed, to facilitate the comprehension of important concepts, and to assist students in evaluating their grasp of these concepts.

- ***Chapter-Opening Vignette.*** Each chapter opens with a brief story that highlights the significance of the topics discussed in the chapter.
- ***Chapter Outline.*** The main sections of each chapter are conveniently listed on the chapter's first page.

- ***Section Summary.*** The content of each major section of text is briefly summarized at the beginning of that section. These opening summaries focus attention on the main ideas developed in a chapter.
- ***Key Points.*** These learning aids appear at the end of each major section in a chapter. They are designed to help students review for exams and to recapitulate the main ideas of the chapter.
- ***Problem-Solving Skills Boxes.*** Each chapter contains a box that guides the student through the analysis and solution of a representative problem. We have chosen a problem that involves important material in the chapter. The box lists the facts and concepts that are relevant to the problem, and then explains how to obtain the solution. Ramifications of the problem and its analysis are discussed in the Student Companion site.
- ***Solve It Boxes.*** Each of these boxes poses a problem related to concepts students encounter as they read the text. The step-by-step solution to each of the problems is presented in the Student Companion site and within WileyPLUS, and for selected problems, it is presented in video format. The two Solve It boxes in each chapter allow students to test their understanding of key concepts.
- ***Basic Exercises.*** At the end of each chapter we present several worked-out problems to reinforce each of the fundamental concepts developed in the chapter. These simple, one-step exercises are designed to illustrate basic genetic analysis or to emphasize important information.
- ***Testing Your Knowledge.*** Each chapter also has more complicated worked-out problems to help students hone their analytical and problem-solving skills. The problems in this section are designed to integrate different concepts and techniques. In the analysis of each problem, we walk the students through the solution step by step.
- ***Questions and Problems.*** Each chapter ends with a set of questions and problems of varying difficulty organized according to the sequence of topics in the chapter. The more difficult questions and problems have been designated with colored numbers. These sets of questions and problems provide students with the opportunity to enhance their understanding of the concepts covered in the chapter and to develop their analytical skills. Also, some of the questions and problems—called GO problems—have been selected for interactive solutions on the Student Companion site and within WileyPLUS. The GO problems are designated with a special icon.
- ***Genomics on the Web.*** Information about genomes, genes, DNA sequences, mutant organisms, polypeptide sequences, biochemical pathways, and evolutionary relationships is now freely available on an assortment of web sites. Researchers routinely access this information, and we believe that students should become familiar with it. To this end, we have incorporated a set of questions at the end of each chapter that can be answered by using the National Center for Biotechnology Information (NCBI) web site, which is sponsored by the U. S. National Institutes of Health.
- ***Appendices.*** These features, found on the Student Companion site, present technical material that is useful in genetic analysis.
- ***Glossary.*** This section of the book defines important terms. Students find it useful in clarifying topics and in preparing for exams.
- ***Answers.*** Answers to the odd-numbered Questions and Problems are given at the end of the text.

ONLINE RESOURCES

WileyPLUS

WileyPLUS is a research-based online environment for effective teaching and learning.

WileyPLUS builds students' confidence because it takes the guesswork out of studying by providing students with a clear roadmap: **what to do, how to do it, if they did it right**. This interactive approach focuses on the following:

Confidence: Research shows that students experience a great deal of anxiety over studying. That's why we provide a structured learning environment that helps students focus on **what to do**, along with the support of immediate resources.

Motivation: To increase and sustain motivation throughout the semester, WileyPLUS helps students learn **how to do it** at a pace that's right for them. Our integrated resources—available 24/7—function like a personal tutor, directly addressing each student's demonstrated needs with specific problem-solving techniques.

Success: WileyPLUS helps to assure that each study session has a positive outcome by putting students in control. Through instant feedback and study objective reports, students know **if they did it right** and where to focus next, so they achieve the strongest result.

With WileyPLUS, our efficacy research shows that students improve their outcomes by as much as one letter grade. WileyPLUS helps students take more initiative, so you'll have greater impact on their achievement in the classroom and beyond.

What do students receive with WileyPLUS?

- The complete digital textbook, saving students up to 60% off the cost of a printed text.
- Interactive problem sets with question assistance, including links to relevant sections in the online digital textbook.
- Immediate feedback and proof of progress, 24/7.
- Integrated, multimedia resources—including animations, video solutions, GO tutorial problems, and much more—that provide multiple study paths and encourage more active learning.

What do instructors receive with WileyPLUS?

- Reliable resources that reinforce course goals inside and outside the classroom.
- The ability to easily identify those students who are falling behind.
- Media-rich course materials and assessment content including—Instructor's Manual, Test Bank, PowerPoint® Slides, Learning Objectives, Solutions Manual, Study Guide, Computerized Test Bank, Pre- and Post-Lecture Quizzes, and much more.

www. wileyplus.com

TEST BANK

The test bank is available on both the Instructor Companion site and within WileyPLUS. The test bank contains approximately 50 test questions per chapter. It is available online as MS Word files and as a computerized test bank. This easy-to-use test-generation program fully supports graphics, print tests, student answer sheets, and answer keys. The software's advanced features allow you to produce an exam to your exact specifications.

LECTURE POWERPOINT PRESENTATIONS

Highly visual lecture PowerPoint presentations are available for each chapter and help convey key concepts illustrated by imbedded text art. The presentations may be accessed on the Instructor Companion site and within WileyPLUS.

PRE- AND POST-LECTURE ASSESSMENT

This assessment tool allows instructors to assign a quiz prior to lecture to assess student understanding and encourage reading, and following lecture to gauge improvement and weak areas. Two quizzes are provided for every chapter.

PERSONAL RESPONSE SYSTEM QUESTIONS

These questions are designed to provide readymade pop quizzes and to foster student discussion and debate in class. Available on the Instructor Companion site and within WileyPLUS.

PRACTICE QUIZZES

Available on the Student Companion site and within WileyPLUS, these quizzes contain 20 questions per chapter for students to quiz themselves and receive instant feedback.

MILESTONES IN GENETICS

The *Milestones* are available on the Student Companion site and within WileyPLUS. Each of them explores a key development in genetics—usually an experiment or a discovery. We cite the original papers that pertain to the subject of the *Milestone*, and we include two *Questions for Discussion* to provide students with an opportunity to investigate the current significance of the subject. These questions are suitable for cooperative learning activities in the classroom, or for reflective writing exercises that go beyond the technical aspects of genetic analysis.

FOCUS ON

Special topics are presented in separate *Focus On* features on the Student Companion site and within WileyPLUS. The material in these features supports or develops concepts, techniques, or skills that have been introduced in the printed text.

SOLVE IT

Solve It boxes provide students with opportunities to test their understanding of concepts as they encounter them in the text. Each chapter poses two Solve It problems; step-by-step explanations of the answers are presented on the book's web site and within WileyPLUS, some in video format. Students can view Camtasia videos, prepared by Dubear Kroening at the University of Wisconsin-Fox Valley. These tutorials enhance interactivity and hone problem-solving skills to give students the confidence they need to tackle complex problems in genetics.

ANIMATIONS

Located within WileyPLUS, these animations illustrate key concepts from the text and aid students in grasping some of the most difficult concepts in genetics. Also included are animations that will give students a refresher in basic biology.

ANSWERS TO QUESTIONS AND PROBLEMS

Answers to odd-numbered Questions and Problems are located at the end of the text for easy access for students. Answers to all Questions and Problems in the text are available only to instructors on the Instructor Companion site and within WileyPLUS.

ILLUSTRATIONS AND PHOTOS

All line illustrations and photos from *Principles of Genetics*, 7th *Edition*, are available on the Instructor Companion site and within WileyPLUS in both jpeg files and PowerPoint format. Line illustrations are enhanced to provide the best presentation experience.

BOOK COMPANION WEB SITE

(www.wiley.com/college/snustad)
This text-specific web site provides students with additional resources and extends the chapters of the text to the resources of the World Wide Web. Resources include:

- **For Students:** practice quizzes covering key concepts for each chapter of the text, flashcards, and the Biology NewsFinder.
- **For Instructors:** Test Bank, PowerPoint Presentations, line art and photos in jpeg and PowerPoint formats, personal response system questions, and all answers to end-of-chapter Questions and Problems.

Acknowledgments

As with previous editions, this edition of *Principles of Genetics* has been influenced by the genetics courses we teach. We thank our students for their constructive feedback on both content and pedagogy, and we thank our colleagues at the University of Minnesota for sharing their knowledge and expertise. Genetics professors at other institutions also provided many helpful suggestions. In particular, we acknowledge the help of the following reviewers:

7TH EDITION REVIEWERS

Gregory C. Booton, The Ohio State University; Kathleen Fitzpatrick, Simon Fraser University; David W. Foltz, Louisiana State University; Elliott S. Goldstein, Arizona State University; Andrew Zelhof, Indiana University; Jianzhi Zhang, University of Michigan

REVIEWERS OF PREVIOUS EDITIONS

Ann Aguano, Manhattan Marymount College; Mary A. Bedell, University of Georgia; Michelle Boissere, Xavier University of Louisiana; Stephen P. Bush, Coastal Carolina University; Jonathan Clark, Weber State University; Sarah Crawford, Southern Connecticut State University; Robert Fowler, San Jose State University; Cheryl Hertz, Loyola Marymount University; Shawn Kaeppler, University of Wisconsin; Todd Kelson, Brigham Young University – Idaho; Xiongbin Lu, University of South Carolina – Columbia; Richard D. Noyes, University of Central Arkansas; Maria E. Orive, University of Kansas; Rongsun Pu, Kean University Valery N. Soyfer, George Mason University; David Starkey, University of Central Arkansas; Frans Tax, University of Arizona; Tzvi Tzfira, University of Michigan; Harald Vaessin, The Ohio State University – Columbus; Sarah VanVickle-Chavez, Washington University in St. Louis; Willem Vermerris, University of Florida; Alan S. Waldman, University of South Carolina – Columbia

Many people contributed to the development and production of this edition. Kevin Witt, Director, and Bonnie Roth, Senior Acquisitions Editor, initiated the project, and the inaugural editorial team of Marian Provenzano, Brian Baker, and Christina Volpe helped to get it underway. Gladys Soto, Production Manager, Marcus Van Harpen, Project Specialist, Janet Wehner, Development Editor, Swathi Chandrasekar, Production Editor, and Carolyn Thompson, Editorial Assistant, worked diligently and thoughtfully to bring the project to completion. Along the way, they enlisted the efforts of SPi-Global who did the copyediting, proofreading, and prepared the index. Mary Ann Price, Senior Photo Editor, obtained several new images for this edition, Elizabeth Swain, Production Editor, helped by providing archived material from previous editions, and Clay Stone, Executive Marketing Manager, developed a plan to get this edition into the hands of readers. Many thanks to all these people for their ideas and their help.

D. Peter Snustad, the lead author of *Principles of Genetics* for so many years, was too ill to contribute directly to this edition; he passed away while it was being written. However, the book still contains much that is Pete's—carefully researched content, thoughtfully designed illustrations, and intriguing questions and problems that could only have been crafted by an accomplished geneticist and esteemed teacher. There is no doubt that the richness of Pete's legacy will continue to be appreciated by all who use this textbook.

With an eye toward the next edition, students, teaching assistants, instructors, and other readers may send comments on this edition to John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ, 07030.

Contents

CHAPTER 7

CHAPTER 8

CHAPTER 9

CHAPTER 12

Translation and the Genetic Code 280

CHAPTER 13

Mutation, DNA Repair, and Recombination 313

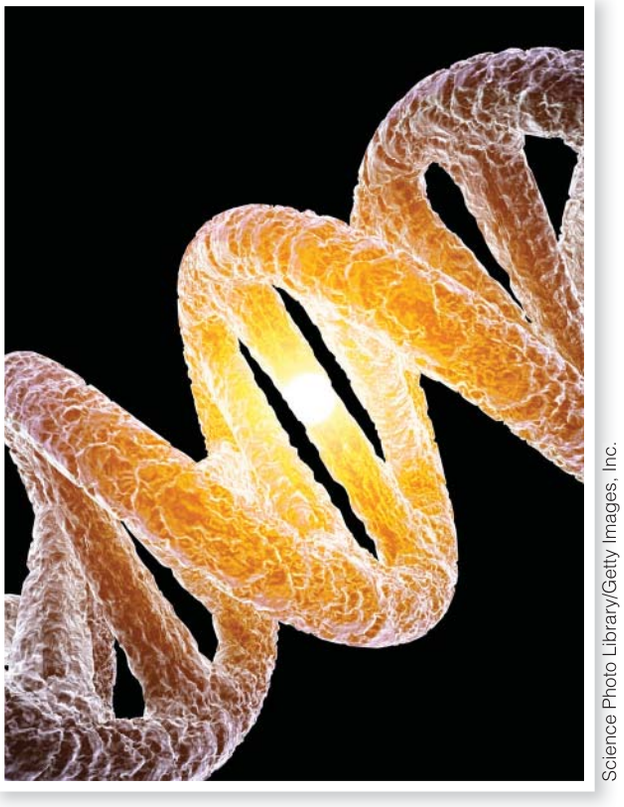
The Science of Genetics

1

CHAPTER OUTLINE

The Personal Genome

Each of us is composed of trillions of cells, and each of those cells contains very thin fibers a few centimeters long that play a major role in who we are, as human beings and as persons. These all-important intracellular fibers are made of DNA. Every time a cell divides, its DNA is replicated and apportioned equally to two daughter cells. The DNA content of these cells—what we call the genome—is thereby conserved. This genome is a master set of instructions, in fact a whole library of information, that cells use to maintain the living state. Ultimately, all the activities of a cell depend on it. To know the DNA is therefore to know the cell, and, in a larger sense, to know the organism to which that cell belongs.



Science Photo Library/Getty Images, Inc.

Computer artwork of deoxyribonucleic acid (DNA).

Given the importance of the DNA, it should come as no surprise that great efforts have been expended to study it, down to the finest details. In fact, in the last decade of the twentieth century a worldwide campaign, the Human Genome Project, took shape, and in 2001 it produced a comprehensive analysis of human DNA samples that had been collected from a small number of anonymous donors. This work—stunning in scope and significance—laid the foundation for all future research on the human genome. Then, in 2007, the analysis of human DNA took a new turn. Two of the architects of the Human Genome Project had their own DNA decoded. The technology for analyzing complete genomes has advanced significantly, and the cost for this analysis is no longer exorbitant. In fact, it may soon be possible for each of us to have our own genome analyzed—a prospect that is sure to influence our lives and change how we think about ourselves.

An Invitation

This book is about genetics, the science that deals with DNA. Genetics is also one of the sciences that has a profound impact on us. Through applications in agriculture and medicine, it helps to feed us and keep us healthy. It also provides insight into what makes us human and into what distinguishes each of us as individuals. Genetics is a relatively young science—it emerged only at the beginning of the twentieth century, but it has grown in scope and significance, so much so that it now has a prominent, and some would say commanding, position in all of biology.

Genetics began with the study of how the characteristics of organisms are passed from parents to offspring—that is, how they are inherited. Until the middle of the twentieth century, no one knew for sure what the hereditary material was. However, geneticists recognized that this material had to fulfill three requirements. First, it had to replicate so that copies could be transmitted from parents to offspring. Second, it had to encode information to guide the development, functioning, and behavior of cells and organisms to which they belong. Third, it had to change, even if only once in a great while, to account for the differences that exist among individuals. For several decades, geneticists wondered what the hereditary material could be. Then in 1953 the structure of DNA was elucidated and genetics had its great clarifying moment. In a relatively short time, researchers discovered how DNA functions as the hereditary material—that is, how it replicates, how it encodes and expresses information, and how it changes. These discoveries ushered in a new phase of genetics in which phenomena could be explained at the molecular level. In time, geneticists learned how to analyze the DNA of whole genomes, including our own. This progress—from studies of heredity to studies of whole genomes—has been amazing.

As experienced geneticists and as teachers, we have written this book to explain the science of genetics to you. As its title indicates, this book is designed to convey the principles of genetics, and to do so in sufficient detail for you to understand them clearly. We invite you to read each chapter, to study its illustrations, and to wrestle with the questions and problems at the end of the chapter. We all know that learning—and research, teaching, and writing too—takes effort. As authors, we hope your effort studying this book will be rewarded with a good understanding of genetics.

This introductory chapter provides an overview of what we will explain in more detail in the chapters to come. For some of you, it will be a review of knowledge gained from studying basic biology and chemistry. For others, it will be new fare. Our advice is to read the chapter without dwelling on the details. The emphasis here is on the grand themes that run through genetics. The many details of genetics theory and practice will come later.

Three Great Milestones in Genetics

Genetics is rooted in the research of Gregor Mendel, a monk who discovered how traits are inherited. The molecular basis of heredity was revealed when James Watson and Francis Crick elucidated the structure of DNA. The Human Genome Project is currently engaged in the detailed analysis of human DNA.

Scientific knowledge and understanding usually advance incrementally. In this book we will examine the advances that have occurred in genetics during its short history—barely a hundred years. Three great milestones stand out in this history: (1) the discovery of rules governing the inheritance of traits in organisms, (2) the identification of the material responsible for this inheritance and the elucidation of its structure, and (3) the comprehensive analysis of the hereditary material in human beings and other organisms.

MENDEL: GENES AND THE RULES OF INHERITANCE

Although genetics developed during the twentieth century, its origin is rooted in the work of *Gregor Mendel* (■ **Figure 1.1**), a Moravian monk who lived in the nineteenth century.

Mendel carried out his path-breaking research in relative obscurity. He studied the inheritance of different traits in peas, which he grew in the monastery garden. His method involved interbreeding plants that showed different traits—for example, short plants were bred with tall plants—to see how the traits were inherited by the offspring. Mendel's careful analysis enabled him to discern patterns, which led him to postulate the existence of hereditary factors responsible for the traits he studied. We now call these factors **genes.**

Mendel studied several genes in the garden pea. Each of the genes was associated with a different trait—for example, plant height, or flower color, or seed texture. He discovered that these genes exist in different forms, which we now call **alleles.** One form of the gene for height, for example, allows pea plants to grow more than 2 meters tall; another form of this gene limits their growth to about half a meter.

Mendel proposed that pea plants carry two copies of each gene. These copies may be the same or different. During reproduction, one of the copies is randomly incorporated into each sex cell or gamete. The female gametes (eggs) unite with the male gametes (sperm) at fertilization to produce single cells, called zygotes, which then develop into new plants. The reduction in gene copies from two to one during gamete formation and the subsequent restoration of two copies during fertilization underlie the rules of inheritance that Mendel discovered.

Mendel emphasized that the hereditary factors—that is, the genes—are discrete entities. Different alleles of a gene can be brought together in the same plant through hybridization and can then be separated from each other during the production of gametes. The coexistence of alleles in a plant therefore does not compromise their integrity. Mendel also found that alleles of different genes are inherited independently of each other.

James King-Holmes/Photo Researchers, Inc.

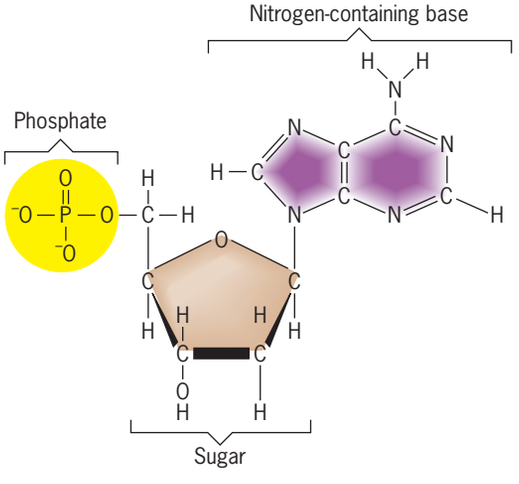
■ **FIGURE 1.1** Gregor Mendel.

These discoveries were published in 1866 in the proceedings of the Natural History Society of Brünn, the journal of the scientific society in the city where Mendel lived and worked. The article was not much noticed, and Mendel went on to do other things. In 1900, 16 years after he died, the paper finally came to light, and the science of genetics was born. In short order, the type of analysis that Mendel pioneered was applied to many kinds of organisms, and with notable success. Of course, not every result fit exactly with Mendel's principles. Exceptions were encountered, and when they were investigated more fully, new insights into the behavior and properties of genes emerged. We will delve into Mendel's research and its applications to the study of inheritance, including heredity in humans, in Chapter 3, and we will explore some ramifications of Mendel's ideas in Chapter 4. In Chapters 5–7 we will see how Mendel's principles of inheritance are related to the behavior of chromosomes—the cellular structures where genes reside.

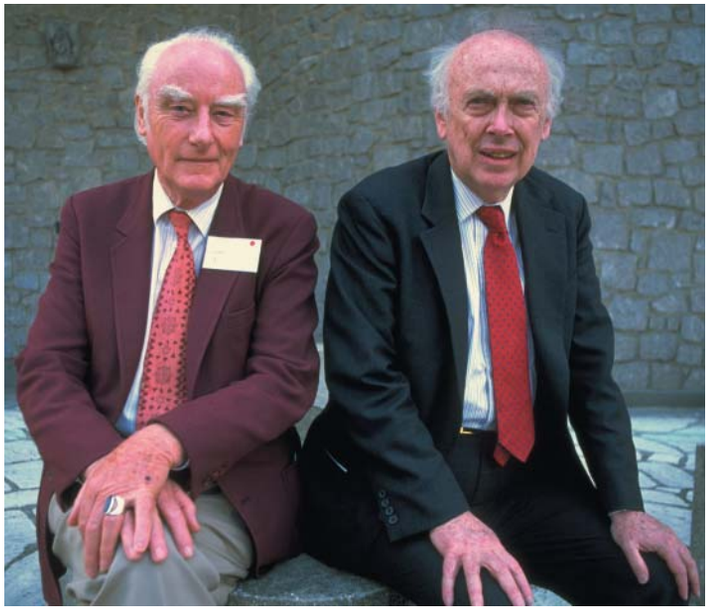
WATSON AND CRICK: THE STRUCTURE OF DNA

The rediscovery of Mendel's paper launched a plethora of studies on inheritance in plants, animals, and microorganisms. The big question on everyone's mind was "What is a gene?" In the middle of the twentieth century, this question was finally answered. Genes were shown to consist of complex molecules called **nucleic acids.**

Nucleic acids are made of elementary building blocks called **nucleotides** (■ **Figure 1.2**). Each nucleotide has three components: (1) a sugar molecule; (2) a phosphate molecule, which has acidic chemical properties; and (3) a nitrogen-containing molecule, which has slightly basic chemical properties. In **ribonucleic acid,** or RNA, the constituent sugar is ribose; in **deoxyribonucleic acid,** or DNA, it is deoxyribose. Within RNA or DNA, one nucleotide is distinguished from another by its nitrogen-containing base. In RNA, the



■ **FIGURE 1.2** Structure of a nucleotide. The molecule has three components: a phosphate group, a sugar (in this case deoxyribose), and a nitrogen-containing base (in this case adenine).



Perrin Pierre/©Corbis.

■ **FIGURE 1.3** Francis Crick and James Watson.

four kinds of bases are adenine (A), guanine (G), cytosine (C), and uracil (U); in DNA, they are A, G, C, and thymine (T). Thus, in both DNA and RNA there are four kinds of nucleotides, and three of them are shared by both types of nucleic acid molecules.

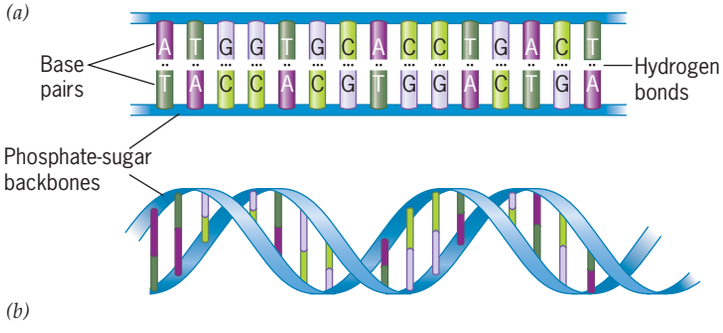
The big breakthrough in the study of nucleic acids came in 1953 when *James Watson* and *Francis Crick* (■ **Figure 1.3**) deduced how nucleotides are organized within DNA. Watson and Crick knew that the nucleotides are linked, one to another, in a chain. The linkages are formed by chemical interactions between the phosphate of one nucleotide and the sugar of another nucleotide. The nitrogen-containing bases are not involved in these interactions. Thus, a chain of nucleotides consists of a phosphate-sugar backbone to which bases are attached, one base to each sugar in the backbone. From one end of the chain to the other, the bases form a linear sequence characteristic of that particular chain. This sequence of bases is what distinguishes one gene from another. Watson and Crick proposed that DNA molecules consist of two chains of nucleotides (■ **Figure 1.4*a***). These chains are held together by weak chemical attractions—called hydrogen bonds—between particular pairs of bases; A pairs with T, and G pairs with C. Because of these base-pairing rules, the sequence of one nucleotide chain in a double-stranded DNA molecule can be predicted from that of the other. In this sense, then, the two chains of a DNA molecule are complementary.

A double-stranded DNA molecule is often called a duplex. Watson and Crick discovered that the two strands of a DNA duplex are wound around each other in a helical configuration (■ **Figure 1.4*b***). These helical molecules can be extraordinarily large. Some contain hundreds of millions of nucleotide pairs, and their end-to-end length exceeds 10 centimeters. Were it not for their extraordinary thinness (about a hundred-millionth of a centimeter), we would be able to see them with the unaided eye.

RNA, like DNA, consists of nucleotides linked one to another in a chain. However, unlike DNA, RNA molecules are usually single-stranded. The genes of most organisms are composed of DNA, although in some viruses they are made of RNA. We will examine the structures of DNA and RNA in detail in Chapter 9, and we will investigate the genetic significance of these macromolecules in Chapters 10–12.

THE HUMAN GENOME PROJECT: SEQUENCING DNA AND CATALOGING GENES

If geneticists in the first half of the twentieth century dreamed about identifying the stuff that genes are made of, geneticists in the second half of that century dreamed about ways of determining the sequence of bases in DNA molecules. Near the end of the century, their dreams became reality as projects to determine DNA base sequences in several organisms, including humans, took shape. Obtaining the sequence of bases in an organism's DNA—that is, *sequencing the DNA*—should, in principle, provide the information needed to analyze all that organism's genes. We refer to the collection of DNA molecules that is characteristic of an organism as its **genome.** Sequencing the genome is therefore tantamount to sequencing all the organism's genes—and more, for we now know that some of the DNA does not comprise genes. The function of this nongenic DNA is not always clear; however, it is present in many genomes, and sometimes it is abundant. A Milestone in Genetics: ΦX174, the First DNA Genome Sequenced describes how genome sequencing got started. You can find this account in the Student Companion site.



■ **FIGURE 1.4** DNA, a double-stranded molecule held together by hydrogen bonding between paired bases. (*a*) Two-dimensional representation of the structure of a DNA molecule composed of complementary nucleotide chains. (*b*) A DNA molecule shown as a double helix.

The paragon of all the sequencing programs is the **Human Genome Project,** a worldwide effort to determine the sequence of approximately 3 billion nucleotide pairs in human DNA. As initially conceived, the Human Genome Project was to involve collaborations among researchers in many different countries, and much of the work

was to be funded by their governments. However, a privately funded project initiated by Craig Venter, a scientist and entrepreneur, soon developed alongside the publicly funded project. In 2001 all these efforts culminated in the publication of two lengthy articles about the human genome. The articles reported that 2.7 billion nucleotide pairs of human DNA had been sequenced. Computer analysis of this DNA suggested that the human genome contained between 30,000 and 40,000 genes. More recent analyses have revised the human gene number downward, to around 20,500. These genes have been cataloged by location, structure, and potential function. Efforts are now focused on studying how they influence the myriad characteristics of humans. There is also considerable effort to assess how much one human genome differs from another—that is, how much genetic variability exists in the human species. For more information about this effort, you can read the Focus on The 1000 Genomes Project on the Student Companion site.

Broad Institute/www.genome.gov

■ **FIGURE 1.5** Researchers in a laboratory that performs DNA sequencing.

The genomes of many other organisms—bacteria, fungi, plants, protists, and animals—have also been sequenced. Much of this work has been done under the auspices of the Human Genome Project, or under projects closely allied to it. Initially the sequencing efforts were focused on organisms that are especially favorable for genetic research. In many places in this book, we explore ways in which researchers have used these model organisms to advance genetic knowledge. Current sequencing projects have moved beyond the model organisms to diverse plants, animals, and microbes. For example, the genomes of the mosquito and the malaria parasite that it carries have both been sequenced, as have the genomes of the honeybee, the poplar tree, and the sea squirt. Some of the targets of these sequencing projects have a medical, agricultural, or commercial significance; others simply help us to understand how genomes are organized and how they have diversified during the history of life on Earth.

All the DNA sequencing projects have transformed genetics in a fundamental way. Genes can now be studied at the molecular level with relative ease, and vast numbers of genes can be studied simultaneously. This approach to genetics, rooted in the analysis of the DNA sequences that make up a genome, is called **genomics.** It has been made possible by advances in DNA sequencing technology, robotics, and computer science (■ **Figure 1.5**). Researchers are now able to construct and scan enormous databases containing DNA sequences to address questions about genetics. Although there are a large number of useful databases currently available, we will focus on the databases assembled by the *National Center for Biotechnology Information (NCBI)*, maintained by the U.S. National Institutes of Health. The NCBI databases—available free on the web at http://www.ncbi.nih.gov—are invaluable repositories of information about genes, proteins, genomes, publications, and other important data in the fields of genetics, biochemistry, and molecular biology. They contain the complete nucleotide sequences of all genomes that have been sequenced to date, and they are continually updated. In addition, the NCBI web site contains tools that can be used to search for specific items of interest—gene and protein sequences, research articles, and so on. In Chapter 15, we will introduce you to some of these tools, and throughout this book, we will encourage you to visit the NCBI web site at the end of each chapter to answer specific questions.

KEY POINTS

- *Gregor Mendel postulated the existence of particulate factors—now called genes—to explain how traits are inherited.*
- *Alleles, the alternate forms of genes, account for heritable differences among individuals.*
- *James Watson and Francis Crick elucidated the structure of DNA, a macromolecule composed of two complementary chains of nucleotides.*
- *DNA is the hereditary material in all life forms except some types of viruses, in which RNA is the hereditary material.*
- *The Human Genome Project determined the sequence of nucleotides in the DNA of the human genome.*
- *Sequencing the DNA of a genome provides the data to identify and catalog all the genes of an organism.*

DNA as the Genetic Material

In biology information flows from DNA to RNA to protein.

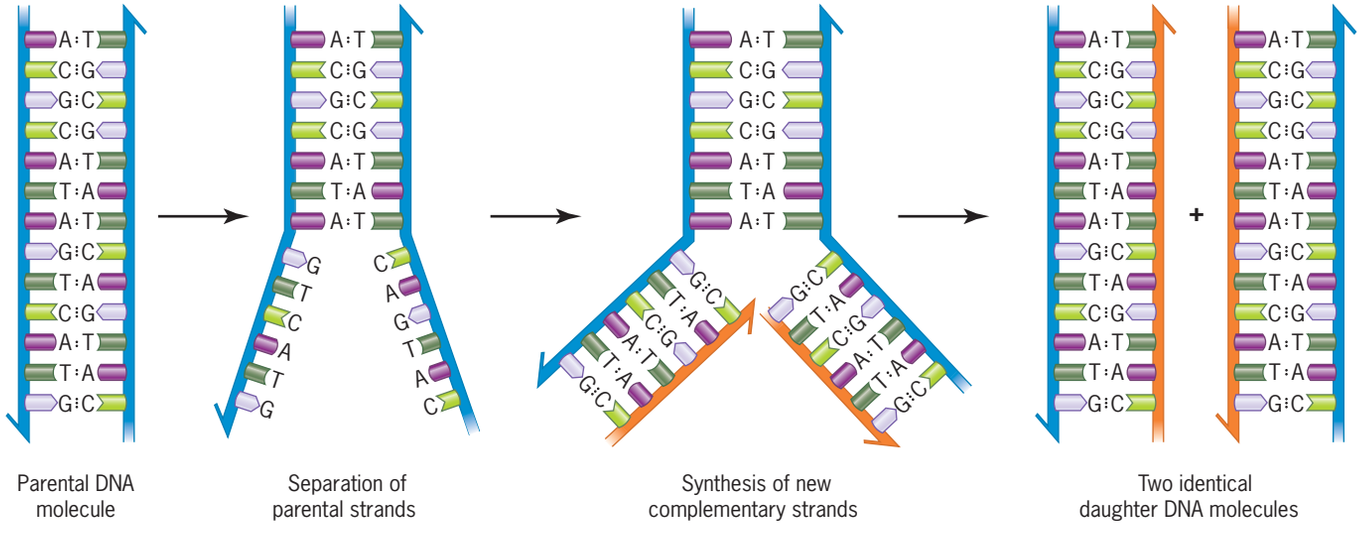
In all cellular organisms, the genetic material is DNA. This material must be able to *replicate* so that copies can be transmitted from cell to cell and from parents to offspring; it must contain *information* to direct cellular activities and to guide the development, functioning, and behavior of organisms; and it must be able to *change* so that over time, groups of organisms can adapt to different circumstances.

DNA REPLICATION: PROPAGATING GENETIC INFORMATION

The genetic material of an organism is transmitted from a mother cell to its daughters during cell division. It is also transmitted from parents to their offspring during reproduction. The faithful transmission of genetic material from one cell or organism to another is based on the ability of double-stranded DNA molecules to be replicated. DNA replication is extraordinarily exact. Molecules consisting of hundreds of millions of nucleotide pairs are duplicated with few, if any, mistakes.

The process of DNA replication is based on the complementary nature of the strands that make up duplex DNA molecules (■ **Figure 1.6**). These strands are held together by relatively weak hydrogen bonds between specific base pairs—A paired with T, and G paired with C. When these bonds are broken, the separated strands can serve as templates for the synthesis of new partner strands. The new strands are assembled by the stepwise incorporation of nucleotides opposite to nucleotides in the template strands. This incorporation conforms to the base-pairing rules. Thus, the sequence of nucleotides in a strand being synthesized is dictated by the sequence of nucleotides in the template strand. At the end of the replication process, each template strand is paired with a newly synthesized partner strand. Thus, two identical DNA duplexes are created from one original duplex.

The process of DNA replication does not occur spontaneously. Like most biochemical processes, it is catalyzed by enzymes. We will explore the details of DNA replication, including the roles played by different enzymes, in Chapter 10.



■ **FIGURE 1.6** DNA replication. The two strands in the parental molecule are oriented in opposite directions (see arrows). These strands separate and new strands are synthesized using the parental strands as templates. When replication is completed, two identical double-stranded DNA molecules are produced.

GENE EXPRESSION: USING GENETIC INFORMATION

DNA molecules contain information to direct the activities of cells and to guide the development, functioning, and behavior of the organisms that comprise these cells. This information is encoded in sequences of nucleotides within the DNA molecules of the genome. Among cellular organisms, the smallest known genome is that of *Mycoplasma genitalium:* 580,070 nucleotide pairs. By contrast, the human genome consists of 3.2 billion nucleotide pairs. In these and all other genomes, the coding information contained within the DNA is organized into the units called genes. An *M. genitalium* has 485 genes, whereas a human sperm cell has around 20,500. Each gene is a stretch of nucleotide pairs along the length of a DNA molecule. A particular DNA molecule may contain thousands of different genes. In an *M. genitalium* cell, all the genes are situated on one DNA molecule—the single chromosome of this organism. In a human sperm cell, the genes are situated on 23 different DNA molecules corresponding to the 23 chromosomes in the cell. Most of the DNA in *M. genitalium* comprises genes, whereas most of the DNA in humans does not—that is, most of the human DNA is noncoding. We will investigate the composition of genomes in many places in this book, especially in Chapter 15.

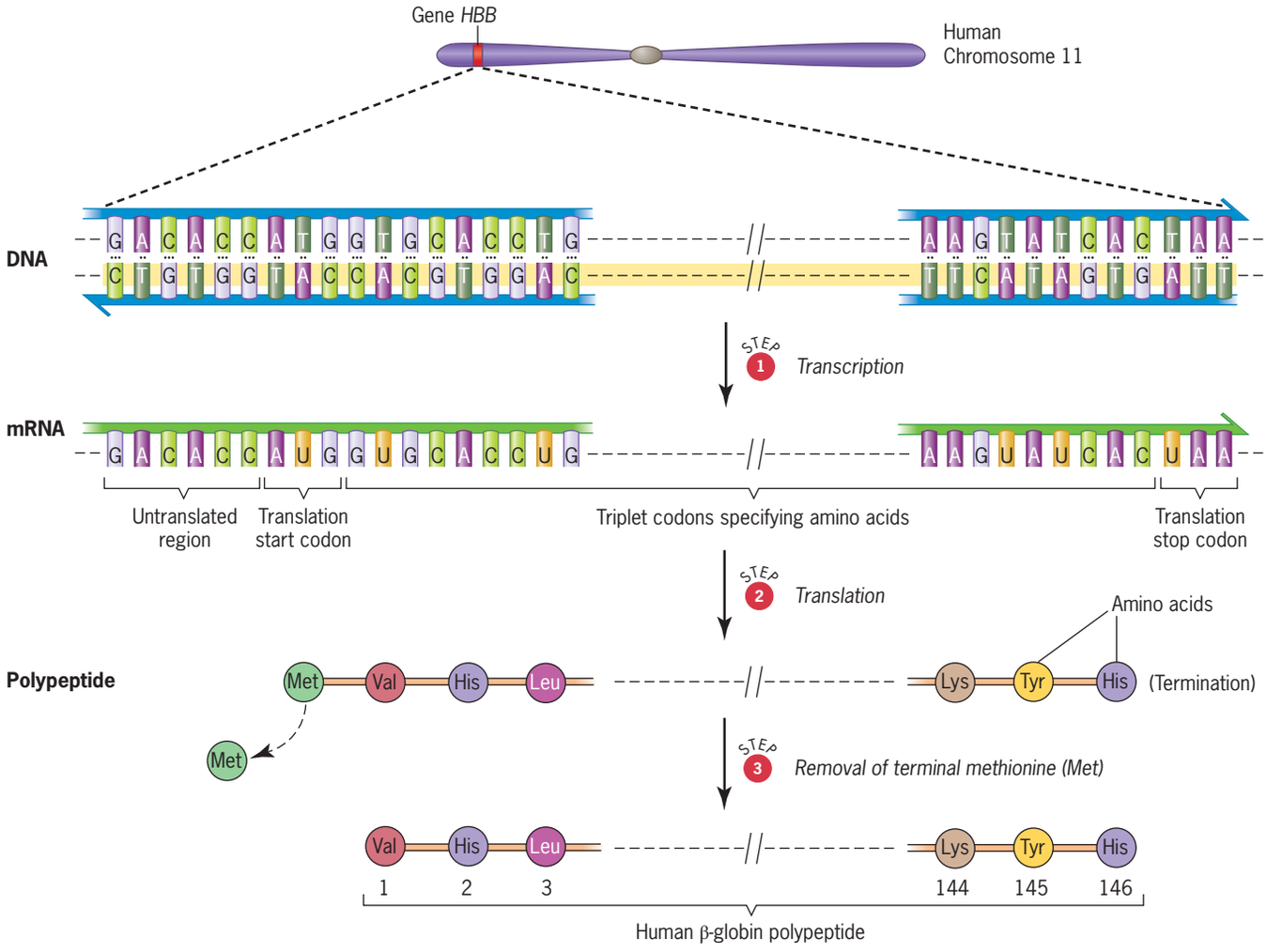
How is the information within individual genes organized and expressed? This question is central in genetics, and we will turn our attention to it in Chapters 11 and 12. Here, suffice it to say that coding genes contain the instructions for the synthesis of proteins. Each protein consists of one or more chains of amino acids. These chains are called **polypeptides.** The 20 different kinds of amino acids that occur naturally can be combined in myriad ways to form polypeptides. Each polypeptide has a characteristic sequence of amino acids. Some polypeptides are short—just a few amino acids long—whereas others are enormous—thousands of amino acids long.

The sequence of amino acids in a polypeptide is specified by a sequence of elementary coding units within a gene. These elementary coding units, called **codons,** are triplets of adjacent nucleotides. A typical gene may contain hundreds or even thousands of codons. Each codon specifies the incorporation of an amino acid into a polypeptide. Thus, the information encoded within a gene is used to direct the synthesis of a polypeptide, which is often referred to as the gene's product. Sometimes, depending on how the coding information is utilized, a gene may encode several polypeptides; however, these polypeptides are usually all related by sharing some common sequence of amino acids.

The expression of genetic information to form a polypeptide is a two-stage process (■ **Figure 1.7**). First, the information contained in a gene's DNA is copied into a molecule of RNA. The RNA is assembled in stepwise fashion along one of the strands of the DNA duplex. During this assembly process, A in the RNA pairs with T in the DNA, G in the RNA pairs with C in the DNA, C in the RNA pairs with G in the DNA, and U in the RNA pairs with A in the DNA. Thus, the nucleotide sequence of the RNA is determined by the nucleotide sequence of a strand of DNA in the gene. The process that produces this RNA molecule is called **transcription,** and the RNA itself is called a **transcript.** The RNA transcript eventually separates from its DNA template and, in some organisms, is altered by the addition, deletion, or modification of nucleotides. The finished molecule, called the **messenger RNA** or simply **mRNA,** contains all the information needed for the synthesis of a polypeptide.

The second stage in the expression of a gene's information is called **translation.** At this stage, the gene's mRNA acts as a template for the synthesis of a polypeptide. Each of the gene's codons, now present within the sequence of the mRNA, specifies the incorporation of a particular amino acid into the polypeptide chain. One amino acid is added at a time. Thus, the polypeptide is synthesized stepwise by reading the codons in order. When the polypeptide is finished, it dissociates from the mRNA, folds into a precise three-dimensional shape, and then carries out its role in the cell. Some polypeptides are altered by the removal of the first amino acid, which is usually methionine, in the sequence.

We refer to the collection of all the different proteins in an organism as its **proteome.** Humans, with around 20,500 genes, may have hundreds of thousands of different proteins



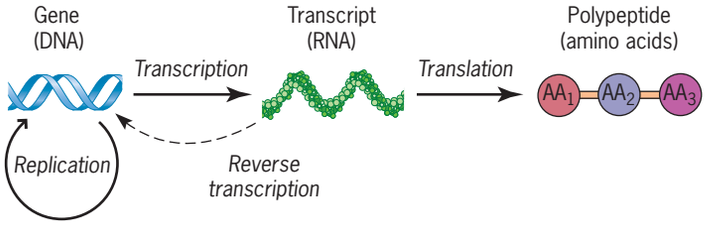
■ **FIGURE 1.7** Expression of the human gene *HBB* coding for the β-globin polypeptide of hemoglobin. During transcription (step 1), one strand of the *HBB* DNA (here the bottom strand shown highlighted) serves as a template for the synthesis of a complementary strand of RNA. After undergoing modifications, the resulting mRNA (messenger RNA) is used as a template to synthesize the β-globin polypeptide. This process is called translation (step 2). During translation each triplet codon in the mRNA specifies the incorporation of an amino acid in the polypeptide chain. Translation is initiated by a start codon, which specifies the incorporation of the amino acid methionine (met), and it is terminated by a stop codon, which does not specify the incorporation of any amino acid. After translation is completed, the initial methionine is removed (step 3) to produce the mature β-globin polypeptide.

in their proteome. One reason for the large size of the human proteome is that a particular gene may encode several different, but related, polypeptides, and these polypeptides may combine in complex ways to produce different proteins. Another reason is that proteins may be produced by combining polypeptides encoded by different genes. If the number of genes in the human genome is large, the number of proteins in the human proteome is larger.

The study of all the proteins in cells—their composition, the sequences of amino acids in their constituent polypeptides, the interactions among these polypeptides and among different proteins, and, of course, the functions of these complex molecules—is called **proteomics.** Like genomics, proteomics has been made possible by advances in the technologies used to study genes and gene products, and by the development of computer programs to search databases and analyze amino acid sequences.

From all these considerations, it is clear that information flows from genes, which are composed of DNA, to polypeptides, which are composed of amino acids, through

an intermediate, which is composed of RNA (■ **Figure 1.8**). Thus, in the broad sense, the flow of information is DNA → RNA → polypeptide, a progression often spoken of as the *central dogma of molecular biology*. In several chapters we will see circumstances in which the first part of this progression is reversed—that is, RNA is used as a template for the synthesis of DNA. This process, called *reverse transcription*, plays an important role in the activities of certain types of viruses, including the virus that causes acquired immune deficiency syndrome, or AIDS; it also profoundly affects the content and structure of the genomes of many organisms, including the human genome. We will examine the impact of reverse transcription on genomes in Chapter 15, and in Chapter 21 on the Instructor Companion site.



■ **FIGURE 1.8** The central dogma of molecular biology showing how genetic information is propagated (through DNA replication) and expressed (through transcription and translation). In reverse transcription, RNA is used as a template for the synthesis of DNA.

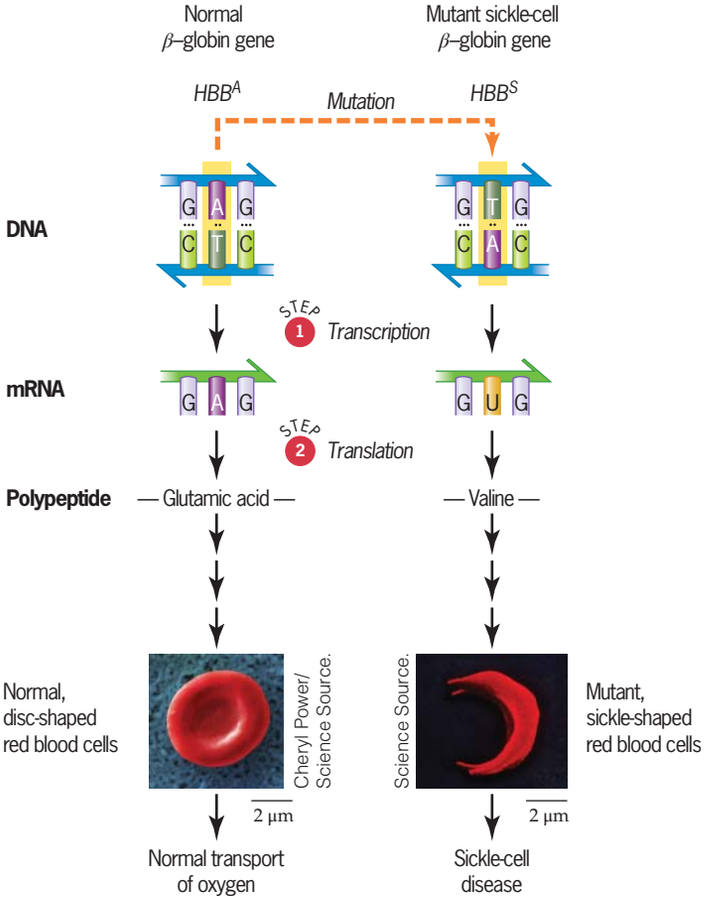
It was once thought that all or nearly all genes encode polypeptides. However, recent research has shown this idea to be incorrect. Many genes do not encode polypeptides; instead, their end products are RNA molecules that play important roles within cells. We will explore these RNAs and the genes that produce them in Chapters 11, 15 and 18.

MUTATION: CHANGING GENETIC INFORMATION

DNA replication is an extraordinarily accurate process, but it is not perfect. At a low but measurable frequency, nucleotides are incorporated incorrectly into growing DNA chains. Such changes have the potential to alter or disrupt the information encoded in genes. DNA molecules are also sometimes damaged by electromagnetic radiation or by chemicals. Although the damage induced by these agents may be repaired, the repair processes often leave scars. Stretches of nucleotides may be deleted or duplicated, or they may be rearranged within the overall structure of the DNA molecule. We call all these types of changes **mutations.** Genes that are altered by the occurrence of mutations are called mutant genes.

Often mutant genes cause different traits in organisms (■ **Figure 1.9**). For example, one of the genes in the human genome encodes the polypeptide known as β-globin. This polypeptide, 146 amino acids long, is a constituent of hemoglobin, the protein that transports oxygen in the blood. The 146 amino acids in β-globin correspond to 146 codons in the β-globin gene. The sixth of these codons specifies the incorporation of glutamic acid into the polypeptide. Countless generations ago, in the germ line of some nameless individual, the middle nucleotide pair in this codon was changed from A:T to T:A, and the resulting mutation was passed on to the individual's descendants. This mutation, now widespread in some human populations, altered the sixth codon so that it specifies the incorporation of valine into the β-globin polypeptide. This seemingly insignificant change has a deleterious effect on the structure of the cells that make and store hemoglobin—the red blood cells. People who carry two copies of the mutant version of the β-globin gene have sickle-shaped red blood cells, whereas people who carry two copies of the nonmutant version of this gene have disc-shaped red blood cells. The sickle-shaped cells do not transport oxygen efficiently through the body. Consequently, people with sickle-shaped red blood cells develop a serious disease, so serious in fact that they may eventually die from it. This sickle-cell disease is therefore traceable to a mutation in the β-globin gene. We will investigate the nature and causes of mutations like this one in Chapter 13.

The process of mutation has another aspect—it introduces variability into the genetic material of organisms. Over time, the mutant



■ **FIGURE 1.9** The nature and consequence of a mutation in the gene for human β-globin. The mutant gene (*HBB^S* top right) responsible for sickle-cell disease resulted from a single base-pair substitution in the β-globin gene (*HBB^A* top left). Transcription and translation of the mutant gene produce a β-globin polypeptide containing the amino acid valine (center right) at the position where normal β-globin contains glutamic acid (center left). This single amino acid change results in the formation of sickle-shaped red blood cells (bottom right) rather than the normal disc-shaped cells (bottom left). The sickle-shaped cells cause a severe form of anemia.