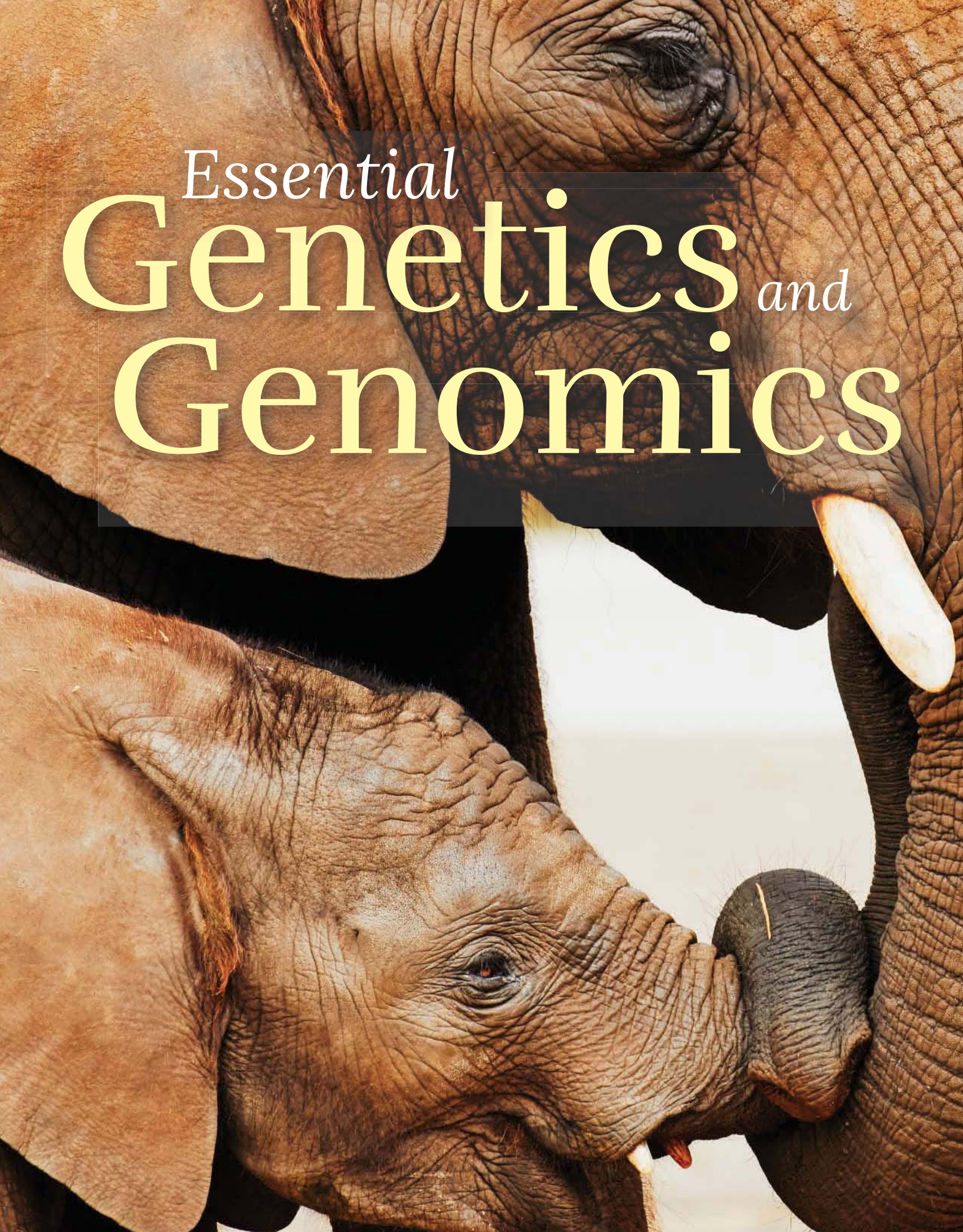


Essential Genetics *and* Genomics

Give Students the Choice

We're happy to offer you and your students ***Essential Genetics and Genomics, Seventh Edition*** in varying formats to best meet individual preferences, whether it be print, digital only, or a combination.

Option 1 – Print + Digital

Essential Genetics and Genomics, Seventh Edition WITH Navigate 2 Advantage Access

Each new print copy of ***Essential Genetics and Genomics, Seventh Edition*** includes Navigate 2 Advantage Access FREE and unlocks a comprehensive and interactive eBook, a *Readiness Assessment* to ensure students are prepared for learning and applying introductory genetics, a study center with links to useful and relevant internet resources, student practice activities and assessments, a full suite of instructor resources, and learning analytics reporting tools. Students can purchase the printed text with digital access at **go.jblearning.com/EssentialGenetics**

ISBN: 978-1-284-15245-6

Standalone Navigate 2 Advantage Access for *Essential Genetics and Genomics, Seventh Edition*

Students can purchase Navigate 2 Advantage Access for ***Essential Genetics and Genomics, Seventh Edition***

QUESTIONS? Visit **go.jblearning.com/FindMyRep** or call 1-800-832-0034 to speak to your dedicated account manager.

Essential Genetics *and* Genomics

SEVENTH EDITION

Daniel L. Hartl, PhD
Harvard University

JONES & BARTLETT
LEARNING

World Headquarters
Jones & Bartlett Learning
5 Wall Street
Burlington, MA 01803
978-443-5000
info@jblearning.com
www.jblearning.com

Jones & Bartlett Learning books and products are available through most bookstores and online booksellers. To contact Jones & Bartlett Learning directly, call 800-832-0034, fax 978-443-8000, or visit our website, www.jblearning.com.

Substantial discounts on bulk quantities of Jones & Bartlett Learning publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones & Bartlett Learning via the above contact information or send an email to specialsales@jblearning.com.

15275-3

Production Credits
VP, Product Management: Amanda Martin
Director of Product Management: Laura Pagluica
Product Specialist: Audrey Schwinn
Product Assistant: Loren-Marie Durr
Production Manager: Dan Stone
Senior Production Editor, Navigate: Jessica deMartin
Digital Products Manager: Jordan McKenzie
Digital Products Specialist: Angela Dooley
Marketing Manager: Lindsay White
Manufacturing and Inventory Control Supervisor: Amy Bacus
Composition: codeMantra U.S. LLC
Cover Design: Kristin E. Parker
Text Design: Kristin E. Parker
Rights & Media Specialist: John Rusk
Media Development Editor: Troy Liston
Cover Image (Title Page): © ktsdesign/Shutterstock
Printing and Binding: LSC Communications
Cover Printing: LSC Communications

Library of Congress Cataloging-in-Publication Data
Names: Hartl, Daniel L., author.
Title: Essential genetics and genomics / Daniel L. Hartl.
Other titles: Essential genetics
Description: Seventh edition. | Burlington, MA : Jones & Bartlett Learning, [2018] | Preceded by Essential genetics / Daniel L. Hartl. 6th ed. 2014. |
Includes bibliographical references and index.
Identifiers: LCCN 2018042898 | ISBN 9781284152456 (pbk.) | ISBN 9781284152685 (e-ISBN)
Subjects: | MESH: Genetic Phenomena | Genomics
Classification: LCC QH430 | NLM QU 500 | DDC 572.8/6—dc23
LC record available at https://lccn.loc.gov/2018042898

6048

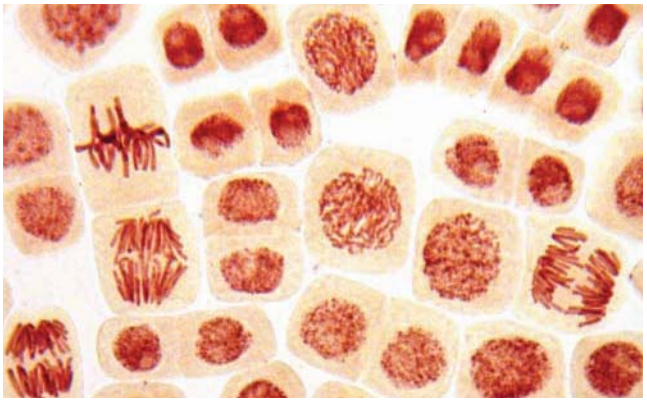
Printed in the United States of America
22 21 20 19 18 10 9 8 7 6 5 4 3 2 1

Brief Contents

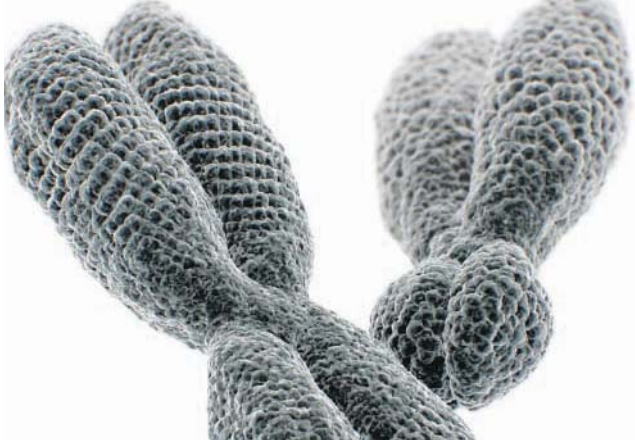
Contents

© ktsdesign/Shutterstock

© Heiko Kueverling/Shutterstock

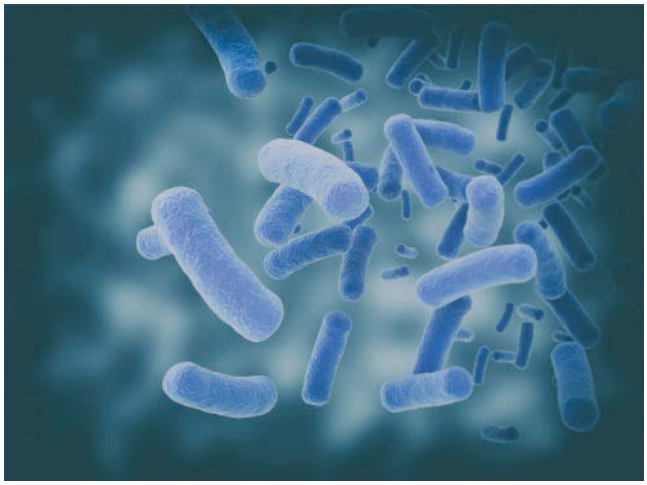


© Dimarion/Shutterstock

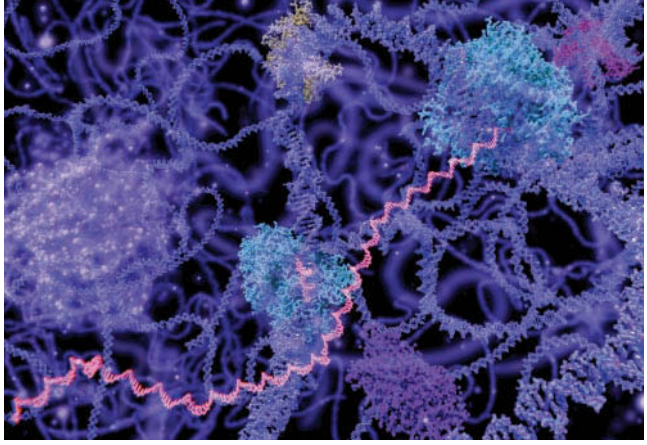


© Iaremenko Sergii/Shutterstock

© H.Tanaka/Shutterstock



© Jezper/Shutterstock



© Juan Gaertner/Shutterstock

© Revers/Shutterstock

© Vchal/ShutterStock

© Valentina Razumova/Shutterstock


© Robert Mcgillivray/Shutterstock


© Photographee.eu/Shutterstock

© Hobbit/Shutterstock

© Yuganov Konstantin/Shutterstock

Preface

A good teacher aims to *un*cover a subject, not *cover* it. So said a wise former teacher of mine. In revising *Essential Genetics and Genomics, Seventh Edition*, I've tried to heed this advice. To *un*cover a subject means to expose, exhibit, unveil. To help *un*cover genetics, this new edition is:

- **Streamlined**, with *emphasis on concepts* illuminated by vivid example and stripped of extraneous detail
- **Focused,** with *Learning Objectives* stated explicitly at the beginning of each chapter
- **Skills oriented**, with *Stop & Think* problems inserted at strategic points in the text to enhance the reading experience and encourage higher-order, analytical thinking.

The brevity of the text meets the needs of the shorter, less comprehensive introductory course of one semester or quarter. The choice of topics is intended to help students master the following scientific competencies:

- Understand the basic processes of gene transmission, mutation, expression, and regulation.
- Analyze genetic processes using numerical relationships including ratios and proportions.
- Learn to formulate genetic hypotheses in a statistical framework, work out their consequences, and test the results against observed data.
- Develop basic skills in problem solving, including single-concept exercises, those requiring the application of several concepts in logical order, and numerical problems requiring some arithmetic for solution.
- Interpret genetic data and make valid inferences to reveal the underlying causes.
- Acquire an appreciation of current trends in genetics, as well as the social and historical context in which genetics has developed.

Scientific competency is the desired learning outcome of any course in a curriculum in STEM (science, technology, engineering, mathematics). Genetics is an excellent subject for achieving scientific competency. It is integrative over a broad territory, ranging from molecular biology to evolutionary genetics. It is also quantitative, using concepts from probability, statistics, and computational biology. Fortunately, students come to a course in genetics highly motivated because of media reports about the human genome and genetic risk factors for disease, as well as many social and ethical controversies related to genetics such as direct-to-consumer and over-the-counter genetic testing, genetic privacy, cloning, stem-cell research, and genetically modified organisms. The challenges for the instructor are to sustain this motivation and to help students acquire the skills and habits of thought that constitute scientific competency.

What's New in the *Seventh Edition*?

This seventh edition has been completely revised and updated. Each chapter has been thoroughly reworked. Important new methods and findings have been added, including **synthetic bacteria**, **higher-order chromatin structure**, **high-throughput genomic sequencing methods**, **personalized medicine**, and **CRISPR/Cas9 gene editing**. As new material has been added, an equal or greater amount of nonessential or outdated material has been deleted. Several of the chapters have been reorganized to allow smooth integration of the new material. The entire text has been condensed, clarified, and updated.

Major revisions and additions in the *Seventh Edition* include the following:

- Chapter 1 includes a new section emphasizing that most common traits are actually **complex traits** affected by multiple interacting genetic factors as well as environment. This principle includes most **common diseases**, which are influenced by multiple genetic risk factors and lifestyle choices. The section on genomes and proteomes has been updated to include **Syn3.0**, the first living, multiplying bacterial cell whose genome was created entirely by chemical synthesis.

- Chapter 2 includes a new section emphasizing that **genes affect traits at multiple levels** (molecular, cellular, developmental, morphological, and behavioral), and that in many cases the traits that are affected appear to be unrelated until the underlying biology is understood.
- Chapter 3 now includes discussion of the **epigenetic specification of the kinetochore**, in which a specialized histone (CENPA) replaces histone 3 in **centromeric nucleosomes** that helps recruit kinetochore-associated proteins leading to the assembly of the mature kinetochore to which spindle fibers attach.
- Chapter 4 has a much shortened and sharper discussion of the **principal types of genetic variation** with emphasis on single nucleotide polymorphisms (SNPs), copy number variations (CNVs), and simple tandem repeats (STRs). The update also includes a summary of the results of **genome-wide association studies (GWAS)** to detect genetic risk factors for common disorders and genetic factors affecting complex traits.
- Chapter 5 has been significantly shortened and streamlined.
- Chapter 6 includes a completely reorganized and simplified discussion of DNA replication updated to include the **trombone model of replication** showing how synthesis of the leading and lagging strands is coordinated. The section on **massively parallel sequencing** has been expanded and moved to Chapter 10.
- Chapter 7 contains a clearer description of how bacterial cells are brought together in conjugation, and unnecessary detail on genetic fine structure of the rII gene in bacteriophage T4 has been eliminated.
- Chapter 8 makes good use of the **Stop & Think** feature to reinforce fundamental concepts and processes of transcription and translation.
- Chapter 9 includes a major new section on how chromatin is organized into higher-order structures composed of **topologically associating domains (TADs)**, **insulators**, and **compartments**. The section on **RNA interference** and **long noncoding RNAs** has been updated, and the material on programmed DNA rearrangements has been removed because it is less generally applicable than once thought to be.
- A completely revised and reorganized Chapter 10 includes a summary of the latest **high-throughput DNA sequencing technology** including **reversible terminators**, **ion-torrent sequencing**, **single-molecule sequencing**, and **nanopore sequencing**. It also includes a major new section on **personalized medicine (precision medicine)** as well as **direct-to-consumer genetic services** and **over-the-counter genetic testing kits**. This section points out the potential values of these approaches but also emphasizes their associated **ethical, legal, and social implications**. Finally, Chapter 10 contains a wholly new section on **CRISPR-Cas9 technology for genetic engineering** showing how CRISPR-Cas9 functions as a sort of immune system in bacteria, how the CRISPR-Cas9 molecules are used to create targeted knockout mutations, and how they are used in DNA editing to change the genome sequence in a predetermined manner. Methods of CRISPR-Cas9 use in insects, mice, and plants are also discussed.
- Chapter 11 contains a slightly expanded section on the use of **epistasis in the analysis of switch-regulation pathways**, and some unnecessary details have been omitted such as minutiae of genetic control of yeast mating type.
- New to Chapter 12 is a major new section on estimates of the **rate of base-substitution mutation in humans** as determined by genome sequencing of parental and offspring genomes. The mutation rate increases steadily with father's age but not with mother's age, and we discuss why this finding is completely consistent with the reproductive biology of males and females. I've also deleted the section on the "ClB method" for detecting mutations in *Drosophila*, as this is mainly of historical interest.
- Chapter 13 puts greater emphasis on the connection between genetic control of the **cell cycle and cancer**, and it has been extensively revised, many of the illustrations simplified, and dispensable details eliminated.
- Chapter 14 contains a shorter, streamlined section on **molecular phylogenetics**.
- What's new in Chapter 15 is a major new section on **genome-wide association studies (GWAS)** focusing on the usually **large number of genes affecting complex traits** and their typically **small individual effects**. Each genetic risk factor for a disease usually increases disease risk by only a small amount. This generalization underlines the importance of proper interpretation of direct-to-consumer (DTC) and over-the-counter (OTC) genetic testing. The chapter also includes a new discussion of **physiological epistasis and statistical epistasis**

and explains why genes can exhibit a great deal of physiological epistasis at the molecular, cellular, and organismal levels without showing any substantial statistical epistasis at the population level. A classic method for estimating the number of genes affecting quantitative traits, based on differences between means of inbred lines and the genetic variance, has been deleted because it is obsolete and usually results in absurdly small estimates.

Chapter Organization

Each chapter begins with a set of **Learning Objectives** to orient students toward the knowledge and skills they should focus on. Explicit **Learning Objectives** help students to:

- *Identify* what they should know or be able to do as a result of their study.
- *Focus* on the knowledge they should have acquired from studying the chapter.
- *Guide* the student to identify key concepts and use them at a variety of learning levels including comprehension, application, analysis, and synthesis.
- *Highlight the skills* they should acquire through practice problems of various types.

Each chapter has an opening paragraph that gives an overview of what is to come, illustrates the subject with engaging examples, and shows how the material is connected to genetics as a whole. The section and subsection **Headings** are in the form of complete sentences that encapsulate the main message. The text makes liberal use of **Numbered** and **Bulleted Lists** to aid students in organizing their learning, as well as **Key Concepts** set apart from the main text to emphasize important principles. A feature called **Stop & Think** recognizes that assessments in real time are critical to reinforce understanding. Each chapter also includes the **Human Connection**. This special feature highlights a research paper in human genetics that reports a key experiment or raises important social, ethical, or legal issues. Each **Human Connection** has a brief introduction of its own, explaining the importance of the experiment and the context in which it was carried out. At the end of each chapter is a complete **Chapter Summary** in the form of bullet points highlighting the most important concepts.

Each chapter also includes several different types and levels of **Problems**, including concept, synthesis, and discussion questions in the form of **Issues and Ideas**, a guide to problem solving called **Solutions: Step By Step**, and application and analysis problems designated **Concepts In Action**.

LEARNING OBJECTIVES

- To understand how genetic information is stored in the base sequence of DNA. For a given sequence of bases in a transcribed strand of protein-coding DNA, you will specify the sequence of bases in the corresponding region of messenger RNA and the sequence of amino acids in the protein. For a mutation in which a specified base is replaced with another, you will deduce the resulting mRNA and protein sequence.
- To realize that enzymes work in sequence in a metabolic pathway. Given a linear metabolic pathway for an essential nutrient, you will deduce which intermediates will restore the ability to grow mutant strains that are defective for any of the enzymes in the pathway. Conversely, using data that specify which intermediates in a linear metabolic pathway restore the ability of mutants to grow, you will infer the order of the enzymes and intermediates in the pathway.
- To learn that genetic complementation is the operational definition of a gene. Given data on the complementation or lack of complementation among all pairs of a set of mutations affecting a biological process, you will sort the mutations into complementation groups, each corresponding to a different gene.

At the end of the book are **Answers** to even-numbered problems, a complete **Glossary** of key terms in genetics, and a compilation of frequently used **Word Roots** that will help students to understand key genetic terms and make them part of their vocabulary. Answers to odd-numbered problems will be available to instructors.

Contents

The organization and number of chapters in the *Seventh Edition* have been retained because they appeal to the majority of instructors who teach genetics. An important feature is the presence of an introductory chapter providing a broad overview of the gene: what it is, what it does, how it changes, how it evolves. Today, most students learn about DNA in grade school or high school. In my teaching, I have found it rather artificial to pretend that DNA does not exist until the middle of the term. The introductory chapter, therefore, serves to connect the more advanced concepts that students are about to learn with what they already know. It also serves to provide each student with a solid framework for integrating the material that comes later. Throughout

KEY CONCEPT

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

each chapter, there is a balance between challenge and motivation, between observation and theory, and between principle and concrete example. Molecular and classical genetics are integrated throughout, and the principles of human genetics are interwoven with the entire fabric of the book. On the other hand, the book is also liberally supplied with examples from animals and plants, especially model organisms. Several points related to organization and coverage should be noted:

CHAPTER SUMMARY

- Inherited traits are affected by genes.
- Genes are composed of the chemical deoxyribonucleic acid (DNA).
- DNA replicates to form (usually identical) copies of itself.
- DNA contains a code specifying what types of enzymes and other proteins are made in cells.
- DNA occasionally mutates, and the mutant forms specify altered proteins.
- A mutant enzyme is an "inborn error of metabolism" that blocks one step in a biochemical pathway for the metabolism of small molecules.
- Genetic analysis of mutants of the fungus *Neurospora* unable to synthesize an essential nutrient led to the one gene–one enzyme hypothesis.
- Different mutations in the same gene can be identified by means of a complementation test, in which the mutants are brought together in the same cell or organism. Mutations in the same gene fail to complement one another, whereas mutations in different genes show complementation.
- Most traits are complex traits affected by multiple genes as well as by environmental factors.
- Organisms change genetically through generations in the process of biological evolution.
- Because of their common descent, organisms share many features of their genetics and biochemistry.

Chapter Summary: Summary of overall concepts discussed in chapter.

ISSUES AND IDEAS

- What special feature of the structure of DNA allows each strand to be replicated without regard to the other?
- What does it mean to say that a strand of DNA specifies the structure of a molecule of RNA?
- What types of RNA participate in protein synthesis, and what is the role of each type of RNA?
- What is meant by the phrase *genetic code*, and how is the genetic code relevant to the translation of a polypeptide chain from a molecule of messenger RNA?
- What is meant by the term *genetic analysis*, and how is genetic analysis exemplified by the work of Beadle and Tatum using *Neurospora*?
- What is a complementation test, and what is it used for in genetic analysis?

Issues and Ideas: Questions asking for genetic principles to be restated in the student's own words.

SOLUTIONS: STEP BY STEP

PROBLEM 1 In the human gene for the beta chain of hemoglobin, the oxygen-carrying protein in the red blood cells, the first 30 nucleotides in the protein-coding region are as shown here.

3′-TACCACGTGGACTGAGGACTCCTCTTCAGA-5′

(a) What is the sequence of the partner strand?
(b) If the DNA duplex of this gene were transcribed from left to right, what is the base sequence of the RNA across this part of the coding region?
(c) What is the sequence of amino acids in this part of the beta-globin polypeptide chain?
(d) In the mutation responsible for sickle-cell anemia, the red T indicated is replaced with an A. The mutant is present at relatively high frequency in some human populations because carriers of the gene are more resistant to falciparum malaria than are noncarriers. What is the amino acid replacement associated with this mutation?

SOLUTION. (a) The partner strand is deduced from the rule that A pairs with T and G pairs with C; however, keep in mind that the paired DNA strands have opposite polarity (that is, their 5′-to-3′ orientations are reversed). (b) The RNA strand is synthesized in the 5′-to-3′ direction, which means that the template DNA strand is transcribed in the 3′-to-5′ direction, which happens to be the same left-to-right orientation of the strand shown above. The base sequence is deduced from the usual base-pairing rules, except that A in DNA pairs with U in RNA. (c) The polypeptide chain is translated in successive groups of three nucleotides (each group constituting a codon), starting at the 5′ end of the coding sequence in the RNA and moving in the 5′-to-3′ direction. The amino acid corresponding to each codon can be found in the genetic code table. (d) The change from T to A in the transcribed strand alters a GAG codon into a GUG codon in the RNA transcript, resulting in the replacement of the normal glutamic acid (GAG) with valine (V). The nonmutant duplex, the RNA transcript, and the amino acid sequence are as shown below. The amino acid that is replaced in the sickle-cell mutant is indicated in red.

3′-TACCACGTGGACTGAGGACTCCTCTTCAGA-5′
5′-ATGGTGCACCTGACTCCTGAGGAGAAGTCT-3′

5′-AUGGUGCACCUGACUCCUGAGGAGAAGUCU-3′
MetValHisLeuThrProGluGluLysSer

Solutions: Step by Step: A section that demonstrates problems worked in full, explaining step by step a path of logical reasoning that can be followed to analyze the problem.

CONCEPTS IN ACTION: PROBLEMS FOR SOLUTION

1.1 Prior to the Avery, MacLeod, and McCarty experiment, what features of cells and chromosomes were already known that could have been interpreted as evidence that DNA is an important constituent of the genetic material?

1.2 In the early years of the twentieth century, why did most biologists and biochemists believe that proteins were probably the genetic material?

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

1.4 What are three principal structural differences between RNA and DNA?

1.5 A region along an RNA transcript contains no U. What base will be missing in the corresponding region of the template strand of DNA?

1.6 When the base composition of a DNA sample from the bacterium *Salinicoccus roseus* was determined, 23.6 percent of the bases were found to be guanine. The DNA of this organism is known to be double stranded. What is the percentage of adenine in its DNA?

1.7 DNA extracted from a certain virus has the following base composition: 15 percent adenine, 25 percent thymine, 20 percent guanine, and 40 percent cytosine. How would you interpret this result in terms of the structure of the viral DNA?

1.8 A duplex DNA molecule contains 532 occurrences of the dinucleotide 5′-GT-3′ in one or the other of the paired strands. What other dinucleotide is also present exactly 532 times?

1.9 A repeating polymer with the sequence

5′-GAUGAUGAUGAU . . .-3′

was found to produce only two types of polypeptides in a translation system that uses cellular components but not living cells (called an *in vitro* translation system). One polypeptide consisted of repeating Asp and the other of repeating Met. How can you explain this result?

1.10 If one strand of a DNA duplex has the sequence 5′-GTCAT-3′, what is the sequence of the complementary strand. (Write the answer with the 5′ end at the left.)

1.11 Consider a region along one strand of a double-stranded DNA molecule consists of tandem repeats of the trinucleotide 5′-CTA-3′, so that the sequence in this strand is 5′-CTACTACTACTA . . .-3′. What is the sequence in the other strand? (Write the answer with the 5′ end at the left.)

1.12 Part of the protein-coding region in a gene has the base sequence 3′-ACAGCATAAACGTTC-5′. What is the sequence of the partner DNA strand?

1.13 If the DNA sequence in Problem 1.12 is the template strand that is transcribed in the synthesis of messenger RNA, would it be transcribed from left to right or from right to left? What base sequence would this region of the RNA contain?

1.14 What amino acid sequence would be synthesized from the messenger RNA region in Problem 1.12?

1.15 If a mutation occurs in the DNA sequence in Problem 1.12 in which the red C is replaced with T, what amino acid sequence would result?

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

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

5′-CGCUUACCACAUGUCGCGAAC-3′

Concepts in Action: Problems for Solution: Problems that require the student to reason using genetic concepts. The problems make use of a variety of formats, and many require some numerical calculation.

Chapter 1 is an overview of genetics designed to bring students with disparate backgrounds to a common level of understanding. This chapter enables classical genetics, molecular genetics, evolutionary genetics, and genomics to be integrated throughout the rest of the book. Included in Chapter 1 are the basic concepts of genetics: genes as regions of DNA that function through transcription and translation, that change by mutation, and that affect organisms through inborn errors of metabolism. Chapter 1 also explains that most traits are actually complex traits affected by multiple genetic and environmental factors, and it introduces genomics and proteomics.

Chapters 2 through 5 are the core of Mendelian genetics, including segregation and independent assortment, the chromosome theory of heredity, mitosis and meiosis, linkage and chromosome mapping, tetrad analysis in fungi, and chromosome mechanics. An important principle of genetics, too often ignored or given inadequate treatment, is that of the complementation test and how complementation differs from segregation or other genetic principles. Chapter 4 expands on the use of molecular markers in genetics, because these are the principal types of genetic markers in use today.

Chapter 6 deals with DNA, including the details of DNA structure and replication. It also discusses how basic research that revealed the molecular mechanisms of DNA replication ultimately led to such important practical applications as DNA hybridization analysis, DNA sequencing, and the polymerase chain reaction. These examples illustrate the value of basic research in leading, often quite unpredictably, to practical applications.

Chapter 7 deals with the principles of genetics in prokaryotes, beginning with the genetics of mobile DNA, plasmids, and integrons, and their relationships to the evolution of multiple antibiotic resistance. There is a thorough discussion of mechanisms of genetic recombination in microbes, including transformation, conjugation, and transduction, as well as a discussion of temperate and virulent bacteriophages.

Chapters 8 through 12 focus on molecular genetics in the strict sense. Chapter 8 examines the details of gene expression, including transcription, RNA processing, and translation.

Chapter 9 is an integrative chapter that deals with genetic mechanisms of regulation, with examples of mechanisms of gene regulation in prokaryotes as well as eukaryotes. Broader aspects of gene regulation that are topics of much current research, such as higher-order chromatin organization, imprinting, and RNAi are included.

Chapter 10 deals with high-throughput genome sequencing and its implications for personalized medicine and the ethical, legal, and social implications of this technology. It also includes basic methods of recombinant DNA, and there is a major new section of CRISPR/Cas9 in DNA editing and its application to genetic engineering.

Chapter 11 examines the genetic control of development with emphasis on models in *C. elegans*, *D. melanogaster*, and *A. thaliana*.

Chapter 12 focuses on mechanisms of mutation and DNA repair, including chemical mutagens.

Chapter 13 stresses cancer from the standpoint of the genetic control of the cell cycle, with emphasis on the checkpoints that, in normal cells, result either in inhibition of cell division or in programmed cell death (apoptosis). Cancer results from a series of successive mutations, usually in somatic cells, which overcome the normal checkpoints that control cellular proliferation.

Chapters 14 and 15 deal with molecular evolution and population genetics. The discussion includes gene trees and species trees and the population genetics of the CCR5 receptor mutation that confers resistance to infection by HIV. It also includes DNA typing in criminal investigations, paternity testing, the effects of inbreeding, and the evolutionary mechanisms that drive changes in allele frequency. The approach to quantitative genetics includes a discussion of how particular genes influencing quantitative traits (QTLs, or quantitative-trait loci) may be identified and mapped by linkage analysis. There is also a section on what has been learned from genome-wide association studies of complex traits in humans, including the identification of QTLs through genetic mapping or studies of candidate genes.

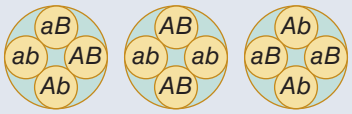
The Student Experience

Stop & Think

A unique feature of this book is found in boxes called **Stop & Think**. These are problems that ask

STOP & THINK 4.3

Yeast cells of genotype *A b* are crossed with those of genotype *a B*. Among 170 unordered tetrads that were analyzed, the following numbers were observed of each type.



Based on these data, what is the map distance between the genes?

a student to pause and think about a concept and apply it to an actual situation. Often these problems use the results of classical experiments to help the student transform a concept from abstract to concrete and carry it from thought to action. Current pedagogy places great emphasis on assessments, and assessments are especially important in genetics because there are many different concepts to be mastered individually as well as in combination. Today's students (and their teachers, too) are often distracted by neighbors, background noise, text messages, email, and all the other disturbances and diversions of the modern world. Three to five **Stop & Think** pauses at strategic points in each chapter encourage students to verify their own understanding of a concept just explained and how to use it to solve an actual problem. The answers are provided at the end of each chapter.

The Human Connection

the human connection in each chapter is our way of connecting to the world of human genetics outside the classroom. All the connections include short excerpts from the original literature of genetics, usually papers, each introduced with a short explanatory

THE HUMAN CONNECTION

Double Trouble

Andrew Fire,[1] SiQun Xu,[1] Mary K. Montgomery,[1] Steven A. Kostas,[1] Samuel E. Driver,[2] and Craig C. Mello[2] (1998)

[1]Carnegie Institution of Washington, Baltimore, Maryland; [2]University of Massachusetts Medical School, Worcester, Massachusetts.

***Potent and Specific Genetic Interference by Double-Stranded RNA in* Caenorhabditis elegans**

Weird and unexpected results began to be reported as soon as it became possible to introduce engineered RNA molecules into organisms. In extreme cases, the engineered RNA prevented the expression of endogenous host genes with sequence homology. At first, it seemed possible that the engineered RNA acted as an antisense inhibitor, in which the introduced RNA undergoes base pairing with the endogenous transcripts and interferes with their function. If this were true, the inhibitory effect of the introduced RNA should be strongly concentration dependent. In this path-breaking paper, the authors show that introduced double-stranded RNA (dsRNA) mediates the inhibitory effects, and that only a few molecules per cell are required. The nematode worm *C. elegans* proved to be ideal for these experiments because, in contrast to some other organisms, dsRNA can be transported from cell to cell and from parent to offspring.

> Experimental introduction of RNA into cells can be used in certain biological systems to interfere with the function of an endogenous gene. . . . Here we investigate the requirements for structure and delivery of the interfering RNA. To our surprise, we found that double-stranded RNA was substantially more effective at producing interference than was either strand individually. . . . Only a few molecules of injected double-stranded RNA were required per affected cell. . . . suggesting that there could be a catalytic or amplification component of the interference process. . . .

> ***To our surprise, we found that double-stranded RNA was substantially more effective at***

Fire and his colleagues looked more closely at this phenomenon, concentrating on the *unc-22* (uncoordinated-22) gene, loss-of-function mutations of which cause severe twitching in the worms. When they injected single-stranded RNA either identical or complementary to *unc-22* mRNA, only minimal interference was observed.

> In contrast, a sense–antisense mixture produced highly effective interference with endogenous gene activity. The mixture was at least two orders of magnitude more effective than either strand alone. . . . The potent interfering activity of the sense–antisense mixture could reflect the formation of double-stranded RNA (dsRNA) or, conceivably, some other synergy between the strands. . . .

The phenotype induced by the introduced RNA was identical to that of conventional loss-of-function mutations of *unc-22*. They concluded by suggesting that RNA interference might be a more general phenomenon.

> Double-stranded RNA could conceivably mediate interference more generally in other nematodes, in other invertebrates, and, potentially, in vertebrates. RNA interference might also operate in plants. . . . Genetic interference by dsRNA could be used by the organism for physiological gene silencing.

A. Fire, et al., *Nature* 391:1998): 806–810.

passage. Many of the connections are excerpts from classic materials, such as Allison's work on the sickle-cell trait and resistance to malaria, but by no means are all the "classic" papers old papers. The pieces are called **the human connection** because each connects the material to something that broadens or enriches its implications for human beings. Some of the connections raise issues of ethics in the application of genetic knowledge, social issues that need to be addressed, or issues related to laboratory animals.

They illustrate other things as well. Because each connection names the place where the research was carried out, the student will see that great science is done in many universities and research institutions throughout the world. In papers that use outmoded or unfamiliar terminology, or archaic gene symbols, I have substituted the modern equivalent to make the material more accessible to the student.

Solutions Step by Step

Each chapter contains a section titled **Solutions: Step by Step** that demonstrates problems worked in full, explaining step by step a path of logical reasoning that can be followed to analyze the problem. The **Solutions: Step by Step** serve as another level of review of the important concepts used in working problems. The solutions also emphasize some of the most common mistakes made by beginning students and give pointers on how students can avoid falling into these conceptual traps.

Levels and Types of Problems

Each chapter provides numerous problems for solution, graded in difficulty, so students can test their understanding. The problems are of two different types:

- ***Issues and Ideas*** ask for genetic principles to be restated in the student's own words; some are matters of definition or call for the application of elementary principles.
- ***Concepts in Action*** are problems that require the student to reason using genetic concepts. The problems make use of a variety of formats, and many require some numerical calculation. The level of mathematics is that of arithmetic and elementary probability as it pertains to genetics. None of the problems uses mathematics beyond elementary algebra. The problems range in difficulty from easy to hard. They are primarily at **Bloom's higher order cognitive level**, and most require **analyzing** data, **evaluating** evidence, or **creating** hypotheses or experiments.

Answers to Problems

The answers to the even-numbered **Concepts in Action** are included in the **Answer** section at the end of the book. The answers are complete; they explain the logical foundation of the solution and lay out the methods. The answers to the remainder of the **Concepts in Action** problems are available with the online instructor's resources and in the optional online Study Guide and Solutions Manual.

5.6 The observation is quite unexpected, because a 47, XXX female would be expected to produce many XX-bearing eggs, and a 47, XYY male would be expected to produce many XY-bearing sperm. Apparently, the extra X chromosome in 47, XXX females, and the extra Y chromosome in 47, XYY males, are eliminated from the nucleus prior to

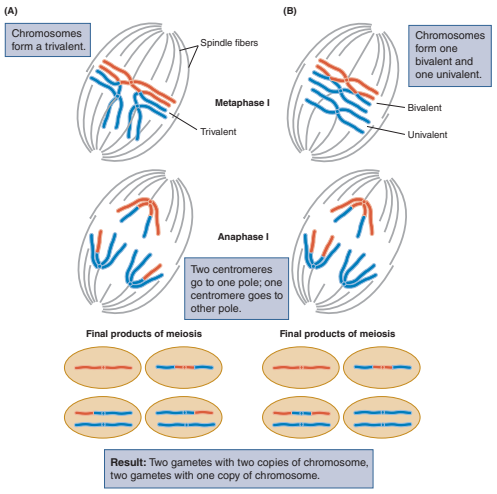
Word Roots and Glossary

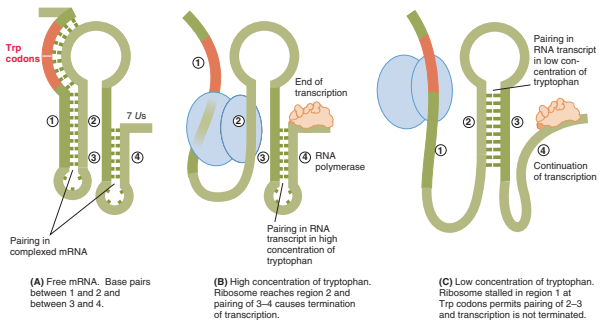
I have included a compilation of **Word Roots** that students find helpful in interpreting and remembering the meaning of technical terms. This precedes the **Glossary** of key words.

ante-	*preceding, before*	antedate, preceding a date
apo-	*former, from*	aporepressor, precursor to repressor
aut-, auto-	*self*	autogenous, self-generated
bi-	*two*	bidirectional, going in two directions

Illustrations

Every chapter is richly illustrated with beautiful graphics in which color is used functionally to enhance the value of each illustration as a learning aid. The illustrations are also heavily annotated with "process boxes" explaining step by step what is happening at each level of the illustration. These labels make the art user friendly, inviting, and maximally informative.

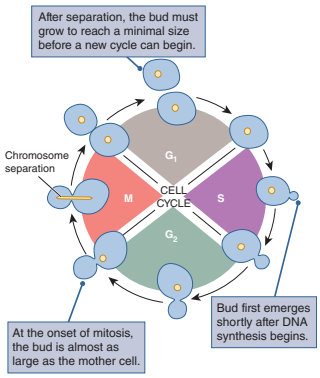




(A) Free mRNA. Base pairs between 1 and 2 and between 3 and 4.

(B) High concentration of tryptophan. Ribosome reaches region 2 and pairing of 3–4 causes termination of transcription.

(C) Low concentration of tryptophan. Ribosome stalled in region 1 at Trp codons permits pairing of 2–3 and transcription is not terminated.



Teaching Tools

Adaptability and Flexibility

There is no compelling reason to start at the beginning and proceed straight to the end of this text. Each chapter is a self-contained unit that stands on its own. This feature gives the book the flexibility to be used in a variety of course formats. Throughout the book, we have integrated classical and molecular principles, so you can begin a course with almost any of the chapters. Most teachers will prefer starting with the overview in Chapter 1, possibly as suggested reading, because it brings every student to the same basic level of understanding. Teachers preferring to cover Mendel early should continue with Chapter 2; those preferring to teach the details of DNA early should continue with Chapter 6. Some teachers are partial to a chromosomes-early format, which would suggest continuing with Chapter 3, followed by Chapters 2 and 4. A novel approach would put genomics first, which could be implemented by continuing with Chapter 10. The writing and illustration programs were designed to accommodate a variety of formats, and we encourage teachers to take advantage of this flexibility to meet their own needs.

Instructor Resources

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

The ***Image Bank in PowerPoint format*** provides all the illustrations and photos (to which Jones & Bartlett Learning owns the copyright or has permission to reprint digitally), inserted into PowerPoint slides. With the Microsoft® PowerPoint program you can quickly and easily copy individual image slides into your existing lecture slides.

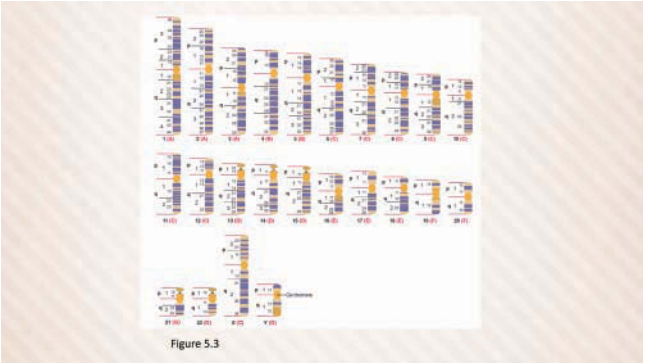


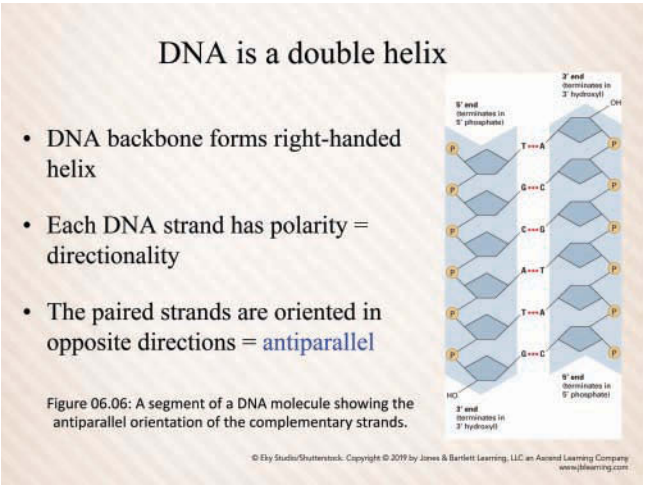
Figure 5.3

A ***Table Bank*** provides images of all of the tables (to which Jones & Bartlett Learning owns the copyright or has permission to reprint digitally), in a PDF file for easy use.

TABLE 9.1 Characteristics of Partial Diploids Containing Several Combinations of *lacI*, *lacO*, and *lacP* Alleles

Genotype	Synthesis of *lac* mRNA	Lac phenotype
1. F′ *lacO*ᶜ *lacZ*⁺/*lacO*⁺ *lacZ*⁺	Constitutive	+
2. F′ *lacO*⁺ *lacZ*⁺/*lacO*ᶜ *lacZ*⁺	Constitutive	+
3. F′ *lacI*⁻ *lacZ*⁺/*lacI*⁺ *lacZ*⁺	Inducible	+
4. F′ *lacI*⁺ *lacZ*⁺/*lacI*⁻ *lacZ*⁺	Inducible	+
5. F′ *lacO*ᶜ *lacZ*⁻/*lacO*⁺ *lacZ*⁺	Inducible	+
6. F′ *lacO*ᶜ *lacZ*⁺/*lacO*⁺ *lacZ*⁻	Constitutive	+
7. F′ *lacI*ˢ *lacZ*⁺/*lacI*⁺ *lacZ*⁺	Uninducible	–
8. F′ *lacI*⁺ *lacZ*⁺/*lacI*ˢ *lacZ*⁺	Uninducible	–
9. F′ *lacP*⁻ *lacZ*⁺/*lacP*⁺ *lacZ*⁺	Inducible	+
10. F′ *lacP*⁺ *lacZ*⁺/*lacP*⁻ *lacZ*⁺	Inducible	+
11. F′ *lacP*⁺ *lacZ*⁻/*lacP*⁻ *lacZ*⁺	Uninducible	–
12. F′ *lacP*⁺ *lacZ*⁺/*lacP*⁻ *lacZ*⁻	Inducible	+

A set of ***Lecture Outlines in PowerPoint format*** provides outline summaries of each chapter. The slide set can be customized to meet your classroom needs.



The ***Test Bank*** contains over 700 test items. There is a mix of factual, descriptive, analytical, and quantitative question types. A typical chapter file contains 20 multiple-choice objective questions, 15 fill-in-the-blank questions, and 15 quantitative problems. Versions easily compatible with most course management software are available to adopting instructors upon request.

5. The R-type strain of *S. pneumoniae* does not cause pneumonia because bacterial cells are
A. Surrounded by a polysaccharide capsule
B. Unable to synthesize a polysaccharide capsule
C. Unable to form colonies
D. Undetectable by the immune system

Ans: B

6. A DNA strand consists of any sequence of four kinds of nucleotides. Suppose there were only 16 different amino acids instead of 20. Which of the following statements would be correct descriptions of the minimal number of nucleotides necessary to create a genetic code?
A. 1
B. 2, provided that chain termination does not require a special codon
C. 3, provided that chain termination does require a special codon
D. 2, no matter how chain termination is accomplished
E. Both B and C

Ans: E

7. tNRA is
A. The major structural material making up ribosomes
B. A molecule that incorporates a specific amino acid into the growing protein when it recognizes a specific group of three bases
C. The major structural component of chromosomes
D. The molecule that carries the genetic information from DNA and is used as a template for protein synthesis
E. The major building block of proteins

Ans: B

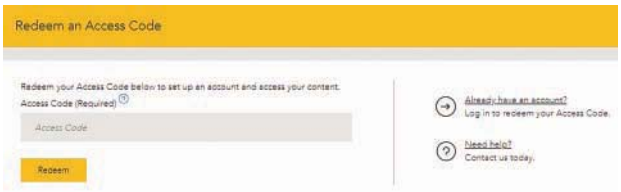
8. rRNA is
A. The major structural material making up ribosomes
B. The molecule that carries the genetic information from DNA and is used as a template for protein synthesis

2

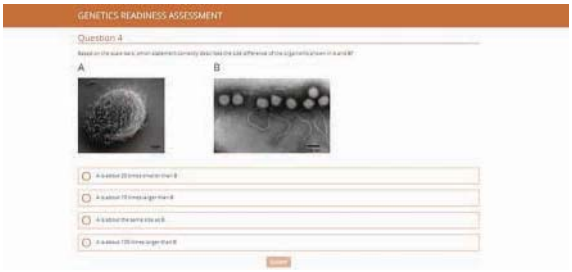
Readiness Assessment and Readiness Review

How ready are you to learn introductory genetics? For the first time we are offering a Readiness Assessment for you to find out. Simply answer the online questions found within Navigate 2 (accessible via the access code in the front of the text*). Once complete, you will be given a score and directed to the color-coded in-text Readiness Review section(s) that will most help you prepare (found at the back of the text). Never has learning genetics been so easy and fun!

STEP 1: Redeem your code.



STEP 2: Take the quiz.



STEP 3: Get your score.

STEP 4: Learn, review, and practice!

Problems

Problem 2.1 Calculate the percentage and ratios of each continent's population to that of the world.

Continent	Proportion (percent)	Calculation	Ratio (Continent : World)	Calculation
Asia				
Africa				
Europe				
North America	0.077 (7.7%)		113	
South America				
Australia/ Oceania				
Antarctica				

Problem 2.2 What is the ratio of the population of Asia to North America?

Problem 2.3 What is the ratio of the area of Asia to North America?

Problem 2.4 What is the ratio of the population of Africa to Europe?

Problem 2.5 What is the ratio of the area of Africa to Europe?

STEP 5: Retake the quiz to check for improvement.

*Access can also be purchased separately. Visit go.jblearning.com/hartl7e to learn more.

Acknowledgments

I am indebted to my colleagues whose advice and thoughts were immensely helpful throughout the preparation of the five editions of this book. These colleagues range from specialists in various aspects of genetics who checked for accuracy or suggested improvement to instructors who evaluated the material for suitability in teaching or sent me comments on the text as they used it in their courses.

Laura Adamkewitz, George Mason University
Jeremy C. Ahouse, Brandeis University
Mary Alleman, Duquesne University
Jennifer Armstrong, Scripps, Pitzer, and Claremont McKenna Colleges
Peter D. Ayling, University of Hull
John C. Bauer, Stratagene, Inc., La Jolla, CA
Anna C. Berkowitz, Purdue University
Mary K. B. Berlyn, Yale University
Thomas A. Bobik, University of Florida
Carrie Baker Brachmann, University of California, Irvine
Jessica Brzyski, Seton Hill University
Colin G. Brooks, The Medical School, Newcastle
Mirella Vasquez Brooks, University of Hawaii at Manoa
Jill A. Buettner, Richland College
Jeffrey J. Byrd, St. Mary's College of Maryland
Susan L. Carney, Hood College
Pierre Carol, Université Joseph Fourier
Domenico Carputo, University of Naples
Sean Carroll, University of Wisconsin
Chris Caton, University of Birmingham
John Celenza, Boston University
Richard W. Cheney, Jr., Christopher Newport University
Alan C. Christiensen, University of Nebraska, Lincoln
Michael J. Christoffers, North Dakota State University
Erin Cram, Northeastern University
Christoph Cremer, University of Heidelberg
Marion Cremer, Ludwig Maximilians University
Thomas Cremer, Ludwig Maximilians University
Leslie Dendy, University of New Mexico
John W. Drake, National Institute of Environmental Health Sciences, Research Triangle Park, NC
Stephen J. D'Surney, University of Mississippi
Kathleen Dunn, Boston College
Chris Easton, State University of New York
David Eisenmann, University of Maryland, Baltimore County
Wolfgang Epstein, University of Chicago
Silviu Faitar, D'Youville College
Brian E. Fee, Manhattan College
Gyula Ficsor, Western Michigan University
Robert G. Fowler, San Jose State University
David W. Francis, University of Delaware
Gail Gasparich, Towson University
Elliott S. Goldstein, Arizona State University
Ruth Grene, Virginia Tech
Patrick Guilfoile, Bemidji State University
Jeffrey C. Hall, Brandeis University
Mark L. Hammond, Campbell University
Randall K. Harris, William Carey University
Steven Henikoff, Fred Hutchinson Cancer Research Center, Seattle, WA
Charles Hoffman, Boston College
Ivan Huber, Fairleigh Dickinson University
Kerry Hull, Bishop's University
Lynn A. Hunter, University of Pittsburgh
Richard Imberski, University of Maryland
Bradley J. Isler, Ferris State University
Diana Ivankovic, Anderson University
Joyce Katich, Monsanto, Inc., St. Louis, MO
Jeane M. Kennedy, Monsanto, Inc., St. Louis, MO
Jeffrey King, University of Berne
Tobias A. Knoch, German Cancer Research Center, Heidelberg, Germany
Laszlo Kovacs, Missouri State University
Yan B. Linhart, University of Colorado
K. Brooks Low, Yale University
Sally A. MacKenzie, Purdue University
Gustavo Maroni, University of North Carolina
Jeffrey Mitton, University of Colorado
Sara Morris, Biology MPCC, North Platte, NE
Robert K. Mortimer, University of California
Gisela Mosig, Vanderbilt University
John R. Nambu, Florida Atlantic University
Steve O'Brien, National Cancer Institute

Kevin O'Hare, Imperial College, London
Michael V. Osier, Rochester Institute of Technology
Catherine A. Palmer, Portland State University
Ronald L. Phillips, University of Minnesota
Jennifer R. Powell, Gettysburg College
Robert Pruitt, Purdue University
Peggy Redshaw, Austin College
Pamela Reinagel, California Institute of Technology
Susanne Renner, University of Missouri
Lynn S. Ripley, University of Medicine and Dentistry of New Jersey
Andrew J. Roger, Dalhousie University, Halifax
Moira E. Royston, St. Joseph's College, NY
Kenneth E. Rudd, National Library of Medicine, Bethesda, MD
Mary Russell, Kent State University at Trumbull
Thomas F. Savage, Oregon State University
Joseph Schlammadinger, University of Debrecen
Brian W. Schwartz, Columbus State University
David Shepard, University of Delaware
Alastair G. B. Simpson, Dalhousie University
Leslie Smith, National Institute of Environmental Health Sciences, Research Triangle Park, NC
Charles Staben, University of Kentucky
Julie Dangremond Stanton, Washington State University
Johan H. Stuy, Florida State University
David T. Sullivan, Syracuse University
Jeanne Sullivan, West Virginia Wesleyan College
Millard Susman, University of Wisconsin
Fusheng Tang, University of Arkansas, Little Rock
Irwin Tessman, Purdue University
James H. Thomas, University of Washington
Michael Thomas, Idaho State University
Jan Trybula, State University of New York, College at Potsdam
Michael Tully, University of Bath
L.K. Tuominen, John Carroll University
David Ussery, The Technical University of Denmark
George von Dassow, Friday Harbor Laboratories, Friday Harbor, WA
Denise Wallack, Muhlenberg College
Kenneth E. Weber, University of Southern Maine
Tamara Western, Okanagan University College
Taek H. You, Campbell University

Finally, a very special thanks goes to geneticist Elena R. Lozovsky of Harvard University, who contributed substantially to the instructional materials at the end of each chapter, as well as to the instructor and student supplements. Elena's help and support is very gratefully acknowledged. I also wish to acknowledge the superb art, production, and editorial staff at Jones & Bartlett Learning who helped make this book possible: Matt Kane, Audrey Schwinn, Loren-Marie Durr, John Rusk, Troy Liston, Dan Stone, and Kristin Parker. I am also grateful to the many people, acknowledged in the legends of the figures, who contributed photographs, drawings, and micrographs from their own research and publications. Every effort has been made to obtain permission to use copyrighted material and to make full disclosure of its source. We are grateful to the authors, journal editors, and publishers for their cooperation. Any errors or omissions are wholly inadvertent and will be corrected at the first opportunity.

We would also like to thank and acknowledge Dr. Bruce Cochrane for his work updating the instructor's resources and Dr. Rebecca Reiss for her work on the assessments, including the online Readiness Assessment and the Readiness Review found in the back of the book.

Dan Hartl

About the Author

Daniel L. Hartl is Higgins Professor of Biology at Harvard University, a Professor of Immunology and Infectious Diseases at the Harvard T. H. Chan School of Public Health, and a Senior Associate Member of the Broad Institute of M.I.T. and Harvard. He is a member of the National Academy of Sciences and the American Academy of Arts and Sciences. Hartl received his B.S. degree and Ph.D. from the University of Wisconsin and carried out postdoctoral research at the University of California at Berkeley. His research interests include molecular genetics, genomics, molecular evolution, and population genetics.

For the Student

Special features designed to help in mastering the material are emphasized in the **Preface.** In my experience, students who struggle in genetics and genomics do so for two reasons. They may lack effective reading and study habits, or they may fail to self-assess. By self-assessment I mean asking yourself whether you understand a concept well enough to express it in your own words, and whether you understand it well enough to use in solving problems. Here are some pointers for improving both study skills and problem solving.

Tips for Learning Concepts

- Go to class, take notes by hand (in telegraphic style, abbreviating as needed), and copy your notes in complete sentences and legible handwriting as soon as possible thereafter. Words written by hand are retained in memory better than words typed on a keyboard.
- Plan 30–35 minute reading sessions, but only when you are not tired or distracted. Most people find that their attention begins to wane after 30–35 minutes of intense concentration.
- Read attentively. Find a quiet, clean, well-lighted place and turn off your laptop and smartphone.
- Start by skimming what you think you can cover in your reading session, including a preliminary look at the illustrations; this is your drone's eye view of the terrain that will help keep you oriented.
- Look up unfamiliar words in the glossary.
- Reread difficult sections and make handwritten notes of the key points.
- Highlight, underline, or better yet summarize the key concepts in your own words. The textbook is designed to help in recognizing these: key terms are in **boldface,** key points are highlighted with **bulleted lists,** and key concepts are set off and labeled **Key Concept.**
- Take a break from reading to solve the problems in **Stop & Think.** They are designed to help you assess whether you have understood the concepts you've just read well enough to apply them.

Tips for Problem Solving

- Make use of the **Solutions: Step By Step.** These guide you through the reasoning used to solve the major types of problems arising from the concepts in each chapter.
- Don't start working a problem until you're sure you understand what is being asked.
- Use the glossary if necessary to understand the key terms in a problem.
- Start with some easy problems to gain self-confidence.
- Once you know how to solve a certain type of problem, don't spend time on similar ones that you already know how to solve. It's problems that you don't immediately know how to solve that you really learn from.
- Don't rush. Haste makes waste, as the saying goes—and when you rush you are more likely to make stupid mistakes.
- Break a complex problem into smaller parts that you can attack individually, and use the parts you understand as leverage to get at the more difficult parts.
- Don't give up! Never, never, never! You may have to attack a problem from two or three different angles before you find yourself on a productive track.
- Don't work backwards from the answer. There's an old adage that "if you know where you're going you can find a way to get there." The problem is that the way you find to "get there" may

use completely messed-up logic. What's worse, you will have trained your neurons to use the wrong logic. Learning a concept is hard enough, unlearning one that you misunderstand is harder still.

- Work in small groups with other students if you can, but be sure you understand the reasoning behind any answer your group comes up with. Sometimes a fellow student can explain a concept more clearly than your instructor.

Representation of DNA sequencing.

CHAPTER 1

The Genetic Code of Genes and Genomes

LEARNING OBJECTIVES

- To understand how genetic information is stored in the base sequence of DNA. For a given sequence of bases in a transcribed strand of protein-coding DNA, you will specify the sequence of bases in the corresponding region of messenger RNA and the sequence of amino acids in the protein. For a mutation in which a specified base is replaced with another, you will deduce the resulting mRNA and protein sequence.
- To realize that enzymes work in sequence in a metabolic pathway. Given a linear metabolic pathway for an essential nutrient, you will deduce which intermediates will restore the ability to grow mutant strains that are defective for any of the enzymes in the pathway. Conversely, using data that specify which intermediates in a linear metabolic pathway restore the ability of mutants to grow, you will infer the order of the enzymes and intermediates in the pathway.
- To learn that genetic complementation is the operational definition of a gene. Given data on the complementation or lack of complementation among all pairs of a set of mutations affecting a biological process, you will sort the mutations into complementation groups, each corresponding to a different gene.

Genetics is worth studying for many reasons, ranging from applications in medicine, agriculture, public health, and conservation biology to ongoing research in cell biology, development, neuroscience, and evolution. Genetics also deals with some of the great philosophical themes of human life and individual differences:

> A spermatozoon plunges headlong into an ovum, and immediately a long-term project is set in motion. The cells proliferate at a dizzying rate, clustering, diversifying. Out of that initial, infinitesimal particle will grow a beating heart, hands, fingernails, hair, glands, and a brain with the power to think of itself . . . But from time to time nature, too, gets things wrong, and so you'll have six fingers on one hand or one leg shorter than the other, or else she may construct a brain incapable of understanding the simplest things.
>
> From *Death in August* by Marco Vichi, translated by Stephen Sartarelli (2011, Pegasus Books, p. 123)

There is indeed a developmental plan encoded in our DNA that makes each of us a member of the human species. But still we differ from one another. Besides rare anomalies like six fingers, we differ from one another in many everyday, observable characteristics, or traits, like hair color, eye color, skin color, height, weight, and personality. Some of these traits differ because of heredity, others because of culture. The color of your eyes results from biological inheritance, but the native language you learned as a child results from cultural inheritance. Many traits are influenced jointly by biological inheritance and environmental factors or lifestyle choices. How much you weigh is determined in part by your inheritance but also in part by how much food you eat, its nutritional content, and your exercise habits.

Genetics is the study of biologically inherited traits, including traits that are influenced in part by the environment. **Genomics** is the study of all the genes in an organism to understand their molecular organization, function, interaction, and evolutionary history. The fundamental concept of genetics and genomics is that:

KEY CONCEPT

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

The existence of genes and the rules governing their transmission from generation to generation were first articulated by Gregor Mendel in 1866. Mendel's formulation of inheritance was in terms of the abstract rules by which genes (he called them "factors") are transmitted from parents to offspring. His objects of study were garden peas, with variable traits like pea color and plant height. The foundation of genetics as a molecular science also dates back to the 1860s when Friedrich Miescher discovered a new type of weak acid, abundant in the nuclei of white blood cells, which turned out to be what we now call **DNA (deoxyribonucleic acid)**. For many years, the biological function of DNA was unknown, and no role in heredity was ascribed to it.

In this book, you will learn a lot about genes and genomes. You will learn what constitutes a gene and how it works in physiology and development, in health and disease. You will also learn how genomes are organized and the activities of different genes coordinated in space and time. If you know nothing about genetics, you will be brought up to speed. And if you already know something, you will see it in a different light. There are lots of details, but try not to get so tangled up in them that you lose sight of how genetics can help you understand the great themes—birth, consciousness, death—that make the details worth knowing.

1.1 DNA is the molecule of heredity.

The importance of the cell nucleus in inheritance became clear in the 1870s, when the nuclei of the male and female reproductive cells were observed to fuse in the process of fertilization. The next major advance was the discovery of **chromosomes**, thread-like objects inside the nucleus that become visible in the light microscope when stained with certain dyes. Chromosomes exhibit a characteristic "splitting" behavior, in which each daughter cell formed by cell division receives an identical complement of chromosomes. More evidence for the importance of chromosomes was provided by the observation that, whereas the number of chromosomes in each cell differs from one biological species to the next, the number of chromosomes is nearly always constant within the cells of any particular species. These features of chromosomes were well understood by about 1900, and they made it seem likely that chromosomes were the carriers of the genes.

By the 1920s, several lines of indirect evidence suggested a close relationship between chromosomes, and DNA. Microscopic studies with special stains showed that DNA is present in chromosomes. Various types of proteins are present in chromosomes, too. But whereas most of the DNA in cells of higher organisms is present in chromosomes, and the amount of DNA per cell is constant, the amount and kinds of proteins and other large molecules differ greatly from one type of cell to another. The indirect evidence for DNA as

the genetic material was unconvincing, because crude chemical analyses had suggested (erroneously, as it turned out) that DNA lacked the chemical diversity needed in a genetic substance. The favored candidate for the genetic material was protein, because proteins were known to be an exceedingly diverse collection of molecules. Proteins therefore became widely accepted as the genetic material, and DNA was thought to provide only the structural framework of chromosomes. Any researcher who hoped to demonstrate that DNA was the genetic material had a double handicap. Such experiments had to demonstrate not only that DNA *is* the genetic material but also that proteins are *not* the genetic material. Some of the experiments regarded as decisive in implicating DNA are described in this section.

Genetic traits can be altered by treatment with pure DNA.

One type of bacterial pneumonia in mammals is caused by strains of *Streptococcus pneumoniae* able to synthesize a gelatinous capsule composed of polysaccharide (complex carbohydrate). This capsule surrounds the bacterium and protects it from the defense mechanisms of the infected animal; thus it enables the bacterium to cause disease. When a bacterial cell is grown on solid medium, it undergoes repeated cell divisions to form a visible clump of cells called a **colony**. The enveloping capsule makes the size of each colony large and gives it a glistening or smooth (S) appearance (**FIGURE 1.1**). Certain strains of *S. pneumoniae*, however, are unable to synthesize the capsular polysaccharide, and they form small colonies that have a rough (R) surface. The R strains do not cause pneumonia; lacking the capsule, these bacteria are inactivated by the immune system of the host. Both types of bacteria "breed true" in the sense that the progeny formed by cell division have the capsular type of the parent, either S or R.

When mice are injected with living R cells or with S cells that have been killed with extreme heat, the animals remain healthy. However, in 1928 Frederick Griffith showed that when mice are injected with a *mixture* of living R cells and heat-killed S cells, they often die of pneumonia (**FIGURE 1.2**). Bacteria isolated from blood samples of the dead mice produce S cultures with a capsule typical of the injected S cells, even though the injected S cells had been killed by heat. Evidently, the injected material from the dead S cells includes a substance that can enter living R bacterial cells and give them the ability to synthesize the S-type capsule. In other words, the R bacteria can be changed—or undergo **transformation**—into S bacteria, and the new characteristics are inherited by descendants of the transformed bacteria.

Griffith's transformation of *Streptococcus* was not in itself definitive, but in 1944 the chemical substance responsible for changing the R cells into S cells was identified as DNA. In a milestone experiment, Oswald Avery, Colin MacLeod, and Maclyn McCarty showed that the substance causing the transformation of R cells into S cells was DNA. In preparation for the experiment, they had to develop chemical procedures for obtaining DNA in almost pure form from bacterial cells, which had not been done before. When they added DNA isolated from S cells to growing cultures of R cells, they observed that a few S-type cells were produced. Although the DNA preparations contained traces of protein and RNA (ribonucleic acid, an abundant cellular macromolecule chemically related to DNA), the transforming activity was not altered by treatments that destroy either protein or RNA. However, treatments that destroy DNA eliminated

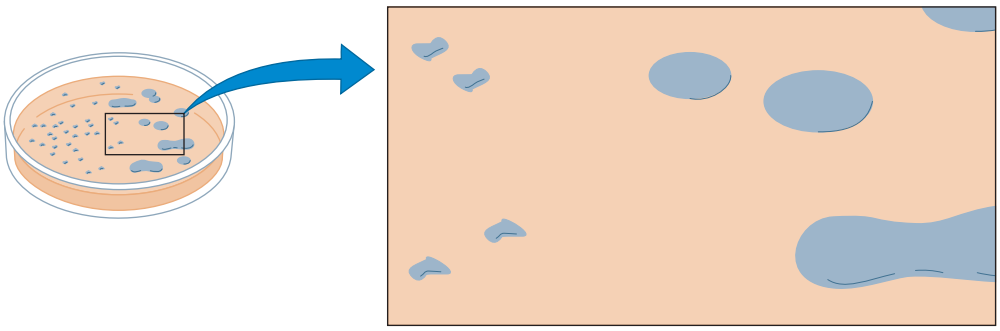


FIGURE 1.1 Colonies of *Streptococcus pneumoniae*. The small colonies on the left are from a rough (R) strain, and the large colonies on the right are from a smooth (S) strain. The S colonies are larger because of the capsule on the S cells.

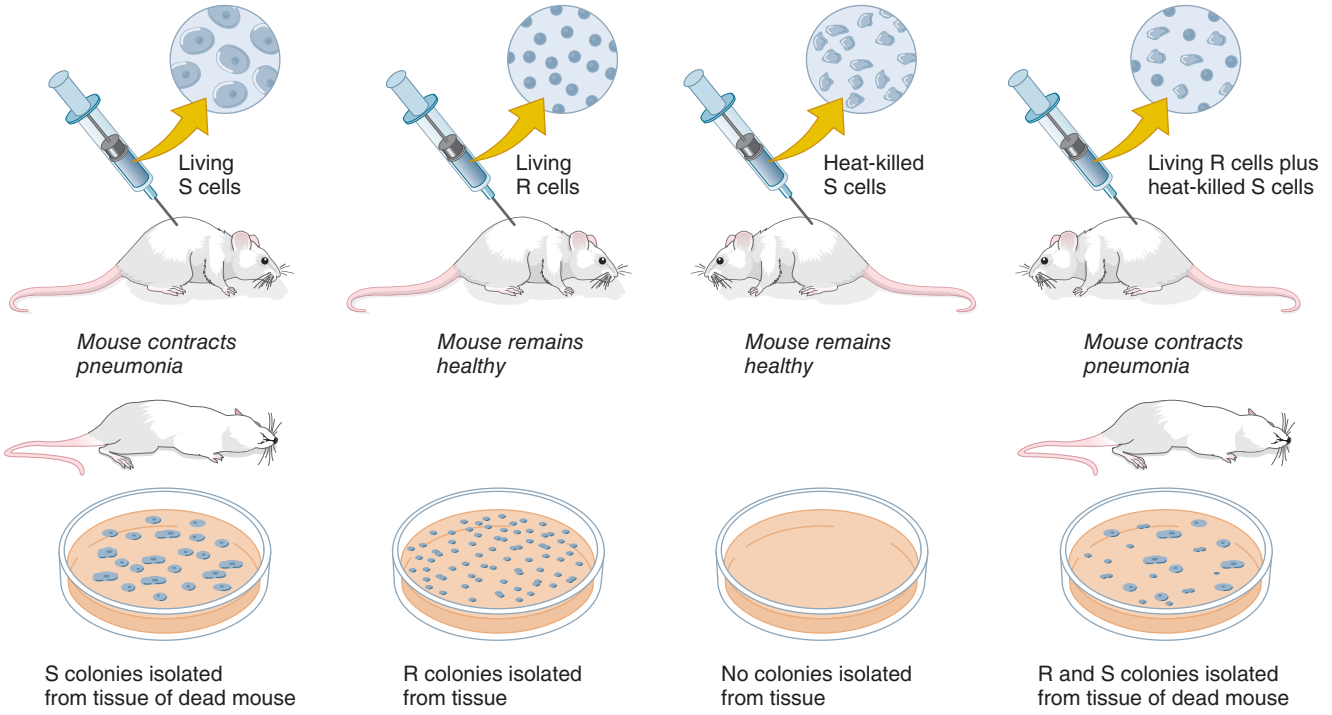


FIGURE 1.2 Griffith's experiment demonstrating bacterial transformation. A mouse remains healthy if injected with either the nonvirulent R strain of *S. pneumoniae* or heat-killed cell fragments of the usually virulent S strain. R cells in the presence of heat-killed S cells are transformed into the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

the transforming activity (**FIGURE 1.3**). These experiments implied that the substance responsible for genetic transformation was the DNA of the cell—and hence that DNA is the genetic material.

Transmission of DNA is the link between generations.

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

The intricate color patterns on butterfly wings demonstrate the complexity that can evolve in developmental processes.

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

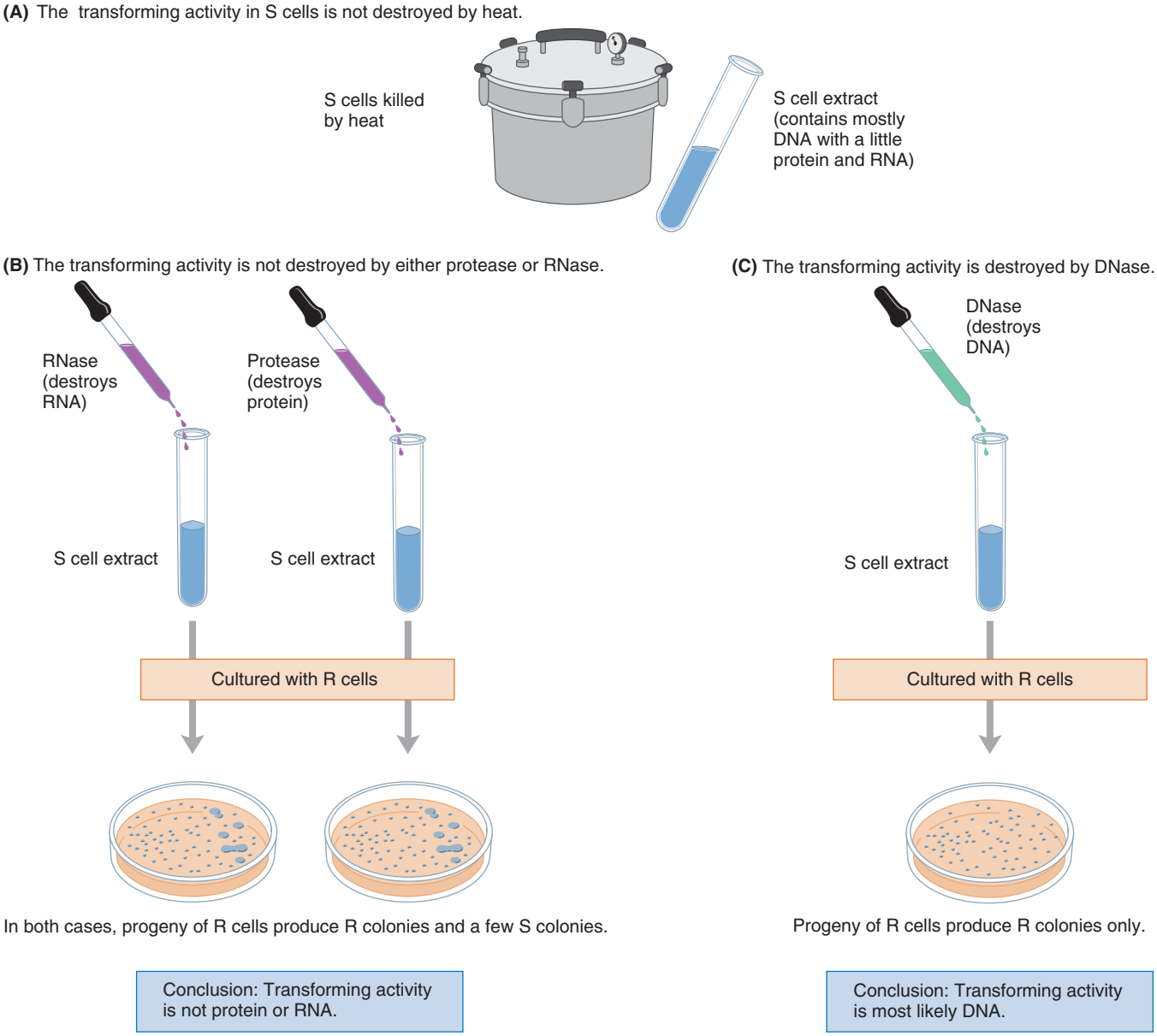


FIGURE 1.3 A diagram of the experiment demonstrating that DNA is the active material in bacterial transformation. (A) Purified DNA extracted from heat-killed S cells can convert some living R cells into S cells, but the extract may still contain undetectable traces of protein and/or RNA. (B) The transforming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

and then collecting the phage progeny. They obtained other particles containing labeled proteins in the same way, using medium that included ^{35}S (a radioactive isotope of sulfur).

In the experiments summarized in **FIGURE 1.4**, nonradioactive *E. coli* cells were infected with phage labeled with *either* ^{32}P (part A) *or* ^{35}S (part B) in order to follow the DNA and proteins separately. Infected cells were separated from unattached phage particles by centrifugation, resuspended in fresh medium, and then swirled violently in a kitchen blender to shear attached phage material from the cell surfaces. This treatment was found to have no effect on the subsequent course of the infection, which implies that the genetic material must enter the infected cells very soon after phage attachment. The kitchen blender turned out to be the critical piece of equipment. Other methods had been tried to tear the phage heads from the bacterial cell surface, but nothing had worked reliably. Hershey later explained, "We tried various grinding arrangements, with results that weren't very encouraging. When Margaret McDonald loaned us her kitchen blender, the experiment promptly succeeded."

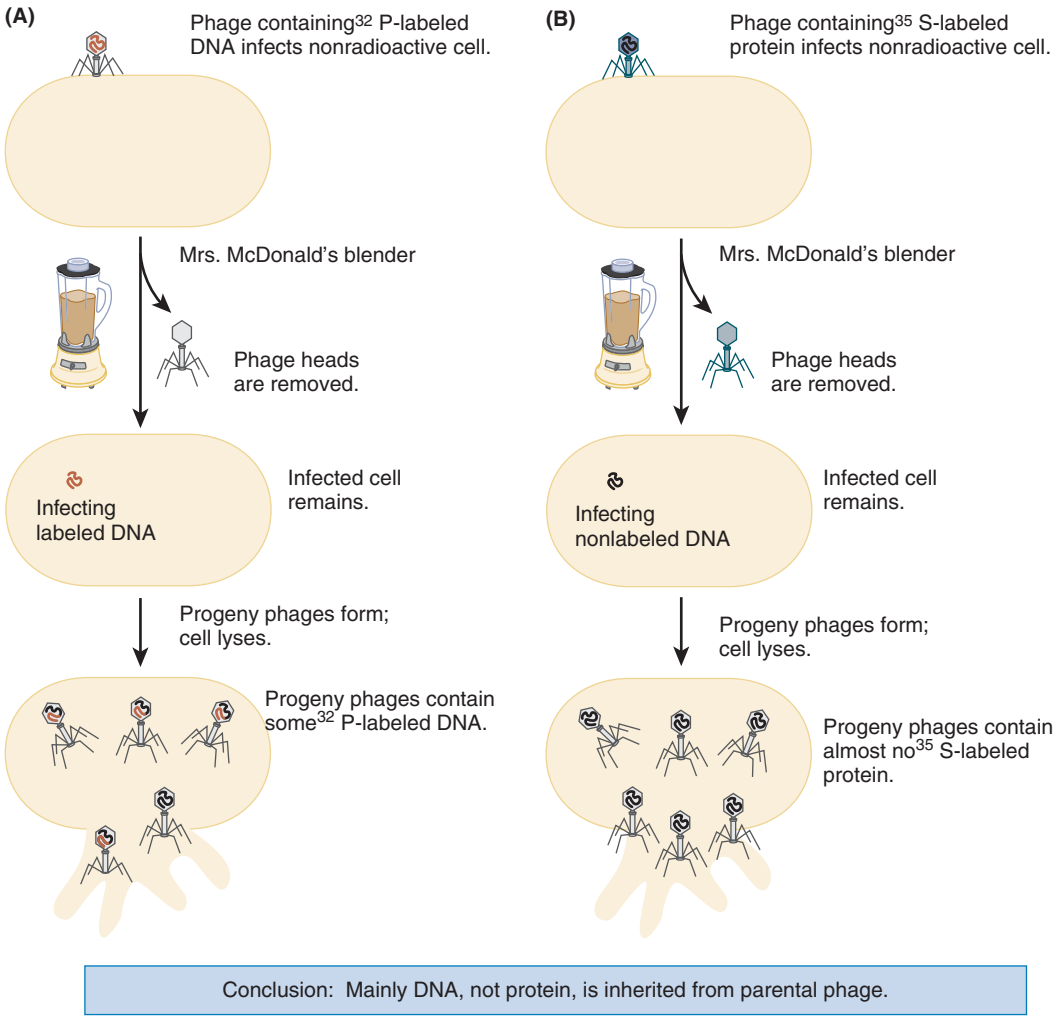


FIGURE 1.4 The Hershey–Chase ("blender") experiment, which demonstrated that DNA, not protein, is responsible for directing the reproduction of phage T2 in infected *E. coli* cells. (A) Radioactive DNA is transmitted to progeny phage in substantial amounts. (B) Radioactive protein is transmitted to progeny phage in negligible amounts.

After the phage heads were removed by blending, the infected bacteria were examined. Most of the radioactivity from ^{32}P-labeled phage was found to be associated with the bacteria, whereas only a small fraction of the ^{35}S radioactivity was present in the infected cells. The retention of most of the labeled DNA, contrasted with the loss of most of the labeled protein, implied that a T2 phage transfers most of its DNA, but very little of its protein, to the cell it infects. The critical finding (Figure 1.4) was that about 50 percent of the transferred ^{32}P-labeled DNA, but less than 1 percent of the transferred ^{35}S-labeled protein, was inherited by the *progeny* phage particles. Hershey and Chase interpreted this result to mean that the genetic material in T2 phage is DNA.

The transformation experiment and the Hershey–Chase experiment are regarded as classic demonstration that genes consist of DNA. At the present time, the equivalent of the transformation experiment is carried out daily in many research laboratories throughout the world, usually with bacteria, yeast, or animal or plant cells grown in culture. These experiments indicate that DNA is the genetic material in these organisms as well as in phage T2.

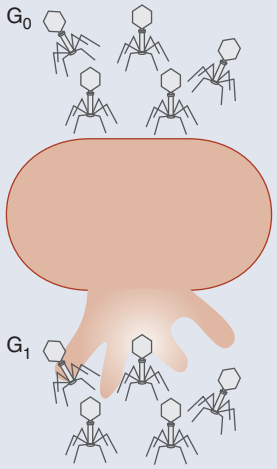
KEY CONCEPT

There are no known exceptions to the generalization that DNA is the genetic material in all cellular organisms.

It is worth noting, however, that in a few types of viruses, the genetic material consists of the other type of nucleic acid called RNA.

STOP & THINK 1.1

In this diagram of the Hershey–Chase experiment, G_0 represents the original population of bacteriophage with radioactive DNA. The progeny bacteriophage (G_1) showed half the amount of radioactivity in their DNA as the G_0 bacteriophage did.



(a) If the G_1 bacteriophage were used to infect bacteria, what fraction of the original G_0 radioactivity would be present in their progeny (the G_2 bacteriophage)?
(b) Suppose that after each cycle of infection, the progeny bacteriophage are used to initiate the next cycle of infection. What fraction of the original G_0 radioactivity would be present in the G_5 bacteriophage?

1.2 The structure of DNA is a double helix composed of two intertwined strands.

Even after it was shown that genes consist of DNA, many questions remained. How is the DNA in a gene duplicated when a cell divides? How does the DNA in a gene control a hereditary trait? What happens to the DNA when a mutation (a change in the DNA) takes place in a gene? Important clues to the answers to these questions emerged from the discovery of the three-dimensional structure of the DNA molecule itself. This structure is discussed next.

A central feature of double-stranded DNA is complementary base pairing.

In the early 1950s, a number of researchers began to try to understand the detailed molecular structure of DNA. The first essentially correct three-dimensional structure of the DNA molecule was proposed in 1953 by James Watson and Francis Crick at Cambridge University. The structure was dazzling in its elegance and revolutionary in suggesting how DNA duplicates itself, controls hereditary traits, and undergoes mutation. Even while their tin sheet and wire model of the DNA molecule was still incomplete, Crick announced in his favorite pub that, "We have discovered the secret of life."

In the Watson–Crick structure, DNA consists of two long chains of subunits twisted around one another to form a double-stranded helix. The double helix is right-handed, which means that as one looks along the barrel, each chain follows a clockwise path as it progresses. You can see the right-handed coiling in part A of FIGURE 1.5 if you imagine yourself looking up into the structure from the bottom: The smaller spheres outline the "backbone" of each individual strand, and they coil in a clockwise direction. The subunits of each strand are **nucleotides**, each of which contains any one of four chemical constituents called **bases**. The four bases in DNA are

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

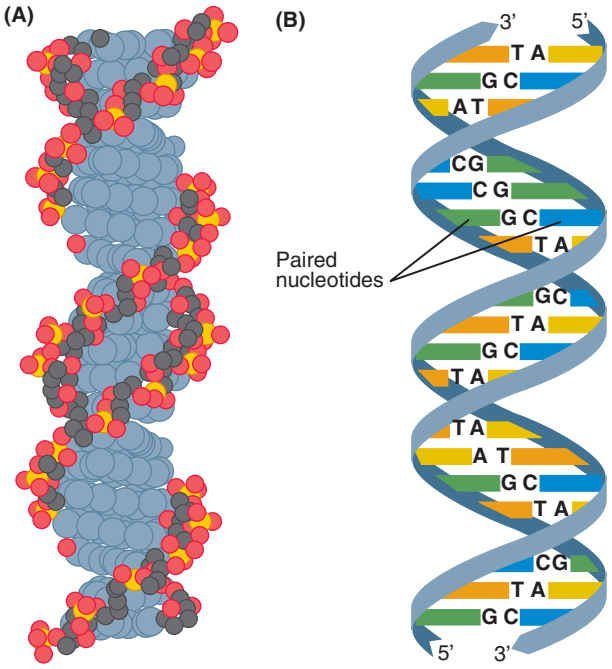


FIGURE 1.5 Molecular structure of a DNA double helix. (A) A "space-filling" model, in which each atom is depicted as a sphere. (B) A diagram highlighting the helical backbones on the outside of the molecule and the stacked A—T and G—C base pairs inside.

The chemical structures of the nucleotides and bases need not concern us at this point. A key point for our present purposes is that the bases in the double helix are paired as shown in Figure 1.5, part B. That is,

KEY CONCEPT

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

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

KEY CONCEPT

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

This principle explains how only four bases in DNA can code for the huge amount of information needed to make an organism. It is the linear order or *sequence* of bases along the DNA that encodes the genetic information, and the sequence is completely unrestricted.

The complementary pairing is also called *Watson–Crick base pairing*. In the three-dimensional structure (Figure 1.5, part A), the base pairs are represented by the larger spheres filling the interior of the double helix. The base pairs lie almost flat, stacked on top of one another perpendicular to the long axis of the double helix, like pennies in a roll. When discussing a DNA molecule, biologists frequently refer to the individual strands as **single-stranded DNA** and to the double helix as **double-stranded DNA** or **duplex DNA**.

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

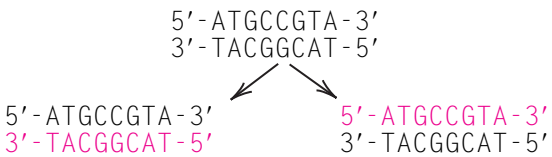
Beyond the most optimistic hopes, knowledge of the structure of DNA immediately gave clues to its function:

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

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

In replication, each parental DNA strand directs the synthesis of a new partner strand.

"It has not escaped our notice," wrote Watson and Crick, "that the specific base pairing we have postulated immediately suggests a copying mechanism for the genetic material." The copying process in which a single DNA molecule becomes two identical molecules is called **replication**. The replication mechanism that Watson and Crick had in mind is illustrated in **FIGURE 1.6**. The strands of the original (parent) duplex separate, and each individual strand serves as a pattern, or **template**, for the synthesis of a new strand (replica). The replica strands are synthesized by the addition of successive nucleotides in such a way that each base in the replica is complementary (in the Watson–Crick pairing sense) to the base across the way in the template strand. Although the mechanism in Figure 1.6 is simple in principle, it is a complex process that is fraught with geometric problems and requires a variety of enzymes and other proteins. The end result of replication is that a single double-stranded molecule becomes replicated into two copies with identical sequences:



Here the bases in the newly synthesized strands are shown in red. In the duplex on the left, the top strand is the template from the parental molecule and the bottom strand is newly synthesized; in the duplex on the right, the bottom strand is the template from the parental molecule and the top strand is newly synthesized.

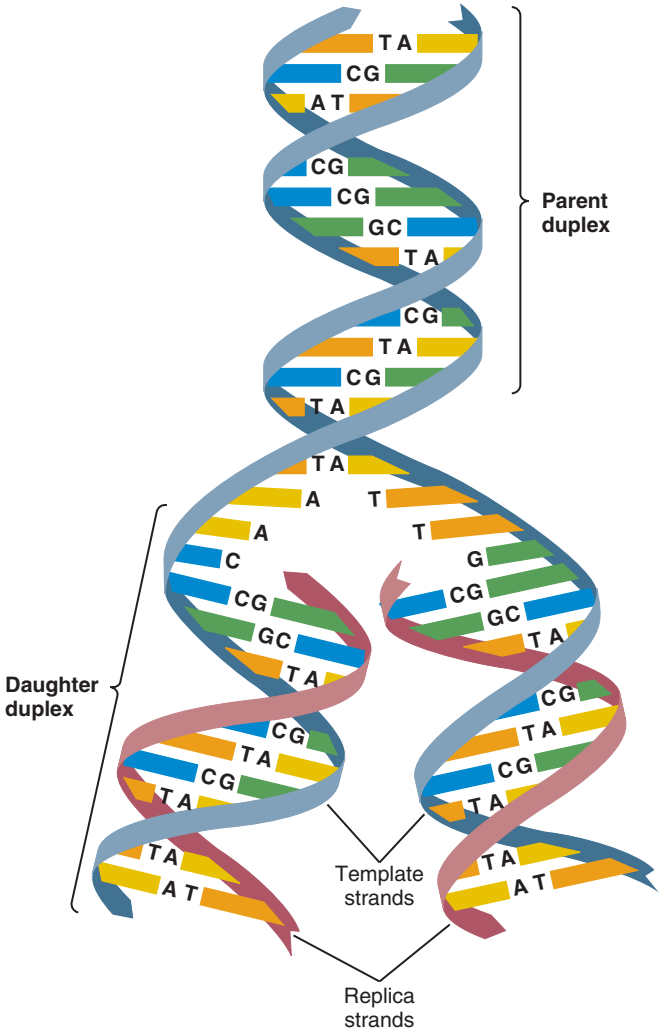


FIGURE 1.6 Replication in a long DNA duplex as originally proposed by Watson and Crick. The parental strands separate, and each parental strand serves as a template for the formation of a new daughter strand by means of A—T and G—C base pairing.

STOP & THINK 1.2

Shown here is part of the base sequence in one strand in a DNA duplex undergoing replication.

5′-TAGCAAAAATAGC-3′

What is the base sequence in the daughter strand?

1.3 Genes affect organisms through the action of proteins.

One of the important principles of molecular genetics is that genes exert their effects on organisms indirectly. For most genes, the genetic information contained in the nucleotide sequence specifies a particular type of *protein*. Proteins control the chemical and physical processes of cells known as **metabolism**. Many proteins are **enzymes**, a term introduced in 1878 to refer to the biological catalysts that accelerate biochemical reactions. Enzymes are essential for the breakdown of organic molecules, generating the chemical energy needed for cellular activities; they are also essential for the synthesis of small molecules and for their assembly into larger molecules and complex cellular structures.

Although the fundamental connection between genes and proteins was not widely appreciated until the 1940s, the first evidence for a relationship came much earlier. The pioneering observations were made by Archibald Garrod, a British physician, who studied genetic diseases caused by inherited defects in metabolism. He concluded that an inherited defect in metabolism results from an inherited defect in an enzyme. The key observations on which Garrod based this conclusion are summarized in the following sections.

Enzyme defects result in inborn errors of metabolism.

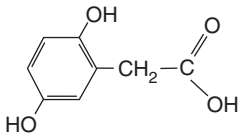
In 1908 Garrod gave a series of lectures in which he proposed this fundamental hypothesis about the relationship between enzymes and disease:

KEY CONCEPT

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

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

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



This is a conventional chemical representation in which each corner of the hexagon represents a carbon atom, and hydrogen atoms attached to the ring are not shown. The six-carbon ring is called a *phenyl* ring. An early name for homogentisic acid was *alkapton*—hence the name *alkaptonuria*. Even though alkaptonuria is rare, with an incidence of about one in 200,000 people, it was well known even before Garrod studied it. The disease itself is relatively mild, but it has one striking symptom: The urine of the patient turns black because

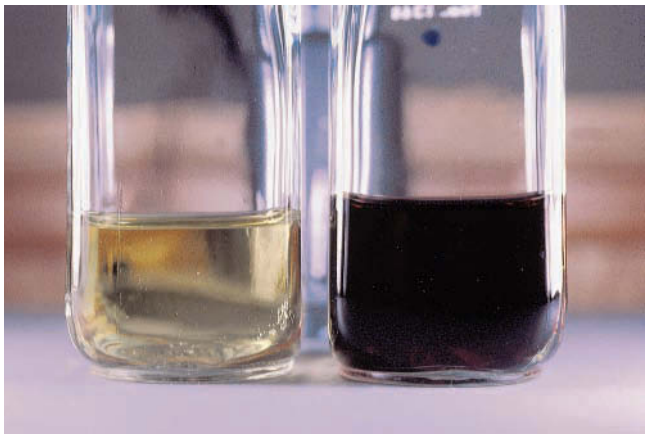


FIGURE 1.7 Urine from a person with alkaptonuria turns black because of the oxidation of the homogentisic acid that it contains.

Courtesy Daniel De Aguiar.

of the oxidation of homogentisic acid (**FIGURE 1.7**). This is why alkaptonuria is also called *black urine disease*. The passing of black urine can hardly escape being noticed. One case was described in the year 1649:

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

Garrod was primarily interested in the biochemistry of alkaptonuria, but he took note of family studies that indicated that the disease was inherited as though it were due to a defect in a single gene. As to the biochemistry, he deduced that the problem in alkaptonuria was the patients' inability to break down the phenyl ring of six carbons that is present in homogentisic acid. Where does this ring come from? Mammals are unable to synthesize it and must obtain it from their diet. Garrod proposed that homogentisic acid originates as a breakdown product of two amino acids, phenylalanine and tyrosine, which also contain a phenyl ring. An **amino acid** is one of the "building blocks" from which proteins are made. Phenylalanine and tyrosine are constituents of normal proteins. The scheme that illustrates the relationship between the molecules is shown in **FIGURE 1.8**. Any such sequence of biochemical

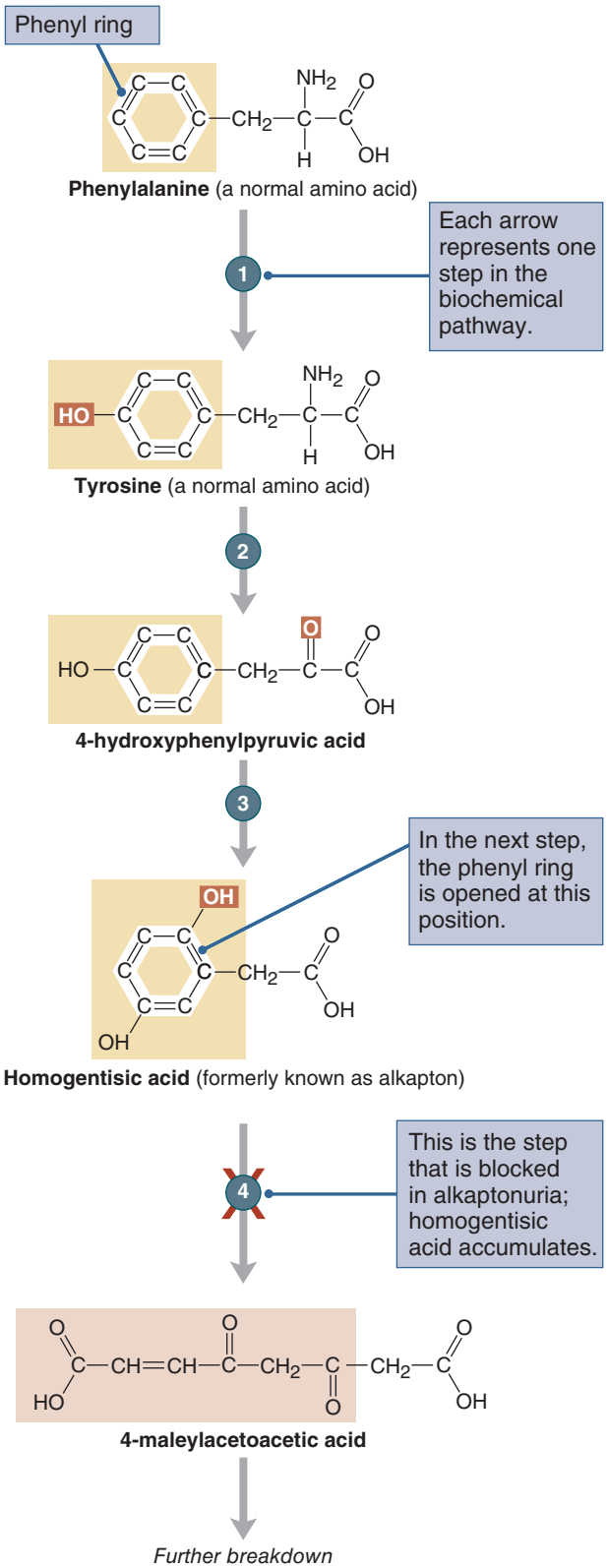


FIGURE 1.8 Metabolic pathway for the breakdown of phenylalanine and tyrosine. Each step in the pathway, represented by an arrow, requires a particular enzyme to catalyze the reaction. The key step in the breakdown of homogentisic acid is the breaking open of the phenyl ring.

reactions is called a **biochemical pathway** or a **metabolic pathway**. Each arrow in the pathway represents a single step depicting the transition from the "input" or **substrate molecule**, shown at the tail of the arrow, to the "output" or **product molecule**, shown at the tip. Biochemical pathways are usually oriented either vertically with the arrows pointing down, as in Figure 1.8, or horizontally, with the arrows pointing from left to right. Garrod did not know all of the details of the pathway in Figure 1.8, but he did understand that the key step in the breakdown of homogentisic acid is the breaking open of the phenyl ring and that the phenyl ring in homogentisic acid comes from dietary phenylalanine and tyrosine.

What allows each step in a biochemical pathway to occur? Garrod's insight was to see that each step requires a specific enzyme to catalyze the reaction and allow the chemical transformation to take place. Persons with an inborn error of metabolism, such as alkaptonuria, have a defect in one step of a metabolic pathway because they lack a functional enzyme for that step. When an enzyme in a pathway is defective, the pathway is said to have a **block** at that step. One frequent result of a blocked pathway is that the substrate of the defective enzyme accumulates. Observing the accumulation of homogentisic acid in patients with alkaptonuria, Garrod proposed that there must be an enzyme whose function is to open the phenyl ring of homogentisic acid and that this enzyme is missing in these patients. Discovery of all the enzymes in the pathway in Figure 1.8 took a long time. The enzyme that opens the phenyl ring of homogentisic acid was not actually isolated until 50 years after Garrod's lectures. In normal people it is found in cells of the liver. Just as Garrod had predicted, the enzyme is defective in patients with alkaptonuria.

The pathway for the breakdown of phenylalanine and tyrosine, as it is understood today, is shown in **FIGURE 1.9**. In this figure the emphasis is on the enzymes rather than on the structures of the **metabolites**, or small molecules, on which the enzymes act. As Garrod would have predicted, each step in the pathway requires the presence of a particular enzyme that catalyzes that step. Although Garrod knew only about alkaptonuria, in which the defective enzyme is homogentisic acid 1,2-dioxygenase, we now know the clinical consequences of defects in the other enzymes. Unlike alkaptonuria, which is a relatively benign inherited disease, the others are very serious. The condition known as **phenylketonuria (PKU)** results from the absence of (or a defect in) the enzyme **phenylalanine hydroxylase (PAH)**. When this step in the pathway is blocked, phenylalanine accumulates. The excess phenylalanine is broken down into harmful metabolites that cause defects in myelin formation that damage a child's developing nervous system and lead to severe mental retardation.

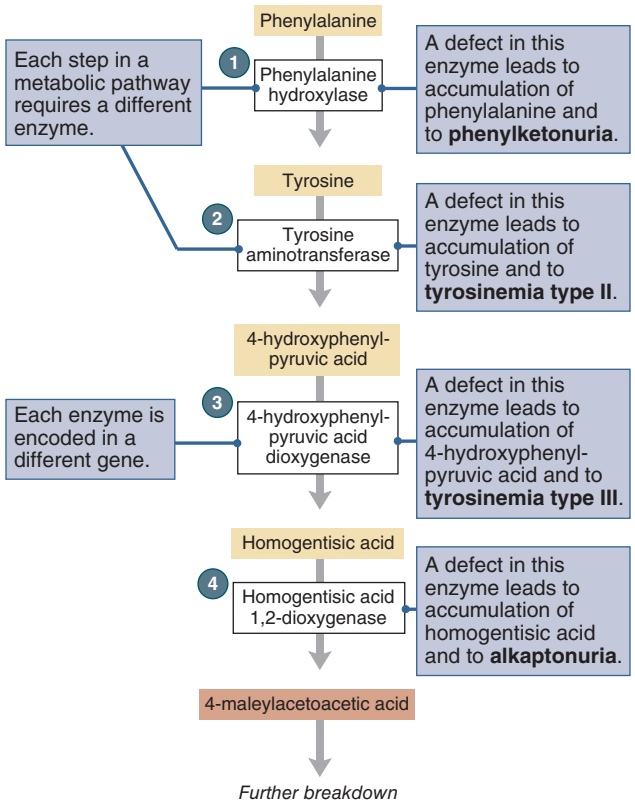


FIGURE 1.9 Inborn errors of metabolism in the breakdown of phenylalanine and tyrosine. A different inherited disease results when each of the enzymes is missing or defective. Alkaptonuria results from a defective homogentisic acid 1,2-dioxygenase, phenylketonuria from a defective phenylalanine hydroxylase.

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

In the metabolic pathway in Figure 1.9, defects in the breakdown of tyrosine or of 4-hydroxyphenylpyruvic acid lead to types of tyrosinemia. These are also severe diseases. Type II is associated with skin lesions and mental retardation, type III with severe liver dysfunction.

A defective enzyme results from a mutant gene.

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

The genes for the enzymes in the biochemical pathway in Figure 1.9 have all been identified and the nucleotide sequence of the DNA determined. In the following list, and throughout this text, we use the typographical convention that the names of *genes* are printed in *italic* type, whereas gene products are printed in regular type. In Figure 1.9 the numbers 1 through 4 correspond to the following genes and enzymes:

1. The gene *PAH* on the long arm of chromosome 12 encodes phenylalanine hydroxylase (PAH).
2. The gene *TAT* on the long arm of chromosome 16 encodes tyrosine aminotransferase (TAT).
3. The gene *HPD* on the long arm of chromosome 12 encodes 4-hydroxyphenylpyruvic acid dioxygenase (HPD).
4. The gene *HGD* on the long arm of chromosome 3 encodes homogentisic acid 1,2-dioxygenase (HGD).

Genetic analysis led to the one gene–one enzyme hypothesis.

Garrod's thinking was far ahead of his time, and his conclusions about inborn errors of metabolism were largely ignored. The influential experiments connecting genes with enzymes were carried out in the 1940s by George W. Beadle and Edward L. Tatum using a filamentous fungus *Neurospora crassa*, commonly called red bread mold, an organism they chose because both genetic and biochemical analysis could be done with ease. In these experiments they identified new mutations that each caused a block in the metabolic pathway for the synthesis of some needed nutrient and showed that each of these blocks corresponded to a defective enzyme needed for one step in the pathway. The experimental approach, now called **genetic analysis**, was important because it solidified the link between genetics and biochemistry. Equally as important, the experimental approach is widely applicable to understanding any complex biological process, ranging from the genetic control of the cell cycle or cancer to that of development or behavior. For this reason the methods of genetic analysis warrant a closer examination.

N. crassa grows in the form of filaments on a great variety of substrates including laboratory medium containing only inorganic salts, a sugar, and one vitamin. Such a medium is known as a **minimal medium** because it contains only the nutrients that are essential for growth of the organism. The filaments consist of a mass of branched threads separated into interconnected, multinucleate compartments allowing free interchange of nuclei and cytoplasm. Each nucleus contains a single set of seven chromosomes. Beadle and Tatum recognized that the ability of *Neurospora* to grow in minimal medium implied that the organism must be able to synthesize all of the other small molecules needed for growth, such as amino acids. If the biosynthetic pathways needed for growth are controlled by genes, then a mutation in a gene responsible for synthesizing an essential nutrient would be expected to render a strain unable to grow unless the strain were provided with the nutrient.

These ideas were tested in the following way. Spores of nonmutant *Neurospora* were irradiated with either x-rays or ultraviolet light to produce mutant strains with various nutritional requirements. The isolation of a set of mutants affecting any biological process, in this case metabolism, is called a **mutant screen**. In the initial step for identifying mutants, summarized in **FIGURE 1.10**, the irradiated spores (purple) were used in crosses with an untreated strain (green). Ascospores produced by the sexual cycle in fruiting bodies were individually germinated in **complete medium**, a complex medium enriched with a variety of amino acids, vitamins, and other substances expected to be essential metabolites whose synthesis could be blocked by a mutation. Even those ascospores containing a new mutation affecting synthesis of an essential nutrient would be expected to germinate and grow in complete medium.

To identify which of the irradiated ascospores contained a new mutation affecting the synthesis of an essential nutrient, spores from each culture were transferred to minimal medium (**FIGURE 1.11,** Part A). The vast majority of cultures yielded spores that could grow on minimal medium; these cultures lacked any new mutation of the desired type and were discarded. The cultures that were kept were the small number producing spores unable to grow on minimal medium, because these were mutant cultures that contained a new mutation blocking the synthesis of some essential nutrient.

THE HUMAN CONNECTION

One Gene, One Enzyme

George W. Beadle and Edward L. Tatum (1941)
Stanford University, Stanford, California

Genetic Control of Biochemical Reactions in Neurospora

How do genes control metabolic processes? The suggestion that genes control enzymes was made very early in the history of genetics, most notably by the British physician Archibald Garrod in his 1903 book, *Inborn Errors of Metabolism*. Nevertheless, the precise relationship between genes and enzymes was still uncertain. Perhaps each enzyme is controlled by more than one gene, or perhaps each gene contributes to the control of several enzymes. The classic experiments of Beadle and Tatum showed that the relationship is usually remarkably simple: One gene codes for one enzyme. Their pioneering experiments united genetics and biochemistry, and for the "one gene, one enzyme" concept, Beadle and Tatum were awarded a Nobel Prize in 1958 (Joshua Lederberg shared the prize for his contributions to microbial genetics). Because we now know that some enzymes contain polypeptide chains encoded by two (or occasionally more) different genes, a more accurate statement of the principle is "one gene, one polypeptide." Beadle and Tatum's experiments also demonstrate the importance of choosing the right organism. *Neurospora* had been introduced as a genetic organism only a few years earlier, and Beadle and Tatum realized that they could take advantage of this organism's ability to grow on a simple medium composed of known substances.

> ***"These preliminary results appear to us to indicate that the approach may offer considerable promise as a method of learning more about how genes regulate development and function."***

Beadle and Tatum's work was published in 1941 and can be found at the reference at the end of this feature. In it, they point out the limitations of starting with the physiological basis of a trait (such as black urine disease) and attempting to determine its genetic basis. First, these analyses are limited to traits in which the variants are nonlethal. Second, the variants must have visible effects. To get around these problems, Beadle and Tatum turned the problem on its head.

> [These limitations] have led us to investigate the general problem of the genetic control of development and metabolic reactions by reversing the ordinary procedure . . . [by setting out] to determine if and how genes control known biochemical reactions . . . If the organism must be able to carry out a certain chemical reaction to survive on a given medium, a mutant unable to do this will obviously be lethal on this medium. . . . [It can be] studied, however, if it will grow on a medium to which has been added the essential product of the genetically blocked reaction. . . .

Thus, rather than starting with observed differences in traits among individuals, Beadle and Tatum started by generating mutations (in their case, mutations resulting from x-irradiation of *Neurospora* cells), then identified the mutations that were lethal on minimal medium but not on medium supplemented with the normal product of the mutated gene. This experimental approach ranks among the most important experimental tool of genetic analysis.

G. W. Beadle and E. L. Tatum, Genetic control of biochemical reactions in *Neurospora*. *Proc. Natl. Acad. Sci. USA* 27 (1941): 499–506.

Spores from each mutant culture were then transferred to a series of media to determine whether the mutation results in a requirement for a vitamin, an amino acid, or some other substance. In the example illustrated in Figure 1.11, Part B, the mutant strain requires one (or possibly more than one) amino acid, because a mixture of all amino acids added to the minimal medium allows growth. Because the proportion of irradiated cultures with new mutations was very small, only a negligible number of cultures would contain two or more new mutations that had occurred simultaneously.

For nutritional mutants requiring amino acids, further experiments testing each of the amino acids

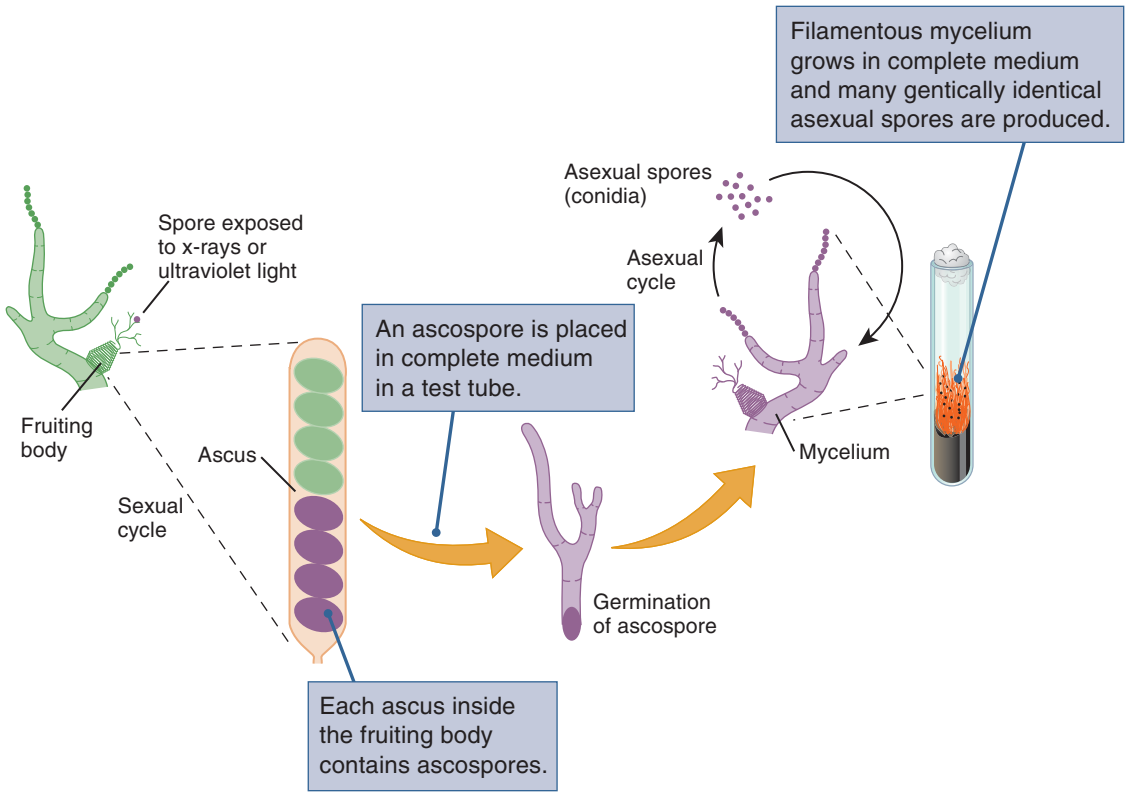


FIGURE 1.10 Beadle and Tatum obtained mutants of the filamentous fungus *Neurospora crassa* by exposing asexual spores to x-rays or ultraviolet light. The treated spores were used to start the sexual cycle in fruiting bodies. After any pair of cells and their nuclei undergo fusion, meiosis takes place almost immediately and results in eight sexual spores (ascospores) included in a single ascus. These are removed individually and cultured in complete medium. Ascospores that carry new nutritional mutants are identified later by their inability to grow in minimal medium.

individually usually revealed that only one amino acid was required to be added to minimal medium to support growth. In Figure 1.11, Part C, the mutant strain requires the amino acid arginine. Even in the 1940s some of the possible intermediates in amino acid biosynthesis had been identified. These were recognized by their chemical resemblance to the amino acid and by being present at low levels in the cells of organisms. In the case of arginine, two candidates were ornithine and citrulline. All mutants requiring arginine were, therefore, tested in medium supplemented with either ornithine alone or citrulline alone (Figure 1.11, Part D). One class of arginine-requiring mutants, designated Class I, was able to grow in minimal medium supplemented with either ornithine, citrulline, or arginine. Other mutants, designated Class II, were able to grow in minimal medium supplemented with either citrulline or arginine but not ornithine. A third class, Class III, was able to grow only in minimal medium supplemented with arginine.

The types of arginine-requiring mutants illustrate the principle of genetic analysis as applied to metabolic pathways. The basic principle is that

KEY CONCEPT

If a strain with a mutant enzyme that blocks a particular step in a linear metabolic pathway can grow when an intermediate is added to the growth medium, it means that the location of the intermediate in the pathway is *downstream* of the enzymatic step that is blocked.

This principle makes intuitive sense because, if the intermediate were upstream of the metabolic block, then adding the intermediate to the growth medium would not allow growth, because conversion of the intermediate would still be blocked at the point of the mutant enzyme.

Application of the principle to the linear pathway for arginine biosynthesis is shown in **FIGURE 1.12**, where arginine is the end product starting with some precursor metabolite, and ornithine and citrulline are intermediates in the pathway. The mutants imply the order of the intermediates shown because

- Mutants in Class I are able to grow in the presence of either ornithine or citrulline, which means that

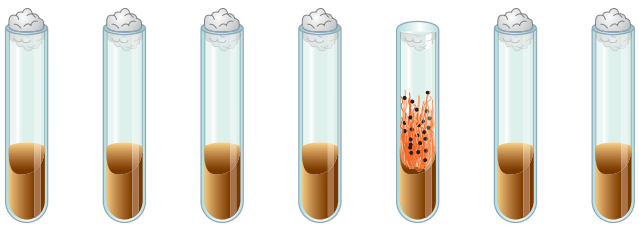


FIGURE 1.11 (A) Mutant spores can grow in complete medium but not in minimal medium. (B) Each new mutant is tested for growth in minimal medium supplemented with a mixture of nutrients. (C) Mutants that can grow on minimal medium supplemented with amino acid are tested with each amino acid individually. (D) Mutants unable to grow in the absence of arginine are tested with likely precursors of arginine.

both ornithine and citrulline are downstream of any of the enzymes blocked in Class I mutants.

- Mutants in Class II are able to grow in the presence of citrulline but not ornithine, which means that citrulline is located downstream of the enzymatic block in Class II mutants and that ornithine is upstream of the metabolic block in Class II mutants.
- Mutants in Class III are unable to grow in the presence of either citrulline or ornithine, which means that these intermediates are upstream of any of the enzymatic steps blocked in Class III mutants.

The structure of the pathway in Figure 1.12 was further confirmed by the observations that Class III mutants accumulate citrulline and Class II mutants accumulate ornithine. Ultimately, direct biochemical experiments demonstrated that the inferred enzymes were actually present in nonmutant strains but defective in mutant strains.

STOP & THINK 1.3

Suppose you do a mutant screen for *Neurospora* mutants unable to grow on minimal medium unless they are supplemented with an amino acid we'll call "A." Based on their molecular structures, you surmise that two molecules, X and Y, are intermediates in the biochemical pathway for the synthesis of A, but you are unsure which of the following pathways may be correct:

(a) X → Y → A

(b) Y → X → A

You find two classes of mutants that require A for growth.

Class 1 grows on minimal medium supplemented with A but not with X or Y.

Class 2 grows on minimal medium supplemented with A or X but not with Y.

Which of the pathways **(a)** or **(b)** do these data support?

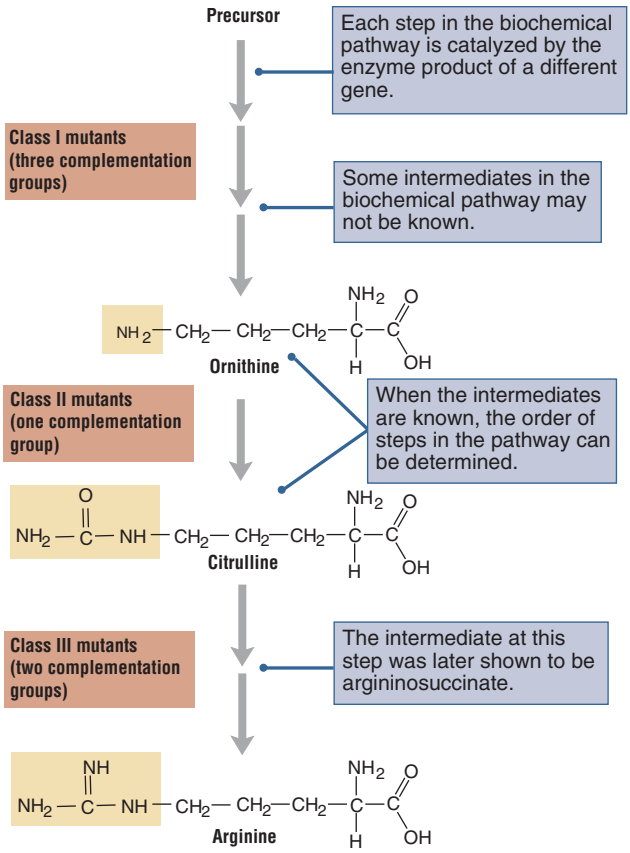


FIGURE 1.12 Metabolic pathway for arginine biosynthesis inferred from genetic analysis of *Neurospora* mutants.

Mutant screens sometimes isolate different mutations in the same gene.

Beadle and Tatum were fortunate to study metabolic pathways in a relatively simple organism in which each gene specifies a single enzyme, a relation often called the **one gene–one enzyme hypothesis**. In such a situation, genetic analysis of the mutants reveals a great deal more about the metabolic pathway than merely the order of the intermediates. By classifying each mutation according to the particular gene it is in and grouping all the mutations in each gene together, each set of mutations and, therefore, each individual gene, correspond to one enzymatic step in the metabolic pathway. In Figure 1.12, for example, Class I includes mutations in any of three different genes, which implies that there are three steps in the pathway between the precursor

and ornithine. Similarly, Class III comprises mutations in either of two different genes, which implies that there are two steps in the pathway between the citrulline and arginine. However, Class II consists of mutations in only one gene, which implies only one step in the pathway between ornithine and citrulline.

Mutations that have defects in the same gene are identified by means of a **complementation test**, in which two mutations are brought together into the same cell. In most multicellular organisms (and even some sexual unicellular organisms), the usual way to do this is by means of a mating. When two parents, each carrying one of the two mutations, are crossed, fertilization brings the reproductive cells containing the two mutations together, and through ordinary cell division each cell in the resulting offspring carries one copy of each mutant gene. In *Neurospora* this procedure does not work because nuclear fusion is followed almost immediately by the formation of ascospores, each of which has only one set of chromosomes.

Complementation tests are nevertheless possible in *Neurospora* owing to the multinucleate nature of the filaments. Certain strains, including those studied by Beadle and Tatum, have the property that when the filaments from two mutant organisms come into physical contact, the filaments fuse and the new filament contains multiple nuclei from both of the participating partners. This sort of hybrid filament is called a **heterokaryon**, and it contains mutant forms of both genes. The word roots of the term *heterokaryon* mean "different nuclei." (A list of the most common word roots used in genetics can be found at the end of the book.)

When a heterokaryon formed from two nutritional mutants is inoculated into minimal medium, it may grow or it may fail to grow. If it grows in minimal medium, the mutant genes are said to undergo **complementation**, and this result indicates that the mutations are in different genes. On the other hand, if the heterokaryon fails to grow in minimal medium, the result indicates **noncomplementation**, and the two mutations are inferred to be in the same gene.

The inferences from complementation or noncomplementation emerge from the logic illustrated in **FIGURE 1.13**. Here the multinucleate filament is shown, and the mutant nuclei are color coded according to which of two different genes (red or purple) is mutant. The thick red and purple horizontal lines represent the proteins encoded in the mutant nuclei, and the × represents a defect in the protein resulting from a mutation in the corresponding gene.

Part A depicts the situation in which the mutant strains have mutations in different genes. In the heterokaryon, the red nuclei produce mutant forms of the red protein and normal forms of the purple protein, whereas the purple nuclei produce mutant forms of the purple protein and normal forms of the red protein. The result is that the red/purple heterokaryon has normal forms of both the red and purple proteins. It also has mutant forms of both proteins, but these do not matter. What matters is that the normal proteins allow the heterokaryon to grow on minimal medium because all needed nutrients can be synthesized. In other words, the normal purple gene in the red nucleus complements the defective purple gene in the purple nucleus, and the other way around. The logic of complementation is captured in the ancient nursery rhyme "Jack Sprat could eat no fat / His wife could eat no lean / And so between the two of them / They licked the platter clean," because each partner makes up for the defect in the other.

Part B in Figure 1.13 shows a heterokaryon formed between mutants with defects in the same gene, in this case purple. Both of the purple nuclei encode a normal form of the red protein, but each purple nucleus encodes a defective purple protein. When the nuclei are together, two different mutant forms of the purple protein are produced, and so the biosynthetic pathway that requires the purple protein is still blocked, and the heterokaryon is unable to grow in minimal medium. In other words, the mutants 2 and 3 in Figure 1.14 fail to complement, and so they are judged to have mutations in the same gene.

The following principle underlies the complementation test.

KEY CONCEPT

The Principle of Complementation: A complementation test brings two mutant genes together in the same cell or organism. If this cell or organism is nonmutant, the mutations are said to *complement* one another and it means that the parental strains have mutations in *different* genes. If the cell or organism is mutant, the mutations fail to complement one another, and it means that the parental mutations are in the *same* gene.

A complementation test identifies mutations in the same gene.

In the mutant screen for *Neurospora* mutants requiring arginine, Beadle and Tatum found that mutants in different classes (Class I, Class II, and Class III in Figure 1.12) always complemented one another. This result makes sense, because the genes in each class encode enzymes that act at different levels between the known intermediates. However, some of the mutants

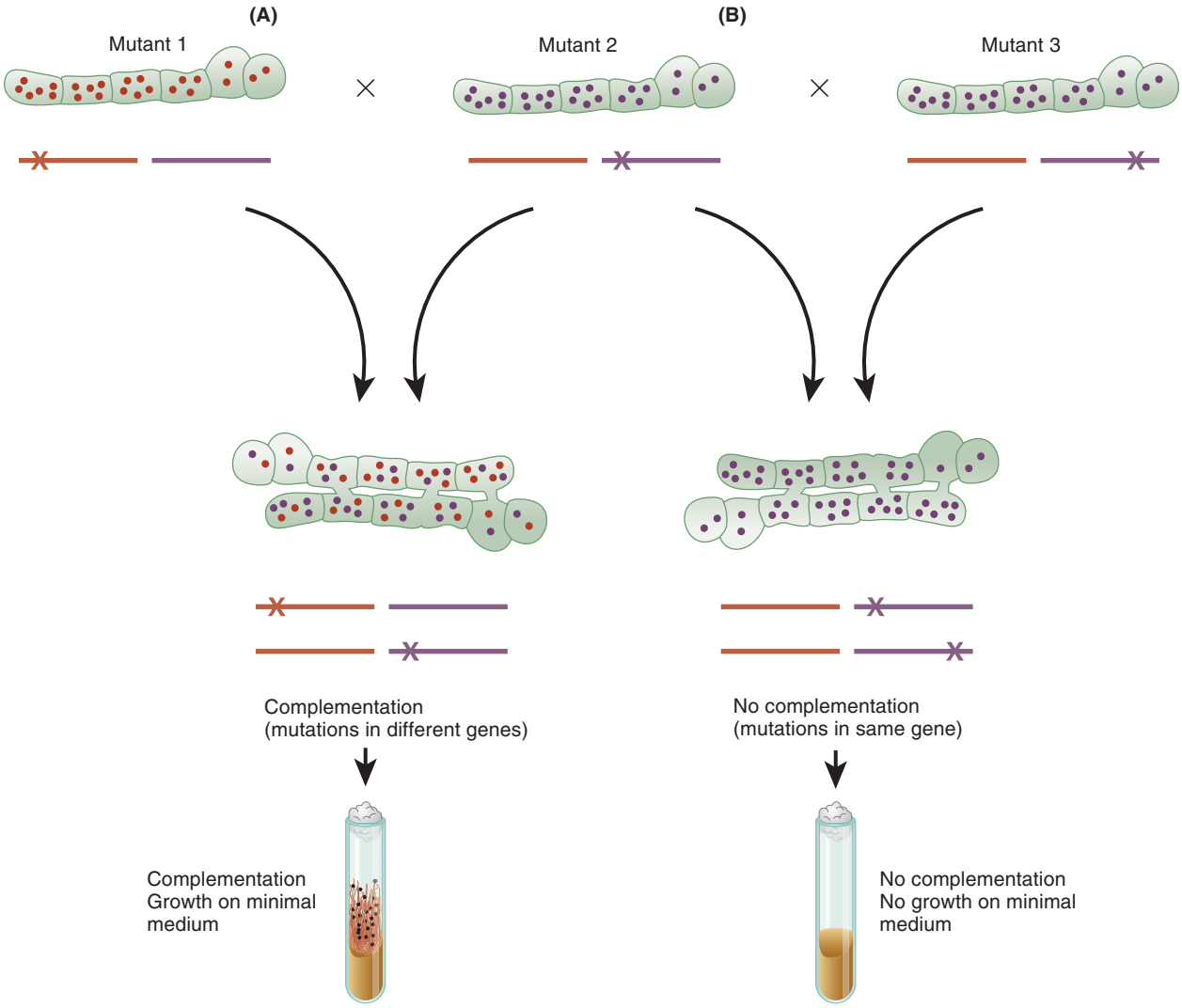


FIGURE 1.13 Molecular interpretation of a complementation test using heterokaryons to determine whether two mutant strains have mutations in different genes (A) or mutations in the same gene (B). In (A) each nucleus contributes a nonmutant form of one or the other polypeptide chain, and so the heterokaryon is able to grow in minimal medium. In (B) both nuclei contribute a mutant form of the same polypeptide chain; hence, no nonmutant form of that polypeptide can be synthesized and the heterokaryon is unable to grow in minimal medium.

in Class I failed to complement others in Class I, and some in Class III failed to complement others in Class III. These results allow the number of genes in each class to be identified.

To illustrate this aspect of genetic analysis, we consider six mutant strains in Class III. These strains were taken in pairs to form heterokaryons and their growth on minimal medium assessed. The data are shown in **FIGURE 1.14,** Part A. The mutant genes in the six strains are denoted *x*1, *x*2, and so forth, and the data are presented in the form of a matrix in which + indicates growth in minimal medium (complementation) and − indicates lack of growth in minimal medium (lack of complementation). The diagonal entries are all −, which reflects the fact that two copies of the identical mutation cannot show complementation. The pattern of + and − signs in the matrix indicate that mutations *x*1 and *x*5 fail to complement one another; hence, *x*1 and *x*5 are mutations in the same gene. Likewise, mutations *x*2, *x*3, *x*4, and *x*6 fail to complement one another in all possible pairs; hence, *x*2, *x*3, *x*4, and *x*6 are all mutations in the same gene (but a different gene from that represented by *x*1 and *x*5).

Data in a complementation matrix can conveniently be analyzed by arranging the mutant genes in the form of a circle as shown in Figure 1.14, Part B. Then, for each possible pair of mutations, connect the pair by a straight line if the mutations *fail* to complement (− signs in part A). According to the principle of complementation, these lines connect mutations that are in the same gene. Each of the groups of noncomplementing mutations is called a **complementation group**. As we have seen, each

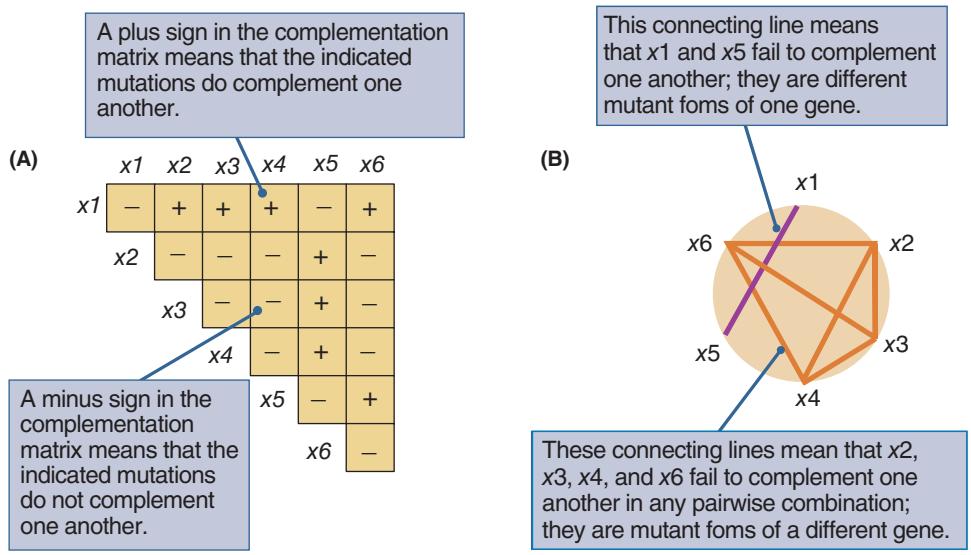


FIGURE 1.14 (A) Results of complementation tests. (B) To interpret the results, arrange the mutations in a circle. Connect by a straight line any pair of mutations that fail to complement—that is, that yield a mutant heterokaryon. Any pair of mutations connected by a straight line are mutations in the same gene, and are more than likely mutations at different nucleotide sites in the gene. This example shows two complementation groups, each of which represents a single gene needed for arginine biosynthesis.

complementation group defines a gene, so the complementation test actually provides the geneticist's operational definition:

KEY CONCEPT

A *gene* is defined experimentally as a set of mutations that make up a single complementation group. Any pair of mutations within a complementation group fail to complement one another.

The mutations in Figure 1.14, therefore, represent two genes, mutation of any one of which results in the inability of the strain to convert citrulline to arginine. On the basis of the one gene–one enzyme hypothesis, which is largely true for metabolic enzymes in *Neurospora*, the pathway from citrulline to arginine in Figure 1.12 must comprise two steps with an unknown intermediate in between. This intermediate was later found to be argininosuccinate. Likewise, Class I mutants defined three complementation groups; hence, there are three enzymatic steps from the precursor to ornithine. These intermediates were also soon identified. Finally, Class II mutations all failed to complement one another, and the finding of only one complementation group means that there is but a single enzymatic step that converts ornithine to citrulline.

STOP & THINK 1.4

Among mutations affecting a metabolic pathway in *Neurospora*, one class of mutants blocks the conversion of W into Z. These mutants can grow on minimal medium when supplemented with Z, but they can't when supplemented with W. You carry out complementation tests with six such mutants (*m1, m2, ..., m6*) and find the complementation matrix shown here.

	m1	m2	m3	m4	m5	m6
m1	–	+	+	–	+	+
m2		–	–	+	+	–
m3			–	+	+	–
m4				–	+	+
m5					–	+
m6						–

(a) How many different genes are indicated by these results?
(b) If each gene codes for a different enzyme in the pathway, how many enzymatic steps are there in the conversion of W into Z?

Genetic analysis can be applied to the study of any complex biological process.

The type of genetic analysis pioneered by Beadle and Tatum is immensely powerful for identifying the genetic control of complex biological processes. Their approach lays out a systematic path—a sort of recipe—for gene discovery. First, decide what process you want to study. Next figure out what characteristics mutant organisms with a disruption in that process would display. Then do a mutant screen for mutants showing these characteristics. Carry out complementation tests to find out how many different genes that you have identified. And finally, find out what the products of those genes are, what they do, how they interact with each other, and in what order they function.

Beadle and Tatum themselves analyzed many metabolic pathways for a wide variety of essential nutrients, but their experiments were especially important in deciphering pathways of amino acid biosynthesis. Their findings over just a few years are said to have "contributed more knowledge of amino acid biosynthetic pathways than had been accumulated during decades of traditional study." They were awarded the 1958 Nobel Prize in Physiology or Medicine for their research, and in the intervening years at least nine more Nobel Prizes in Physiology or Medicine were awarded in which genetic analysis carried out along the lines of Beadle and Tatum played a significant role. Here is a list, with quotations from the official citations of the Nobel Foundation.

- 1958—George Beadle and Edward Tatum "for their discovery that genes act by regulating definite chemical events." (It was in doing literature research for his Nobel Prize Lecture that Beadle discovered Garrod's earlier work and brought it to the world's attention.)
- 1965—François Jacob, André Lwoff, and Jacques Monod "for their discoveries concerning genetic control of enzyme and virus synthesis"
- 1995—Edward B. Lewis, Christiane Nüsslein-Volhard, and Eric F. Wieschaus "for their discoveries concerning the genetic control of early embryonic development"
- 2000—Leland H. Hartwell, Tim Hunt, and Sir Paul Nurse "for their discoveries of key regulators of the cell cycle"
- 2002—Sydney Brenner, H. Robert Horvitz, and John E. Sulston "for their discoveries concerning genetic regulation of organ development and programmed cell death"
- 2007—Mario R. Capecchi, Martin J. Evans, and Oliver Smithies "for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells"
- 2009—Elizabeth H. Blackburn, Carol W. Greider, and Jack W. Szostak "for the discovery of how chromosomes are protected by telomeres and the enzyme telomerase"
- 2013—James E. Rothman, Randy W. Schekman, and Thomas C. Südhof "for their discoveries of machinery regulating vesicle traffic, a major transport system in our cells"
- 2015—Tomas Lindahl, Paul Modrich, and Aziz Sancar "for mechanistic studies of DNA repair"
- 2017—Jeffrey C. Hall, Michael Rosbash, and Michael W. Young "for their discoveries of molecular mechanisms controlling the circadian rhythm"

1.4 Genes specify proteins by means of a genetic code.

The Beadle and Tatum experiments established that a gene specifies the structure of an enzyme but left open the issue of how this happens. We now know that the relationship between genes and proteins is indirect. The genetic information that specifies a protein is actually contained in the sequence of bases in DNA in a manner analogous to letters printed on a strip of paper. In a region of DNA that directs the synthesis of a protein, the genetic code for the protein is contained in only one strand, and it is decoded in a linear order. The result of protein synthesis is a polypeptide chain, which consists of a linear sequence of amino acids connected end to end. Each polypeptide chain folds into a characteristic three-dimensional configuration that is determined by its particular sequence of amino acids. A typical protein is made up of one or more polypeptide chains. Many proteins function as enzymes that participate in metabolic processes such as amino acid biosynthesis.

One of the DNA strands directs the synthesis of a molecule of RNA.

The details of how genes code for proteins were not understood until the 1960s, and an outline of the process is shown in **FIGURE 1.15**. The decoding of the genetic information takes place in two distinct steps known as *transcription* and *translation*. The indirect route of information transfer

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

is known as the **central dogma** of molecular genetics. The term *dogma* means "set of beliefs"; it dates from

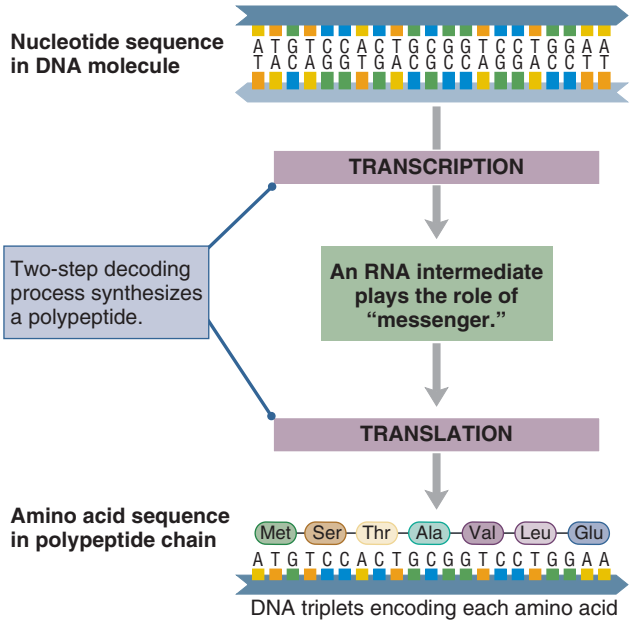


FIGURE 1.15 DNA sequence coding for the first seven amino acids in a polypeptide chain. The DNA sequence specifies the amino acid sequence through a molecule of RNA that serves as an intermediary "messenger." Although the decoding process is indirect, the net result is that each amino acid in the polypeptide chain is specified by a group of three adjacent bases in the DNA. In this example, the polypeptide chain is that of phenylalanine hydroxylase (PAH).

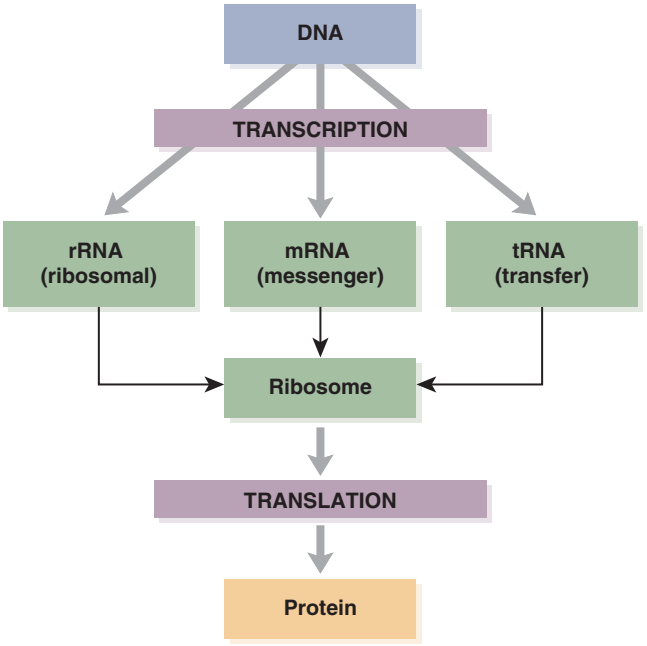


FIGURE 1.16 The "central dogma" of molecular genetics: DNA codes for RNA, and RNA codes for proteins. The DNA → RNA step is transcription, and the RNA → protein step is translation.

the time the idea was first put forward as hypothesis. Since then the "dogma" has been confirmed experimentally, but the term persists. The main concept in the central dogma is that DNA does not code for protein directly but rather acts through an intermediary molecule of **ribonucleic acid (RNA)**. The structure of RNA is similar to, but not identical with, that of DNA. There is a difference in the sugar (RNA contains the sugar **ribose** instead of deoxyribose), RNA is usually single stranded (not a duplex), and RNA contains the base **uracil (U)** instead of thymine (T), which is present in DNA. Three types of RNA take part in the synthesis of proteins:

- A molecule of **messenger RNA (mRNA)**, which carries the genetic information from DNA and is used as a template for polypeptide synthesis. In most mRNA molecules, a relatively high proportion of the nucleotides actually code for amino acids. For example, the mRNA for phenylalanine hydroxylase is 2400 nucleotides in length and codes for a polypeptide of 452 amino acids; in this case, more than 50 percent of the length of the mRNA codes for amino acids.
- Three types of **ribosomal RNA (rRNA)**, which are major constituents of the cellular particles called **ribosomes** on which polypeptide synthesis takes place.
- A set of about 45 **transfer RNA (tRNA)** molecules, each of which carries a particular amino acid as well as a three-base recognition region that base-pairs with a group of three adjacent bases in the mRNA. As each tRNA participates in translation, its amino acid becomes the terminal subunit of the growing polypeptide chain. A tRNA that carries methionine is denoted $tRNA^{Met}$, one that carries serine is denoted $tRNA^{Ser}$, and so forth. (Because there are more than 20 different tRNAs, but only 20 amino acids, some amino acids can be attached to any of several tRNAs.)

The central dogma illustrated in **FIGURE 1.16** is the fundamental principle of molecular genetics because it summarizes how the genetic information in DNA becomes expressed in the amino acid sequence in a polypeptide chain.

KEY CONCEPT

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

The manner in which genetic information is transferred from DNA to RNA is shown in **FIGURE 1.17**.

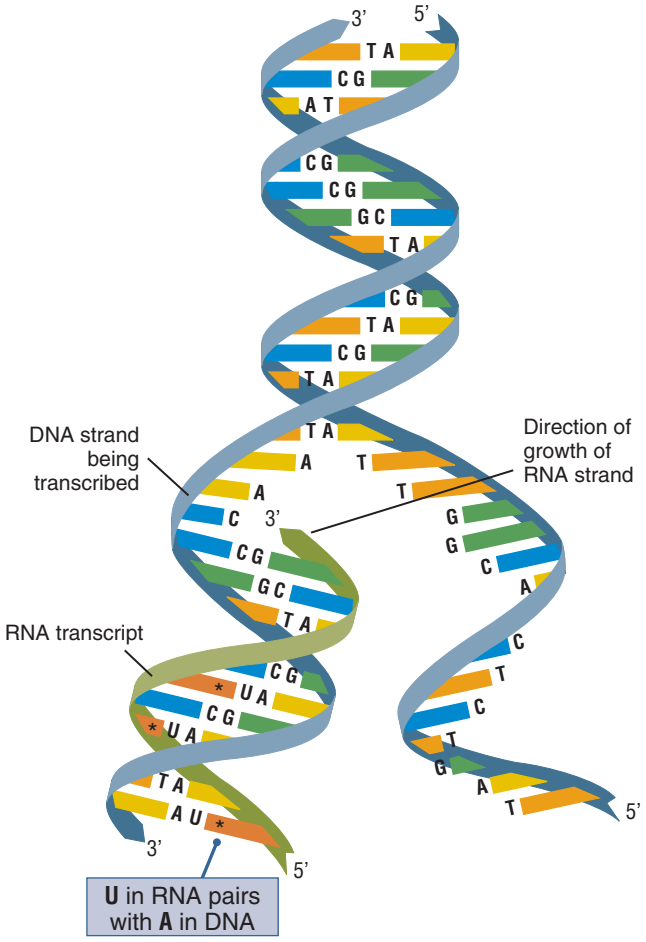
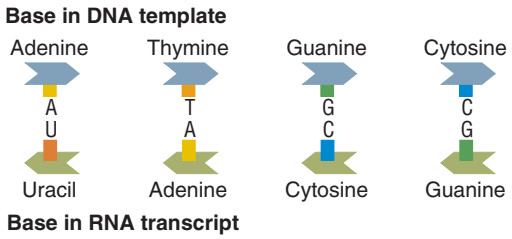


FIGURE 1.17 Transcription is the production of an RNA strand that is complementary in base sequence to a DNA strand. In this example, a DNA strand is being transcribed into an RNA strand at the bottom left. Note that in an RNA molecule, the base U (uracil) plays the role of T (thymine) in that it pairs with A (adenine). Each A–U pair is marked with an asterisk.

The DNA opens up, and one of the strands is used as a template for the synthesis of a complementary strand of RNA. The process of making an RNA strand from a DNA template is **transcription**, and the RNA molecule that is made is the **transcript**. The base sequence in the RNA is complementary (in the Watson–Crick pairing sense) to that in the DNA template, except that U (which pairs with A) is present in the RNA in place of T. The rules of base pairing between DNA and RNA are summarized below.



Like DNA, an RNA strand also exhibits polarity; its 5′ and 3′ ends are determined by the orientation of the nucleotides. The 5′ end of the RNA transcript is synthesized first, and in the RNA–DNA duplex formed in transcription, the polarity of the RNA strand is opposite to that of the DNA strand. Each gene includes particular nucleotide sequences that initiate and terminate transcription. The RNA transcript made from any gene begins at an initiation site in the template strand, which is located "upstream" from the amino acid coding region, and ends at a termination site, which is located "downstream" from the amino acid coding region. For any gene, the length of the RNA transcript is very much smaller than the length of the DNA in the entire chromosome. For example, the transcript of the *PAH* gene for phenylalanine hydroxylase is 90,000 nucleotides in length, but the DNA in all of chromosome 12 is about 130,000,000 nucleotide pairs. In this case, the length of the *PAH* transcript is less than 0.1 percent of the length of the DNA in the chromosome. A different gene in chromosome 12 would be transcribed from a different region of the DNA molecule in chromosome 12, but the transcribed region would again be small in comparison with the total length of the DNA in the chromosome.

A molecule of RNA directs the synthesis of a polypeptide chain.

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

The role of tRNA in translation is illustrated in **FIGURE 1.18** and can be described as follows:

KEY CONCEPT

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

The tRNA molecules used in translation do not line up along the mRNA simultaneously as shown in Figure 1.18. The process of translation takes place on a ribosome, which combines with a single mRNA and moves along it in steps, three nucleotides at a time (codon by codon). As each new codon comes into place, the next tRNA binds with the ribosome, and the growing end of the polypeptide chain becomes attached to the amino acid on the tRNA. In this way, each tRNA in turn serves temporarily to hold the polypeptide chain as it is being synthesized. As the polypeptide chain is transferred from each tRNA to the next in line, the tRNA that previously held the polypeptide is released from the ribosome. The polypeptide chain elongates one amino acid at a step until any one of three particular codons specifying "stop" is encountered. At this point, synthesis of the chain of amino acids is finished, and the polypeptide chain is released from the ribosome.

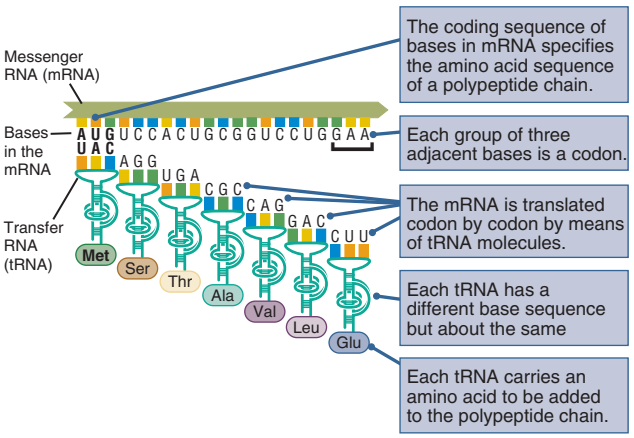


FIGURE 1.18 The role of messenger RNA in translation is to carry the information contained in a sequence of DNA bases to a ribosome, where it is translated into a polypeptide chain. Translation is mediated by transfer RNA (tRNA) molecules, each of which can base-pair with a group of three adjacent bases in the mRNA. Each tRNA also carries an amino acid; when it is brought to the ribosome by base pairing, its amino acid becomes the growing end of the polypeptide chain.

The genetic code is a triplet code.

Figure 1.18 indicates that the mRNA codon AUG specifies methionine (Met) in the polypeptide chain, UCC specifies Ser (serine), ACU specifies Thr (threonine), and so on. The complete decoding table is called the **genetic code**, and it is shown in **TABLE 1.1**. For any codon, the column on the left corresponds to the first nucleotide in the codon (reading from the 5′ end), the row across the top corresponds to the second nucleotide, and the column on the right corresponds

TABLE 1.1 The Standard Genetic Code

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

Codon — Three-letter and single-letter abbreviations

to the third nucleotide. The complete codon is given in the body of the table, along with the amino acid (or "stop") that the codon specifies. Each amino acid is designated by its full name as well as by a three-letter abbreviation and a single-letter abbreviation. Both types of abbreviations are used in molecular genetics. The code in Table 1.1 is the "standard" genetic code used in translation in the cells of nearly all organisms.

In addition to the 61 codons that code only for amino acids, there are 4 codons that have specialized functions:

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

How the genetic code table is used to infer the amino acid sequence of a polypeptide chain may be illustrated using phenylalanine hydroxylase again, in particular the DNA sequence coding for amino acid numbers 1 through 7. The DNA sequence is

```
5′-ATGTCCACTGCGGTCCTGGAA-3′
3′-TACAGGTGACGCCAGGACCTT-5′
```

This region is transcribed into RNA in a left-to-right direction, and because RNA grows by the addition of successive nucleotides to the 3′ end (Figure 1.17), it is the bottom strand that is transcribed. The nucleotide sequence of the RNA is that of the top strand of the DNA, except that U replaces T, so the mRNA for amino acids 1 through 7 is

```
5′-AUGUCCACUGCGGUCCUGGAA-3′
```

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

```
5′-AUGUCCACUGCGGUCCUGGAA-3′
   MetSerThrAlaValLeuGlu
```

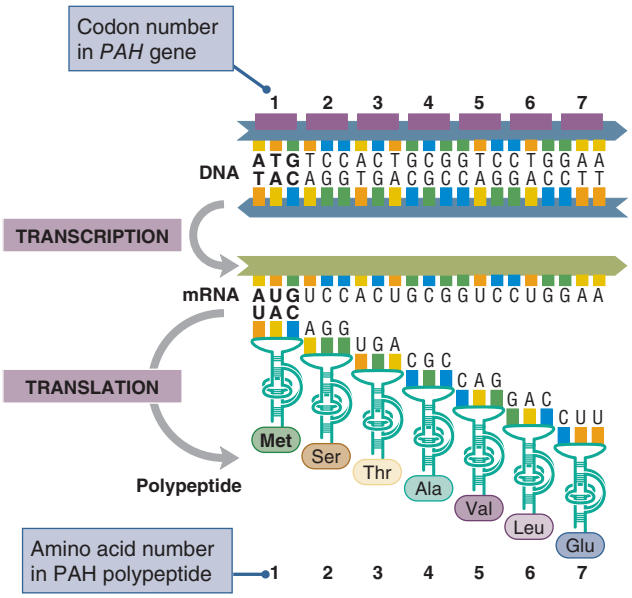


FIGURE 1.19 The central dogma in action. The DNA that encodes PAH serves as a template for the production of a messenger RNA, and the mRNA, in turn, serves to specify the sequence of amino acids in the PAH polypeptide chain through interactions with the tRNA molecules.

or, in terms of the single-letter abbreviations,

```
5′-AUGUCCACUGCGGUCCUGGAA-3′
   M  S  T  A  V  L  E
```

The full decoding operation for this region of the *PAH* gene is shown in **FIGURE 1.19**. In this figure, the initiation codon AUG is highlighted because some patients with PKU have a mutation in this particular codon. As might be expected from the fact that AUG is the initiation codon for polypeptide synthesis, cells in patients with this particular mutation fail to produce any of the PAH polypeptide. Mutation and its consequences are considered next.

1.5 Genes change by mutation.

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

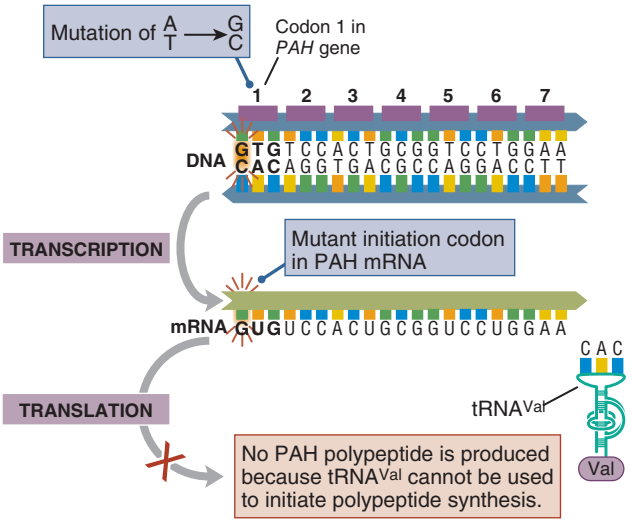


FIGURE 1.20 The M1V mutant in the *PAH* gene. The methionine codon needed for initiation mutates to a codon for valine. Translation cannot be initiated, and no PAH polypeptide is produced.

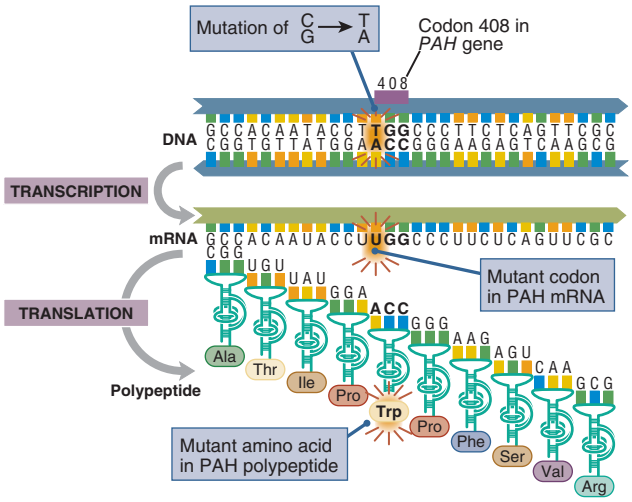


FIGURE 1.21 The R408W mutant in the *PAH* gene. Codon 408 for arginine (R) is mutated into a codon for tryptophan (W). The result is that position 408 in the mutant PAH polypeptide is occupied by tryptophan rather than by arginine. The mutant protein has no PAH enzyme activity.

organism that exhibits the effects of the mutation—for example, an inborn error of metabolism.

DNA from patients from all over the world who have phenylketonuria has been studied to determine what types of mutations are responsible for the inborn error. There are a large variety of mutant types. More than 400 different mutations have been described. In some cases part of the gene is missing, so the genetic information to make a complete PAH enzyme is absent. In other cases the genetic defect is more subtle, but the result is still either the failure to produce a PAH protein or the production of a PAH protein that is inactive. In the mutation shown in **FIGURE 1.20**, substitution of a G—C base pair for the normal A—T base pair at the very first position in the coding sequence changes the normal codon AUG (Met) used for the initiation of translation into the codon GUG, which normally specifies valine (Val) and cannot be used as a "start" codon. The result is that translation of the PAH mRNA cannot occur, so no PAH polypeptide is made. This mutant is designated M1V because the codon for M (methionine) at amino acid position 1 in the PAH polypeptide has been changed to a codon for V (valine). Although the M1V mutant is quite rare worldwide, it is common in some localities, such as in Québec province in Canada.

One PAH mutant that is quite common is designated R408W, which means that codon 408 in the PAH polypeptide chain has been changed from one coding for arginine (R) to one coding for tryptophan (W). This mutant is one of the four most common in cases of PKU among European Caucasians. The molecular basis of the mutation is shown in **FIGURE 1.21**. In this case, the first base pair in codon 408 is changed from a C—G base pair into a T—A base pair. The result is that the PAH mRNA has a mutant codon at position 408; specifically, it has UGG instead of CGG. Translation does occur in this mutant because everything else about the mRNA is normal, but the result is that the mutant PAH carries a tryptophan (Trp) instead of an arginine (Arg) at position 408 in the polypeptide chain. The consequence of the seemingly minor change of one amino acid is very drastic, because the mutant PAH has no enzyme activity and so is unable to catalyze its metabolic reaction. In other words, the mutant PAH protein is complete but inactive. With PAH, as with other proteins, some amino acid replacements result in a polypeptide chain that is unable to fold properly. The incorrectly folded polypeptides are digested by proteases in the cell, which recycles the amino acids for use in the synthesis of other proteins.

STOP & THINK 1.5

Suppose you discover a novel mutant form of the enzyme phenylalanine hydroxylase (PAH) with the amino acid replacement R408Q—that is, the amino acid arginine (R) normally found at position 408 in the protein is replaced in the mutant by the amino acid glutamine (G). What single nucleotide substitution in the DNA coding for arginine at position 408 would result in R408Q?

1.6 Most traits are complex traits affected by multiple genetic and environmental factors.

Inborn errors of metabolism illustrate the general principle that genes code for proteins and that mutant genes code for mutant proteins that can result in inherited diseases such as phenylketonuria. But few people have even met anyone with phenylketonuria because its frequency is only about 1 in 10,000 individuals. An inherited trait like phenylketonuria, which is due to mutations in a single gene, is called a **simple Mendelian trait** because it occurs in families according to simple genetic ratios first discovered by Gregor Mendel (Chapter 2). Although about 2000 simple Mendelian diseases have been described, each of them is quite rare in the human population as a whole (although some individual diseases are more common in particular subgroups).

Most traits that you will encounter in everyday life are not simple Mendelian traits. The most commonly encountered diseases include heart disease, diabetes, kidney disease, autism, and bipolar disorder. Almost everyone knows somebody who is affected with one or more of these conditions, often a family member. Each of these common conditions occurs in about 1/100 individuals—at least 100 times more frequent than a typical simple Mendelian disorder.

These common diseases are examples of **complex traits** because their causation is a complex interplay between multiple genetic and environmental factors. A disease in which causation is complex is affected by genetic factors, but each genetic factor, acting alone, does not determine presence of the disease. Each genetic factor is a **risk factor** that increases the chance that an individual carrying the gene will manifest the disease. Each risk factor may have a relatively small effect, but the risk factors are cumulative.

Complex traits are also affected by environmental factors and lifestyle choices. An individual may have multiple genetic risk factors for heart disease, for example, but still delay the onset of the disease, minimize its severity, or avoid it altogether with lifestyle choices like eating a healthy diet, getting regular exercise, and not smoking. Conversely, an individual with few genetic rick factors for heart disease may nevertheless come down with the disease owing to poor lifestyle choices in diet, exercise, and tobacco use.

Not all complex traits are diseases. Most commonly observed differences among individuals are due to variation in complex traits. Height and weight are two prominent examples. Both traits are affected by multiple genetic and environmental factors acting together. In the case of weight, obvious environmental factors are diet and exercise, but weight is also affected by the cumulative impact of at least 700 known genetic factors, each of small effect.

1.7 Evolution means continuity of life with change.

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

KEY CONCEPT

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

Groups of related organisms descend from a common ancestor.

Organisms share a common set of similar genes and proteins because they evolved by descent from a common ancestor. The process of **evolution** takes place when a population of organisms gradually changes in genetic composition through time. Evolutionary changes in genes and proteins result in differences in metabolism, development, and behavior among organisms, which allows them to become progressively better adapted to their environments. From an evolutionary perspective, the unity of fundamental molecular processes in organisms alive today reflects inheritance from a distant common ancestor in which the molecular mechanisms were already in place.

Not only the unity of life but also many other features of living organisms become comprehensible from an evolutionary perspective. For example, the interposition of an RNA intermediate in the basic flow of genetic information from DNA to RNA to protein makes sense if the earliest forms of life used RNA for both genetic information and enzyme catalysis. The importance of the evolutionary perspective in understanding aspects of biology that seem pointless or needlessly complex is summed up in a famous aphorism

of the evolutionary biologist Theodosius Dobzhansky: "Nothing in biology makes sense except in the light of evolution."

Biologists distinguish three major kingdoms of organisms:

1. ***Bacteria*** This group includes most bacteria and cyanobacteria (formerly called blue-green algae). Cells of these organisms lack a membrane-bounded nucleus and mitochondria, are surrounded by a cell wall, and divide by binary fission.

2. ***Archaea*** This group was initially discovered among microorganisms that produce methane gas or that live in extreme environments, such as hot springs or pools with high salt concentrations. They are widely distributed in more normal environments as well. Superficially resembling bacteria, the cells of archaeans show important differences in the manner in which their membrane lipids are chemically linked. The machinery for DNA replication and transcription in archaeans resembles that of eukaryans, whereas their metabolism strongly resembles that of bacteria. DNA sequence analysis indicates that about half of the genes found in the kingdom Archaea are unique to this group.

3. ***Eukarya*** This group includes all organisms whose cells contain an elaborate network of internal membranes, a membrane-bounded nucleus, and mitochondria. Their DNA is present in the form of linear molecules organized into true chromosomes, and cell division takes place by means of mitosis. The eukaryotes include plants and animals as well as fungi and many single-celled organisms, such as amoebae and ciliated protozoa.

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

The molecular unity of life is seen in comparisons of genomes.

The totality of DNA in a cell, nucleus, or organelle is called its **genome**. When used with reference to a species of organism, for example in phrases such as "the human genome," the term genome is defined as the DNA present in a normal reproductive cell.

Modern methods for sequencing DNA are so rapid and efficient that the complete DNA sequence is known for hundreds of different species of organisms. These include the genomes of multiple representatives of many groups of organisms, including extinct human ancestors sequenced from DNA extracted from fossil bones. **TABLE 1.2** shows a small sample of sequenced genomes. Genome size is given in megabases (Mb), or millions of base pairs.

The organism denoted syn3.0 is very special in that it is not a naturally occurring organism, and we will discuss it later. Among the naturally occurring organisms in Table 1.2 are two bacteria: *Mycoplasma mycoides* is notable for its small genome and limited number of genes; *Escherichia coli* is more typical of bacteria in size and gene number. Both bacteria have

TABLE 1.2 Comparison of Genomes

Organism	Genome size, Mb[a] (approximate)	Number of genes (approximate)
syn3.0 (synthetic DNA bacterium)	0.5	473
Mycoplasma mycoides (causes bovine pneumonia)	1.2	985
Escherichia coli (common colon bacterium)	4.6	4000
Saccharomyces cerevisiae (baker's yeast)	12	6000
Caenorhabditis elegans (soil nematode)	100	20,000
Drosophila melanogaster (fruit fly)	180	16,000
Arabidopsis thaliana (mouse-ear cress)	135	28,000
Mus musculus (laboratory mouse)	2500	25,000
Homo sapiens (human being)	3000	25,000

[a]Millions of base pairs.

about the same density of genes—about one gene per Mb of DNA.

Baker's yeast (*Saccharomyces cerevisiae*) is a single-celled eukaryote. It has a genome size of 12 Mb organized into 16 chromosomes containing about 6000 genes. The gene density is about one gene per 2 Mb—twice that of typical bacteria.

Caenorhabditis elegans, Drosophila melanogaster, and the diminutive flowering plant *Arabidopsis thaliana* are complex, multicellular eukaryotes notable for their relatively small genome size, which is still substantially larger than that of baker's yeast. Their gene number is also larger than that of yeast, but only by a factor of 3–4, with an average gene density of one gene per 5–10 Mb. (Not all insects have a genome as small as that of *Drosophila;* the genome size of the mountain grasshopper, *Podisma pedestris*, is about 100 times larger than that of *Drosophila*, but it has about the same number of genes. Such paradoxes of genome size are discussed in Chapter 6.)

The genomes of mouse and human are much larger than those of the other multicellular eukaryotes in Table 1.2, yet they have about the same number of genes. This means that the gene density in mouse and human is much reduced, to roughly one gene per 100 Mb. The reduced gene density is reflected in the fact that only about 1.5 percent of the human genome sequence codes for protein. (About 27 percent of the human genome is present in protein-coding genes, but much of the DNA sequence present in such genes does not actually code for amino acids.)

Which brings us back to syn3.0, the world's first synthetic organism. It was created to identify the minimal set of genes that would enable a bacterial cell to multiply in growth medium containing amino acids and other small-molecule nutrients. Starting with cells of *Mycoplasma mycoides*, researchers at the J. Craig Venter Institute in La Jolla, California systematically knocked out each of the 985 genes to determine which genes were essential for growth. They then chemically synthesized a 0.5-Mb DNA molecule that contained only essential genes. Synthesis of such a large piece of DNA is technically extremely difficult, because long molecules of DNA in solution are fragile and break easily due to mechanical shear. In practice, large molecules have to be synthesized in smaller pieces that must then be assembled in proper order. In the case of syn3.0, the researchers made clever use of living yeast cells to combine the synthetic pieces in the right order and to faithfully replicate the molecule.

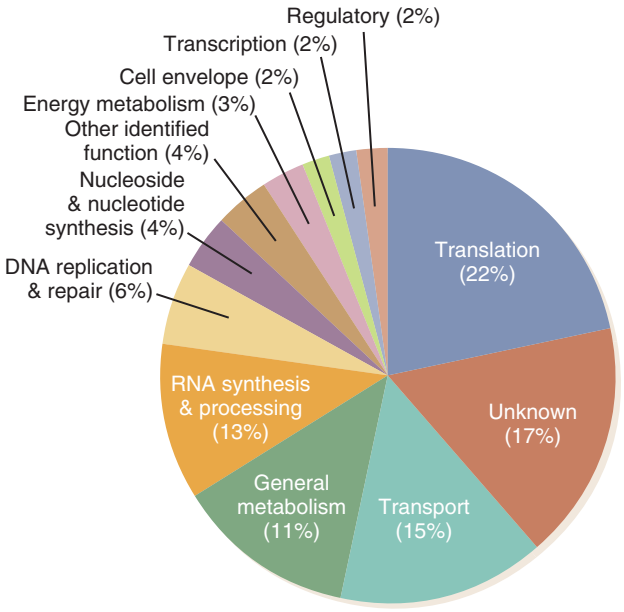


FIGURE 1.22 Functions of the 473 genes in the synthetic bacterium syn3.0.

Data from C. A. Hutchison III et al. *Science* 2016 Mar 25;351(6280):aad6253. doi: 10.1126/science.aad6253.

The completed genome was then tested for viability by being introduced into living *Mycoplasma* cells in which the original DNA had been destroyed.

The project was not without its surprises. There are, for example, more than a few cases in which each of two genes appears to be nonessential, but eliminating both results in cells that grow extremely slowly or not at all. A good analogy is the engines on a twin-engine jet like the Boeing 767: It can fly when one engine is disabled, but not both.

After much work and some trial and error, syn3.0—a viable cell with a synthetic minimal genome comprising 473 genes—was created. The functions of these 473 genes are summarized in **FIGURE 1.22**. As might be expected, the largest numbers of genes function in small-molecule transport, general metabolism, or synthesis and processing of macromolecules. Remarkably, 17 percent of the genes in the minimal genome have no identified function ("unknown"), which means a great deal of biology remains to be discovered. It is also worth noting that a substantial fraction of the genes in syn3.0 have recognizable counterparts in the other organisms listed in Table 1.2, attesting to the molecular unity of life on Earth.

CHAPTER SUMMARY

- Inherited traits are affected by genes.
- Genes are composed of the chemical deoxyribonucleic acid (DNA).
- DNA replicates to form (usually identical) copies of itself.
- DNA contains a code specifying what types of enzymes and other proteins are made in cells.
- DNA occasionally mutates, and the mutant forms specify altered proteins.
- A mutant enzyme is an "inborn error of metabolism" that blocks one step in a biochemical pathway for the metabolism of small molecules.
- Genetic analysis of mutants of the fungus *Neurospora* unable to synthesize an essential nutrient led to the one gene–one enzyme hypothesis.
- Different mutations in the same gene can be identified by means of a complementation test, in which the mutants are brought together in the same cell or organism. Mutations in the same gene fail to complement one another, whereas mutations in different genes show complementation.
- Most traits are complex traits affected by multiple genes as well as by environmental factors.
- Organisms change genetically through generations in the process of biological evolution.
- Because of their common descent, organisms share many features of their genetics and biochemistry.

ISSUES AND IDEAS

- What special feature of the structure of DNA allows each strand to be replicated without regard to the other?
- What does it mean to say that a strand of DNA specifies the structure of a molecule of RNA?
- What types of RNA participate in protein synthesis, and what is the role of each type of RNA?
- What is meant by the phrase *genetic code*, and how is the genetic code relevant to the translation of a polypeptide chain from a molecule of messenger RNA?
- What is meant by the term *genetic analysis*, and how is genetic analysis exemplified by the work of Beadle and Tatum using *Neurospora*?
- What is a complementation test, and what is it used for in genetic analysis?

SOLUTIONS: STEP BY STEP

PROBLEM 1 In the human gene for the beta chain of hemoglobin, the oxygen-carrying protein in the red blood cells, the first 30 nucleotides in the protein-coding region are as shown here.

3′-TACCACGTGGACTGAGGACTCCTCTTCAGA-5′

(a) What is the sequence of the partner strand?
(b) If the DNA duplex of this gene were transcribed from left to right, what is the base sequence of the RNA across this part of the coding region?
(c) What is the sequence of amino acids in this part of the beta-globin polypeptide chain?
(d) In the mutation responsible for sickle-cell anemia, the red T indicated is replaced with an A. The mutant is present at relatively high frequency in some human populations because carriers of the gene are more resistant to falciparum malaria than are noncarriers. What is the amino acid replacement associated with this mutation?

SOLUTION. (a) The partner strand is deduced from the rule that A pairs with T and G pairs with C; however, keep in mind that the paired DNA strands have opposite polarity (that is, their 5′-to-3′ orientations are reversed). (b) The RNA strand is synthesized in the 5′-to-3′ direction, which means that the template DNA strand is transcribed in the 3′-to-5′ direction, which happens to be the same left-to-right orientation of the strand shown above. The base sequence is deduced from the usual base-pairing rules, except that A in DNA pairs with U in RNA. (c) The polypeptide chain is translated in successive groups of three nucleotides (each group constituting a codon), starting at the 5′ end of the coding sequence in the RNA and moving in the 5′-to-3′ direction. The amino acid corresponding to each codon can be found in the genetic code table. (d) The change from T to A in the transcribed strand alters a GAG codon into a GUG codon in the RNA transcript, resulting in the replacement of the normal glutamic acid (GAG) with valine (V). The nonmutant duplex, the RNA transcript, and the amino acid sequence are as shown below. The amino acid that is replaced in the sickle-cell mutant is indicated in red.

3′-TACCACGTGGACTGAGGACTCCTCTTCAGA-5′
5′-ATGGTGCACCTGACTCCTGAGGAGAAGTCT-3′

5′-AUGGUGCACCUGACUCCUGAGGAGAAGUCU-3′
MetValHisLeuThrProGluGluLysSer

PROBLEM 2 The accompanying diagram shows a linear bioynthetic pathway for an essential nutrient designated F in an organism, such as *Neurospora*, able to grow in a minimal medium. Each red letter indicates one intermediate in the pathway, and each blue number indicates a mutant that blocks one step in the pathway.

Make a table in which the columns correspond to the intermediates, arranged in alphabetical order, and the rows correspond to the mutants, arranged in numerical order. In the body of the table, insert a plus sign if the mutant will grow on minimal medium supplemented with the nutrient and a minus sign if the mutant will not grow under these conditions. Assume that all intermediates can be transported into the cell from the growth medium.

SOLUTION. This is a classic type of genetic analysis pioneered by Beadle and Tatum. The principle is that a mutant will grow on any intermediate whose position in the pathway is *downstream* of the metabolic block. Hence, mutant 1 will grow on any intermediate except A, mutant 2 will grow on any intermediate except A or B, and so forth. The complete matrix is as shown. It looks exceptionally simple because both the rows (mutants) and columns (intermediates) are arranged in the same order as their constituents appear in the pathway. Normally this will not be the case.

	A	B	C	D	E	F
1	–	+	+	+	+	+
2	–	–	+	+	+	+
3	–	–	–	+	+	+
4	–	–	–	–	+	+
5	–	–	–	–	–	+

PROBLEM 3 A complementation test is used to sort a set of mutants into groups, each group corresponding to a subset of the mutants that have defects in the same gene. Shown here are the genes (1–5) from the previous problem and 10 mutants (a–j) grouped according to the gene they affect.

1	2	3	4	5
a	*e, g*	*c, b, h, j*	*f*	*d, i*

Gene 1 is represented by mutant *a* only, gene 2 by mutants *e* and *g*, and so forth.

(a) Prepare a square complementation matrix of data, with the rows and columns representing the mutants in alphabetical order. Each entry in the matrix should be a plus sign if the row mutant and the column mutant do show complementation (that is, if they are mutants of different genes) or a minus sign if they do not show complementation (that is, if they are mutants of the same gene).
(b) What is special about the principal diagonal of the matrix? (The principal diagonal is the diagonal that runs from upper left to lower right.) What does this result mean biologically?
(c) What is special about the triangular parts of the matrix above and below the diagonal? What does this result mean biologically?
(d) Prepare a circular diagram of the mutants as discussed in the text, showing which of the mutants form complementation groups.

SOLUTION. (a) The complementation matrix is as shown here. (b) The principal diagonal consists exclusively of minus signs; biologically, this means that a mutant cannot undergo complementation with itself, because two copies of the identical mutation must be in the same gene. (c) The upper and lower triangular matrices are symmetrical, mirror images of one another; biologically, this means that the parent of origin of the mutant makes no difference to whether the mutants undergo complementation. Because of the symmetry of the data matrix, complementation data are often presented only in the form of the upper diagonal. (d) The circular type of the complementation test is also shown. It indicates that the complementation groups are {*a*}, {*b, c, h, j*}, {*d, i*}, {*e, g*}, and {*f*}. The complementation groups are not informative about where the product of each gene acts in the pathway; this information must come from the type of analysis illustrated in the previous problem.

(a–c)

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*
a	–	+	+	+	+	+	+	+	+	+
b	+	–	–	+	+	+	+	–	+	–
c	+	–	–	+	+	+	+	–	+	–
d	+	+	+	–	+	+	+	+	–	+
e	+	+	+	+	–	+	–	+	+	+
f	+	+	+	+	+	–	+	+	+	+
g	+	+	+	+	–	+	–	+	+	+
h	+	–	–	+	+	+	+	–	+	–
i	+	+	+	–	+	+	+	+	–	+
j	+	–	–	+	+	+	+	–	+	–