The Personal Genome

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

Given the importance of the DNA, it should come as no surprise that great efforts have been expended to study it, down to the finest details. In fact, in the last decade of the twentieth century a worldwide campaign, the Human Genome Project, took shape, and in 2001 it produced a comprehensive analysis of human DNA samples that had been collected from a small number of anonymous donors. This work—stunning in scope and significance—laid the foundation for all future research on the human genome. Then, in 2007, the analysis of human DNA took a new turn. Two of the architects of the Human Genome Project had their own DNA decoded. The technology for analyzing complete genomes has advanced significantly, and the cost for this analysis is no longer exorbitant. In fact, it may soon be possible for each of us to have our own genome analyzed—a prospect that is sure to influence our lives and change how we think about ourselves.
Chapter 1  The Science of Genetics

An Invitation

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

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

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

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

Three Great Milestones in Genetics

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

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

MENDEL: GENES AND THE RULES OF INHERITANCE

Although genetics developed during the twentieth century, its origin is rooted in the work of Gregor Mendel (Figure 1.1), a Moravian monk who lived in the nineteenth century.
Mendel carried out his path-breaking research in relative obscurity. He studied the inheritance of different traits in peas, which he grew in the monastery garden. His method involved interbreeding plants that showed different traits—for example, short plants were bred with tall plants—to see how the traits were inherited by the offspring. Mendel's careful analysis enabled him to discern patterns, which led him to postulate the existence of hereditary factors responsible for the traits he studied. We now call these factors genes.

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

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

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

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

**WATSON AND CRICK: THE STRUCTURE OF DNA**

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

Nucleic acids are made of elementary building blocks called nucleotides (Figure 1.2). Each nucleotide has three components: (1) a sugar molecule; (2) a phosphate molecule, which has acidic chemical properties; and (3) a nitrogen-containing molecule, which has slightly basic chemical properties. In ribonucleic acid, or RNA, the constituent sugar is ribose; in deoxyribonucleic acid, or DNA, it is deoxyribose. Within RNA or DNA, one nucleotide is distinguished from another by its nitrogen-containing base. In RNA, the
four kinds of bases are adenine (A), guanine (G), cytosine (C), and uracil (U); in DNA, they are A, G, C, and thymine (T). Thus, in both DNA and RNA there are four kinds of nucleotides, and three of them are shared by both types of nucleic acid molecules.

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

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

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

THE HUMAN GENOME PROJECT: SEQUENCING DNA AND CATALOGING GENES

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

The paragon of all the sequencing programs is the Human Genome Project, a worldwide effort to determine the sequence of approximately 3 billion nucleotide pairs in human DNA. As initially conceived, the Human Genome Project was to involve collaborations among researchers in many different countries, and much of the work
was to be funded by their governments. However, a privately funded project initiated by Craig Venter, a scientist and entrepreneur, soon developed alongside the publicly funded project. In 2001 all these efforts culminated in the publication of two lengthy articles about the human genome. The articles reported that 2.7 billion nucleotide pairs of human DNA had been sequenced. Computer analysis of this DNA suggested that the human genome contained between 30,000 and 40,000 genes. More recent analyses have revised the human gene number downward, to around 20,500. These genes have been cataloged by location, structure, and potential function. Efforts are now focused on studying how they influence the myriad characteristics of humans. There is also considerable effort to assess how much one human genome differs from another—that is, how much genetic variability exists in the human species. For more information about this effort, you can read the Focus on The 1000 Genomes Project on the Student Companion site.

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

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

**KEY POINTS**

- **Gregor Mendel** postulated the existence of particulate factors—now called genes—to explain how traits are inherited.
- **Alleles**, the alternate forms of genes, account for heritable differences among individuals.
- **James Watson and Francis Crick** elucidated the structure of DNA, a macromolecule composed of two complementary chains of nucleotides.
- **DNA** is the hereditary material in all life forms except some types of viruses, in which RNA is the hereditary material.
- **The Human Genome Project** determined the sequence of nucleotides in the DNA of the human genome.
- **Sequencing** the DNA of a genome provides the data to identify and catalog all the genes of an organism.
In all cellular organisms, the genetic material is DNA. This material must be able to \textit{replicate} so that copies can be transmitted from cell to cell and from parents to offspring; it must contain \textit{information} to direct cellular activities and to guide the development, functioning, and behavior of organisms; and it must be able to \textit{change} so that over time, groups of organisms can adapt to different circumstances.

\section*{DNA as the Genetic Material}

In biology information flows from DNA to RNA to protein.

\section*{DNA Replication: Propagating Genetic Information}

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

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

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

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{DNA_replication_diagram.png}
\caption{DNA replication. The two strands in the parental molecule are oriented in opposite directions (see arrows). These strands separate and new strands are synthesized using the parental strands as templates. When replication is completed, two identical double-stranded DNA molecules are produced.}
\end{figure}
DNA as the Genetic Material

Gene expression: using genetic information

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

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

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

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

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

We refer to the collection of all the different proteins in an organism as its proteome. Humans, with around 20,500 genes, may have hundreds of thousands of different proteins
in their proteome. One reason for the large size of the human proteome is that a particular gene may encode several different, but related, polypeptides, and these polypeptides may combine in complex ways to produce different proteins. Another reason is that proteins may be produced by combining polypeptides encoded by different genes. If the number of genes in the human genome is large, the number of proteins in the human proteome is larger.

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

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

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

MUTATION: CHANGING GENETIC INFORMATION

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

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

The process of mutation has another aspect—it introduces variability into the genetic material of organisms. Over time, the mutant...
genes created by mutation may spread through a population. For example, you might wonder why the mutant β-globin gene is relatively common in some human populations. It turns out that people who carry both a mutant and a nonmutant allele of this gene are less susceptible to infection by the blood parasite that causes malaria. These people therefore have a better chance of surviving in environments where malaria is a threat. Because of this enhanced survival, they produce more children than other people, and the mutant allele that they carry can spread. This example shows how the genetic makeup of a population—in this case, the human population—can evolve over time.

**KEY POINTS**

- When DNA replicates, each strand of a duplex molecule serves as the template for the synthesis of a complementary strand.
- When genetic information is expressed, one strand of a gene's DNA duplex is used as a template for the synthesis of a complementary strand of RNA.
- For most genes, RNA synthesis (transcription) generates a molecule (the RNA transcript) that becomes a messenger RNA (mRNA).
- Coded information in an mRNA is translated into a sequence of amino acids in a polypeptide.
- Mutations can alter the DNA sequence of a gene.
- The genetic variability created by mutation is the basis for biological evolution.

**Genetics and Evolution**

Genetics has much to contribute to the scientific study of evolution. As mutations accumulate in the DNA over many generations, we see their effects as differences among organisms. Mendel's strains of peas carried different mutant genes, and so do people from different ancestral groups. In almost any species, at least some of the observable variation has an underlying genetic basis. In the middle of the nineteenth century, Charles Darwin and Alfred Wallace, both contemporaries of Mendel, proposed that this variation makes it possible for species to change—that is, to evolve—over time.

The ideas of Darwin and Wallace revolutionized scientific thought. They introduced an historical perspective into biology and gave credence to the concept that all living things are related by virtue of descent from a common ancestor. However, when these ideas were proposed, Mendel's work on heredity was still in progress and the science of genetics had not yet been launched. Research on biological evolution was stimulated when Mendel's discoveries came to light at the beginning of the twentieth century, and it took a new turn when DNA sequencing techniques emerged at the century's end. With DNA sequencing we can see similarities and differences in the genetic material of diverse organisms. On the assumption that sequences of nucleotides in the DNA are the result of historical processes, it is possible to interpret these similarities and differences in a temporal framework. Organisms with very similar DNA sequences are descended from a recent common ancestor, whereas organisms with less similar DNA sequences are descended from a more remote common ancestor. Using this logic, researchers can establish the historical relationships among organisms (Figure 1.10). We call these relationships a phylogenetic tree, or more simply, a phylogeny, from Greek words meaning “the origin of tribes.”

Today the construction of phylogenetic trees is an important part of the study of evolution. Biologists use the burgeoning DNA sequence data from the genome projects and other research ventures, such as the U.S. National Science Foundation's...
“Tree of Life” program, in combination with anatomical data collected from living and fossilized organisms to discern the evolutionary relationships among species. We will explore the genetic basis of evolution in Chapter 20, and in Chapter 24 on the Instructor Companion site.

**KEY POINTS**

- Evolution depends on the occurrence, transmission, and spread of mutant genes in groups of organisms.
- DNA sequence data provide a way of studying the historical process of evolution.

### Levels of Genetic Analysis

Genetic analysis is practiced at different levels. The oldest type of genetic analysis follows in Mendel’s footsteps by focusing on how traits are inherited when different strains of organisms are hybridized. Another type of genetic analysis follows in the footsteps of Watson and Crick and the army of people who have worked on the various genome projects by focusing on the molecular makeup of the genetic material. Still another type of genetic analysis imitates Darwin and Wallace by focusing on entire populations of organisms. All these levels of genetic analysis are routinely used in research today. Although we will encounter them in many different places in this book, we provide brief descriptions of them here.

#### CLASSICAL GENETICS

The period prior to the discovery of the structure of DNA is often spoken of as the era of classical genetics. During this time, geneticists pursued their science by analyzing the outcomes of crosses between different strains of organisms, much as Mendel had done in his work with peas. In this type of analysis, genes are identified by studying the inheritance of trait differences—tall pea plants versus short pea plants, for example—in the offspring of crosses. The trait differences are due to the alternate forms of genes. Sometimes more than one gene influences a trait, and sometimes environmental conditions—for example, temperature and nutrition—exert an effect. These complications can make the analysis of inheritance difficult.

The classical approach to the study of genes can also be coordinated with studies on the structure and behavior of chromosomes, which are the cellular entities that contain the genes. By analyzing patterns of inheritance, geneticists can localize genes to specific chromosomes. More detailed analyses allow them to localize genes to specific positions within chromosomes—a practice called chromosome mapping. Because these studies emphasize the transmission of genes and chromosomes from one generation to the next, they are often referred to as exercises in transmission genetics. However, classical genetics is not limited to the analysis of gene and chromosome transmission. It also studies the nature of the genetic material—how it controls traits and how it mutates. We present the essential features of classical genetics in Chapters 3–8.

#### MOLECULAR GENETICS

With the discovery of the structure of DNA, genetics entered a new phase. The replication, expression, and mutation of genes could now be studied at the molecular level. This approach to genetic analysis was raised to a new level when it became possible to sequence DNA molecules easily. Molecular genetic analysis is rooted in the study of DNA sequences. Knowledge of a DNA sequence and comparisons to other DNA sequences allow a geneticist to define a gene chemically. The gene’s internal components—coding sequences, regulatory sequences, and noncoding sequences—can be identified, and the nature of the polypeptide encoded by the gene can be predicted.
But the molecular approach to genetic analysis is much more than the study of DNA sequences. Geneticists have learned to cut DNA molecules at specific sites. Whole genes, or pieces of genes, can be excised from one DNA molecule and inserted into another DNA molecule. These “recombinant” DNA molecules can be replicated in bacterial cells or even in test tubes that have been supplied with appropriate enzymes. Milligram quantities of a particular gene can be generated in the laboratory in an afternoon. In short, geneticists have learned how to manipulate genes more or less at will. This artful manipulation has allowed researchers to study genetic phenomena in great detail. They have even learned how to transfer genes from one organism to another. We present examples of molecular genetic analysis in many chapters of this book.

**POPULATION GENETICS**

Genetics can also be studied at the level of an entire population of organisms. Individuals within a population may carry different alleles of a gene; perhaps they carry different alleles of many genes. These differences make individuals genetically distinct, possibly even unique. In other words, the members of a population vary in their genetic makeup. Geneticists seek to document this variability and to understand its significance. Their most basic approach is to determine the frequencies of specific alleles in a population and then to ascertain if these frequencies change over time. If they do, the population is evolving. The assessment of genetic variability in a population is therefore a foundation for the study of biological evolution. It is also useful in the effort to understand the inheritance of complex traits, such as body size or disease susceptibility. Often complex traits are of considerable interest because they have an agricultural or a medical significance. We discuss genetic analysis at the population level in Chapters 19 and 20, and in Chapter 24 on the Instructor Companion site.

**KEY POINTS**

- In classical genetic analysis, genes are studied by following the inheritance of traits in crosses between different strains of an organism.
- In molecular genetic analysis, genes are studied by isolating, sequencing, and manipulating DNA and by examining the products of gene expression.
- In population genetic analysis, genes are studied by assessing the variability among individuals in a group of organisms.

**Genetics in the World: Applications of Genetics to Human Endeavors**

Modern genetic analysis began in a European monastic enclosure; today, it is a worldwide enterprise. The significance and international scope of genetics are evident in today’s scientific journals, which showcase the work of geneticists from many different countries. They are also evident in the myriad ways in which genetics is applied in agriculture, medicine, and many other human endeavors all over the world. We will consider some of these applications in Chapters 14–16, and 19. Some of the highlights are introduced in this section.

**GENETICS IN AGRICULTURE**

By the time the first civilizations appeared, humans had already learned to cultivate crop plants and to rear livestock. They had also learned to improve their crops and livestock by selective breeding. This pre-Mendelian application of genetic principles had telling effects. Over thousands of generations, domesticated plant and animal species
came to be quite different from their wild ancestors. For example, cattle were changed in appearance and behavior (Figure 1.11), and corn, which is descended from a wild grass called teosinte (Figure 1.12), was changed so much that it could no longer grow without human cultivation.

Selective breeding programs—now informed by genetic theory—continue to play important roles in agriculture. High-yielding varieties of wheat, corn, rice, and many other plants have been developed by breeders to feed a growing human population. Selective breeding techniques have also been applied to animals such as beef and dairy cattle, swine, and sheep, and to horticultural plants such as shade trees, turf grass, and garden flowers.

Beginning in the 1980s, classical approaches to crop and livestock improvement were supplemented—and in some cases, supplanted—by approaches from molecular genetics. Detailed genetic maps of the chromosomes of several species were constructed to pinpoint genes of agricultural significance. By locating genes for traits such as grain yield or disease resistance, breeders could now design schemes to incorporate particular alleles into agricultural varieties. These mapping projects have been carried on relentlessly and for a few species have culminated in the complete sequencing of the genome. Other crop and livestock genome sequencing projects are still in progress. All sorts of potentially useful genes are being identified and studied in these projects.

Plant and animal breeders are also employing the techniques of molecular genetics to introduce genes from other species into crop plants and livestock. This process of changing the genetic makeup of an organism was initially developed using test species such as fruit flies. Today it is widely used to augment the genetic material of many kinds of creatures. Plants and animals that have been altered by the introduction of foreign genes are called GMOs—genetically modified organisms. BT corn is an example. Many corn varieties now grown in the United States carry a gene from the bacterium
The development and use of GMOs has stirred up controversy worldwide. For example, African and European countries have been reluctant to grow BT corn or to purchase BT corn grown in the United States. Their reluctance is due to several factors, including the conflicting interests of small farmers and large agricultural corporations, and concerns about the safety of consuming genetically modified food. There is also a concern that BT corn might kill nonpest species of insects such as butterflies and honeybees. Advances in molecular genetics have provided the tools and the materials to change agriculture profoundly. Today, policymakers are wrestling with the implications of these new technologies.

**GENETICS IN MEDICINE**

Classical genetics has provided physicians with a long list of diseases that are caused by mutant genes. The study of these diseases began shortly after Mendel's work was rediscovered. In 1909 Sir Archibald Garrod, a British physician and biochemist, published a book entitled *Inborn Errors of Metabolism*. In this book Garrod documented how metabolic abnormalities can be traced to mutant alleles. His research was seminal, and in the next several decades, a large number of inherited human disorders were identified and cataloged. From this work, physicians have learned to diagnose genetic diseases, to trace them through families, and to predict the chances that particular individuals might inherit them. Today some hospitals have professionals known as genetic counselors who are trained to advise people about the risks of inheriting or transmitting genetic diseases. We will discuss some aspects of genetic counseling in Chapter 3.

Genetic diseases like the ones that Garrod studied are individually rather rare in most human populations. For example, among newborns, the incidence of phenylketonuria, a disorder of amino acid metabolism, is only one in 10,000. However, mutant genes also contribute to more prevalent human maladies—heart disease and cancer, for example. In Chapter 19 we will explore ways of assessing genetic risks for complex traits such as the susceptibility to heart disease, and in Chapter 23 on the Instructor Companion site we will investigate the genetic basis of cancer.

Advances in molecular genetics are providing new ways of detecting mutant genes in individuals. Diagnostic tests based on the analysis of DNA are now readily available. For example, a hospital lab can test a blood sample or a cheek swab for the presence of a mutant allele of the *BRCA1* gene, which strongly predisposes its carriers to develop breast cancer. If a woman carries the mutant allele, she may be advised to undergo a mastectomy to prevent breast cancer from occurring. The application of these new molecular genetic technologies therefore often raises difficult issues for the people involved.

Molecular genetics is also providing new ways to treat diseases. For decades diabetics had to be given insulin obtained from animals—usually pigs. Today, perfect human insulin is manufactured in bacterial cells that carry the human insulin gene. Vats of these cells are grown to produce the insulin polypeptide on an industrial scale. Human growth hormone, previously isolated from cadavers, is also manufactured in bacterial cells. This hormone is used to treat children who cannot make
sufficient amounts of the hormone themselves because they carry a mutant allele of the growth hormone gene. Without the added hormone, these children would be affected with dwarfism. Many other medically important proteins are now routinely produced in bacterial cells that have been supplied with the appropriate human gene. The large-scale production of such proteins is one facet of the burgeoning biotechnology industry. We will explore ways of producing human proteins in bacterial cells in Chapter 16.

Human gene therapy is another way in which molecular genetic technologies are used to treat diseases. The strategy in this type of therapy is to insert a healthy, functional copy of a particular gene into the cells of an individual who carries only mutant copies of that gene. The inserted gene can then compensate for the faulty genes that the individual inherited. To date, human gene therapy has had mixed results. Efforts to cure individuals with cystic fibrosis (CF), a serious respiratory disorder, by introducing copies of the normal \textit{CF} gene into lung cells have not been successful. However, medical geneticists have had some success in treating immune system and blood cell disorders by introducing the appropriate normal genes into bone marrow cells, which later differentiate into immune cells and blood cells. We will discuss the emerging technologies for human gene therapy and some of the risks involved in Chapter 16.

**GENETICS IN SOCIETY**

Modern societies depend heavily on the technology that emerges from research in the basic sciences. Our manufacturing and service industries are built on technologies for mass production, instantaneous communication, and prodigious information processing. Our lifestyles also depend on these technologies. At a more fundamental level, modern societies rely on technology to provide food and health care. We have already seen how genetics is contributing to these important needs. However, genetics impacts society in other ways too.

One way is economic. Discoveries from genetic research have initiated countless business ventures in the biotechnology industry. Companies that market pharmaceuticals and diagnostic tests, or that provide services such as DNA profiling, have contributed to worldwide economic growth. Another way is legal. DNA sequences differ among individuals, and by analyzing these differences, people can be identified uniquely. Such analyses are now routinely used in many situations—to test for paternity, to convict the guilty and to exonerate the innocent of crimes for which they are accused, to authenticate claims to inheritances, and to identify the dead. Evidence based on the analysis of DNA is now commonplace in courtrooms all over the world.

But the impact of genetics goes beyond the material, commercial, and legal aspects of our societies. It strikes the very core of our existence because, after all, DNA—the subject of genetics—is a crucial part of us. Discoveries from genetics raise deep, difficult, and sometimes disturbing existential questions. Who are we? Where do we come from? Does our genetic makeup determine our nature? our talents? our ability to learn? our behavior? Does it play a role in setting our customs? Does it affect the ways we organize our societies? Does it influence our attitudes toward other people? Will knowledge about our genes and how they influence us affect our ideas about morality and justice, innocence and guilt, freedom and responsibility? Will this knowledge change how we think about what it means to be human? Whether we like it or not, these and other probing questions await us in the not-so-distant future.

- **Discoveries in genetics are changing procedures and practices in agriculture and medicine.**
- **Advances in genetics are raising ethical, legal, political, social, and philosophical questions.**
Basic Exercises

Illustrate Basic Genetic Analysis

1. How is genetic information expressed in cells?

Answer: The genetic information is encoded in sequences in the DNA. Initially, these sequences are used to synthesize RNA complementary to them—a process called transcription—and then the RNA is used as a template to specify the incorporation of amino acids in the sequence of a polypeptide—a process called translation. Each amino acid in the polypeptide corresponds to a sequence of three nucleotides in the DNA. The triplets of nucleotides that encode the different amino acids are called codons.

2. What is the evolutionary significance of mutation?

Answer: Mutation creates variation in the DNA sequences of genes (and in the nongenic components of genomes as well). This variation accumulates in populations of organisms over time and may eventually produce observable differences among the organisms. One population may come to differ from another according to the kinds of mutations that have accumulated over time. Thus, mutation provides the input for different evolutionary outcomes at the population level.

Testing Your Knowledge

Integrate Different Concepts and Techniques

1. Suppose a gene contains 10 codons. How many coding nucleotides does the gene contain? How many amino acids are expected to be present in its polypeptide product? Among all possible genes composed of 10 codons, how many different polypeptides could be produced?

Answer: The gene possesses 30 coding nucleotides. Its polypeptide product is expected to contain 10 amino acids, each corresponding to one of the codons in the gene. If each codon can specify one of 20 naturally occurring amino acids, among all possible gene sequences 10 codons long, we can imagine a total of $20^{10}$ polypeptide products—a truly enormous number!

Questions and Problems

Enhance Understanding and Develop Analytical Skills

1.1 In a few sentences, what were Mendel’s key ideas about inheritance?

1.2 Both DNA and RNA are composed of nucleotides. What molecules combine to form a nucleotide?

1.3 Which bases are present in DNA? Which bases are present in RNA? Which sugars are present in each of these nucleic acids?

1.4 What is a genome?

1.5 The sequence of a strand of DNA is ATGGCCGTC. If this strand serves as the template for DNA synthesis, what will be the sequence of the newly synthesized strand?

1.6 A gene contains 141 codons. How many nucleotides are present in the gene’s coding sequence? How many amino acids are expected to be present in the polypeptide encoded by this gene?

1.7 The template strand of a gene being transcribed is CTTGCCAGT. What will be the sequence of the RNA made from this template?

1.8 What is the difference between transcription and translation?

1.9 RNA is synthesized using DNA as a template. Is DNA ever synthesized using RNA as a template? Explain.

1.10 The gene for α-globin is present in all vertebrate species. Over millions of years, the DNA sequence of this gene has changed in the lineage of each species. Consequently, the amino acid sequence of α-globin has also changed in these lineages. Among the 141 amino acid positions in this polypeptide, human α-globin differs from shark α-globin in 79 positions; it differs from carp α-globin in 68 and from cow α-globin in 17. Do these data suggest an evolutionary phylogeny for these vertebrate species?

1.11 Sickle-cell disease is caused by a mutation in one of the codons in the gene for β-globin; because of this mutation the sixth amino acid in the β-globin polypeptide is a valine instead of a glutamic acid. A less severe disease is caused by a mutation that changes this same codon to one specifying lysine as the sixth amino acid in the β-globin polypeptide. What word is used to describe the two mutant forms of this gene? Do you think that an individual carrying these two mutant forms of the β-globin gene would suffer from anemia? Explain.
1.12 Hemophilia is an inherited disorder in which the blood-clotting mechanism is defective. Because of this defect, people with hemophilia may die from cuts or bruises, especially if internal organs such as the liver, lungs, or kidneys have been damaged. One method of treatment involves injecting a blood-clotting factor that has been purified from blood donations. This factor is a protein encoded by a human gene. Suggest a way in which modern genetic technology could be used to produce this factor on an industrial scale. Is there a way in which the inborn error of hemophilia could be corrected by human gene therapy?


You might enjoy using the NCBI web site to explore the Human Genome Project. Click on About NCBI and then on Outreach and Education. From there click on Recommended Links to get information about the Human Genome Project.