Foundations of Biology

Samuel M. Scheiner and Kayla I. Scheiner

Copyright (c) 2014 by S. M. Scheiner and K. I. Scheiner. This work is made available under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 license (http://creativecommons.org/licenses/by-nc-nd/3.0/)

Front cover
American egret, Gainsville, Florida. Photograph by Samuel Scheiner
Chapter 3
Genetics

Attend any family reunion or holiday gathering and certain similarities shared among the attendees quickly become obvious. The same is true across the entire spectrum of life. Relatives resemble each other in both appearance and behavior; offspring look like their parents and siblings act like each other. That resemblance is determined by information, and the way information is transferred. This information is at the heart of life itself, and those similarities form the basis of how the great diversity of life on earth arose from the simple envelopes of self-replicating molecules proposed by the Replication First and Metabolism First hypotheses (see Chapter 1). This holds true for the differences as well as the similarities; after all, the resemblances we see are rarely perfect, and sometimes relatives may look and act nothing like each other. As explained in Chapter 1, information exists in the ordered complexity of living systems, particularly in the order of the base-pairs in a DNA molecule, just like information exists in the specific ordering of the letters in a sentence. In this chapter, we explore how the information systems of organisms work, how that information is transmitted from one generation to another, and how that information is used.

Living systems come from other living systems, resulting in a tree of life and connections among all species (see Figure 4.2), and it is the transmission of information that is responsible for that continuity. Evolution requires variation (see Chapter 4), and in this chapter we will look at where that variation comes from. The concepts presented in this chapter are framed in general terms of information, rather than in specific terms of DNA, RNA, and so forth. The reasons are two-fold. First, the theory of genetics presented here could apply to any living system, even ones based on other biomolecules, as might exist on other planets. Second, even for life on Earth, there are many variants in the details of how the genetic system works. The theory is written to encompass all of those variants, and in this chapter we will explore the major variants.

The first question we must address is a seemingly very basic one: What is information? One way to think about information is that it is a sort of symbol that represents something else, such as the way the letter “m” represents a particular sound. The more types of symbols, the more possible information. But information is more than each individual symbol – it is also how those pieces are combined to create new information. This creation of new information through new combinations is an example of an emergent property (see Chapter 1).

The Basics of Genetic Structure and Duplication

The gene is the fundamental unit of information. While genes are composed of parts that themselves act as information, it is the gene itself that turns information into a biological molecule that determines the characteristics of organisms. Our understanding of what a gene is and how it functions in information storage, transmission, and usage has grown considerably in the past 150 years.

Genes are, in turn, assembled into larger structures called chromosomes. Those chromosomes come in two forms: closed circles and open strings. Of the three domains of life (see Figure 4.2), the Bacteria and Archaea have closed circles and the Eukaryota have open strings. Before we explore how the information contained in these genes and chromosomes are transmitted and used, we must first examine the basics of their structure.
Figure 3.1
(A) The four types of DNA units joined as along a DNA molecule. The sugar-phosphate portion is the same for all types. They differ in the structure of the base. The paired bases are held together by hydrogen bonds, which are weaker than the chemical bonds that hold the rest of the molecule together. (Source: Wikipedia) (B) The DNA molecule is twisted, forming a double helix. (Created by Richard Wheeler, Source: Wikipedia)

The basic building-block of the chromosome is deoxyribonucleic acid, DNA. While all DNA units contain one part that is the same for all types, the sugar-phosphate portion (Figure 3.1), they differ in their bases. There are four types of bases that result in four types of DNA units: adenine (A), guanine (G), thymine (T), and cytosine (C). These bases are strung together in long strands that link up the sugar-phosphate portions, with the bases sticking out the side. A DNA chromosome consists of a series of two of these strands arranged so that the bases of one strand match up with the bases of the other. The form of the bases is such that A and T fit together, while G fits with C. The structure of the chromosome is somewhat like a ladder with the sugar-phosphate portion forming the side poles and the base pairs forming the rungs. The ladder is twisted, resulting in a structure known as a double-helix (Figure 3.1B). The double-helix strand is then further wound around a protein forming a much larger structure, the chromosome (Figure 3.2). The packing is so efficient that if all of the DNA contained in one human cell were stretched out end-to-end, it would be 1.8 meters long.

The structure of DNA is essential to transmission of information from parent to offspring, including from a parental cell when it divides to create two daughter cells. During that process, the chromosomes can first be seen to duplicate, then to be split up so that each daughter cell receives one copy of each chromosome. The double-helix structure is physically unzipped, or split down the middle, opening up the A-T and G-C pairs. The
new DNA units – previously synthesized by the cell – are then lined up with each strand so that each of the two strands becomes the template for a new, matching strand. Because the bases are complementary, A with T and G with C, the end result are two chromosomes, each a duplicate of the original (Figure 3.3). The process of chromosome duplication and sorting into the daughter cells is known as mitosis (Figure 3.4).

Living systems can carry their information in multiple places. All cells have at least one chromosome that is the primary information carrier. In Bacteria and Archaea, besides the main chromosome, a cell might additionally contain one or more small, circular pieces of DNA called plasmids containing just a few genes which will also get replicated and passed along to daughter cells. Eukaryotes have multiple chromosomes. In addition, nearly all eukaryotic cells have specialized substructures that are responsible for processing energy called mitochondria, and plants have specialized structures for capturing light energy called chloroplasts. Both mitochondria and chloroplasts have circular chromosomes similar to those of Bacteria. We look at both in detail in Chapter 5.
Figure 3.4
The duplication of chromosomes can be seen during the process of cell division when a parental cell gives rise to daughter cells. During that process, the chromosomes duplicate and split up so that each daughter cell received one copy of each chromosome. (A) Chromosome duplication in Eukaryotes which have linear chromosomes contained in a specialized structure called the nucleus (see Chapter 5). (From The Science Primer, NCBI, NIH, Source: Wikipedia) (B) Chromosome duplication in Bacteria and Archaea which have circular chromosomes that are attached to the cell membrane. (Created by Ecoddington14, Source: Wikipedia)
Box 3A
Critical Experiment: The Structure of DNA

The discovery and elucidation of the structure and duplication process of DNA came about through performing a series of experiments and building various models. Early in the 20th century it was recognized that chromosomes were the structures that stored and transmitted genetic information. Analyses revealed that chromosomes contained both DNA and proteins (Figure 3.2), but nothing was known about its structure. Initially it was thought that the proteins were the information carriers and the DNA was just there for structural integrity. This was because it was known that DNA contained just four types of bases (Figure 3.1), while proteins contained twenty different kinds of amino acids (see Box 1A). This greater variability of the basic units of proteins was thought to permit it to hold more complex information.

The first critical breakthrough came with experiments designed to explore the method by which information is transmitted, and ultimately demonstrated that DNA was the information carrier. In one of those experiments, Alfred D. Hershey and Martha Chase of the Carnegie Laboratory of Genetics used a very simple system, a bacterium and a virus that infected it. Hershey and Chase knew about previous experiments that seemed to show that DNA was the information carrier, but which had not yet convinced other scientists. They thought that they could demonstrate conclusively that it was the DNA, rather than amino acids, that serves as the main carrier of information in living organisms. In their experiment, the virus they used consisted of a small piece of DNA packaged inside of a protein, and had the advantage that the DNA and the protein were separate structures, not tangled up together like most chromosomes.

They exploited the fact that DNA and proteins have a key difference in their chemical make-up: DNA contains phosphorus, but no sulfur, while some amino acids – and thus proteins – contain sulfur, but no amino acids contain phosphorus. They grew batches of bacteria and viruses in broth that contained either radioactive phosphorus or radioactive sulfur, which would be absorbed by the bacteria and viruses, effectively labeling their proteins and DNA.

Using a centrifuge, they could then separate the viruses from the bacteria so that they had pure samples of the virus that were labeled with either the radioactive phosphorus or radioactive sulfur. These radioactively-labeled viruses were placed in new batches of bacteria that had no radioactivity. After just a short
time, the solution was shaken up and spun in a centrifuge so that any parts of the virus remaining outside the bacteria would be washed away, and only the parts of the virus that entered the bacterial cells were left. They then measured the radioactivity of those cells and found radioactivity only in the cells exposed to the phosphorus-labeled viruses. The viral protein did not enter the cell. Further, if the viruses were allowed to reproduce, the new viruses contained radioactive phosphorus. These studies, published in 1952, convinced scientists that DNA was the information carrier.

The British scientist Francis Crick and the American scientist James D. Watson are credited with determining the structure of DNA. They met each other while both were working at Cambridge University in the early 1950s. Although credit is usually given exclusively to them, two clues from other scientists were critical in finally determining the structure of DNA. The first clue came from Erwin Chargaff of Columbia University who published an analysis of the DNA of many species in 1950, in which he found that the amount of adenine always matched the amount of thymine and the amount of guanine always matched the amount of cytosine. The second clue came from a process of photographing the DNA molecules using X-rays. Rosalind Franklin had taken such photographs using samples prepared by Maurice Wilkins. Unlike ordinary photographs, this type of X-ray photograph results in a series of darker and lighter spots that require skilled knowledge to interpret.

There is still controversy over how Watson and Crick obtained copies of those photographs, whether the photographs were made publicly available, whether Wilkins had shown them to Watson on a visit to Wilkin’s laboratory, or whether Watson had simply seen them during that visit. What is not in dispute is that Watson and Crick realized that the photographs indicated that the structure of DNA was helical, something that Franklin and Wilkins did not. Armed with all of this information and using wire models, Watson and Crick were able to deduce the structure of DNA. For this discovery, they and Wilkins
The Theory of Genetics

The fundamental principles of the theory of genetics (Table 3.1) trace back to the pioneering work of Gregor Mendel in the 19th century (Box 3B). However, unlike the theory of evolution (see Chapter 4), no single person set forth this theory; rather, the principles were developed by many people over the course of the 20th century. One such critical development was the determination of the structure of DNA in the middle of that century (Box 3A). Since that time the focus of most genetics studies has been the molecular details of the structure of genetic material, how it is replicated, and what regulates the usage of that information. As with other aspects of biology, discoveries in genetics have led to further developments in other fields as well; for example, the discoveries in genetics during the first half of the 20th century were particularly important for our understanding of the process of evolution. It was bringing together that theory with the developing theory of genetics that lead to the development of the today’s understanding of the principles of the theory of evolution as the Modern Synthesis (see Chapter 4).

Today, new technological advances are driving our deepening understanding of genetic mechanisms. For example, primary to the understanding of genetic systems is knowledge of the sequence of base pairs in DNA molecules. The cost and speed of determining those sequences continues to drop by an order of magnitude about every five years. The original effort to determine the sequence of the DNA of a human - which consists of 3 billion base pairs - was begun in 1990, took 13 years, and cost about 3 billion dollars. There is now talk that by 2015, you will be able to have your personal DNA sequence for only $10,000. However, knowing what that sequence means and how it influences your life will take a bit longer, and will come through the application of the principles of the theory of genetics.

Table 3.1. The fundamental principles of the theory of genetics

<table>
<thead>
<tr>
<th>Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Offspring resemble their parents.</td>
</tr>
<tr>
<td>2. The information system requires an error correction system.</td>
</tr>
<tr>
<td>3. The information system must be capable of producing new information.</td>
</tr>
<tr>
<td>4. The imperfections of error correction create new information.</td>
</tr>
<tr>
<td>5. The exchange and recombination of information create new information.</td>
</tr>
<tr>
<td>6. Random processes play an important role in the information system.</td>
</tr>
<tr>
<td>7. Information usage must be robust to changes.</td>
</tr>
<tr>
<td>8. Information usage is context dependent.</td>
</tr>
<tr>
<td>9. The properties of information systems are the result of evolution.</td>
</tr>
</tbody>
</table>

Resemblance of Relatives

Understanding the information in your DNA sequence begins with an understanding of why you look like your relatives (Table 3.1, principle 1). The first steps were the working out of the rules by which the characteristics of parents were manifest in their offspring, and were first explained in multicellular organisms such as plants, mammals, insects, and fungi. This is not surprising as those are the most visible species and the easiest to manipulate and measure. Most of these early studies focused on characteristics such as form, color, and size. Technological advances made the measurement of molecular characteristics easier and easier, and by the middle of the 20th century much of the focus shifted to single-celled organisms, especially Bacteria and their viruses. Those
organisms had much smaller chromosomes than those of Eukaryotes and so were easier to sequence and study. Now, with modern sequencing technologies, the focus has shifted back to Eukaryotes.

The earliest geneticists were confronted with trying to explain two types of characteristics. One type consisted of discrete categories, such as the color of the eyes of fruit flies which could be red, brown, white and so forth (Figure 3.9). The other type was continuous, such as height or weight. We will explore each in turn, and then show how the two apparently discordant types can be explained by a single theory.

**Discrete traits**

In order to explore discrete traits, we start with a simple example, flower color, and consider the case of a species that shows three different colors: red, pink and white. We then perform an experiment designed to model how this trait is passed from parents to offspring. First, we mate individuals with red flowers with others that also have red flowers. We then plant the seeds and grow up the offspring and find that they all have red flowers. We do the same with white-flowered individuals and get a similar result, with all of their offspring having white flowers. Then we mate an individual with red flowers with one with white flowers, and find that their offspring have pink flowers, as if they had taken the colors of the two parents and averaged them together. We then take two pink-flowered individuals and mate them, and their offspring are a mixture of red, pink and white in proportions 1/4, 1/2 and 1/4 respectively. How can we explain these results?

First, we start by postulating that flower color is determined by a gene (C). We assume that there are “red” versions of this gene (that we will indicate as $C_R$), and “white” versions of this gene ($C_W$). Different versions of the same gene are called alleles. Red-flowered individuals have the $C_R$ allele which they pass on to their offspring who then also have red flowers. The same is true for white flowers. But where do the pink flowers come from?

We can explain both the pink flowers and the variation in their offspring if we postulate that each individual has two copies of the flower color gene and one copy is passed along to a given offspring. A red-flowered individual has a genetic make-up of $C_R C_R$. When it mates it produces reproductive cells, called gametes, that contain one copy of that gene ($C_R$). Another red-flowered individual does the same. Those gametes combine to form a new individual, again with two copies of the gene ($C_R C_R$). This system of segregation of the two copies of the gene into separate gametes explains both why it takes two individuals to reproduce and how the amount of information or genetic material stays the same from one generation to the next.

Pink-flowered individuals are now easy to explain. They receive a $C_R$ allele from one parent and a $C_W$ allele from the other, resulting in a genetic makeup of $C_R C_W$. If each allele contributes to half of the color, the resulting individual has a color that is
intermediate between the two parents. Next, when we cross two pink-flowered individuals, we have offspring with four possible sets of alleles (Figure 3.5). In this example, parent 1 produces gametes containing either the \(C^r\) or \(C^w\) allele. To keep track of them we will designate them as \(C_{r1}\) and \(C_{w1}\). Parent 2 does the same: \(C^r_{2}\) and \(C^w_{2}\). So we end up with the following offspring: \(C^r_{r1}C^r_{r2}\), \(C^r_{r1}C^w_{w2}\), \(C^w_{w1}C^r_{r2}\) and \(C^w_{w1}C^w_{w2}\). If the alleles of a given type are identical no matter which parent they come from, we end up with 1/4 of the offspring being \(C^r_{r1}C^r_{r2}\), 1/2 being \(C^r_{r1}C^w_{w2}\) and 1/4 being \(C^w_{w1}C^w_{w2}\), thus explaining the observed proportions of flower colors in the offspring. These observations form the basis of Mendel’s Law of Segregation (Table 3.2).

Table 3.2. The propositions that result in Mendel’s Laws of Segregation and Assortment

<table>
<thead>
<tr>
<th>Proposition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>There are alternate forms of genes.</td>
</tr>
<tr>
<td>2.</td>
<td>An individual carries two copies of each gene. Those copies may be the same form or alternate forms.</td>
</tr>
<tr>
<td>3.</td>
<td>An individual passes one copy of a gene to each gamete and an offspring receives one copy from each parent.</td>
</tr>
<tr>
<td>4.</td>
<td>Pairs of genes segregate (separate) during gamete formation and the fusion of gametes at fertilization recreates gene pairs. (Law of Segregation)</td>
</tr>
<tr>
<td>5.</td>
<td>Different pairs of genes segregate independently during gamete formation. (Law of Assortment)</td>
</tr>
</tbody>
</table>

Figure 3.6
Similar to the previous example, here flower color is determined by one gene with two variants \(C^r\) and \(C^w\). Individuals with two copies of the \(C^r\) allele have red flowers and those with two copies of the \(C^w\) allele have white flowers. However, because the \(C^r\) allele is dominant over the \(C^w\) allele because individuals with one copy of each have red flowers. (Created by Benutzer: Magnus Manske, Source: Wikipedia)

One of the experiments performed by Mendel involved flower color. Unlike the example shown in Figure 3.5, here there are only two colors, red and white. When a red-flowered individual is mated with a white-flowered individual, all of the offspring have red flowers. Rather than look like an intermediate of the two parents, the offspring look like just one of them. If two of those individuals are mated with each other, 3/4 of the offspring have red flowers and 1/4 have white flowers (Figure 3.6). These results can be explained by Mendel’s Law of Segregation (see main text) with one critical difference, the way in which individuals with mixed alleles appeared. The red-flowered parent had alleles \(C^rC^r\), the white-flowered parent had alleles \(C^wC^w\) and their red-flowered offspring had alleles \(C^rC^w\). In this case, the red allele is dominant over the white allele. Dominance is a type of emergent property since the flower color is not a simple averaging of the effects of each separate allele.

The Law of Segregation matches what we know about chromosomes. In most
Eukaryotes, the chromosomes come in pairs. In this case, we refer to such individuals as being diploid, which comes from the Greek word “diplos” meaning double. Individuals with only one copy of each chromosome are called haploid, while those with more than two copies are call polyploid. For example, humans are diploid with 23 pairs of chromosomes that are distinguishable by their length (Figure 3.7).

A given pair contains the same sets of genes, but may have different alleles. When gametes are formed, one member of the pair goes to one gamete and the other to the other gamete (we look at the process in more detail later in this chapter). Thus, most Eukaryotes have two types of doubles in their DNA. A single chromosome consists of a double-helix of two complementary strands, and those chromosomes come in pairs. We emphasize, though, that this is not a universal pattern. Some viruses consist of only a single strand of DNA, and many species have only a single chromosome or may have chromosomes that come in three, four or more copies. We will come back to this issue later in the chapter.

The second great insight of Mendel has to do with the inheritance of two traits at a time. Consider the following example: peapods can come in a variety of colors and shapes. Assume that there are three pea colors: green, yellow-green, and yellow. Similarly, there are three pea shapes: smooth, partially wrinkled, and very wrinkled. We start with two types of parents, one with smooth yellow seeds and one with very wrinkled green seeds. These are mated and the offspring all have partially wrinkled, yellow-green seeds. Two of those individuals are mated. Among their offspring are all nine possible combinations of seed colors and shapes in the following proportions: 1/16 smooth and yellow, 2/16 smooth and yellow-green, 1/16 smooth and green, 2/16 partially wrinkled and yellow, 4/16 partially wrinkled and yellow-green, 2/16 partially wrinkled and green, 1/16 very wrinkled and yellow, 2/16 very wrinkled and yellow-green, and 1/16 very wrinkled and green (Figure 3.8). These results can be explained if there are two genes each with two alleles, one gene determining seed color (C), with yellow (CY) and green (CG) alleles, and one gene determining seed shape (S) with smooth (SS) and wrinkled (SW) alleles. The parents have genotypes SSSCICY and SWSWCGCG, and their offspring have genotype SSSWCYCG (genotype refers to the genetic makeup of an individual). Those individuals can produce four types of gametes, each carrying one copy of the color-determining gene and the shape-determining gene: SSCY, SSCG, SWCY and SWCG. Their offspring each get one gamete from each parent, resulting in 16 possible combinations of genotypes and 9 possible phenotypes.

![Figure 3.7](image-url) 
The 23 pairs of chromosomes of a human male. Masleness is determined by the one copy of the X chromosome and one copy of the Y chromosome that can be seen in the lower right corner of the figure. The chromosomes are shown following duplication, just before they would be split between the daughter cells. (Photo from the National Human Genome Research Institute, NIH, Source: Wikipedia)
Figure 3.8
The process of independent assortment of two genes, one that determines seed pod color and one that determines seed pod shape. There are 16 possible genotypes that give rise to 4 possible phenotypes. (Source: K. Scheiner)
Box 3B
Gregor Johann Mendel

Born Johann Mendel – taking the name Gregor upon entering the monastery – Mendel’s pioneering work in genetics became the basis of our modern understanding of the process of inheritance and was a critical precursor to the development of evolutionary theory in the 20th century (see Chapter 4). A modest man, he was unlike most modern scientists in not publicly advocating his ideas. Yet his theories of inheritance set the stage for the green revolution in agriculture in the 20th century and today’s revolution in genetic technologies.

Mendel was born July 20, 1822 to an ethnic German family in Heinzendorfbei Odrau, Austrian Silesia in the Austrian Empire (now called Hynčice in the Czech Republic). He grew up with his parents, Anton and Rosine Mendel, and two sisters – one older and one younger. They lived on a farm that had been in the Mendel family for 130 years, which might have sparked his interest in botany. Growing up he worked as a gardener and studied beekeeping, and attended the Philosophical Institute in Olomouc from 1840 to 1843. It was the recommendation of his physics teacher, Friedrich Franz that lead him to enter the Augustinian Abbey of St. Thomas in Brno in 1843, where he took the name Gregor.

His scientific career flourished in the monastery. While a monk he attended the University of Vienna from 1851 to 1853, where he studied zoology, botany, chemistry and physics, and began his experiments with pea plants. Inspired by his university professors and monastic colleges, he began researching variation in plants, cultivating and testing 29,000 plants between 1856 and 1863. With the encouragement of his abbot, Cyrill Franz Napp, he began his experiments in the monastery garden; Napp himself had a lively interest in the sciences and plant cultivation. These experiments resulted in the Law of Segregation and Law of Independent Assortment, later called Mendel’s Laws of Inheritance (Table 3.2).

Despite the enormous impact Mendel’s experiments would have, at the time he presented his work he was met with overwhelming ambivalence. He read his preliminary paper, Experiments on Plant Hybridization, at two meetings of the Natural History Society of Brünn in Moravia in 1865, publishing them the following year in the Natural History Society’s Proceedings. Over the following 35 years, Mendel’s work would be cited only 3 times. Contemporary writings say that audiences received Mendel’s lectures courteously, but with blank incomprehension. No one had previously attempted to use mathematical and statistical analyses to interpret the results of biological inquiry, and at the time, most biologists held with the idea of blending inheritance, such as Darwin’s attempts to explain inheritance via pangenesis (which were, ultimately, unsuccessful).

Mendel himself was a relatively shy person, and might not have presented his groundbreaking results with the necessary emphasis and stress. Although he sent out forty special reprints of his papers to various botanists and biologists known to be interested in the hybridization of plants, most seem to have ignored it or paid scant attention. One of those recipients was Charles Darwin, although it appears that Darwin never read the work as the pages of the copy in his library are uncut. If he had read them, Darwin might have been able to reconcile Mendel’s ideas with his own theory of evolution, something which did not happen for another 60 years. However, as that reconciliation required the work of several brilliant mathematicians, skills not evidenced by Darwin, it is likely that he would have failed to make the connections. It is impossible to know. The only recipient to correspond with Mendel was Carl Wilhelm von Nägeli, one
of the most highly acclaimed botanists of the time. It appears, however, that he only glanced at the work. Although it dealt with 355 cross-bred strains and 12,980 resultant hybrids, Nägeli described it as “incomplete” and urged Mendel to continue his investigations. Unfortunately, Nägeli advised Mendel to begin using hawkweed, a member of the sunflower family that reproduces asexually, a fact not appreciated at the time. This meant any results would be uninformative and contrary to the predictions of Mendel’s theories, since hawkweed’s genetic information is transferred exclusively through the maternal line. When the experiments failed, the disappointing results combined with the previous audience reactions led Mendel to become frustrated with his investigations.

Mendel tried expanding his scientific inquiries into the animal kingdom using honeybees, but was stymied there as well. It was difficult to control the mating behaviors of the queens, and thus impossible to generate a clear picture of their heredity. He also managed to create a hybrid strain of bees that was so vicious it had to be destroyed. When he was elevated to abbot in 1868, his scientific work largely ended. Although he regretted the loss of its pursuit, his time became consumed by his new administrative responsibilities, such as a dispute with the civil government over their attempt to impose taxes on religious institutions. The truncation of a scientific career in the morass of administrative duties stymies academic scientists even today. Despite the lack of appreciation for his work in his lifetime, in 1883, a matter of months before his death Mendel remarked “My scientific studies have afforded me great gratification, and I am convinced that it will not be long before the whole world acknowledges the results of my work.” In less than 20 years he was proven correct in what is now called the rediscovery of Mendel. At about 1900, several scientists working on problems of inheritance and (unknowingly) performing similar experiments, independently came upon his work, including Hugo de Vries, Carl Correns, Erich von Tschermak and William Bateson. In May of 1900, while on the train going to present a paper on heredity to the Royal Horticultural Society, Bateson read Mendel’s actual paper for the first time. He immediately incorporated Mendel’s laws into his lecture.

Until then the predominant view on patterns of heredity was that the traits of offspring were a blend of those of its parents, a view influenced by Darwin’s theories of slow, continuous evolutionary change. Those who held this view were termed biometricians. In the 1890s the biggest opposition to that school of thought came from Bateson, who pushed the idea of discontinuous variation, an idea that fit well with Mendel’s notions of discrete genes that could segregate independently. (Bateson also coined the term “genetics”.) The dichotomy between the established biometric school of thought and Mendel’s newly rediscovered laws raged in the first few decades of the 20th century, with biometricians claiming the weight of statistic and mathematical rigor, and Mendelians claiming a better and deeper understanding of biology. The conflict was finally resolved by R. A. Fisher (see Box 2D) who developed a mathematical model from which many discrete genes of small effect could result in continuous variation in a trait.

One controversy has dogged Mendel’s legacy, the claim that he fudged his data to make them better fit his theories. This issue was first raised in 1936 by Fisher who published a re-analysis of some of Mendel’s data. Because of the way that Mendel assessed the genotype of his pea plants, such as the experiment shown in Figure 3.6, chance events should have resulted in ratios that were slightly less than the expected ones. While Fisher thought that the data were falsified, he attributed the act to a supposed assistant, rather than Mendel himself. There the issue lay until 1965, the centenary of Mendel’s original publication. A series of papers that re-examined the issue resulted in a vigorous debate that continued for the next 40 years. We now conclude that this is probably not a case of scientific malpractice or fraud. Mendel’s results may have been
the result of confirmation bias; Mendel might have detected an approximate 3:1 ratio in a small sample size early on in his experiments, then collected more data until his results conformed to a more exact ratio. Some suggest that he may have censored his results. The seven traits he studied occur on separate chromosomes, which is extremely unlikely if they were chosen at random. It is also possible that the way he scored some of the traits was more accurate than that assumed by Fisher. Finally, additional analyses of some of the data examined by Fisher show that the results closely match the ratios that Fisher predicts. While Mendel carefully reported these somewhat deviant results, he still interpreted them as supporting his theories. Given the absence in Mendel’s day of the kinds of statistical analyses developed in the 20th century, such interpretations were justified. So, there is no evidence that Mendel was anything less than scrupulous and honest in his scientific work.
**Continuous traits**

While some characteristics of organisms come in discrete types, many characteristics show continuous variation—traits that are measured on a spectrum, such as height or weight. Consider a trait such as height in a plant at the end of the growing season. In an experiment, we would choose pairs of plants, pollinate one individual with pollen from the other in each pair, and cover the flowers to prevent any other plant from pollinating that individual. Thus, when we collect the seeds at the end of the growing season, we know exactly who the parents were. Before the parental plants die, we measure their heights. Then we plant the seeds, let them grow, and measure the heights of the new plants at the end of the next growing season.

We then plot the heights of the offspring against the heights of their parents (Figure 3.9). In this case, we find that taller parents tend to produce taller offspring. Because the trait being measured does not sort into discrete categories, we must measure this tendency using a statistical technique called correlation. By plotting the data on a graph comparing the parents’ and offsprings’ height, we can draw a line through the middle of those points; once the mathematical slope of the line has been calculated, we determine how closely the offsprings’ height matches that of their parents. If offspring always exactly matched their parents, the correlation between parental height and offspring height would be 1.0, and if there were no relationship, the correlation would be 0.0. In our example, the correlation is 0.41, and the slope of the line is 0.78; there is a resemblance, but some offspring are taller than their parents, while others are shorter. This correlation is one measure of **heritability**, the amount of resemblance among relatives that is due to shared genes. Offspring tend to resemble their parents and their siblings because the **phenotype** (physical characteristics) of an individual is determined in part by its genotype and an individual receives its genes from its parents and shares those genes with its siblings.

![Figure 3.9](image-url)

Plot of offspring height against mean height of the parents of plants grown in a greenhouse. The heritability of this trait (the slope of the line) is 0.78. (Unpublished data courtesy of Ann Evans.)

However, comparing offspring trait values against parental values is only one measure of the heritability of a trait. In the example above, because we used information from both parents, the slope is exactly equal to the heritability. If we had measured the trait in only one parent, we would have information about only half of the genes being contributed to the offspring, and the slope would be one-half the heritability.

The other common way of measuring heritability is to measure the correlation among siblings. Suppose we took two seeds from each of many plants. We could germinate the seeds and grow the pairs of siblings, measure their heights, and construct a graph much like Figure 3.8, except that now the axes would be the heights of the two
siblings, and each point would represent a sibling pair. Again, the slope would measure heritability, with the exact relationship depending on whether the individuals shared both parents or just one parent. We can do such an analysis with cousins or with any individuals that are related as long as we know their relationships. Nor are we restricted to using pairs of individuals. Various statistical techniques can be used to measure heritability in groups of related individuals with different degrees of relatedness.

During the first decades of the 20th century there were heated disagreements between geneticists that focused on discrete traits and those that focused on continuous traits. At the time it was not clear that continuous traits were controlled by the same sorts of genetic factors that controlled discrete traits. These differences related to arguments about whether evolution occurred through gradual change or by large, discrete changes. These disagreements were resolved by R. A. Fisher (see Box 2D), who showed that continuous variation and its heritability could be explained by assuming that a continuous trait was determined by very many genes, each of which had just a small effect on the phenotype.

There is a critical distinction between the heritability of a trait - how closely offspring resemble their parents - and whether that trait has a genetic basis. Heritability requires that phenotypic differences among individuals be due, at least in part, to the genetic information those individuals carry. In Box 3C we describe a case in which height is genetically determined. In that example, some individuals have a genotype of AA, some Aa, and some aa - height, therefore, is at least partially related to the information passed down from an individual’s parents. Instead, imagine that all individuals in the population have the same genotype, AA. Assume, however, that height also depends on the amount of nitrogen in the soil. If the population is growing in a field that varies in soil nitrogen from spot to spot, then individuals will differ in height. However, none of those phenotypic differences will be due to genotypic differences. Consider what would happen if we were to measure these plants, collect their seeds, and grow the offspring in that same field. By chance the seeds from a tall individual that grew in nitrogen-rich soil would end up in a variety of soils, some nitrogen rich and some nitrogen poor, and thus would have both tall and short offspring. The resulting correlation between parental height and offspring height would be 0 and the heritability of height in that population would be 0. Yet there is still a gene in that population that determines height. In this case, the heritability of height is zero because phenotypic variation in height is due to variation in an environmental factor, not to variation in the gene for height.

This example also demonstrates that the heritability of a trait depends on the frequencies of its alleles in the population. When the frequency of A is 1.0 – all individuals have the AA genotype – the heritability of the trait is 0. Thus, heritability estimates for the same trait can differ among populations, or in the same population measured at different times. Heritability estimates are always specific to the population and environment in which they are measured.

Heritability values tell us whether there is genetic variation for a trait in a population, and if so, whether there is just a little variation or a lot of variation. In terms of evolution, the amount of genetic variation may impose a constraint on evolution (see Chapter 4). If there is no genetic variation, the constraint is strong. No matter how much natural selection there is on a trait, there will be no genetic response; the potential for evolution requires variation in a trait before change can occur. If there is a little bit of genetic variation, the constraint is weak; there will be a genetic response, but it will be small, and evolution will proceed slowly. If there is a lot of genetic variation, there is almost no constraint on evolution.
Box 3C
A Simple Genetic System and the Resemblance of Relatives

This example shows how heritability as measured as a correlation between the phenotype of parents and offspring can be related to the effects of individual genes. Although this example is based on just a single gene, the same principles hold no matter how many genes affect a trait. Consider a simple genetic system in which plant height is determined by a single diploid gene. We assume that individuals with genotype AA are tall (100 cm) and those with genotype aa are short (20 cm).

Case 1: Strict Additivity
If individuals with genotype Aa have a phenotype that is exactly intermediate between AA and aa individuals, then genetic variation is strictly additive. In this case, Aa individuals would be intermediate in height (60 cm tall). Because the effects of the alleles are strictly additive, we can predict the phenotypes of the offspring of a cross. If both parents are tall, the cross will be AA x AA, and all offspring will be tall. If both parents are short, the cross will be aa x aa, and all offspring will be short. If one parent is tall and the other short (AA x aa), all offspring will be 60 cm tall (Aa). That is also the height that we get by averaging the parental phenotypes; the mean offspring phenotype equals the mean value of the parents’ phenotypes. If one parent is 100 cm tall and the other is 60 cm tall (AA x Aa), half the offspring will be 100 cm tall and half will be 60 cm tall. Again, the mean value of the parents’ phenotypes, 80 cm, exactly equals the mean value of the offspring phenotypes. Note that, for this cross, no parent or offspring is actually at the mean height; the mean is a descriptor of the group, not a property of any particular individual. A graph of mean parental phenotype against mean offspring phenotype (part A of the accompanying figure) has a slope of 1.0. That is, the heritability of this trait is 1.0, because we can perfectly predict the mean offspring phenotype from our knowledge of the parental phenotypes.

Case 2: Dominance
Now assume that A is dominant to a, such that Aa individuals are also 100 cm tall. In this case, predicting offspring phenotypes becomes more difficult. If both parents are short, all offspring will be short. But if both parents are tall, their genotypes could be both Aa, both AA, or one AA and the other Aa. In the latter two instances, all offspring will be 100 cm tall. But if both parents are Aa, then 1/4 of the offspring will be aa and will be short. The mean offspring phenotype will be 80 cm (3/4 x 100 + 1/4 x 20)
x 20), even though the mean phenotype of the parents was 100 cm. If we assume that the two alleles exist at equal frequencies in our population, then a graph of mean parental phenotype against mean offspring phenotype will have a slope of 0.67 (part B of the figure). The heritability of the trait is less than 1.0 because some of the genetic variation is nonadditive due to the dominance relationship. In other words, some offspring differ phenotypically from their parents because of the effects of dominance; if they do not inherit the dominant allele from either of their parents, they do not resemble their parents. Thus, the exact heritability of a trait in a population depends on both the degree of dominance and allele frequencies in the population.

**Fidelity of information transmittal**

The resemblance of parents and offspring is explained in the double-stranded nature of DNA and its replication process. The double strands accomplish two tasks at once, efficient replication and keeping the chromosome true to its template (Table 3.1, principle 2). During cell replication, each daughter cell must end up with all of the information contained in the parental cell. The DNA duplication process (Figure 3.3) results in two new chromosomes in a single operation. Those chromosomes are easy to separate; a fact which becomes important when it comes time to pass the new chromosome to the daughter cell. If the DNA molecule consisted of just a single strand, the two halves would tend to stick together through the same forces that hold the base pairs together. Also, a DNA molecule is very long and if unpaired, all of those "sticky" nucleic acids would tend to match up with bases elsewhere on the same strand, creating a tangled mess. Double-strandedness solves the templating problem while also keeping the information in a form that is easy to replicate and store.

Double-strandedness is also one way in which the information can be kept from changing. Occasionally, a chemical reaction will happen to the base pairs on one of the two strands changing its composition. When that occurs, specialized proteins will come along and repair the strand using the information from the complementary strand. Thus, the two strands act as “back up” copies of each other. We discuss these processes in more detail in the follow section.

**New Information**

**Capability for new information**

Living systems are the product of natural selection, which requires genotypic and phenotypic variation (see Chapter 4). If species could not evolve, they would go extinct; the constantly changing environment would eventually reach a point that survival and reproduction of that species would not longer be possible. Thus, the persistence of life requires the generation of variation in the form of new information (Table 3.1, principle 3). Of course, all species eventually go extinct; their evolution allows them to persist for far longer than they would otherwise.

The genetic system creates new information through a variety of processes which can be collected under two broad headings: mutation and recombination (Table 3.1, principles 4 and 5). **Mutation** is a change in the sequence of the DNA molecule. In the next section we explore the many ways that this can happen. **Recombination** is, as the name implies, the bringing together of DNA sequences in new combinations.

In general, mutations create new information thereby increasing the amount of genetic variation, although not all genetic variation results in phenotypic variation (as we will see later in this chapter). There are a few types of mutations that involve the loss of DNA and can even act to decrease variation. Recombination is different in that it is equally likely to increase as decrease variation. However, it produces new variation at a faster rate than mutation. Thus, both processes create new information in different ways.
and at different rates.

The rates at which mutation and recombination happen are determined by evolution and natural selection. The basic chemistry of DNA molecules is such that mutations constantly occur. But the rate at which those mutations get passed along to the next generation is controlled, in part, by the efficiency and accuracy of genetic repair mechanisms. Similarly, the rate of recombination is under the control of the organism through a variety of mechanisms that are described below. If the mutation and recombination rates are very low, then extinction eventually occurs. But if those rates are too high, offspring tend not to resemble their parents. Although the environment is always changing, it is not changing that quickly. For the most part, offspring that resemble their parents tend to do better at surviving and reproducing. Thus, natural selection will hone the mutation and recombination rates.

Natural selection has resulted in organisms that have both some background mutation rate, that is, a tendency for mutations to appear under normal circumstances, and in at least some species an increase in mutation rate under stressful conditions. If organisms never mutated, genetic variation would eventually disappear; evolution would cease, and life would likely go extinct as the environment changed and species did not. Only those forms of life that continue to mutate avoid extinction.

**Mutation**

The concept of mutation encompasses a wide variety of types of changes in DNA sequence. The most obvious is a simple change in the identity of one of the base pairs, for example if one of the strands changed from a sequence of ATTCCG to ATACCG, with a corresponding change in the matching strand. Some changes swap the order of base pairs, so that the previous sequence could change to ACCTTG, the middle four bases flipping position. Such flips can include stretches of DNA as long as tens of thousands of base pairs. Some changes involve deletions, for example changing ATTCCG to ATCG (again, the deleted piece can be very long). Some changes involve moving pieces of DNA, for example from one chromosome to another (Box 3D). Some changes involve creating duplications; the previous sequence could be changed to ATTCCTTCGG. Such duplications could be just a few base pairs, consist of very long stretches of DNA, or even involve the duplication of whole chromosomes. In some cases the entire set of chromosomes is duplicated, so that an organism that had 8 chromosomes would now have 16, and so on. The duplication of whole genes and chromosomes can be an important source of new gene functions. Once there are two copies of a gene, one can continue to perform its original function while the other can evolve a new function. We describe how genes function later in this chapter.
Box 3D
Barbara McClintock

While science can sometimes seem like an all-boys’ club, there have been many women who have made substantial contributions to the scientific edifice. Born on June 16, 1902, Barbara McClintock stands as one of the world’s most distinguished geneticists, discovering the phenomenon of the movement of genes from one chromosomal location to another, and demonstrating the roles of chromosome structures in the conservation of genetic information. She is a Nobel Laureate, having received the Nobel Prize for Physiology and Medicine in 1983.

Born the third of four children in Hartford, Connecticut, from the time she was three until she started school she lived with her aunt and uncle in Massachusetts, in order to alleviate the financial burden on her parents. Her father, Thomas Henry McClintock, was a physician, and it took him several years to successfully establish a practice so that he could comfortably support his family. When he was finally able to do so in 1908, the McClintocks moved to semi-rural Brooklyn, New York. Barbara was solitary, independent and a tomboy from early on; she had a good relationship with her father, but her relationship with her mother was plagued with strife. During high school she discovered her love of science and wanted to attend Cornell University, but her mother resisted the idea of higher education for her daughters, believing it would make them unmarriageable. Fortunately, her father intervened on her behalf, and she entered Cornell in 1919.

McClintock enrolled in the College of Agriculture where she studied botany, and received a BS degree in 1923. As an undergraduate, her interest in genetics was sparked when she took a course in the subject taught by C. B. Hutchinson, a plant breeder and geneticist. Hutchinson was impressed by McClintock’s keen interest in genetics, and invited her to participate in his graduate genetics course, even though she was still an undergraduate. McClintock would later point to that invitation as the reason she continued in genetics.

McClintock remained at Cornell as a graduate student from 1923 to 1927, and then as an instructor until 1933 McClintock. She was instrumental in assembling a group that studied maize (corn) and the then-new field of cytogenetics, which focused on the structure and function of chromosomes, bringing together plant breeders and cytologists. McClintock’s work focused on developing ways to visualize and characterize maize chromosomes, techniques which are still taught today. By studying chromosomes’ morphology, McClintock was able to link specific chromosomes to groups of traits that are inherited together, triggering a surge of interest in maize cytogenetics.

In 1930 she became the first person to describe the interaction of chromosomes during meiosis (Figure 3.4), and the next year she and a graduate student, Harriet Creighton, proved a link between the process during meiosis which results in the physical reassortment of chromosomal pairs and the recombination of genetic traits. She observed that when chromosomes were recombined, the resulting phenotype would result in the inheritance of new traits; until then, it was only hypothesized that genetic recombination was possible during meiosis, although there was genetic evidence that it occurred.

During the summers of those years, she worked with geneticist Lewis Stadler at the University of Missouri - Columbia, who introduced her to the use of X-rays to gener-
ate mutations. Using mutagenized maize, McClintock discovered ring chromosomes, which form when the ends of a single chromosome fuse together after radiation damage. From this, she hypothesized the existence of a structure on the chromosome tip that normally ensures the stability of the genetic material.

Unfortunately, McClintock was unhappy at the University of Missouri because of how she was treated by the other faculty, for example being excluded from faculty meetings. Early in 1941 she was invited by the Director of the Department of Genetics at Cold Spring Harbor to spend the summer there, and used the opportunity to take a leave of absence from Missouri. In December of the following year, she was offered a research position at the Carnegie Institute of Washington’s Department of Genetics Cold Spring Harbor Laboratory, where she continued her work with the breakage-fusion-bridge cycle, using it as a substitute for X-rays as a tool for mapping new genes.

By this time she was receiving wide recognition for her groundbreaking work. In 1939, at the age of 37, she was elected vice-president of the Genetics Society of America, and in 1945 she became its first woman president. In 1944 she became the third woman in history to be elected to the National Academy of Sciences. By 1951 she had received the Achievement Award of the Association of University Women, and been awarded two honorary degrees. She was undoubtedly one of the best-respected cytogenetics at that time; however, in that same year this respect was to be tested with her theory about the nature of the gene.

In the summer of 1944 at Cold Spring Harbor McClintock began a series of systematic studies on the mechanisms of mosaic color patterns of maize seed and the unstable inheritance of mosaicism. She identified two new dominant and interacting genes she named Dissociator (Dd) and Activator (Ac), finding that Dd did not only cause the chromosome to break, but had a variety of effects on neighboring genes when Ac was also present. By 1948 McClintock had discovered that Dd and Ac could move on the chromosome; the effects of this transposition could be observed via the changing patterns of coloration in maize kernels over generations of controlled crosses. McClintock described the relationship between the two loci through microscopic analysis, concluding that Ac controls transposition of Dd, and the movement of Dd is accompanied by breakage of the chromosome. The transposition of Dd in different cells is random, moving in some but not others, thus producing color mosaicism. The size of the colored spot on the seeds is determined by the stage of seed development reached during dissociation, and the transposition of Dd is determined by the number of Ac copies in the cell.

Over the next two years McClintock developed a theory by which these mobile elements regulated genes by inhibiting and modulating action. She referred to Dd and Ac as “controlling units” (later “controlling elements”) to distinguish them from genes. Hypothesizing that gene regulation explained how complex multicellular organisms made of cells with identical genomes have cells with different functions, this theory challenged
the concept of the genome as a static set of instructions passed between generations. It would not be until a few years later, after the structure of DNA was discovered (see Box 3A) that the more complex nature of genes began to be revealed.

McClintock reported her findings in a 1950 paper, and the next year at Cold Spring Harbor’s annual genetics symposium. Her paper addressed a critical missing piece in the understanding of genes, how their expression was regulated. Unfortunately for McClintock, she conflated the action of Ac as a controller of transposition with its potential role as a controller of gene expression. The combination of a conceptually difficult idea that challenged accepted wisdom along with an extremely long and dense verbal (and later written) presentation, meant that the reaction of her peers was mostly to ignore her. While McClintock was greatly admired as an experimentalist, her theory seemed to go to far beyond her data. It did not help that she aligned herself with an even more extreme iconoclast, Richard Goldschmidt, who gave the keynote address at that same symposium.

Ironically, McClintock is now primarily remembered for what to her was a side issue, the transposition of genes. Her major point was that one had to considered the context of a gene to understand how it was regulated. While the details of her theory were wrong, she was correct that more than the expressed part was critical. However, because she presented these ideas in review articles and non-peer-reviewed venues and because it went against the current theory it was not appreciated at the time. In hindsight, however, we can see that she was a pioneer in understanding gene regulation. Even in the 1960s, when the first models of gene regulation were worked out in bacteria, McClintock was not credited with the fundamental concept of regulation of expression in one part of the chromosome by elements in other parts (because the mechanism of that regulation was so different from the one she had earlier described). McClintock took a philosophical view on it, however, writing in 1973 on her decision to cease publishing by 1953 that “One must await the right time for conceptual change.”

McClintock officially retired from her position at the Carnegie Institution in 1967, but continued working with graduate students and colleagues in Cold Spring Harbor Laboratory as scientist emerita. Although she published little after the early 1950s, her previous achievements continued to gain her honors. She was awarded the National Medal of Science by Richard Nixon in 1971, and Cold Spring Harbor named a building after her in 1973. In 1981 she was the one of the first recipients of a MacArthur Foundation Grant (called “genius grants”) and awarded the Albert Laster Award for Basic Medical Research. She received the Wolf Prize in Medicine and the Thomas Hunt Morgan medal by the Genetics Society of America the same year. In 1982 she was awarded the Louisa Gross Horwitz Prize from Columbia University. All of this was capped in the following year when she received the Nobel Prize for discovering mobile genetic elements.

She remained a regular presence in Cold Spring Harbor community, giving talks on mobile genetic elements and the history of genetics research for the benefit of younger scientists. She died in Huntington, New York on September 2nd, 1992 at the age of 90. Today McClintock is widely held up as a role model for girls in children’s literature. On May 4, 2005 the United States Postal Service issued their “American Scientists” commemorative postage stamp series, which featured McClintock.

It is common to think of mutations as having a dramatic effect on the phenotype (Figure 3.10A and 3.10B). But mutations can also have very subtle effects (Figure 3.10C), so small that they are only apparent in very close contrast to the species as a whole. Some mutations may even have no effect at all on the phenotype. Other mutations, such as deletions, may destroy the function of a gene, which can result in extreme or even fatal effects on the organism. Duplications of whole genes may have no
immediate effect on the phenotype, but may allow the eventual evolution of a new function in one of the two copies. Further, in order to produce a permanent change in a population rather than a single individual, a mutation must occur in gametes or gamete-producing cells.

Figure 3.10
Mutations can have a variety of effects on organisms from large and striking to small and subtle. (A) Typically *Drosophila melanogaster* (fruit flies) have red eyes and yellow bodies (lower left in photo), but various mutations that affect pigment formation results in flies with a variety of eye and body colors. This is an example of a qualitative difference. (Source: Wikipedia) (B) A single mutation also can have a large effect on a quantitative (continuous) trait. The mouse on the left has a mutation that prevents the production of a hormone that controls the amount of fat in its body. (Source: Wikipedia) (C) The plant *Arabidopsis thaliana* typically has small oval leaves. Mutations in some genes can create small changes in size or shape, while others create wildly misshapen leaves. (Photo courtesy of Hirokazu Tsukaya)
Mutations are relatively rare events. Mutations of single base pairs occur at a frequency of 10–8 to 10–10 per base pair per generation. For the average-sized gene, mutations occur at a frequency of 10–5 to 10–7 per gene per generation. On average, in a given population for most species, one individual in ten has a new mutation somewhere in its genome each generation. New technologies that allow the measurement of variation in DNA are quickly adding to our understanding of mutation rates. At any one time there are likely to be many variants of a given gene even though mutations of a per gene or per individual basis are rare, since most species consist of millions of individuals. It is the accumulation of a lot of small mutations that provides much of the variation upon which evolution depends. Usually mutations with large effects on the phenotype are harmful, while mutations with very small effects can be either harmful or beneficial. Despite nearly 90 years of trying to determine the frequency of small and large, or harmful and beneficial mutations, this is still an intense area of research.

Mutation is never purposeful. Mutations do not occur because they can make individuals or species better adapted to their environment. Rather, both beneficial and harmful mutations occur all the time and are then winnowed out by natural selection. We often observe mutations only after many of the harmful mutations have been eliminated because the individual with that mutation dies or produces few offspring. Thus, it can seem as if all mutations are beneficial. During the 1980s some scientists claimed to demonstrate that Bacteria placed in conditions where they required specific mutations to survive—the ability to use unusual food sources—purposely mutated to better survive under those conditions. Others showed, though, that the results could be explained entirely because only bacteria with the correct mutations were ever seen.

Organisms under stressful conditions have an increased rate of mutation, and such an increased mutation rate can increase the chances that beneficial mutations will appear. But there is a distinct difference between a mechanism that simply increases mutation rate and one that directs those mutations to be best suited for the individual’s environment.

Although mutation is not purposeful or directed, it does not mean that all mutations are equally likely. Random does not mean that something occurs with equal frequencies (see Chapter 1). This non-equality of the occurrence of mutations happens in several ways. For example, because of their chemical similarity (Figure 3.1) it is much more likely that an A will change to a G and vice versa, than either will change to a C or T. Similarly swaps between C and T are more likely. Nor is the likelihood of mutation the same at all base pairs. There are mutational “hotspots” on chromosomes where base pair changes can be 10 times more likely than at other places. Similarly, some types or locations of deletions or duplications are more likely than others.

Figure 3.11
Frog with an extra hind leg due to a mutation caused by pollution.

Certain types of phenotypic changes are more likely to happen. For the most part, the development of an organism from fertilization to fully formed individual is highly regulated (see Chapter 6) so that small changes in that process are more likely than large changes and some large changes are more likely than others. For example, terrestrial vertebrate animals have four limbs. Mutations that change the length of those limbs, making them longer or shorter, happen frequently. Just consider the relatively long legs and arms of a
professional basketball player as compared to that of an average person. Eliminating limbs entirely is also possible (e.g., the hind limbs of whales, see Figure 4.1). But, there are no six-limbed terrestrial vertebrates, indicating that such a change in development is either extremely difficult or impossible. Although there have been instances of vertebrates born with extra limbs due to mutations (Figure 3.11), these limbs are often misshapen and not functional, showing that the developmental limitations on the kinds of mutations that are possible can provide important directionality to the course of evolution.

Exchange and recombination

Recombination is the process by which the genetic material of one individual is mixed with that of another. In Bacteria and Archaea this process occurs in a haphazard fashion. In some species, when a cell dies and ruptures, its DNA can get taken up by other cells. Most of the time, the other cell simply uses the individual nucleotides to build new copies of its own DNA molecules. Occasionally, however, a stretch of DNA from one individual can get incorporated into the chromosome of another. Sometimes this exchange occurs when a virus picks up a piece of its host’s DNA during replication (Box 3A). This sort of exchange of genetic material can occur among very distantly related species, including Eukaryotes, Bacteria and Archaea. How often this occurs among Bacteria and Archaea is not known and currently a hotly debated issue. This type of genetic exchange can act like an extreme form of mutation, bringing genes with entirely new functions to a species. However, this sort of movement is limited to cellular functions (see Chapter 5). A cat is not going to suddenly grow a dog’s tail. In some cases, the exchange process among Bacteria is somewhat less haphazard. Some species have specialized structures and special enzymes for transferring DNA from one individual to another (Figure 3.12). Such transfers, however, still occur only occasionally.

The process of DNA exchange and recombination became regularized with the evolution of Eukaryotes. Recombination occurs during the process of meiosis during which a diploid individual produces haploid gametes (Figure 3.13). Even those Eukaryotes that are haploid for nearly their entire life will mate and produce diploid cells that have at least a brief existence. Recombination occurs during the process of meiosis during which a diploid individual produces haploid gametes (Figure 3.13).

The process of meiosis creates new variation in several ways. For example, if an entire population consists of just red-flowered (CRCR) and white-flowered (CWCW) individuals as in Figure 3.5, the process of gamete formation and mating creates new pink-flowered (CRCW) individuals.

Recombination also allows genes linked in some way to separate, creating greater variation. Consider Mendel’s Law of Independent Assortment (Figure 3.7). Since independent assortment occurs only if genes are located on different chromosomes, if two genes are on the same chromosome, they will very likely end up in the same gamete. The process of recombination unlinks those genes by breaking the DNA chain of each chromosome and creating new links with the other chain. For example, if there were only genes for seed shape and seed color (Figure 3.8) and there was no recombination, then the only types of gametes that would exist would have either the SSCY combination of genes or the SWCG combination. The only types of offspring would be either SSSCYC, SWSWCGCG or SSSWCYCG. Tracing patterns of recombination has even helped scientists uncover the structure and function of genetic material itself. Only if there was recombination could there be combinations such as SSSWCYCY or SSWWCYCG. One of the confirmations that genes were carried on chromosomes came from the work of Thomas Hunt Morgan while studying patterns of recombination in fruit flies (Box 3E).
From the perspective of evolution, recombination is important because it creates new combinations of genes which can increase the rate of evolution. Consider two different mutations (A and B) that each increase the fitness of an organism. An individual has mutation A. Since mutations are rare, it is highly unlikely that its offspring would also have mutation B. But, if that individual mated with an individual that carried mutation B, then many of their offspring would carry both mutations and be more fit. Hybridization, matings between individuals of different species, is an extreme example of this process of bringing together very different genes to form new combinations.
Box 3E
Thomas Hunt Morgan

Thomas Hunt Morgan helped create the modern science of genetics, both because of the ideas that he fostered and because of the methods that he developed to test those ideas. Yet, he was initially skeptical of the entire field and later largely abandoned it to return to his first love, the development of marine organisms.

Morgan was born on September 25, 1866, in Lexington, Kentucky, the eldest son of an old southern aristocratic family. Thomas could trace his family deep into the history of the region; he was the nephew of Confederate General John Hunt Morgan, and the great-grandson of the first millionaire west of the Allegheny Mountains, John Wesley Hunt.

As with many biologists, Morgan showed an interest in the study of life before the age of 10, collecting birds and their eggs as well as fossils. He continued to pursue these interests at a time in America when a career in the sciences was unusual, receiving a B.S. degree in biology in 1886 from the University of Kentucky, the only person in his class to do so. That summer he visited the seashore laboratory of Alpheus Hyatt at Annisquam, Massachusetts, starting him on his life-long interest in the biology of marine organisms. When he began his graduate studies at Johns Hopkins University, he brought with him is newfound interest, beginning a project on the development of sea spiders. Much of this work was done at the Marine Biological Laboratory at Woods Hole, Massachusetts, a place that inspired many biologists (see Boxes 5B and 7D). After receiving his Ph.D. in 1890, he was hired as a professor of zoology at Bryn Mawr College, Johns Hopkins’ sister school (the schools were segregated by gender, like nearly all colleges and universities of the time). He stayed until 1904. For the next 24 years, he took up a position as professor of Experimental Zoology at Columbia University, then became a professor of biology at the California Institute of Technology, Pasadena, where he was also the director of the William G. Kerckhoff Laboratory.

During the year he received his Ph.D. he received a fellowship to visit Europe, spending a great deal of time in Naples, Italy. There he met Hans Driesch, a German biologist noted for his work in embryology, with whom he later collaborated. Perhaps as a result of this association, he turned to experimental embryology after returning from his trip. Although he would later become deeply involved with genetic research using the fruit fly, *Drosophila*, throughout his life he worked on the problem of developmental stability using the sea squirt, *Ciona*, as a model organism. Sea squirts are closely related to the group that gave rise to vertebrates and, thus, much studied as a glimpse into the origins of vertebrate traits. He was associated with Woods Hole continuously from 1902 onward, actively taking part in biological expeditions to the Bahamas and Jamaica.

As a student Morgan was boldly critical, skeptical and of a very independent judgment. His work in experimental embryology and regeneration won him a high reputation during his early career, becoming the president of the American Morphological Society in 1900. He was an intent worker, impatient of unnecessary interruptions, and a prolific and sometimes hasty writer. Although he did no genetics work prior to 1905, his early papers show considerable distrust of Mendelian laws. Even up to 1909, after working himself with mice and others in his department began working on *Drosophila*, Morgan...
did not clearly distinguish problems of heredity from those of development. However, it was the work he did with *Drosophila* that eventually brought him around.

Morgan’s groundbreaking work was built upon a long line of scientists interested in genetics. *Drosophila* were first bred in quantity by Charles W. Woodworth, who studied them at Harvard during the winter of 1900-1901. There he suggested to William E. Castle that the flies would be useful for genetic study; Castle and his students used it for studies on the effect of inbreeding. From their activities, Frank E. Lutz became interested and suggested them to Morgan as an experimental animal.

Throughout his career, Morgan always insisted that his work was that of a team. He assembled a group headed by Alfred H. Sturtevant, Hermann J. Muller and Calvin B. Bridges who carried out experiments in a much larger scale than anyone had attempted before. His laboratory at Columbia University became known as the Fly Room and was the hub of some of the most important work in genetics in the first part of the 20th century. Morgan’s early *Drosophila* papers focused attention on the demonstration that the gene for white eyes was associated with those that determined gender. From these early studies, Morgan’s skepticism about Mendelian genetic laws quickly faded. In a 1911 letter in Science, he put forward the theory of that genes were linked to each other on chromosomes in a linear arrangement. He also discovered that males and females had differences in the chromosomes that determined sex, with females having two large chromosomes, dubbed X and so being XX, while males had one large and one small chromosome, dubbed Y and so being XY. In 1915 Morgan, Sturtevant, Bridges and Muller wrote The Mechanism of Mendelian Heredity, a text that would stand as a seminal work of genetics, and would form the basis of geneticists’ efforts for decades to come.

Because of Morgan’s success with *Drosophila*, it was picked up for genetic studies all over the country and the world, with Columbia at the center of an exchange network of promising mutant strains. When he established the biology division at the California Institute of Technology, he wanted to distinguish it by focusing on genetics and evolution, experimental embryology, physiology, biophysics and biochemistry; he succeeded not only in advancing and expanding biology at CIT, but in spreading his impact throughout the entire field of biology via his efforts with *Drosophila*.

In keeping with the personality that lead to sometimes hasty research publication, Morgan did not keep organized notes on experiments, but would pull envelopes and scrap paper out of his pockets when examining ongoing and completed experiments. His colleague Tyler kept an eye on Morgan’s experiments, and probably helped to keep the chaos at bay. Morgan, for all his apparent absent-mindedness, was extremely passionate about experimentation and distasteful of speculation, believing only what could be proven. He was known for his sardonic wit; in 1909 during his speech at the American Breeders Association, he was critical of genetics, saying, “In the modern interpretation of Mendelism, facts are being transformed into factors at a rapid rate. If one factor will not explain the facts, then two are involved; if two prove insufficient, three will sometimes work out....” As fate would have it, a year later he discovered the gene that controls white eye pigmentation in *Drosophila*. This is an excellent example of the fact that real scientists change course when confronted by facts.

In 1904 Morgan married Lilian Vaughn Sampson, a former research student of his at Bryn Mawr and a frequent associate in his lab work. They eventually had a son and three daughters. While at Cal Tech, he lived in a comfortable ranch house on the north side of Kerckhoff Labs, and would give a General Biology Seminar at 7:30 PM on Tuesdays. He would open the seminar by commenting on stories in the New York Times with a scientific bent, making fun of human gullibility. He would introduce the speaker for the night, then sit in the front row next to his wife, and usually fell asleep within two
sentences. He wife would nudge him awake, and he would be refreshed and usually had acute questions for the speaker at the end.

Morgan died on December 4, 1945, leaving behind a legacy writ large in genes. He was highly decorated during his lifetime, becoming a foreign member of the British Royal Society in 1919, receiving the Darwin Medal in 1924, the Copley Medal in 1939, and finally the Nobel Prize in 1933. When he established the biology division at Cal Tech, he wanted to distinguish it by focusing on genetics and evolution, experimental embryology, physiology, biophysics and biochemistry; he succeeded not only in advancing and expanding biology at Cal Tech, but in spreading his impact throughout the entire field of biology via his efforts with Drosophila. Morgan’s work laid the foundation of the science of genetics and the theoretical foundation for the mechanism of evolution via natural selection, fields that are still seeing advances today.
Random processes

Mutation and recombination are an important source of contingent effects in living systems (Table 3.1, principle 6). As discussed previously, mutations occur randomly; so, while they are an important part of new variation for the evolutionary process, when and where they occur is unpredictable. Cougars are found in North America and tigers in Asia because the mutations that led to each of these species appeared on one continent and not the other.

The random nature of mutation has direct implications for human health. For example, humans and the influenza virus are in a continual arms race. The human immune system produces molecules that are able to attack the virus when it enters your body (see Chapter 6). Your body recognizes that a particle is a flu virus because of the proteins in the virus’s outer envelop, and your body has the ability to continually produce new molecules to match those proteins. Once a particular molecule is produced your body continues to produce them for several years and they remain circulating in your blood. This is important because those molecules can attack an entering virus particle immediately; you only get sick when a virus enters your body that is not recognized by your immune system. In that case, the virus has time to multiply and attack your body for several days before your body can mount a counterattack. Vaccination keeps you from getting sick by prompting your body to make the antivirus molecules before the virus enters your body.

It takes about six months for flu vaccine production to go from the start to the point of having the tens of millions of doses necessary to protect the U.S. population. So, it is important to know well ahead of time which influenza strains will be circulating in the population. The problem is that the influenza virus is continually mutating, changing the composition of the proteins in its outer envelop. Those mutations are unpredictable, to some extent. Most years the mutations create small changes in the proteins and so the chance of the virus making you very sick is small. Based on past patterns of mutation, those small changes are somewhat predictable. Scientists can use mathematical models to decide which form of the vaccine should be produced – even if the vaccine is not an exact match, being close to a match is enough to give your body the head start it needs to mount its defense. In some years, though, the mutations create much larger changes. While scientists know that these mutations with larger effects have a small chance of happening each year, exactly when they will occur and what the nature of the mutation will be is much harder to predict. In those years the vaccine will not match the strains and many people will end up sick with the flu.

The independent assortment of genes on different chromosomes is the best example of truly random processes in biology. To see why this must be so, consider the following thought experiment. In some species, not all of the products of meiosis go on to produce gametes. For example, in human females only one of the four gametes goes on to become an egg. Imagine that there was a gene that had two alleles and one was able to increase its chances of ending up in that egg. Even a very small advantage in getting passed along to the next generation would mean that this allele would quickly become the only type in the population. Alleles of any other genes that are linked to that allele would also be quickly fixed, and alleles linked with the unfavored allele would quickly become much rarer in a population. In fact, the entire meiotic machinery is geared to suppressing the possibility of such skewing of gene segregation during meiosis. The end result is that unlinked genes end up in gametes in the proportions expected by random chance.
Information Usage
In the previous sections we described how the genetic system accounts for resemblances among relatives in Eukaryotes, especially complex multicellular organisms. Although the link between genotype and phenotype is much simpler for single-celled organisms, the problem of the resemblance of relatives was first considered and solved for those more complex organisms. This is another example of the advancement of technology leading to further biological discovery, albeit in a somewhat ironically backwards fashion. While the greater number and duplication of chromosomes in Eukaryotes results in a much more complex information system, Bacteria and Archaea are much simpler to study, because much of the phenotype involves the direct expression of the information stored in the DNA; there is only a single chromosome, and thus one copy of each gene.

Transcription and translation
Transcription and translation is the two-stage process (Figure 3.14) by which the information encoded in the DNA molecule is used. First, the information contained in the DNA molecule is copied to a RNA molecule (transcription). One of the two DNA strands is used as a template and through a process similar to DNA replication, a complementary strand of RNA is produced. The RNA strand consists of A, G and C bases just like DNA. The one difference is that thymine (T) is replaced with a closely related base, uracil (U).

![Figure 3.14](image)
The process of transcription copies the information contained in the DNA molecule to an RNA molecule. That information is then used to build a protein in the process of translation using other types of RNA molecules. (Created by Madeleine Price Ball, Source: Wikipedia)

Next, the information in that RNA molecule is used to create a protein (translation). The correspondence of the sequences of bases in the RNA molecule and the amino acids in the protein molecule is called the genetic code. Thus, the genetic code is how the information in the genotype is translated into the phenotype.

The genetic code is built on triplets of bases called codons. That is, a set of three bases corresponds to one amino acid. With four different bases (A, G, C, and U), there are 64 different possible three base combinations. However, because there are only 20 different amino acids, nearly all amino acids are coded for by more than one combination of bases (Figure 3.15). For example, tyrosine is coded for by two sequences: UAU and UAC. In this case, the first two bases are the same and only the third base differs. Thus, the genetic code is redundant as a given amino acid is coded for by more than one possible combination of bases. The use of only 4 bases to code for 20 amino acids by triplet codons is an example of an emergent property because the grouping of the individual bases into codons creates new information in the same way that words are combinations of letters.
The redundancy of the genetic code is part of the robustness of the information system (Table 3.1, principle 7). Earlier we mentioned that mutations between A and G and between C and T are much more likely than the other possible types. When an amino acid is coded for by just two different codons, the two codons have in their third position bases that are either A and G or C and T. So, the most likely mutations in third-position bases will not result in the resulting protein and the phenotype of the organism. For example, consider arginine. A mutation from AGA to AGG still codes for arginine. Because of this lack of change, we speak of silent mutations. Whether different triplets that code for the same amino acid are truly equivalent and silent is an active area of research.

Diplody further increases the error tolerance of the system. A mutation that simply changes one base into another will likely have a very small change in the properties of the resulting protein. But many types of mutation can completely destroy the functioning of the protein, such as deletion, changes in the order of the sequence, or a change in the base sequence that causes translation to halt before the entire protein is
produced. However, a diploid organism has two copies of each gene, so even if one copy does not produce a functioning protein, the other will. Some forms of dominance will result in the dominant form of the gene producing the functioning protein, thus preserving its function in the phenotype.

The phenotype of an organism is not just determined by whether a protein is produced or not, or by the form of the protein; it is also determined by the relative amount of protein produced. This can be regulated at many different steps in the process: the rate of transcription or translation can be sped up, slowed down, or shut off entirely. The rate of transcription is controlled by the DNA sequence adjacent to the region that is transcribed. These parts of the sequence are referred to as noncoding, regulatory regions. In translation, typically many copies of a protein will be made from a single RNA molecule, so the amount of protein can also be regulated by how long that RNA molecule lasts before it is broken down. Cells contain a complex machinery for this regulation (see Chapter 5).

In humans, only about 2% of the information in the DNA ends up as proteins. Typically in Eukaryotes much of the DNA that is transcribed into RNA is not translated. Some of this RNA is spliced out of the strand before translation. Other stands of RNA are part of the translation machinery, or combine with proteins to regulate their function. Yet other RNA strands directly interact with DNA or the proteins that provide structure to the chromosomes and help regulate the rate of transcription. Until recently, scientists generally considered all of these untranslated portions of the genome to be “junk.” The more we learn about the regulation of information usage, the more we have come to understand that even parts of the DNA not directly coding for proteins help determine an organism’s phenotype.

The relationship between the genotype and phenotype is far more complex than it might first appear. Mendel’s original experiments generally involved characteristics that were controlled by one or maybe two genes, making it easy to discover their effects on a particular aspect of the phenotype. However, as the biochemistry of the cell was worked out in the early decades of the 20th century, it was realized that genes coded for enzymes, which in turn were responsible for an organism’s characteristics. In the 1940s, the American scientists George Beadle and Edward Tatum put forward the “one gene-one enzyme” model, for which they were later awarded the Nobel Prize. Genes were no longer seen as having a direct effect on the outward appearance of the organism, but a direct role in the production of the cellular machinery that eventually was responsible for the structure and function of the organism (see Chapters 5 and 6).

**Direction of information flow**

In the process of transcription and translation, information flows in one direction - from DNA to RNA to proteins. The idea that information flow is only in one direction was termed the **central dogma** by Francis Crick (Box 3A). For several decades it was thought that this unidirectional flow was universal. We know now of several important exceptions.

The first exception is a flow of information from RNA to DNA. The information in viruses can be stored and transferred as either DNA or RNA. When some RNA viruses enter a cell, they will get translated back into DNA. In that form they can be incorporated in the chromosome of the host cell. If that cell is one that produces gametes, the virus can even be passed from parent to offspring. The most infamous of these viruses is the agent that causes AIDS, the human immunodeficiency virus (HIV). The incorporation of HIV into human chromosomes is why it is so difficult to cure an infected person and eradicate the disease. To eliminate the virus, a person’s own cells must be killed. After initial infection, the virus can lay dormant in the chromosome for several years,
multiplying as the cells divide, before it eventually emerges and makes the infected person sick.

The second exception involves changes in the chemical structure of the DNA molecule (other than base pair sequence changes). Cytosine (in Eukaryotes) and adenine (in Bacteria) will sometimes have a hydrogen atom replaced with a CH3 group. Because CH4 is methane, this process is called methylations. Methylation affects the rate at which genes get transcribed, thus affecting how information is used by the organism, which can change its phenotype. When the cell divides, the methylation changes get passed along to the daughter cells, preserving the information changes. If the methylation occurs in cells that produce gametes, those changes in the DNA chemistry can get passed along to the offspring, sometimes for several generations. Because methylation is controlled by enzymes, this is an instance of information going from proteins to DNA. A heritable change in the information content of a cell is occurring, even if this is not a change in the DNA sequence. It is important to note, however, that such changes are not permanent and likely to be reversed.

Such changes sound like an old idea. It was once believed that organisms could pass along changes that occurred in their phenotype to their offspring, a process termed the inheritance of acquired characteristics. However, this idea was discarded by the middle of the 20th century as the genetic basis of inheritance was worked out. The inheritance of acquired characteristics posited that changes in the traits of an organism through use or disuse (such as bulking up a muscle) would get passed on to its offspring. This is quite different from the idea that a virus can be incorporated into a chromosome and be inherited, or that DNA methylation patterns might be passed on. However, there are many ways in which the environment of an individual, especially as determined by its parents, can affect its phenotype, which is the topic of the next section.

Context dependency

How the information contained in a stretch of DNA gets used depends on its context (Table 3.1, principle 8). That context includes the other DNA sequences surrounding it on the chromosome, the sequence of the DNA in the matching gene on the paired chromosome, other genes, and most notably, the rest of the environment outside the cell. That environment can include other cells within the same organism, other organisms, and the individual’s physical surroundings. Thus, the phenotype of an organism is the result not just of that individual’s genes, but also the environment within which those genes reside and are expressed.

Previously we discussed how flower color might depend on the combined effects of the two paired genes, either as an average of the two (Figure 3.5) or with one gene being completely dominant to the other (Box 3B). Other patterns are also possible. Dominance may be incomplete - for example, the color of the flower might be mostly red (as opposed to a strictly additive genotype in which the color of the offspring is the average of its parents). Sometimes the phenotype of the offspring falls outside the range of the parents entirely. In other cases, both genes may be expressed, such as in the blood type of humans. Humans have blood types O, A, B and AB, a condition that is due to three alleles. The A and B alleles each codes for an enzyme that attaches a different sugar molecule to a protein on the surface of red blood cells, while the O allele produces an enzyme that lacks activity. Because of this, an individual that has either the AA or AO genotype has blood type A. Individuals with blood type O all have an OO genotype, and individuals with an AB genotype have blood type AB.
The products of different genes can interact with each other. The hair color of wild mice is a grayish color (described as agouti) due to bands on individual hairs (Figure 3.16). One of the genes that controls a mouse’s hair color comes in two forms, the dominant form of which is agouti in color. The recessive form lacks the banding pattern, so a mouse with both forms of that gene is black. Another gene also affects hair color, and individuals with both recessive forms have white hair. The effects of the gene causing the white color gene completely override the effects of the agouti gene. Such interacting genes may be on different chromosomes, or even in different parts of the cell. In Eukaryotes, for example, specialized structures called mitochondria and chloroplasts (see Chapter 5) also contain chromosomes and the enzymes coded for by the genes on those chromosomes can interact with the enzymes coded for by the genes in the nucleus.

Some genes code for enzymes that are responsible for regulating the transcription processes; these enzymes can bind to the DNA and prevent the transcription enzymes from doing so. The binding sites are adjacent to the parts of the DNA sequence that are transcribed, thus defining the limits of the information being read, similar to punctuation in a sentence. Changes in the sequence to those stretches of DNA can change the binding ability of both the transcription enzymes and the regulatory enzymes. Changes in DNA sequences both adjacent to a transcribed gene and at other genes affects the rate of transcription and the usage of the information — similar to replacing a period with a semicolon, or vice versa. Transcription rates may be determined by chemical signals that come from other cells in the organism (see Chapter 6).

The phenotype of an organism depends not just on all of its genes, but also the environment. One aspect of that environment is the parents of the organism. Previously we described how changes in methylation patterns on chromosomes can be passed to offspring. Much more common, however, are direct nutritional effects. The size of an egg or a seed will also affect the size of the organism at birth or germination and many types of animals provide food for their offspring after birth. Those behaviors are controlled, in part, by the genes of the parents, so you can think of such parent-offspring effects as interactions between the parents’ genes and the offspring’s genes.

How a gene is expressed can depend on where it is within an organism. All of the cells in your body have the same genes, and the differentiation of those cells into specialized function (e.g., skin, muscles, nerves) is due to differences in gene expression (see Chapters 5 and 6), which can be determined by the external environment. Siamese cats have light color hair on most of their body, but dark colored hair on their extremities because the temperature is cooler. Kittens are all white because the mother’s womb is warm, and their extremities darken with age. You can affect this darkening process by raising the cats at warmer or cooler temperatures.

The disease sickle cell anemia provides an example of two types of context dependency, the identity of the paired gene and the environment. Anemia is a condition of having a deficiency of red blood cells leading to breathlessness and weakness due to a lack of oxygen, which is carried in red blood cells by the protein hemoglobin. In humans, a variant of that molecule causes the red blood cell to form a characteristic bent or sickle shape (Figure 3.17). For simplicity, we will label the allele that gives rise to the
normal variant A, and the allele that gives rise to the sickle variant B. If a person has two copies of the sickle allele (BB), a severe loss of red blood cells in that individual can lead to an increased likelihood of dying in childhood. A person with just one copy (AB), however, will be anemic but will be able to live into adulthood.

![Image of sickle cells](image)

**Figure 3.17**
Some of the red blood cells of individuals with sickle cell disease have a distinctive flattened or sickle-shaped appearance. Those cells are rigid and unable to pass through narrow capillaries restricting the blood supply, and are rapidly broken down in the spleen leading to anemia. (Photo from OpenStax College, Source: Wikipedia)

Another context is equally important, the environment. Malaria is caused by a single-celled parasite that is transmitted from the blood of one person to another by mosquitoes. If a person with the AA genotype gets infected, they can become very ill. But a person with the AB genotype is resistant to infection by the malaria parasite. Malaria is not found in temperate or polar regions of the world, but is very common throughout the tropics. So, for someone living in the far north who never gets infected, the hemoglobin characteristic of “malaria resistance” is never expressed. However, if that same individual were to travel to the tropics and become infected, she would be resistant to the disease. When living in the far north, individuals with AA and AB genotypes have identical phenotypes with respect to avoiding becoming ill with malaria because the parasite is absent. But in the tropics, AB individuals are much less likely to become ill with malaria, so in that environment the information content of “malaria resistance” is expressed or used.

The environment can also influence the heritability of continuous traits. Our previous discussion of heritability assumed that differences among individuals with different genotypes do not depend on their environment. That is, we assume that if an individual is 10% taller than another when growing in one environment, it will still be 10% taller in a different environment. But what happens when this assumption does not hold? Suppose, for example, that when a certain plant species is grown under shady conditions, all individuals are short and about the same size, but when it is grown in a sunny spot, some of those individuals are much taller than the others due to genetic differences. In other words, the genetic differences are apparent in some environments, but not in others. These kinds of differences in genetic expression as a function of the environment are referred to as **genotype-environment interactions** (Figure 3.18).

The presence of variation resulting from genotype-environment interactions can have large effects on heritability. Heritability is not simply a result of the genetic differences among individuals: Those genetic differences must result in phenotypic differences. Some kinds of genetic differences among individuals never result in phenotypic differences; for example, some types of variation in regions of the DNA that are not translated into proteins. In other cases, whether genetic differences result in phenotypic differences depends on the environment. When variation in genotype-environment interactions is present, the amount of expressed genetic variation may differ among environments. In the example given above, the plants grown in the shade were all of similar height; in other words, phenotypic differences were minimized in that environment. If the heritability of height were measured only in the shade, we would conclude that it
was very low because the amount of phenotypic variation would be low. On the other hand, if heritability were measured only in the sunny environment, it would be larger. Evolution would be constrained in the shady environment because of a lack of heritable variation.

Variation within an individual’s genetic makeup leads to variation among individuals. Some of that variation is deterministic, e.g., individuals that differ in their genetic makeup. Some of that variation is context dependent, e.g., individuals that differ because they reside in different environments. Finally, some of that variation is due to contingent events, coming about through random events in metabolic processes or development. Individuals with different forms of a gene may be phenotypically identical in one environment and different in another, and those environmental triggers may have a random component. Thus, the usage of information in living systems depends on complex interactions that lead to emergent properties and a role for contingency.

**Evolution of the Information System**

The genetic system came about through a process of evolution (Table 3.1, principle 9). Several key events occurred during that process. The first event was the appearance of a molecule that could act as a carrier of information. As described in Chapter 1, it is likely that the first information carrier was RNA, rather than DNA. It is likely that the origins of the genetic code trace to this period as RNA molecules evolved relationships with amino acids.

The next key event was the replacement of RNA by DNA as the information carrier. This replacement likely occurred because DNA is a more stable molecule that can exist as very long double strands. The double-strand structure for the first time allowed for error corrections using the information in the complementary strand, making the information system more resistant to mutations. The legacy of RNA as the information
carrier can still be seen in our cells. RNA acts as an intermediary so that information in DNA is first transferred to RNA before being translated into proteins and other forms of RNA play various roles in the translation process (Figure 3.14).

At some point the genetic code was set in place. It is possible that the original code consisted of two-base codons. Such a code has only 16 possible codon combinations and lacks the redundancy of the current code (Figure 3.15). The evolution of a three-base code, therefore, both increased the number of possible amino acids being coded for and, again, added robustness to the information system.

Why is the genetic code the way it is? Are the particular associations of codons and amino acids inevitable because of the biochemistry of those molecules? Or, is the code merely a frozen accident – if life were to evolve all over again, would a different code form? This issue is one of current debate among scientists with no clear resolution at the moment. It is clear that the genetic code is not simply random, providing clues to its evolutionary origin. Amino acids with similar chemical properties tend to have similar codons, so that mutations that changed the amino acid composition of a protein would tend to have a small change in its biochemical properties and be less likely to completely disrupt its function. In addition, amino acids made through the same synthetic pathway tend to have the same first base in their codons, which is what you might predict if relationships between specific RNA sequences and amino acids came about through gradual evolution from an original code consisting of two-base codons to one consisting of three-base codons. One way to resolve this debate would be if we were to find life elsewhere in the universe and could examine its genetic code.

The genetic code is not quite universal. While the code shown in Figure 3.15 is found in nearly all organisms, exceptions exist. For example, in most species the codon AUG serves two purposes, coding for methionine and serving as a marker for the start position for translation. In most proteins the methionine is then clipped off the protein. However, in Bacteria and Archaea, UUG and GUG can also serve as start codons. Bacteria in the genus Mycoplasma translate the codon UGA as tryptophan, rather than as a stop signal. Several differences from the standard code are found in mitochondria and chloroplasts, especially a reduction in the code because several codons are no longer used.

In general, we would expect the code to evolve extremely slowly. Once the code is set in place, a change effects not just a single protein, but possibly every protein. Thus, any changes are likely to be highly deleterious. Many of the known variants are replacements of start or stop signals with amino acids, rather than replacements of amino acids with stop signals, or switches of one amino acid for another. One can imagine that the first sorts of changes are less likely to be deleterious than the others. For example, it would be very hard to create a functional protein if a change created a stop codon in the middle of the gene. That some of the most numerous changes are found in mitochondria and chloroplasts can be explained by the fact that they are the result of an ancient symbiosis, so that many of their proteins are now produced by other genes (see Chapter 5).

What is a Gene?

At the beginning of this chapter, we stated that the gene was the fundamental unit of information in living systems. But what is a gene? The concept of the gene has undergone a continual evolution since it was first proposed by Mendel. Initially, a gene was associated with a specific characteristic of an organism such as the color of a pea. With the discovery of linkage, genes on chromosomes were treated like beads on a string. The concept of a gene became associated with a discrete unit that could be localized on a chromosome.
The association of genes with the molecular properties of cells produced the “one gene-one protein” concept. A gene was still thought of as a discrete unit, but now the association between genes and organismal properties was moved back a step. Along with this move was the growing appreciation that organismal characteristics like size, shape and color were often determined by more than a single gene.

The concept of the gene became much more complex in the middle of the 20th century as the structure of DNA and the genetic code was worked out. The gene was no longer a discrete unit, but a long strand. Recombination could occur within a gene as well as among genes. Still, the gene was seen as a single stretch of DNA. That idea had to be modified beginning in the late 1970s with the discovery that a protein was not the straightforward product of a single stretch of DNA. For many genes, it was discovered that after transcription some of the RNA was snipped out of the sequence so that the eventual protein was coded for by several disconnected stretches of DNA. Then scientists discovered that RNAs were sometimes assembled from pieces coming from very different places along the chromosome, not even from a single stretch. In some cases, a single stretch of DNA would result in a RNA molecule that would be associated with many other RNAs in different combinations, so that a single stretch of DNA could be translated as parts of more than one protein. In other cases, it was discovered that a single stretch of DNA could be translated into more than one protein by the simple process of starting the translation offset by one base pair, or by translating both strands of the DNA molecule. Finally, a stretch of DNA could also include parts that were never transcribed, but were essential in regulating the transcription of the adjacent parts. Thus, we have gone from a concept of a gene as a discrete unit associated in a simple fashion with a characteristic of an organism, to a complex, construct that is multipartite with a multifaceted relationship between the DNA and the resulting proteins and on to the characteristics of the organism.

Today scientists no longer have a single concept of the gene. Instead, they use different concepts depending on the circumstances. While such conceptual flexibility can be useful, it also has its pitfalls. For the purposes of modeling the evolutionary process, the concept of the gene as a discrete unit is often convenient (see Box 4C). But such models fail to account for the complex ways that mutation and recombination can occur. More complex models can be built, but are often much less tractable; one result is that computer simulations take much longer to run. The key is deciding when the simpler model is sufficient to provide reasonable predictions.

Technology is now capable of sequencing your genome in just a few days, but that’s only a small part of how your genetic information is expressed. We need more than just the sequence of base pairs, we also want to associate that sequence with functional consequences, the proteins that they produce and the organismal characteristics that they affect. Unfortunately, it would be impossible to work out those associations separately for every stretch of DNA. Instead, sequences of DNA are matched against other, known sequences from other species. Here, the gene concept that is used becomes critical. Do we look for matches against a stretch of DNA that includes the control regions, just the transcribed regions, or just the translated regions? How do we account for cases where RNAs are assembled from pieces transcribed from widely separated stretches of DNA? How do we ascribe a function to a stretch of DNA whose RNA is combined to create more than one protein? All of these issues are combined with the fact that the sequences of base pairs in a stretch of DNA varies among individuals within a species as well as between species. Sometimes those sequence differences result in no change in the protein product, but sometimes they do and those changes in the protein may or may not alter the way it functions (see Box 6D). How different can they be and
still function in a similar manner? Would we be able to recognize these two genes as being related to each other?

Recognizing the complexity of the concept of the gene is important for how biologists pursue their science. It is also a lesson in how all of the concepts presented in this book need to be considered as possibly having multiple meanings that can change over time and depend on the context in which they are used.