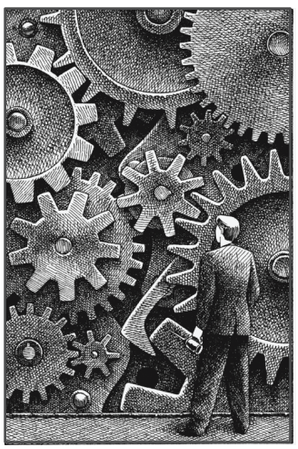
PART TWO

MACHINERY OF EVOLUTION



The great spectacle of biological evolution is one of the most inspiring things a young biologist can be exposed to: the colors of birds, flowers, and insect wings; the vast scale, from virus particles to blue whales; the intricate shapes of shells, bones, and fossils. The diversity of life has a majesty like that of a great symphony, detailed and overpowering. This diversity is the main topic of most introductory biology courses. However, the beauty of life evolving on this small blue planet tends to inspire wonder more than rational thought.

Yet at its core, evolution runs on machinery that is utterly repetitive, implacable, hardly beautiful at all. The workings of evolution resemble the ledgers of a Victorian firm owned by Ebenezer Scrooge more than they do the sheet music of Wolfgang Mozart. Evolution works by scrutinizing populations with a jaundiced eye. The feeble and the barren are discarded with an uncaring brutality. Like a merciless robot, evolution produces life but does not care for it. Individuals die, species go extinct; yet the evolutionary process always goes on.

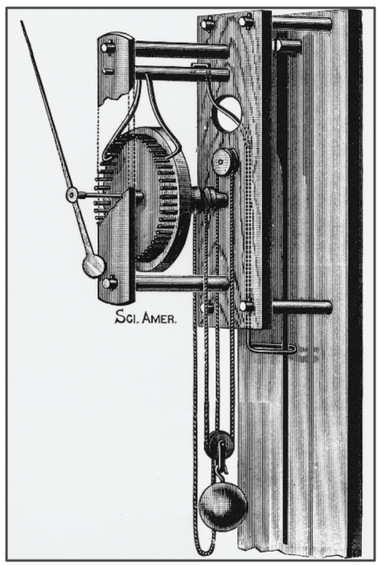
The biggest problem with understanding how evolution works is that it is invisible. You can see organisms right in front of you. With a microscope you can see cells, and you can pick up a white, stringy hunk of DNA that you can see with the naked eye. The processes of ecology are visible in the birth, death, feeding, and decomposition of plants and animals. But evolution is the ghost at the banquet of life. Its machinery is hidden.

There is nothing unusual about scientists looking for the hidden machinery underlying a process. For a century, first atomic physicists, then nuclear physicists, and now particle physicists have been looking for smaller and smaller particles to explain matter. They seek the hidden machinery of matter and energy. In the same way, biologists have sought the hidden machinery of life's evolution.

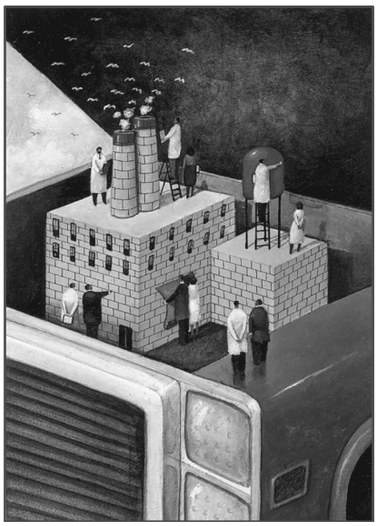
At the core of this machinery is a genetic engine, a whirling thing that contains the information used to produce each generation, the record keeper for the entire process of evolution, containing the fuel on which evolution works. Genetic transmission works with greater precision and reliability than any other strictly biological process. Like the atoms of physics and chemistry, genes are building blocks; without them, life could not exist. To have a solid understanding of evolution, you have to know basic genetics. This we supply in Chapter 3.

But genetics does not determine the direction in which the evolutionary leviathan advances. The bit of machinery that guides evolution, its steering wheel and its accelerator, is natural selection. Natural selection is the discriminator within the evolutionary machine. It is natural selection that chooses among the individual organisms in a population, determining who dies, who has a chance to reproduce, how much they reproduce, and finally the duration of their lives. This selection process ultimately adapts populations to the environments that they inhabit.

Evolution was first "seen" around the year 1800 by people like Jean-Baptiste Lamarck and Erasmus Darwin, who were introduced in Chapter 1. But evolution was not properly understood and analyzed until Charles Darwin discovered natural selection as part of the machinery of evolution. All students of biology need an understanding of natural selection. We supply you with that understanding in Chapter 4.



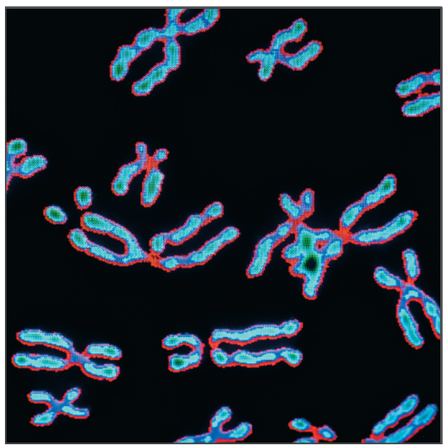
The fundament of the machinery of evolution lies at the molecular level. Life is an interplay of molecules—a process built up from a swamp of physics and chemistry, but not just physics and chemistry. Molecular evolution supplies the foundations for all processes of evolution, because nothing happens in evolution unless the molecular composition of organisms changes. But there is a great inscrutability to molecular evolution; because it can proceed with or without natural selection, it is very difficult to determine whether selection has acted at the molecular level. The ultimate machinery of evolution is DNA, the molecule of heredity. At this deep level of the evolutionary machinery, we can see how evolution plays out—much like we can learn how a computer works by examining the lines of code that make up its programs. Evolution at the molecular level is the topic of Chapter 5.



The ultimate regulators of the evolutionary machinery are like two conveyor belts. One belt brings species into the machinery of evolution; the other belt takes species out of the machinery. Speciation adds species. Extinction eliminates species. For a long time, evolutionists thought of speciation as mysterious and extinction as straightforward. Now we know that speciation is not as inexplicable as we thought, while extinction appears to be considerably more complex than we imagined. The creation and destruction of species are the most dramatic actions of the evolutionary machinery. In Chapter 6 we introduce the machinery of speciation and extinction.

Genetics, natural selection, molecular evolution, speciation, and extinction all play a role in the evolutionary machinery—sometimes in isolation, but often not. In considering speciation, for example, issues of genetics, selection, and molecular evolution all play an important role. We will separate these processes from each other while introducing them; but in nature they normally operate together, like a giant factory with many machines grinding, and spinning, and stamping.

The genetic engine supplies information for development and variation for selection.



3

The Genetic Engine

The evolutionary machinery has genetics as its core. Genetics in turn has two facets. One is the role of genetics as the keeper of the library of information for making each organism, its complete *genome.* Now that the human genome has been completely sequenced, anyone can marvel at the fact that our biological coherence depends on some billions of nucleotide pairs of DNA. Without genetics, life would have to be renewed by some external force in every generation. With the information in the genome, life has momentum, continuing on from parent to offspring. Thanks to this genetic transfer of genomic information from one generation to the next, each offspring is the cumulative product of millions of years of evolution.

Almost as important as the genetic transmission of the genome is the role of genetics in providing *variation*—heterogeneity in the biological characters of the individuals who make up the population. Some of this variation has nothing to do with heredity, and is called *environmental,* though this term includes all nongenetic sources of variation. Environmental variation can be very important. Human learning, for example, is environmental variation. But there is also genetic variation. Such genetic variation is essential to the process of evolution, because it provides the raw material for natural selection. We have already seen how Darwin emphasized the importance of heredity.

Thus genetics faithfully transmits the information built up by evolution over long periods, and it supplies the fuel for evolution—genetic variation. These two roles of genetics are key to the machinery of evolution, and so they are integral to this part of the book. To understand the role of genetics in evolution is a fair start to understanding the evolutionary process as a whole. ❖

How Genetics Works

3.1 Genetics is central to modern biology

The field of genetics was probably the greatest achievement of twentieth-century biology. It is doubtful that modern biology could have been created without genetics. Too many basic questions about life had no good answers before the creation of genetics.

As we have seen, the person who started the genetic revolution was the monk **Gregor Mendel** (1822-1884), one of the most obscure of all the great scientists (Figure 3.1A). It is not clear whether Mendel's genetic theory of inheritance preceded his plant-breeding experiments, or whether he formed his ideas from his data. It is certain that by the time he presented his first papers on genetics, he had both a remarkably clear model as well as garden pea data that beautifully illustrated his theory. These papers were presented locally, in Brno, and circulated to some of the leading botanists of Mendel's day. Darwin had an unread copy of one of Mendel's papers. Yet no leading scientist appreciated the implications of Mendel's work during his lifetime.

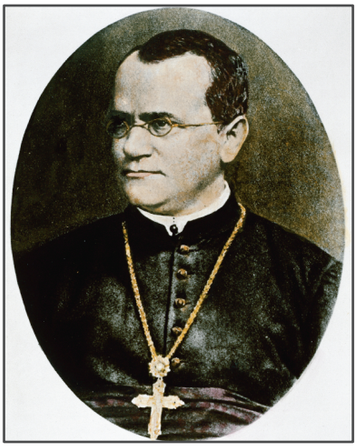


FIGURE 3.1A Gregor Mendel

All that changed in 1900, when several botanists independently rediscovered Mendel's work. Very quickly these scientists and their colleagues abandoned the old models of heredity and formed a new discipline within biology—genetics. Basic discoveries of genetics have since flowed rapidly. A partial list would include the discoveries listed in Table 3.1A.

Even this list does very poor service to more than a century of brilliant work, in which many scientists played major roles. The point is that we can view the twentieth century as a century in which the development of biology has been dominated by the unfolding of the research of geneticists and their allies.

A warning needs to be provided here: It is easy to think about genetics strictly as the triumphs of molecular genetics, starting with the double-helix DNA model of Watson and Crick. Scientific progress had been rampant, however, for the 53 years preceding publication of the Watson-Crick DNA model in the journal *Nature*. In particular, the foundations of

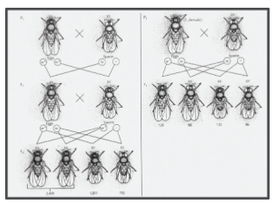


FIGURE 3.1B Thomas Hunt Morgan Morgan was the founder of fruit fly genetics. This is one of his drawings.

genetics have important elements of evolutionary reasoning, beginning in 1908 with the derivation of the Hardy-Weinberg equilibrium, a key concept in population genetics. Molecular genetics would be difficult to sort out without the use of important principles from population genetics.

In this chapter, we will introduce the essential model of Mendelian genetics and then present some of the most basic ideas of population genetics and quantitative genetics.

TABLE 3.1A Abbreviated List of Twentieth-Century Genetic Discoveries

- Mendel's genetic theory of inheritance is rediscovered (1900).
- Genetic linkage between characters is discovered (1911).
- Fisher develops quantitative genetics for continuous characters (1918).
- Muller demonstrates the physical basis of mutation (1928).
- Morgan establishes the chromosomal basis of genetics (1931; *Figure 3.1B*).
- Avery, MacLeod, and McCarty determine that DNA is the hereditary molecule (1944).
- Watson and Crick propose that the DNA double helix codes for genes (1953; *Figure 3.1C*).
- DNA replication is worked out (1958).
- The genetic code is determined (1966).
- Recombinant DNA allows rapid cloning of DNA (1972).
- First rapid gene sequencing is performed (1975).
- Organisms are genetically engineered by the insertion of transposable DNA (1982).
- Polymerase chain reaction (PCR) is used to clone small amounts of DNA (1985).
- The entire human genome is sequenced (2000).

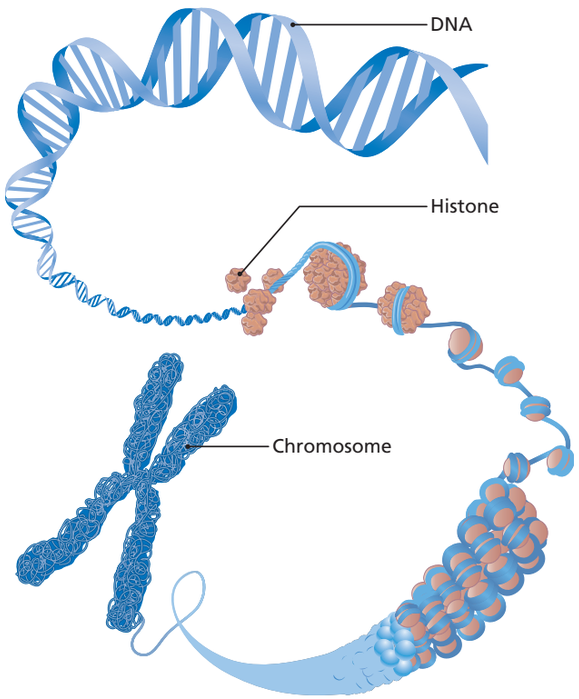
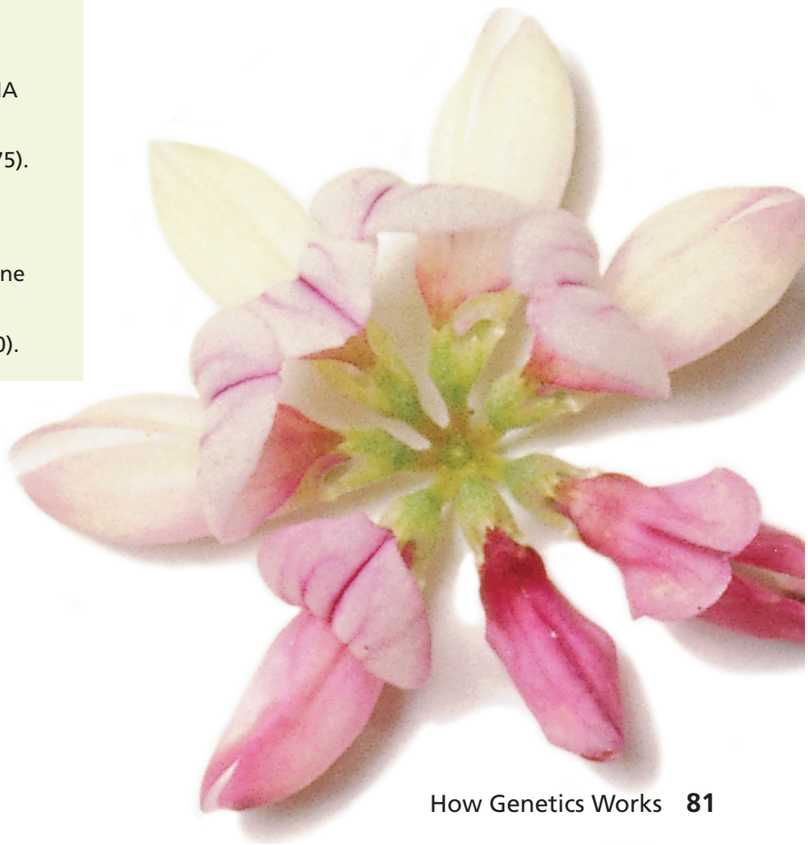


FIGURE 3.1C The Structural Features of DNA



3.2 Reproduction may transmit one or two copies of the hereditary information to the next generation

However the specifics of inheritance work, species vary greatly in the amount of hereditary information that is copied from parent to offspring. There are considerable differences in the size of the genome that each species carries, from the small genomes of viruses to the enormous genomes of some cereal plants and some salamanders. This variation in genome size is examined further in Chapter 5. The rest of the variation in DNA content is in the number of copies of the basic genome that each cell has.

A bit of terminology is important here. An organism or cell that has only one copy of the basic genome is called **haploid** (Figure 3.2A). Most organisms that we are used to seeing are not haploid. However, almost all gametes are haploid, so you may already have some familiarity with haploid cells. Like most animals, humans are **diploid**, possessing two copies of the basic genome (Figure 3.2B). (We are neglecting hereditary differences between male and female, which we will consider later in this chapter.)

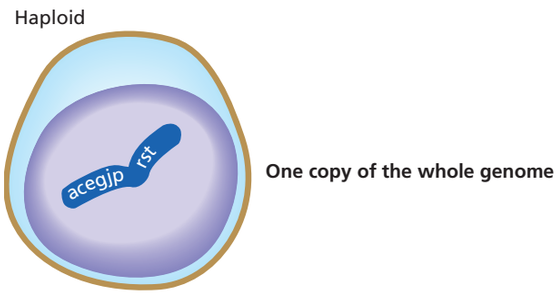


FIGURE 3.2A Haploid Genome

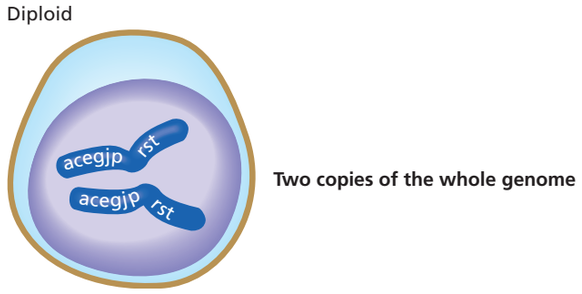


FIGURE 3.2B Diploid Genome

Organisms that have more than two copies of the genome are called **polyploid** (Figure 3.2C). Most of these have higher multiples of two copies, with four and eight copies being common. Some species that consist only of females are triploid, with three haploid genomes. This condition is known in some fish and some lizards. We discuss it further in Chapters 6 and 18. Bacteria are sometimes considered haploid, but it is also common for them to have a variable number of additional copies of their basic genome. This variable ploidy in bacteria is called **meroploidy** .

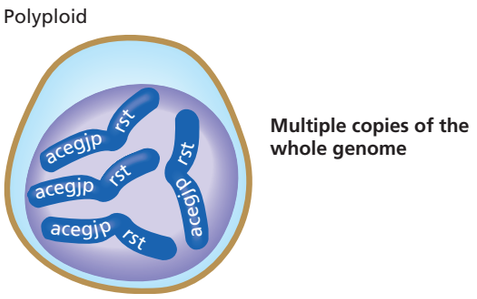


FIGURE 3.2C Polyploid Genome

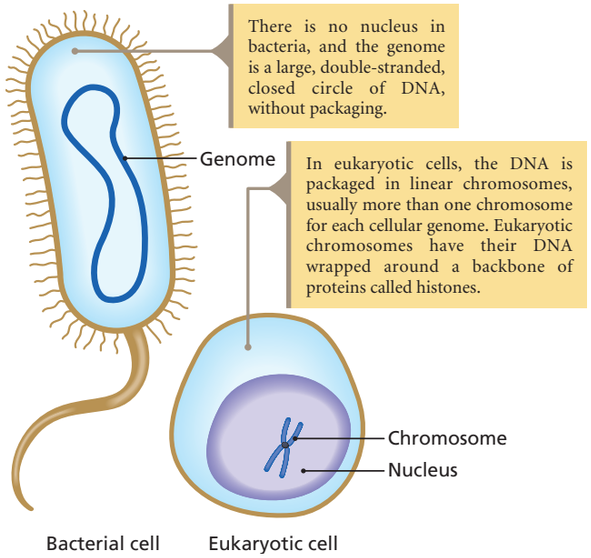


FIGURE 3.2D Genome organization is very different in bacterial cells as compared to eukaryotic cells.

Even though we consider ourselves diploid organisms, let's bear in mind that many organisms, humans included, regularly alternate between haploid and diploid genomes. In plants, the haploid stage of the life cycle can produce a substantial plant. Both mosses and ferns have haploid plants. Angiosperms, on the other hand, have vestigial haploid forms contained within flowers. Usually, haploidy is associated with the production of gametes, anzd fertilization then creates a diploid organism as a result of the union of two haploid cells (the gametes).

Some organisms have specific tissues with odd *ploidy* (number of genome copies). Plant **endosperm**, a tissue involved in fertilization, is triploid. The larvae of fruit flies have salivary gland cells that contain hundreds of copies of the genome—a condition called **polyteny**. Ploidy is thus a complicated business. The tissues and life-cycle stages of animals and plants can vary significantly in number of genome copies. ❖

3.3 Sexual reproduction recombines chromosomes containing many discrete loci

In the vast majority of organisms, Mendelian genetics is the system of inheritance. Numerous bacteria do not have Mendelian genetics, as well as some asexual organisms that retain only vestiges of it. But for most organisms of interest, Mendelian genetics defines inheritance. Alternation of ploidy is a basic feature of the Mendelian system. Another basic feature is the **chromosome**. Think of chromosomes as strings of genetic **loci** that can be defined as locations for particular genes. Each locus contains **alleles**, which are the specific versions of genes, defined by DNA sequences.

During the processes of fertilization, development, and meiosis, each chromosome in the cell will be present in one, two, and four copies. This is indicated in the genetic cycle diagram of Figure 3.3A by n, $2n$, and $4n$, where n is the number of chromosomes in the haploid genome. Combinations of genes are scrambled, or randomized, at three points during this cycle. In this way, Mendelian genetics generates variation, as we will now describe.

First, start with the pool of **gametes** that precedes fertilization in Figure 3.3A. It is a basic tenet of Mendelian genetics, and a demonstrable empirical fact, that during fertilization gametes tend to combine at random with respect to their genetic makeup. (There are some exceptions, but they are minor.) Thus the **zygote** that is formed during fertilization is a random combination of two gametes among many. This process is analogous to a card game in which the "hand" has only two cards, but the deck contains thousands or millions of different cards, the cards being gametes and the hand being the zygote. This random combination of gametes is the first point at which Mendel's genetic combinatorics can generate large amounts of variation, following orderly rules of probability.

The second point in the cycle at which randomizing factors play a key role is the recombination of chromosomes during **meiosis** (Figure 3.3B). Meiosis works with recombination as follows. At the start of meiosis, each of the two chromosomes of the diploid cell produces an additional copy of itself. There are thus four copies of the chromosome present in the cell. Chromosomes have a tendency to break, but once they have broken into pieces, they usually find matching broken ends

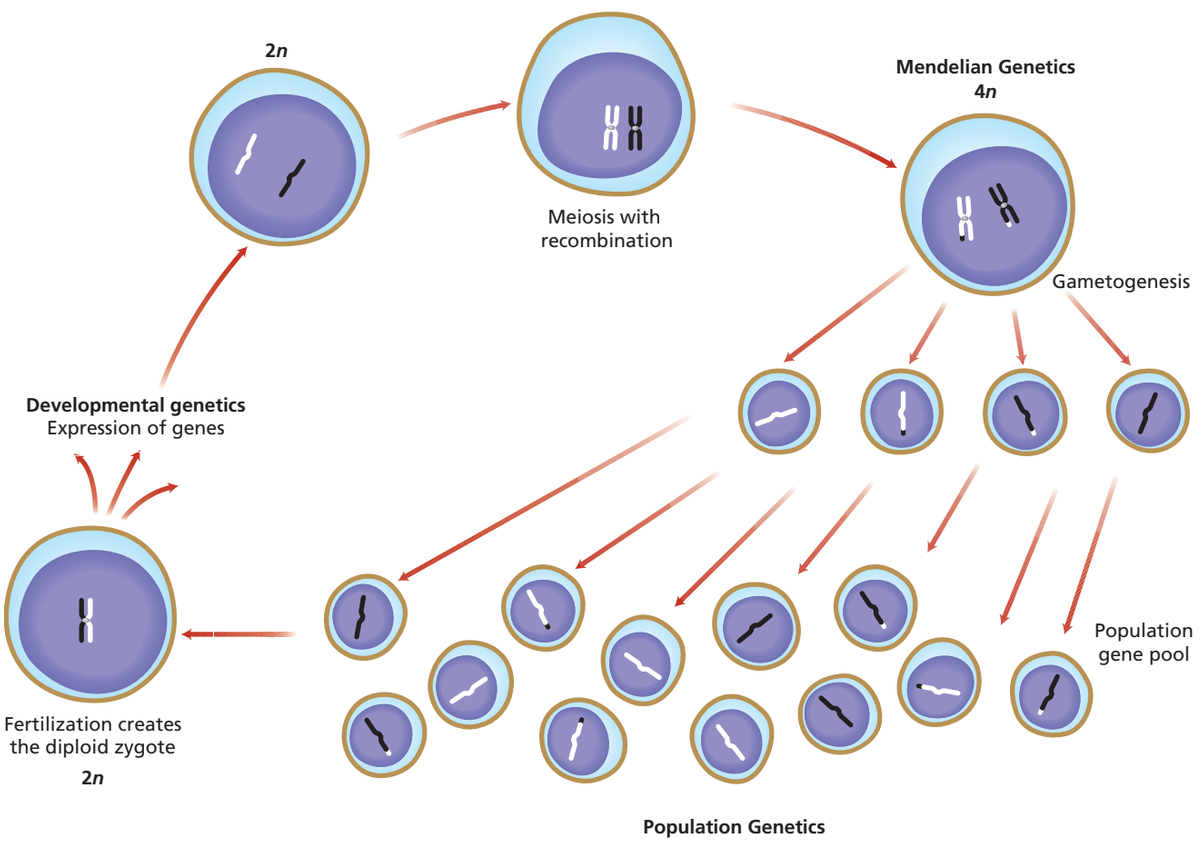


FIGURE 3.3A The Genetic Cycle Different fields emphasize different phases of this cycle: Mendelian genetics, developmental genetics, and population genetics. The black and white structures are the chromosomes; pieces of black or white joined to chromosomes of the opposite color indicate recombination.

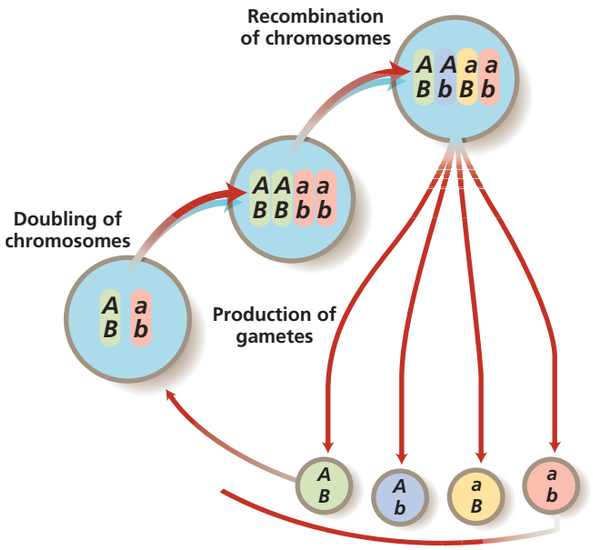


FIGURE 3.3B This close-up of Mendelian genetics shows the recombination and segregation of specific genetic alleles from two loci, both located on a single chromosome. Here the chromosomes are represented by colored oblong shapes with letters on them.

(which are effectively "sticky") and reconstitute themselves. If these broken ends are all from the same chromosome, no genetic recombination occurs. But different chromosomes of the same type can break and rejoin, forming new composite chromosomes that never existed before. In Figure 3.3B we show this process for two genes, each with two alleles. At the *A* locus we show an *A* allele and an *a* allele. Similarly, at the *B* locus the two alleles are *B* and *b*. A chromosome with the *A* and *B* alleles side by side can recombine so that the *A* and *b* alleles are put together on the same chromosome.

Third, with **independent assortment** of chromosomes, different gametes can be produced from the same sets of chromosomes. Recombined chromosomes join in varied combinations to create gametes with combinations of chromosomes that never existed before. (This combination of multiple recombined chromosomes is not shown in Figure 3.3A or Figure 3.3B.) And this process is superimposed on any process of chromosomal recombination that may have occurred. Thus at three points, Mendelian genetics scrambles the alleles of the strings of genetic loci that make up entire genomes. ❖

GENES IN POPULATIONS

3.4 Genes specify phenotypes, then the phenotypes are selected, which changes gene frequencies

For evolution, Mendelian processes define the basic foundations for heredity and variation. From these foundations, the process of evolution creates adaptations, species, new ways of life. Our interest now lies in how the evolutionary machinery uses genes.

First, we need to know the frequencies of alleles and gene combinations. These frequencies reveal the genetic substratum to the evolutionary process. Second, we need to know how genes determine the characteristics of organisms, a process called **gene expression**. Knowing gene expression gives us information about the consequences of having particular genes, in particular combinations. Third, we specifically need information about the consequences of particular genes for survival and fertility, because these consequences generate natural selection, the directional motor of the evolutionary machinery.

Some useful terminology helps to define the operation of evolutionary machinery. We use the term **genotype** to define that part of the genetic makeup of the organism in which we are interested—for instance, a haploid genotype or a diploid genotype. Sometimes the genotype of interest may be a large part of the genome. Sometimes it may be just one Mendelian locus.

From the genotype, in each particular environment, an organism develops. The features of interest for a particular organism are called the **phenotype**. Sometimes we say phenotypes when we mean several different attributes of the organism, such as height or color. But there is really only one phenotype in total, just as there is only one complete genome.

Sometimes the evolution of genes is divorced from the phenotype. It is possible that variation does not matter for natural selection. For example, exactly how your fingerprints curve over your fingertips may not make any difference to your survival or your reproductive success. But this absence of effect is a very important point for the evolution of the genes that affect fingerprints, in that their evolution will then be uncoupled from natural selection. We will consider this situation later in this chapter, and elsewhere.

In studying genes in populations, we are attempting to learn the genotypic components of the evolutionary machinery. Genotypes can determine phenotypes that have consequences for survival or reproduction. Such consequences generate natural selection. Selection then changes allele frequencies, an essential feature of evolution. Allele frequencies define selection, and selection in turn changes these allele frequencies. The process feeds back on itself. (The process is not, however, circular—because selection does not determine earlier allele frequencies.) The phenotypes (P) of one generation undergo selection and other processes, which determine the genotypes (G) of the next generation, which determine the phenotypes of the following generation, and so on.

$$G_1 \rightarrow P_1 \rightarrow G_2 \rightarrow P_2 \rightarrow G_3 \rightarrow P_3 \rightarrow$$

The Population Concept

To monitor the evolutionary process, allele and genotype frequencies are some of the most basic variables that we need to follow. These variables are the particular interest of *population genetics*. Population genetics is based on the Mendelian model for inheritance coupled with the concept of population. This population concept therefore requires some closer attention.

The key concept that allows us to move from genetics to evolution is **population**. Loosely speaking, a population is just members of a sexual species living within easy traveling distance of each other. "Easy traveling distance" strictly relates to the biology of a species. For an albatross, that may be thousands of miles over open ocean, almost regardless of the weather. For a small worm, a lifetime's easy traveling distance may be a few dozen yards. The issue is whether or not organisms can mate with each other. If mating is a reasonable possibility, then two organisms are members of the same population. If mating is unlikely to occur, then they are not. The mating pattern defines the scope of the population in *space*, as shown in Figure 3.4A.

When organisms belong to the same sexually reproducing population, then the fates of their genes are intertwined. They may have ancestors in common. They may later have descendants in common. This is because the Mendelian cycle of fertilization, recombination, and gamete production ensures that genes are shuffled among organisms (Module 3.3). Indeed, whereas organisms may seem to constitute the population, in another sense the animals and plants that we see with our eyes are only the fronts, or masks, for the genes that define these organisms and determine their fates. The continuous transmission of genes from one generation to another defines the scope of the population in *time*, as shown in Figure 3.4B.

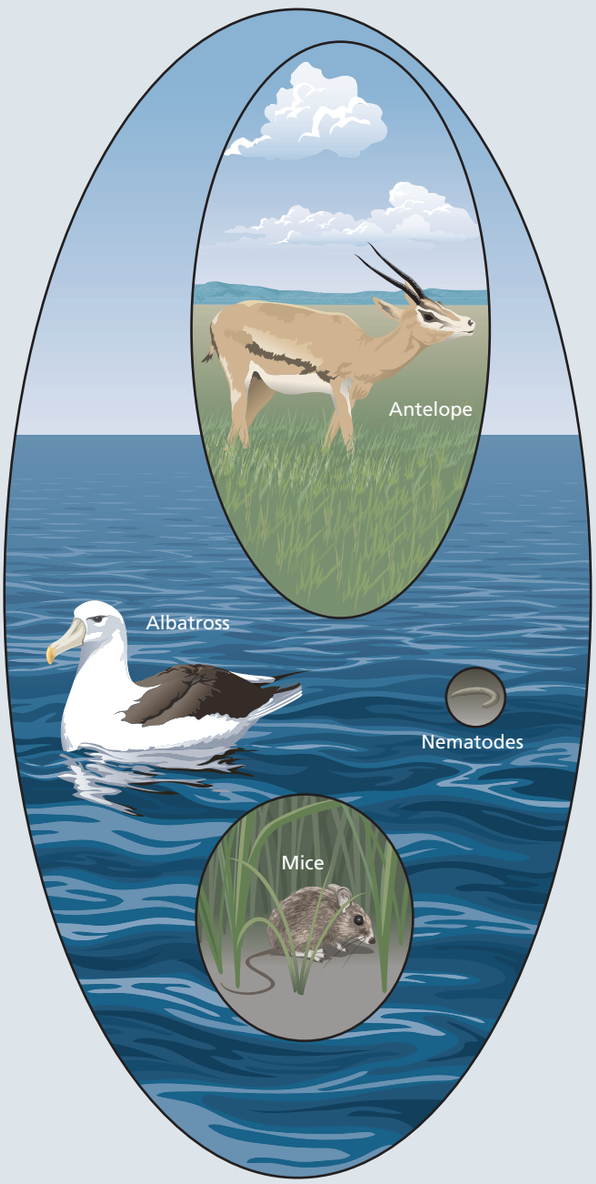


FIGURE 3.4A Populations in Space The animal species is indicated in the ellipse that qualitatively illustrates its approximate range in space.

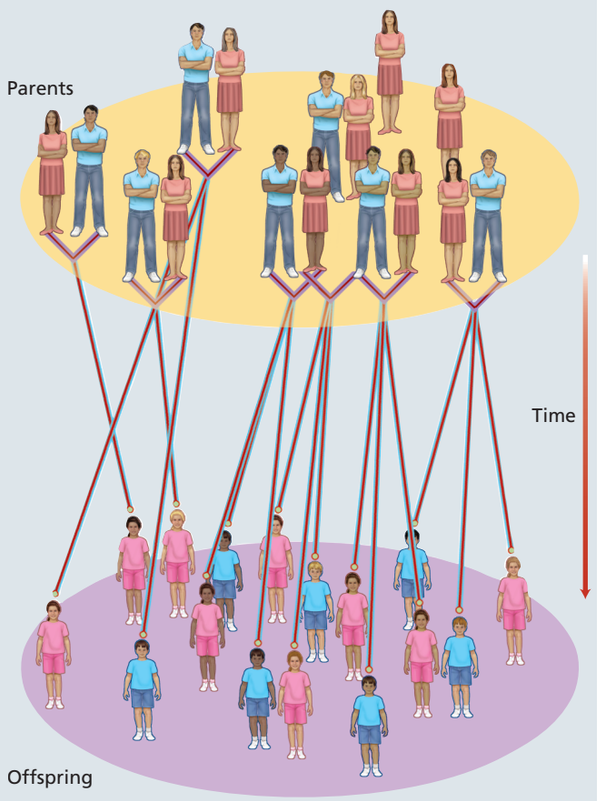


FIGURE 3.4B Populations through Time, from One Generation to the Next. Parents are joined by *v*'s. Lines connect offspring to parents. Note that a man and a woman have other partners with whom they have children. Some other adults don't have any children.

3.5 The evolutionary state of a population is defined by its genotype frequencies

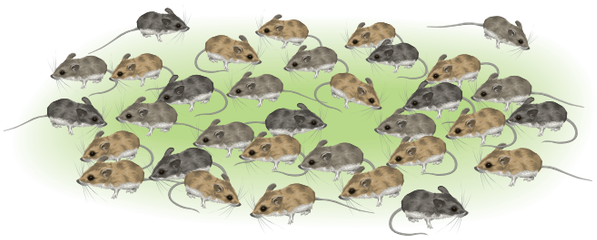
The genes at a locus come in different flavors or alternative forms, which we have referred to as *alleles*. These alleles have different DNA sequences. The different sequences sometimes specify the production of different amino acid sequences in proteins. But sometimes they do not. Instead, they may lead to increases or decreases in the total amount of a particular protein. Or the variant DNA sequence may change when the protein is made, perhaps in response to a temperature change. Another possibility is that the alleles code for RNA that does not make protein. For example, the RNA in ribosomes is used to synthesize proteins, but does not itself code for protein. So allele differences may have quite heterogeneous effects on the molecular and cell biology of organisms.

Mendel and the geneticists who came before the double-helix model of DNA knew none of these molecular details. They were able to detect discrete genetic variants in their breeding experiments. They then treated those variants as alleles, with no knowledge of the molecular biology of the gene.

Genetics thus began without mechanistic detail. This viewpoint probably helped the birth of population genetics. So long as genes were not known biochemically, one of the few things that a scientist could do was to count them in a population. For example, given a population of 124 mice, we might identify three carriers of a fairly rare allele for albinism using the gene sequences present at the locus. Because mice are diploid, we would have to sequence a total of 248 alleles, and 3 albinism alleles out of 248 would give us an albinism allele frequency of 1.2 percent. From data like this, population geneticists reconstruct and predict the evolution of populations. This activity will be our main concern over the next few chapters.

But first we need to be quite sure about how allele frequencies are estimated. In genetics, biologists often do not assay the alleles of an organism directly. Instead, they group phenotypes into distinct genotypic classes. That is, they determine what phenotype is associated with each genotype, and then calculate the allele frequency as shown in Figure 3.5A, which uses the example of a small population of 34 mice. This frequency calculation process is somewhat like accounting. It is the foundation of population genetics, which in turn is the key to understanding the machinery of evolution, since evolution depends on changes in allele frequencies, as we will describe in detail.

These calculations can also be applied to real genetic data as in the following box.




Mice homozygous for allele *a*: 9


Mice heterozygous for allele *a* and *A*: 16


Mice homozygous for allele *A*: 9

The *aa* homozygotes carry 9 × 2 = 18 *a* alleles.
The *Aa* heterozygotes carry 16 × 1 = 16 *a* alleles.
The total number of *a* alleles = 18 + 16 = 34 alleles.

The *AA* homozygotes carry 9 × 2 = 18 *A* alleles.
The *Aa* heterozygotes carry 16 × 1 = 16 *A* alleles.
The total number of *A* alleles = 18 + 16 = 34 alleles.

There are a total of 68 allele copies at this locus.
The frequency of *A* = 0.5. The frequency of *a* = 0.5.

FIGURE 3.5A A Small Mouse Population with Two Alleles, *a* and *A*

Population Genetics of a Human Blood Type

To gather this information on human **blood groups**, the blood from 730 individuals was tested and classified as either type 'M', 'MN', or 'N'. Each of these types corresponds to a single genotype, permitting us to directly estimate allele frequencies as shown in the table. Note that the M blood group corresponds to the homozygous L^ML^M genotype, while the N blood group corresponds to the L^NL^N blood group. Since the allele frequency (L^M) + allele frequency(L^N) = 1, we can use the following results to estimate the frequency of L^N as 1 − 0.18 = 0.82.

Allele Frequencies of a Human Blood Group

Blood Group	Genotype	Number	Frequency
M	L^ML^M	22	0.030
MN	L^ML^N	216	0.296
N	L^NL^N	492	0.674
Total		730	1.00

Frequency of allele L^M = (2 x 22 + 216)/1460 = 0.18

Data from a study of Australian aborigines; published in Ayala and Kiger (1980, p. 603).

3.6 With no selection, allele frequencies do not change in randomly mating large populations

Early in population genetics, the question came up as to whether gene frequencies would have some inherent tendency to change, perhaps as a result of the blind operation of the Mendelian machinery during sexual reproduction. By 1908, G.H. Hardy and W. Weinberg had independently worked out that there was no such tendency for allele frequencies to change, providing there were no perturbing outside forces, like selection or migration. This idea is known as the **Hardy-Weinberg equilibrium,** and we can demonstrate it mathematically.

Understanding this demonstration first requires an understanding of two basic rules of probability:

1. If two *independent* events together cause a third event to occur, then the probability of the third event is the *product* of the probabilities of these two events.

   For example, if a car crash requires you to look away from oncoming traffic *and* it requires you to lose control of your vehicle temporarily, then the probability of the crash is the *product* of these two unlikely (we hope) events. This is the **multiplication rule**.
2. When *either* of two events suffices to cause an outcome, then the probability of this outcome is the *sum* of the probabilities of these individual events.

   Thus if you can crash on an icy mountain highway by *either* ignoring a curve *or* by losing traction on a patch of ice, then the probability of such a crash is the *sum* of these two events. This is the **addition rule**.

To demonstrate the Hardy-Weinberg equilibrium, we need to calculate the consequences of Mendelian genetics for allele frequencies. Thus we need to go through a complete cycle of reproduction, from the parents of the first generation through their production of gametes, fertilization, and the creation of the next generation of adults. This calculation is shown in Figure 3.6A.

It is crucial to keep track of the different combinations of alleles in each mating, their probabilities of occurrence, and the genotypes of the progeny that they produce. If we follow all these variables carefully, then we can see that the frequencies of the two alleles (*A* and *a,* given by *p* and *q* respectively) do not change from one generation to the next. This result also holds true if there are more than two alleles at a locus, but the calculations get much more complicated. In all these calculations, it is an absolute requirement that the population size be large enough so that we can calculate probabilities exactly. When this is not true, additional evolutionary processes arise, as described later in this chapter. Likewise, there must be no biases coming from selection or differences in reproductive success. These are discussed in Chapter 4. From these stipulations, you will realize that the Hardy-Weinberg equilibrium is the simplest case considered by population genetics.

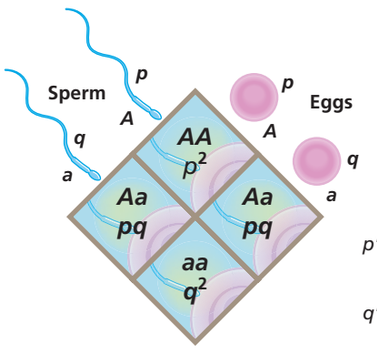
The attractive thing about the Hardy-Weinberg equilibrium is that it allows biologists to calculate the frequencies of the genotypes at a genetic locus from the frequencies of the alleles that make up those genotypes. If you keep track of order, there are four diploid genotypes at a locus: *AA*, *Aa*, *aA*, and *aa*. At Hardy-Weinberg equilibrium, the frequencies of these genotypes are given by the product of the frequencies of their alleles: p^2, pq, qp, and q^2, respectively. At one locus, we do not normally keep track of order, so the frequency of the *Aa* genotype is given as $2pq$, the sum of pq and qp. This enables us to understand evolution in terms of allele frequencies, instead of genotype frequencies.

Start with three possible genotypes: *AA*, *Aa*, and *aa*. Say these genotypes have frequency: *P*, H, and *Q*.

The frequency of *A* is $P + H/2 = p$
The frequency of *a* is $Q + H/2 = q$

Therefore, the frequency of *A* gametes is *p* and the frequency of *a* gametes is *q*.

With random mating, gametes combine by the rules of independent assortment like cards and coins.



$P' = p^2$

$H' = pq + pq = 2pq$

$Q' = q^2$

$p' = P' + H'/2$
$= p^2 + pq = p(p + q)$
$= p$
$q' = 1 - p' = 1 - p = q$

We represent frequencies of alleles and genotypes in the next generation by a prime , so p' indicates the frequency of allele *A* in the next generation.

FIGURE 3.6A Hardy-Weinberg Law Gene frequencies do not change because of genetic segregation alone—nor do genotype frequencies at each locus, once Hardy-Weinberg equilibrium has been achieved.

3.7 When the alleles of different loci are combined randomly, they are in linkage equilibrium

It would be nice if the Hardy-Weinberg equilibrium applied to genotypes that are combinations of loci. But it doesn't. Say we are interested in two loci: *A* and *B*. Suppose there are only two alleles at each of these loci: *A*-*a* and *B*-*b*. Let's also suppose that these alleles have frequencies *p*-*q* and *r*-*s*, respectively. Then we might expect that the frequency of the *AABB* genotype would be p^2r^2, by analogy to the frequency of *AA* being p^2. However, there is no principle like the Hardy-Weinberg law that always applies to calculating the frequency of genotypes across more than one locus.

Genotype frequencies can differ substantially from expectations based on random combination of alleles over loci. However, as we will show later in the chapter, there is a tendency for the frequencies of gametes to evolve so that they *do* come to follow such simple expectations. But this tendency is not as quickly or as reliably expressed as the Hardy-Weinberg equilibrium is.

However, we can calculate what genotype frequencies would be like *if* alleles combined at random when two or more loci are involved. This hypothetical situation is called **linkage equilibrium** in evolutionary biology. Deviation from this ideal of random combination is called **linkage disequilibrium**.

In the display in Figure 3.7A, we calculate a simple example of genotype frequencies when there is linkage equilibrium. Note that the four combinations of the same gamete type (*AB/AB, ab/ab, Ab/Ab, aB/aB*) have half the frequency of the six combinations of different gametes (*AB/ab, Ab/ab, aB/ab, AB/Ab, AB/aB, Ab/aB*). This is just the working out of the laws of probability, not some peculiar biology. The following box explains this in greater detail.

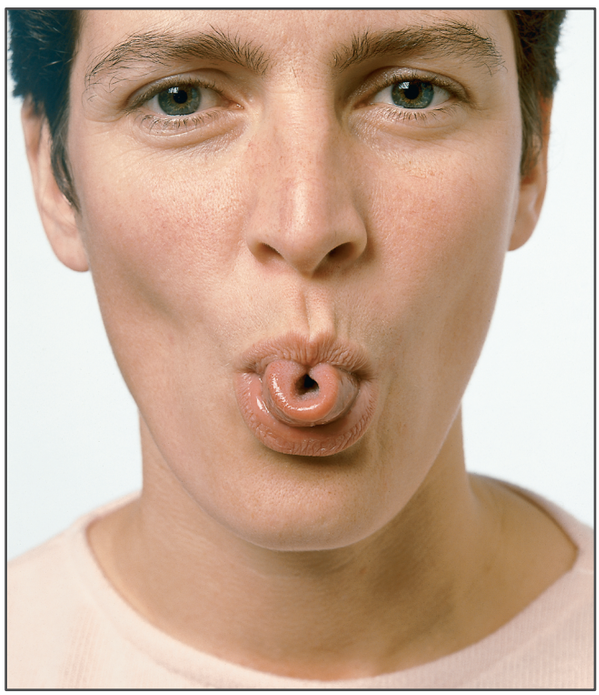


FIGURE 3.7B Can You Do This?

There are many loci in a genome, and many possible alleles at each locus.

These alleles need not be combined randomly, like cards in a shuffled deck. When combined randomly, the alleles are in *linkage* equilibrium. When not combined randomly, the alleles are in *linkage* disequilibrium.

The simplest case involves two loci, each with just two alleles. If each allele has a frequency of 0.5, there are four possible chromosomes, each with frequency 0.25.

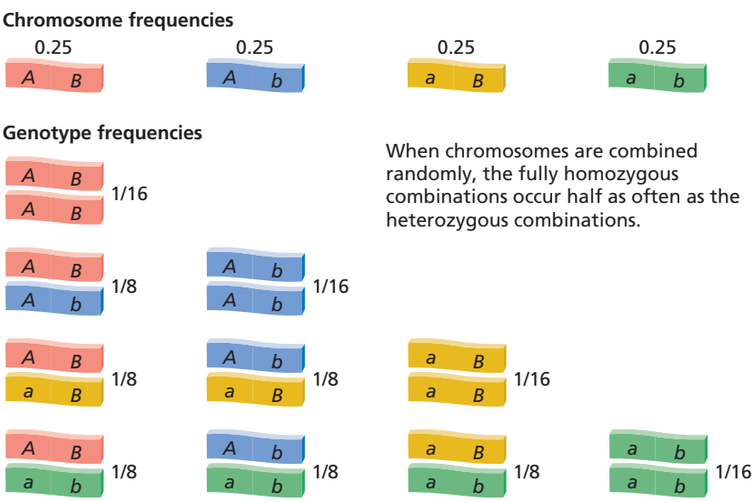


FIGURE 3.7A Linkage Equilibrium or Gamete Phase Equilibrium

Linkage Equilibrium and Disequilibrium

Linkage equilibrium is like Hardy-Weinberg equilibrium. Recall that at Hardy-Weinberg equilibrium, the frequencies of genotypes are given by the product of the frequencies of their alleles: p^2, $2pq$, and q^2.

At linkage equilibrium, the frequencies of the four possible genotypes are as follows, when the frequencies of alleles *B* and *b* at the second locus are *r* and *s*, respectively:

Gamete	*AB*	*Ab*	*aB*	*ab*
Linkage equilibrium frequency	*pr*	*ps*	*qr*	*qs*
Actual gamete frequency	P_{AB}	P_{Ab}	P_{aB}	P_{ab}

As for Hardy-Weinberg equilibrium, there is linkage disequilibrium when P_{AB} is not equal to the product *pr*, and so on for the other gamete frequencies. For this reason, linkage equilibrium is sometimes called "product" equilibrium. Note the parallel between such products in genotype frequencies and the products calculated using the multiplication rule for coin tossing, poker, and so forth. Both reflect independent probabilities, when there is linkage equilibrium or when coins are tossed fairly and cards are dealt by the rules.

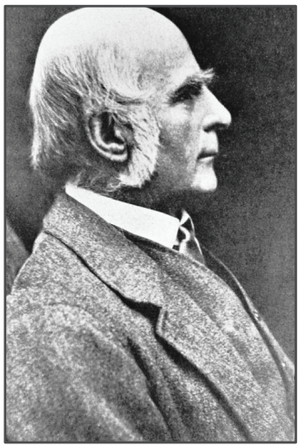
QUANTITATIVE CHARACTERS

3.8 Quantitative characters have to be studied statistically

Most Mendelian genetics focuses on qualitatively distinct characters, like eye color or flower color. More recently the genetic character of choice is the DNA sequence. The DNA sequence is indeed the most fundamental and accurately recorded character of all. It is the genotype itself.

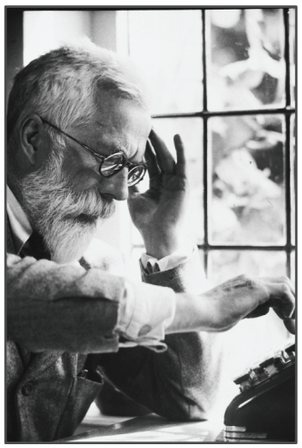
But to study evolution often requires that we study messy characters like fertility, size, and resistance to lethal stress. These characters cannot usually be scored qualitatively. Instead, they need to be evaluated quantitatively. We need special tools to make sense of them.

Those tools come from statistics. Indeed, much of statistics had its historical origins in the work of early evolutionists, like **Francis Galton**, **Karl Pearson**, and **R. A. Fisher** (Figure 3.8A). These evolutionists wanted to understand such characters as intelligence and size, and there were no statistical tools already available. So they created them. Some of the terminology of statistics reflects its origin in the words of evolutionists. One example is the word *regression*, which originally referred to a pattern of inheritance and now refers to a statistical method for calculating the best line to plot through graphical data. This combination of evolutionary interests and statistical tools gave birth to the field of quantitative genetics.



(i)

(ii)



(iii)

FIGURE 3.8A **(i) Francis Galton, (ii) Karl Pearson, and (iii) R. A. Fisher**

Before introducing quantitative genetics, we will explain some elementary statistical ideas. These ideas are the essential tools for understanding patterns of gene expression in populations. (For more detail see the Appendices, following chapter 22.)

What is a **quantitative character**, when compared to a *qualitative* (or "Mendelian") character? It is a character that has no clear categories. It is fairly easy to say that a color is red or blue. It is much harder to say whether a mouse is large or small.

Therefore, we measure quantitative characters. We might have 122 weight records for a laboratory population of 124 mice. (You will lose one or two mice during weighing.) These 122 numbers are our record of the quantitative character body weight in these mice. They are the raw data that we want to make sense of. This is the starting point for further research on quantitative characters.

An important point about the study of quantitative characters is that the conditions of measurement are often more important than they are for typical Mendelian characters. Mendelian characters are not normally changed by minor differences in laboratory handling. The eye color of a laboratory animal will not normally change depending on whether the animal is hot or cold, recently fed or not, and so on. But such factors might be important for a quantitative character. Body weight, for example, will be affected by recent feedings. Quantitative characters are not only harder to measure, but those measurements themselves are not always reliable, unless efforts are made to control and standardize handling.

Once we have a collection of numbers that are measurements of a quantitative character in a group of organisms, what is the next step? In quantitative genetics, we normally calculate two important pieces of statistical information.

First, we usually calculate a **mean** for these measurements. Scientists use different kinds of mean. One is the **median,** which is the value at which half the measurements are above and half are below. For example, we might determine the median height among a group of corn plants. Most of the time, geneticists employ the **arithmetic mean**, which is the sum of all observations divided by their number. For example, the 122 weighed mice will have an arithmetic mean weight.

Means provide crude summaries of the features shared by groups of organisms. For example, the arithmetic mean is used to give a sense of "where" a population is located. If one group of mice has mean weight of 10 grams and another group has a mean weight of 9 grams, we usually say that the first group is heavier than the second. We say this even though the second group probably has members that weigh more than 10 grams.

The second important piece of information that we normally collect concerning a quantitative character is its *variance*. Formally, a **variance** (see Figure 3.8B) is the sum of squared deviations of the observations from the arithmetic mean, divided by the total number of observations. A *deviation* is the difference between an observation and the mean. (These deviations are squared because the sum of all deviations by itself is zero.)

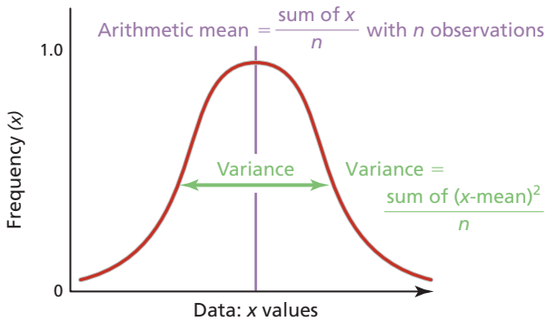


FIGURE 3.8B The variance of a random variable may be estimated as shown.

Just as a mean gives us a sense of "where" a population is with respect to a character, its variance gives us a sense of how dispersed the population is about that location. For example, a group of mice that has a mean weight of 20 grams might have a variance of 1.2, while another group of mice might have a mean weight of 19.5 grams but a variance of 3.8. The second group can be considered more variable.

Using these concepts of mean and variance, quantitative genetics studies quantitative characters. Figure 3.8B summarizes these concepts and gives their formulas. The Appendices give additional information about statistics. ❖

3.9 When environmental (E) and genetic (G) influences on a phenotype (P) are independent, $V_P = V_G + V_E$

Genetic research started with simple characters that could be scored qualitatively. As we have noted, color was a favorite character—the color of eyes and other parts. But geneticists also studied other discrete differences, like wrinkling of peas, loss of hair, dwarfism, and so on. In many cases, these distinctive phenotypic differences were based on differences in the alleles present at single genetic loci. Because Mendel himself performed this kind of research, it is called *Mendelian genetics*.

But most characters of interest to biologists, especially evolutionary biologists, do not vary in this discrete Mendelian fashion, as we have already mentioned. Instead, they vary *continuously*. This means that it is hard to pick out distinct groups. For example, we talk about people as short or tall. But there are no such distinct groups, leaving aside dwarfism. Almost everyone falls within a smeared distribution, in which every height is represented from the very short to the very tall. And height is just one example of a continuously varying character. Others include weight, endurance, hand strength, running speed, resistance to disease, fertility, longevity, and so on.

How can we understand the genetics of such continuously varying characters? The first step is to realize that most characters are determined by both genetic and environmental factors. For much of the twentieth century, a controversy raged over the "nature vs. nurture" issue, particularly with respect to child rearing. In other words, does nature (genes) determine an organism's characteristics, or does nurture (environment)? Scientifically, this controversy is now dead, because almost all biologists qualified to address this issue agree that the answer is that *both* genes and environment are important, not one or the other (Figure 3.9A). This conclusion is embodied in a simple equation, where P refers to the phenotype, G refers to the genotype, and E refers to all other influences, from the physical environment to disease to development:

$$P = G + E$$

This equation summarizes the theoretical starting point for quantitative genetics. It is not, however, always true. As presented in Figure 3.9B, when there is an interaction between genetic and environmental effects, the equation fails. However, it does not fail in such a way that G or E alone determine P. The importance of both components remains.

A very simple statistical law is useful for understanding the action of genes and environments. When we have a variable A that is determined by an equation like $A = B + C$, and B and C are independent, then the variance of A is equal to the sum of the variances of B and C. That is, variances accumulate as additional causal factors are added in. Recall that variances (V) are the averages of the squared deviations from the mean. That is, the more heterogeneity there is for a character, the greater its variance. In the genetic situation, then,

$$V_P = V_G + V_E$$

This equation means that variation in a character like height has both genetic and environmental sources. The **phenotypic variance** (V_P) is equal to the sum of the genetic variance (V_G) and the **environmental variance** (V_E).

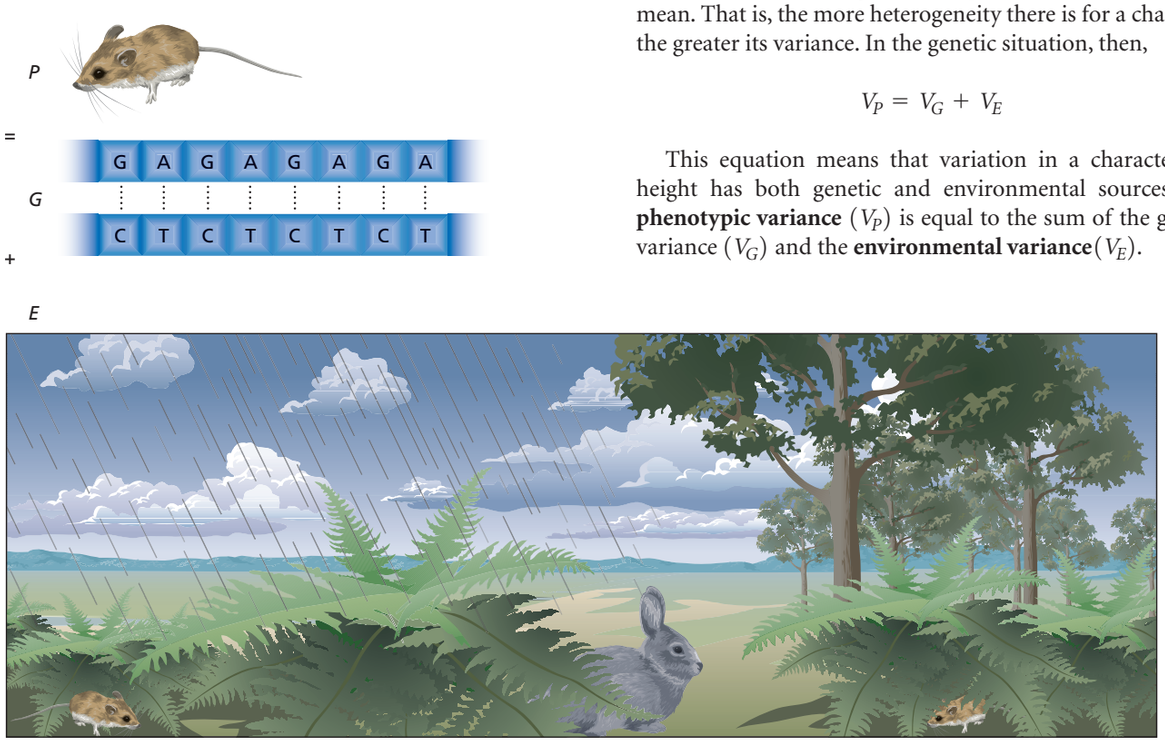


FIGURE 3.9A **Phenotype = Genotype + Environment**

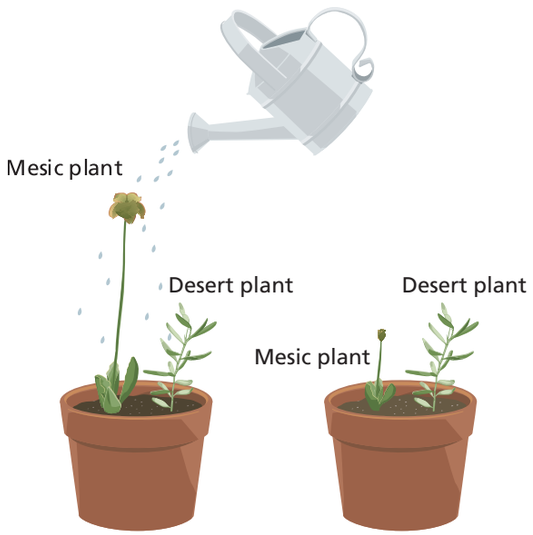


FIGURE 3.9B When Genes and Environment Don't Add Up If the populations of a species evolve in both desert and mesic (not dry) habitat, then plants from these two habitats may respond to a lack of water very differently. In this example, the mesic plants grow very poorly when denied additional water, but desert plants are not affected very much.

3.10 The genes that make up a genotype may determine the phenotype additively or nonadditively

One concern of biologists is how genes determine characters. In developmental genetics, this topic would be called the problem of gene expression. In evolutionary genetics, this topic is the mapping from genotype to phenotype that is half of the basic cycle of the evolutionary process. The topic is also studied in quantitative genetics.

A central difficulty of genetics is that the same allele does not always have the same effect on the phenotype. The effects of an allele may be modulated by other alleles in the genotype. The simplest form of this modulation is called **dominance** (see Figure 3.10A). In genetic dominance, the expression of one allele largely or completely dominates the expression of another allele at the same locus. For example, diploid loci that code for pigmentation usually have dominance of nonwhite alleles over white or albino alleles. This is true in animals as different as mammals and insects. In these cases, having only one copy of an allele for normal pigment is enough to produce an almost normal coloration. This type of genetic dominance was much studied by early geneticists, and is still of interest to geneticists today.

Of greater medical interest is the fact that many human *genetic diseases* also exhibit dominance. This dominance comes in two forms. In one form, alleles causing genetic diseases may do so when they are present in just one copy. This is true of Huntington's disease, for example. Chapter 4 consider these diseases in more detail.

Another form of dominance in human genetic disease is **recessiveness** of the disease-causing allele, so that it takes two copies to cause the disease. Individuals who have only one copy of the disease allele are disease-free "carriers." Recessiveness is a very common pattern, illustrated by such genetic diseases as cystic fibrosis and Tay-Sachs disease. We will discuss these diseases further in Chapter 4, also. Genetic dominance is not just interesting; it can also be deadly.

Alleles that are not at the same genetic locus can also interact. This interaction is called **epistasis**. With epistasis, alleles at loci located far away in the genome, perhaps on another chromosome, can alter patterns of gene expression at another locus. This type of genetic interaction is not as easy to characterize as dominance. The simple thing about dominance is that it involves just two alleles, and they are "across from" each other genetically on matching (homologous) chromosomes. With epistasis, the important allele could be located millions of nucleotides down the chromosome, and there is no simple way to find it. Experiments with laboratory and agricultural organisms have also shown that epistasis can be important in determining such important characters as early survival, growth, and reproduction, as shown in the accompanying box. ❖

Epistasis for Viability (survival to maturity) in *Drosophila*

This experiment measured the viability levels of fruit fly (*Drosophila*) larvae with different genotypes at two loci: alcohol dehydrogenase and alpha-glycerophosphate dehydrogenase. S and F refer to the mobility of the proteins made by their corresponding alleles, S meaning slow, F meaning fast. This particular experiment was performed with alcohol absent. The values shown are average viabilities for the genotypes at the corresponding genotypes that align with the row and column of the viability's position.

Note that in some gene combinations there is superiority of the heterozygote. In other combinations, the heterozygote is not superior. This is epistasis, because the effects of alleles at each of the two loci are modified by the genotype at the other locus. This is an example of "nonadditive" inheritance.

(This experiment by Cavener and Clegg is presented in more detail in many genetics textbooks, including those in the list of readings at the end of the chapter.)

		Alcohol Dehydrogenase Genotype		
		SS	SF	FF
Alpha-glycerophosphate dehydrogenase genotype	SS	0.99	1.06	0.86
	SF	1.08	1.00	0.94
	FF	0.77	1.16	0.75

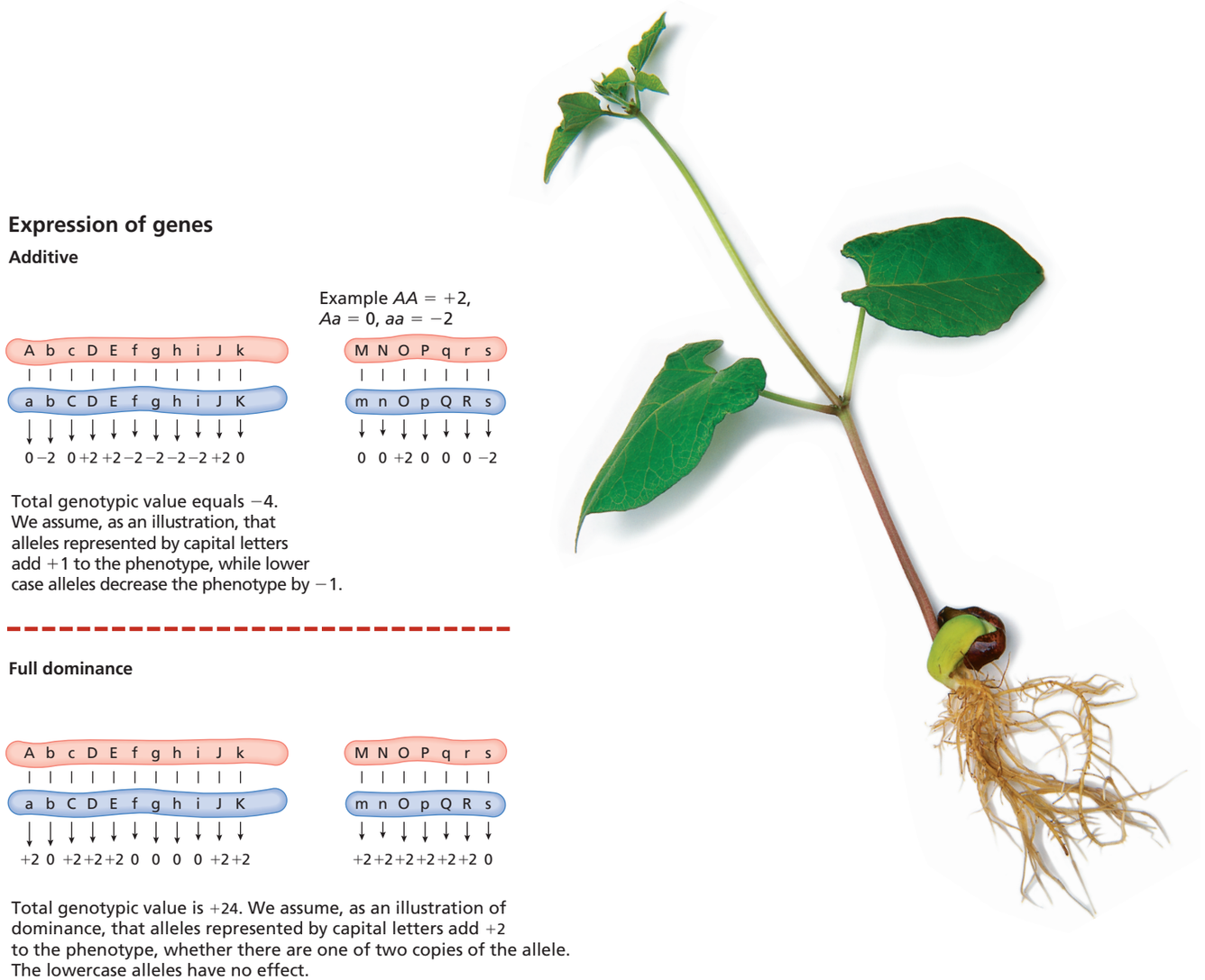


FIGURE 3.10A Additive versus Fully Dominant Inheritance

3.11 The resemblance of relatives is determined by the ratio of the additive genetic variance to the phenotypic variance

What can we predict with variances? What can we learn from them? In fact, variances can be highly revealing. Suppose you wanted to know how much relatives are likely to resemble each other. Consider the simplest genetic situation: no interactions between different alleles, no interaction between genes and environment, and linkage equilibrium. This is the simplest, essentially ideal, case for quantitative genetics. Under these conditions, the genetic var-iance is reduced to the **additive genetic variance,** which is simply the genetic variance when there is no dominance or epistasis. We can define the **heritability** (h^2) of a character as the ratio of the additive genetic variance (V_A) to the phenotypic variance (V_P) that we have already seen:

$$h^2 = V_A/V_P$$

Heritability indicates the relative importance of inheritance in determining quantitative characters. When heritability is zero, inheritance has no importance. When heritability is 1.0, inheritance has overwhelming importance. The surprising thing about heritability is that it has a simple quantitative relationship to the **resemblance between relatives**. Resemblance between relatives means the similarity of biological characters between parent and child, brother and brother, grandparent and grandchild. This is a quantitative similarity. We can represent it graphically by plotting, for example, the average values of parental characters against the values of offspring characters, as shown in Figure 3.11A. In this type of graph, each point represents the quantitative characters that come out of an entire family. The collection of plotted points gives the data for the collection of families that have been studied. This kind of plot is just like any other in science. If we have two variables that are closely and positively related, we expect the points to fall near a rising line.

In Figure 3.11A, we show hypothetical data for adult weight in Old English Sheepdogs, a breed that you may have seen in children's movies, if not in real life. Usually the pattern of data like this is characterized by linear regression of the *y*-dimension data on the *x*-dimension data. (Linear regression, or least-squares linear regression, gives the straight line that comes closest to fitting the scatter of data plotted in two dimensions; see the Appendices.) In this example, the data are plotted with the average weight of the parents on the x-axis and the average weight of the offspring on the y-axis. A linear regression of average offspring weight on the average weight of their parents gives the best straight line for the fit of offspring to parent data. The slope of this straight line measures the strength of the relationship between parent and offspring weight when they are measured at the same age. Genetic theory shows that the slope of this regression is equal to the heritability. Thus characters with higher heritability, like body weight, should have larger slopes relating offspring phenotypes to parental phenotypes. Characters with lower heritabilities, such as most behavioral characters, should have shallower slopes relating offspring to parent, which would mean that genetics are relatively less important in determining such characters. Table 3.11A gives some examples.

It is also interesting that when the data are re-plotted using the weight of just one parent, instead of two, the expected slope falls to half the heritability. This makes sense, intuitively. If we know the phenotype of half the parents, we have half the information needed to predict the phenotype of the offspring. The importance of heritability in the regression scales with the amount of genetic information available.

Size characters tend to have higher heritabilities than do characters closely related to fertility. Analysis of many more characters than those shown here confirms this pattern. This may indicate that natural selection has used up genetic variability for characters like fertility, for the reasons discussed in Chapter 4. ❖

TABLE 3.11A Some Examples of Heritabilities

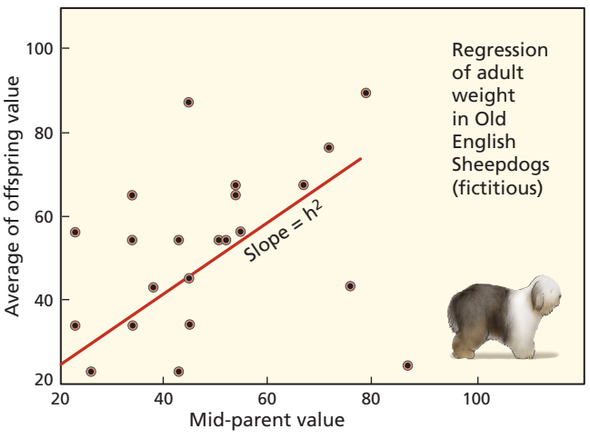
Morphology

Human height: 0.65	Adult weight in cattle: 0.65
Pig growth rate: 0.40	Adult weight in poultry: 0.55

Fertility

Pig litter size: 0.05	Egg production in poultry: 0.10
Mouse litter size: 0.20	Fruit-fly egg production: 0.20

From D. S. Falconer and T.F.C. Mackay, *Introduction to Quantitative Genetics*, 4th ed. (Harlow, Essex: Longman, 1996).



The slope of the line predicting an individual's phenotype from the phenotypes of its relatives has the following values:

h^2 when the corresponding phenotypes of both parents are used to make the prediction

$h^2/2$ when the phenotype of only one parent is used to make the prediction

For characteristics like size and weight, heritability is typically 0.6 – 0.8. For such characters as fertility or running speed, heritability is typically 0.2 – 0.6.

FIGURE 3.11A How Much Will Organisms Resemble Their Relatives? This depends on "heritability" or h^2. When inheritance is additive, $h^2 = V_G/V_P$.

Sex and Recombination

3.12 Population genetics is like shuffling and dealing cards

From an evolutionary standpoint, the important thing about genetic processes is that they shuffle genes, just like cards are shuffled in card games. This shuffling process does not create any new genes, and it is not usually biased. Like the dealer in a card game, the genetic machinery makes new combinations of alleles—like new combinations of cards in "hands"—and thus new genotypes.

From card games, you may know that even with the same 52 cards in a deck, the number of different hands that you might be dealt is extremely large, if there are a lot of cards per hand. Blackjack, which has relatively few cards in play at any one time, may be at the simpler end of the spectrum. But in bridge, where each player gets 13 cards, the likelihood that you will get two hands with exactly the same combination of 13 cards in any one game is vanishingly remote.

In genetics, the number of genes per genotype is far higher than the number of cards in the hands of any card game. Humans have about 20-25 thousand genes that code for important functions, and each of these genes may have many different alleles. If you calculate the possible genotypes, the total number that might occur is very large. By some estimates, there are more possible human genotypes than there are atoms in the universe.

In this sense, the genetics of populations are very big card games indeed, and we have little prospect of winning by "counting cards." In other words, it is unlikely that we can understand life by directly calculating all of its genetic possibilities. Instead, we try to understand how the evolutionary machinery works in some average, or typical, sense. What does the evolutionary machinery normally do?

As mentioned, a notable feature of genetics is that it usually operates without any bias to its shuffling. Because of this, it tends to randomize genotype frequencies, within loci and between loci. Neither of these statements is absolute, however. Genetic processes are sometimes biased, as we will discuss in Chapter 5. And the tendency to randomize genotype frequencies is not absolute. Indeed, sustained inbreeding tends to produce odd combinations of genes, though sustained inbreeding

is not the rule in the natural world. The idea of genetics as randomizing is no absolute law of science, but it is a general rule of thumb.

For example, consider a piece of the genetic machinery that should already be familiar. We have seen in Module 3.6 what one generation of random mating can do to genotype frequencies at a single locus. One generation of random mating leads immediately to the Hardy-Weinberg equilibrium. At this equilibrium, we can calculate the frequency of genotypes from the frequencies of the alleles at the locus. If the frequency of allele A is p, then the probability of A occurring twice in a genotype (AA) is p^2. This is just like the probability of getting two heads in a row when tossing a coin: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. And so on. The Hardy-Weinberg equilibrium is a kind of "even" or "smoothed-out" state, in which probabilities are well behaved. Alleles are randomly associated with each other; this we have already seen.

Figure 3.12A shows two different populations in which the frequencies of alleles A and B are both 0.5, at their two respective loci. In part (i), the improbable case, all the uppercase alleles are in one genotype and all the lowercase alleles are in the other genotype. In part (ii), the probable case, the genotype frequencies reflect the random combination of alleles into genotypes, with no unusual biases or associations. Intuitively, we expect the evolutionary machinery to undermine the improbable pattern of genotype frequencies, and so produce something like the random combinations of alleles of this example. Next we consider exactly how this happens.

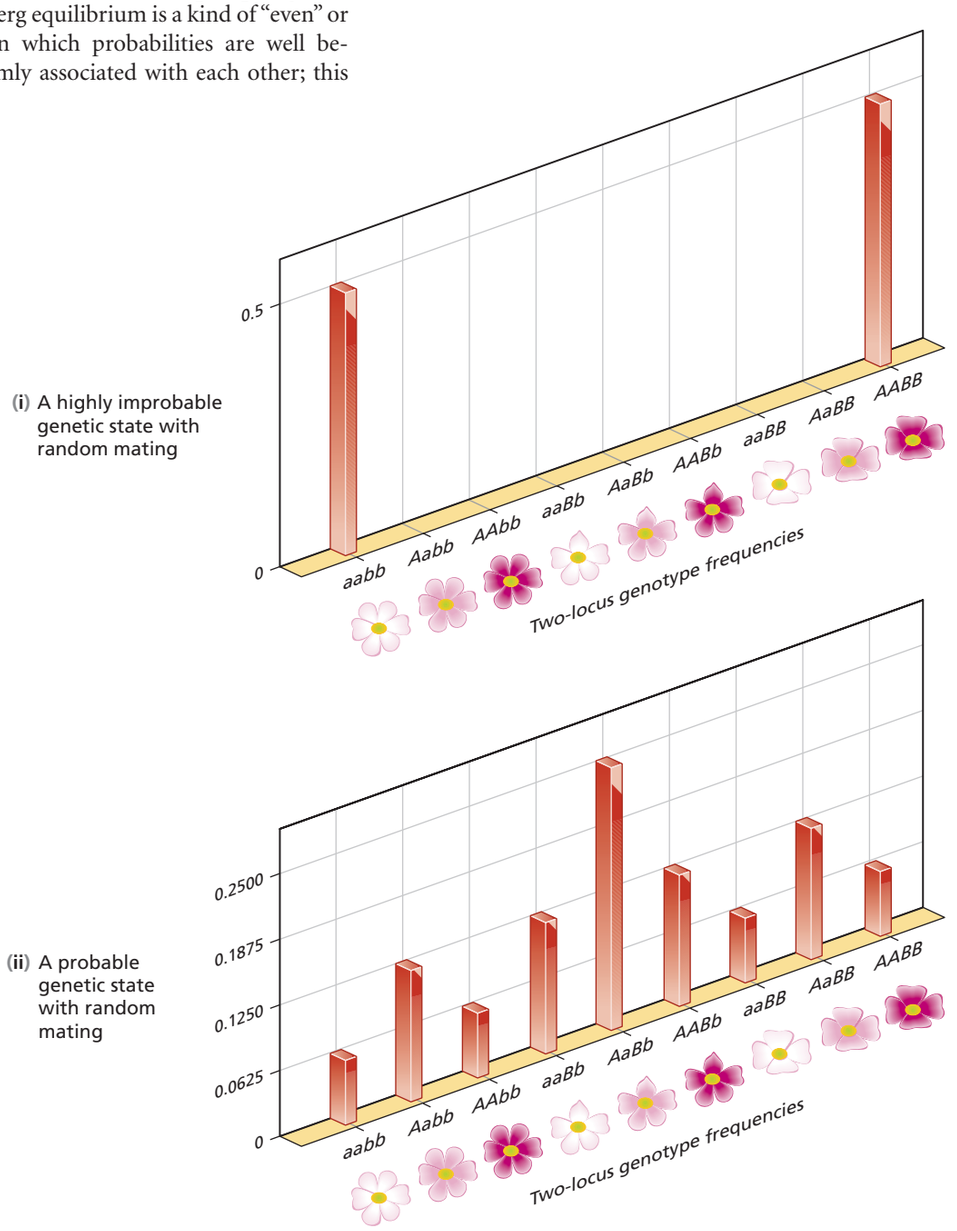


FIGURE 3.12A Two Genetic States, One Improbable and One Probable

3.13 With random mating, sex-chromosome genes that start out at different frequencies move toward the same frequency

One of the simplest cases of genetics randomizing genotypes occurs when there are two sexes. In principle, though rarely in practice, the two sexes can have very different frequencies for alleles located on the **X chromosome**. Let's take a hypothetical example. Imagine a group of female space-travelers whose spaceship crashes on an all-male prison planet. If the females came from a different solar system, which had been colonized long before, then they might have X-chromosome alleles that are totally different from those of the men in the prison. All the women would have one allele, and all the men would have another allele. The interesting question then is, what will happen to the X-chromosome allele frequencies of the two sexes if they start mating with each other and thereby create their own autonomous population on the prison planet?

A key factor is that, after the founding generation, the XY males of each generation will have the X-chromosome allele frequency of the XX females of the previous generation. This occurs because males receive their **Y chromosome** from their father and their X chromosome from their mother. You can see this in Figure 3.13A, which shows the outcomes of all possible matings when there is one X-chromosome locus having just two variant alleles. When mating is random, the male X chromosomes will be a random sample of the X chromosomes of the females of the preceding generation.

But the female case is different. Daughters do not get a Y chromosome from their fathers. Their XX genotype comes from a paternal X chromosome paired with a maternal X chromosome, exactly one from each parent. With random mating, this pair of X chromosomes is a random combination of X chromosomes from both males and females, the two sexes equally represented. Therefore, for daughters, the frequencies of alleles on the X are averages of

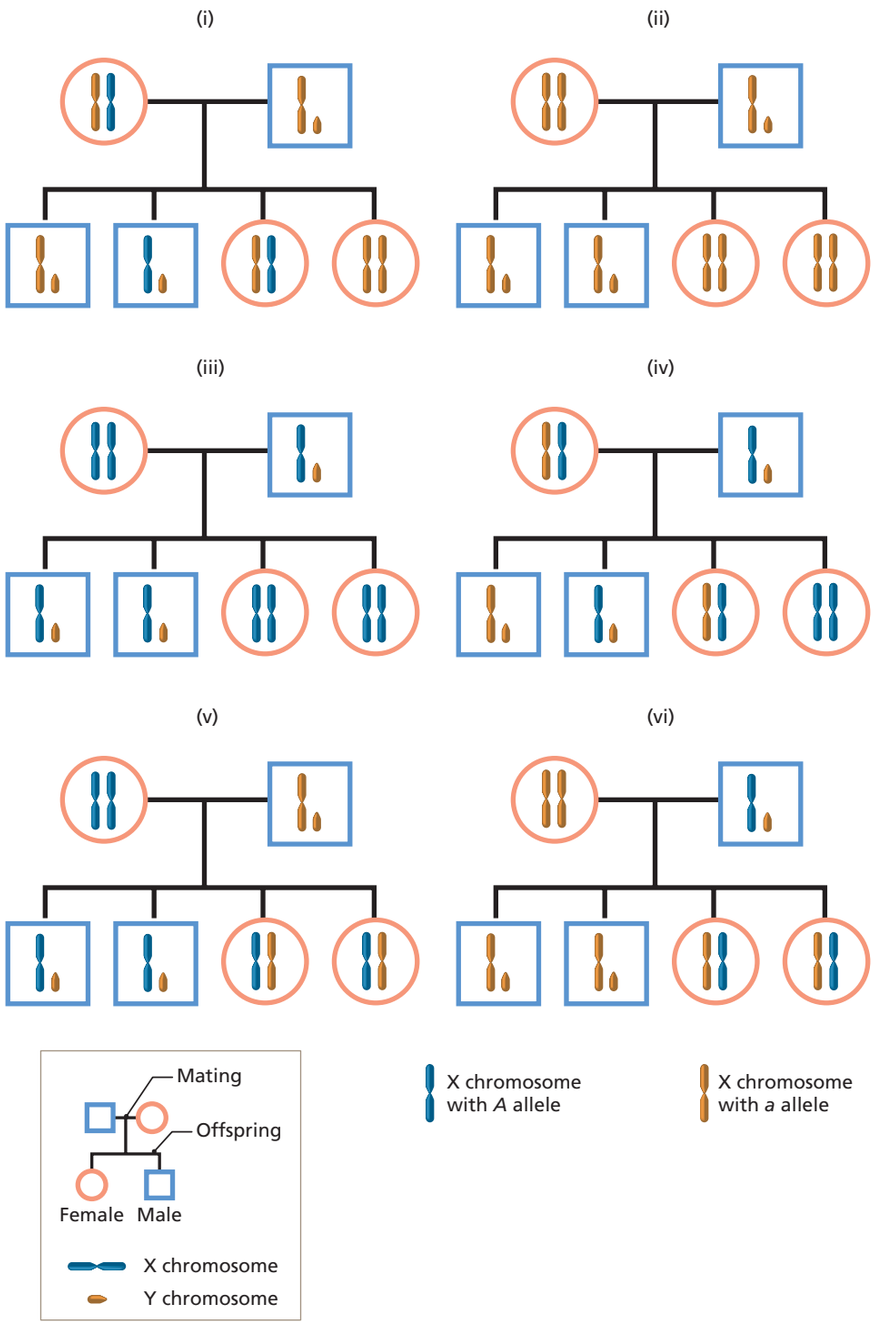


FIGURE 3.13A All Possible Matings When There Is Polymorphism at a Sex-Linked X Chromosome Locus

allele frequencies from both sexes. This averaging tends to reduce the difference between male and female allele frequencies for genes on the X chromosomes, because the average of two different numbers is equidistant between them. (The average of 2 and 8, for example, is 5.) This averaging does not immediately eliminate the difference in allele frequency between the sexes, however. Instead, as Figure 3.13B shows for allele *a*, the difference between them progressively falls, as allele frequencies bounce back and forth between males and females. Eventually, there would be no difference between male and female X-chromosome allele frequencies in the new population on the prison planet.

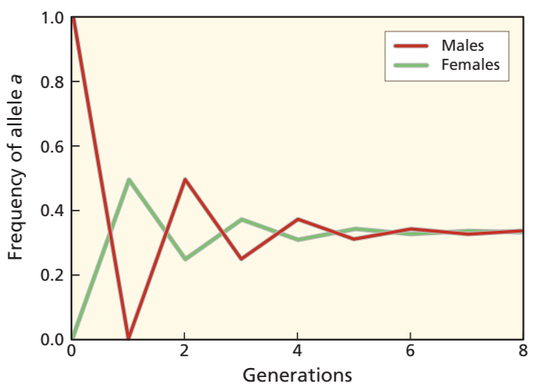


FIGURE 3.13B Allele Frequency Change at a Sex-Linked Locus

3.14 Recombination progressively breaks up nonrandom associations of alleles among loci

A simple analogy for understanding genetics is that genes are organized like sentences in books, where chromosomes are the books. There are 23 pairs of human chromosomes, together containing about 20-25 thousand genes. This means that each chromosome has an average of about 1,000 genes. These genes are not strung along, one after another, with no gaps. At many places along the length of a human chromosome, there are large gaps where the DNA does not code for useful RNA or protein. Because of these large gaps, pairs of chromosomes sometimes recombine their genes without much effect on gene-coding regions.

Let's continue with the book analogy to see how this happens. Imagine two novels *of the same length* being printed at the same time by a book factory. If a printer mixes up the printing plates, then instead of producing copies of *Death's Dishonor* and *Summer Swoon*, they would print some books in which the text comes from *Death's Dishonor* for the first 128 pages, and then *Summer Swoon* for the last 211 pages, as well as the reciprocal switch of *Summer Swoon* for the first 128 pages, followed by *Death's Dishonor* for the last 211 pages. The printer's recombination occurred between pages 128 and 129. At that point, the two stories *cross over*. In the same way, chromosomes of the same type (homologous) can be recombined by breaking in two between genes and then rejoining by crossing over between chromosomes.

Chromosome **recombination** has varied results. First, genetic recombination may make no difference, because pairs of chromosomes may be carrying the same alleles at a genetic locus, as shown in Figure 3.14B. In our two novels, this case is analogous to recombination between two books that begin with exactly the same words, at least up to the point of recombination. If the printer's mix-up takes place before the two stories differ, then it makes no difference.

But when physical recombination occurs and there are different alleles on the recombining chromosomes, the genetic system acts to shuffle the alleles. This shuffling tends to break down unique or unusual genetic combinations, rendering them only as common as they would be if the alleles had combined at random, irrespective of their chromosomal location.

For example, as shown in Figure 3.14A, *A* and *B* alleles could be associated with each other, and *a* and *b* likewise. These alleles would show **coupled phase** association with each other, according to uppercase or lowercase. Similarly, associations between *A* and *b* or *a* and *B*, which are called **repulsion phase** associations, could occur. Such coupled phase associations and repulsion phase associations can contribute to **linkage disequilibrium**. Linkage disequilibrium is measured as the difference between actual genotype frequencies and the frequencies expected from random combination of alleles, as described earlier. Either type of association tends to be broken down by recombination, as shown in Figures 3.14C and 3.14D, reducing linkage disequilibrium.

Figure 3.14E shows this process of randomization among genes undergoing recombination at different rates (r = 50%, 10%, and 1%). The more recombination, the faster the association between alleles disappears. But in all these cases, recombination does eventually destroy nonrandom association, making linkage disequilibrium fall to zero.

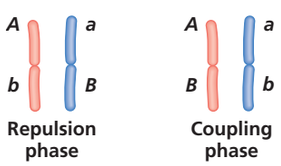


FIGURE 3.14A Coupling and Repulsion Phases of Linkage Here and in the other figures on this page, the color of the chromosome indicates maternal (pink) or paternal (blue) origin of the chromosome, not the allelic composition of the chromosome.

FIGURE 3.14C Recombination matters in coupling phase, when it produces repulsion-phase gametes.

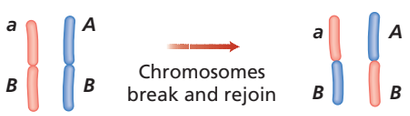


FIGURE 3.14B Physical recombination of chromosomes may not change genetic makeup.

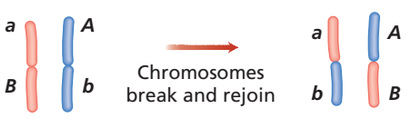


FIGURE 3.14D Recombination matters in repulsion phase, when it produces coupling-phase gametes.

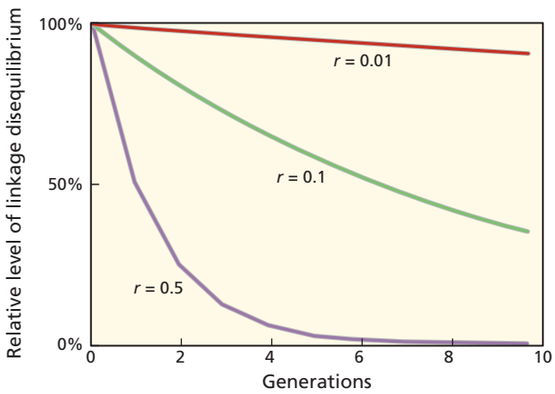


FIGURE 3.14E **Decay in Linkage Disequilibrium through Time**

INBREEDING

3.15 Inbreeding is a bad thing in normally outbreeding natural populations

Inbreeding occurs when related individuals mate. This process is very important for the genetics of families and for the genetics of populations. Inbreeding is not an all-or-nothing event. Instead, the degree of inbreeding is quantitative. The pedigree in Figure 3.15A diagrams a case of brother-sister mating, from which three daughters were produced. The pedigree in Figure 3.15B shows a first-cousin mating that produces two daughters and a son. Both of these pedigrees are examples of inbreeding. The critical point in the definition of inbreeding is whether individuals share alleles that they inherited from a common ancestor. Only if there is biological descent from at least one common ancestor can there be inbreeding. (A stepfather's rape of his stepdaughter is not usually biological inbreeding, even when the law considers it incest.) Having a great-grandfather in common, as may be true of second cousins, is inbreeding even if it is not considered incest. From here on, our discussion will consider inbreeding only, whatever the social or legal conventions concerning incestuous mating.

Within biology, degrees of inbreeding are distinguished according to the probability that two mating individuals have an allele in common from an ancestor. This is a probability that can be calculated, as we will discuss shortly. For now, the important point is that inbreeding varies quantitatively.

Inbreeding has many important effects on the evolutionary process—reducing **heterozygosity**, decreasing genetic variance, and so on. In medicine, however, inbreeding is better known from its effects on health. Many human genetic diseases are caused by recessive alleles. In such cases, the diseases occur only when the patient has two copies of a defective allele, resulting in a failure to produce the normal protein coded for by the genetic locus. Such disorders are called **recessive genetic diseases**. Some of the most common and devastating human genetic diseases, like cystic fibrosis and Tay-Sachs (see Chapter 4), are caused by a lack of normal alleles at a single locus.

Inbreeding greatly increases the frequency of genetic diseases that arise from **homozygosity** (having two copies of the same allele). Natural selection normally keeps disease genes at very low frequencies, so only a few individuals will have even one copy of a gene for a recessive genetic disease. But because there are thousands of loci that can cause recessive genetic disease, even though disease alleles are rare at each locus, each of us carries one copy of a few alleles for recessive genetic diseases among all our loci. Fortunately, these alleles are heterozygous,

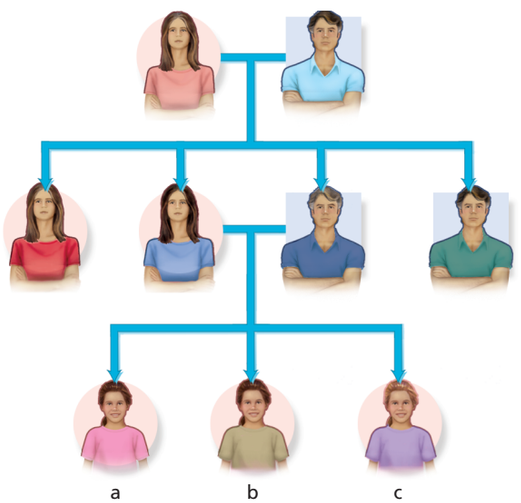


FIGURE 3.15A Females a, b, and c are the inbred offspring of full siblings. On average, they will be very inbred.

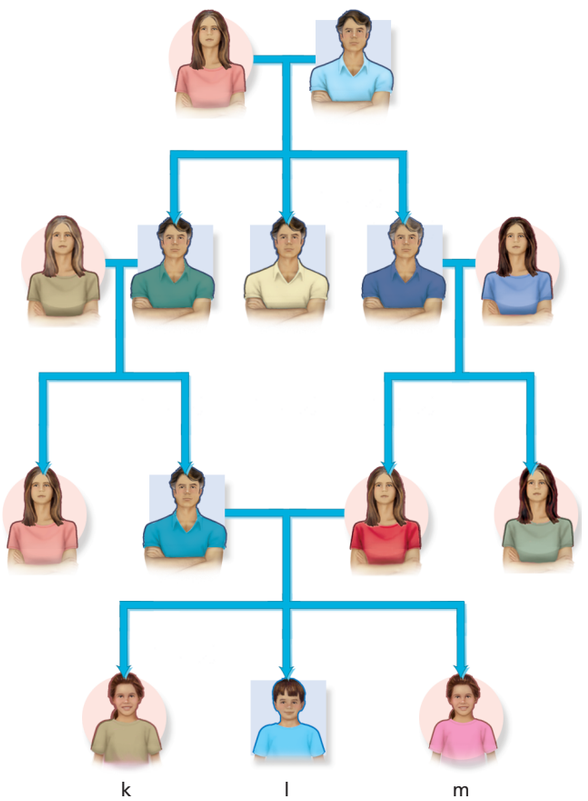


FIGURE 3.15B Females k and m, as well as male l, are the offspring of a first-cousin marriage. As a result, they will be moderately inbred.

so they cause us few medical problems. Statistically, we are unlikely to mate with an individual carrying the same disease alleles, if we are not related to each other. For this reason, most of our offspring are normal regarding genetic disease.

Things are different for inbred individuals. Parents that share a common ancestor are much more likely to have inherited the same genetic disease genes, albeit in heterozygous form. When such related individuals have children, their children are at a much greater risk of being homozygous for one or more alleles conferring a genetic disease. These inbred offspring are more likely to suffer medical complaints and more likely to die young. Such victims of inbreeding also tend to suffer a range of morphological oddities, such as extra toes or fingers, along with intellectual retardation.

How Much Inbreeding Is There in Nature?

Inbreeding is common on farms and in scientific laboratories. But how common is it when humans don't interfere with animal or plant breeding?

It is fairly well established that species with very limited dispersal are often inbred in the sense of having very small populations. Small cave invertebrates, mostly-selfing plants like peas, plants with very local pollination, and selfing worms like nematodes all may have high levels of inbreeding. With such inbreeding comes high levels of homozygosity and accidental differentiation of local subpopulations.

On the other hand, outbreeding is extremely common. Animals and plants that seem like they should be inbred often are not. The humble fruit fly *Drosophila melanogaster*, which doesn't seem like much of a flyer and can't walk very fast, apparently mates over a large area, so that it exchanges genes over hundreds of square miles. Pollen can spread surprising distances, genetically uniting plants over wide areas. This is to say nothing of organisms that obviously disperse and mate over large areas, like most birds and many mammals. We currently do not know enough about how common inbreeding is relative to outbreeding. This is an important question, because the human impact on the environment accumulates unchecked. We do not know if we are causing the extinction of distinct populations or merely reducing the abundance of a species that disperses widely. If species mate widely, then their homozygosity will be much less than it would be if the species are broken up into local breeding populations. (This is considered further in Module 3.20.) Less homozygosity, in turn, should reduce inbreeding depression. With less inbreeding depression, endangered species should be less likely to go extinct.

3.16 The degree of inbreeding can be calculated from the probability that parents share alleles from a common ancestor

Because inbreeding is one of the most important processes of population genetics, it is convenient that biologists can calculate the degree of inbreeding fairly easily. These calculations revolve around the probability that related individuals will have offspring that are homozygous. The most important thing in these calculations is to keep track of who is mating with whom. Here we will assume that we know the exact truth about every individual's parentage.

Figure 3.16A shows what happens when two half-siblings mate. Half-siblings have only one parent in common—in this case, their mother. We can consider one locus at a time, because we will be calculating average probabilities for the genetics of this locus. Many loci may be made homozygous by inbreeding, but the average frequency of homozygosity will be close to the results of the one-locus calculation.

We need to follow the alleles that each half-sib inherits from its mother, so we label the mother's alleles a_1 and a_2 in Figure 3.16A. We can ignore the genotypes of the fathers when we can assume that they are not related to each other or to the mother. For this reason, the alleles that come from the fathers are represented by dashes.

Mendelian genetics is like a card game, as we remarked earlier in the chapter. When the mother has a son with one father, the son has a probability of $\frac{1}{2}$ of getting allele a_1 from his mother and a probability of $\frac{1}{2}$ of getting allele a_2 from her. The same thing is true of the daughter by the other father.

Under these conditions, there are two scenarios by which the half-sibs can have an allele in common: Each receives a copy of a_1 or each receives a copy of a_2. The probability of the first scenario is $\frac{1}{4}$, because the chance of each of them receiving an a_1 is $\frac{1}{2}$, and these two genetic transmission events are independent of each other. Therefore, we multiply their probabilities together to get the probability that both occur: $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Likewise, the probability of the second scenario, in which the two half-sibs have the a_2 allele in common, is also $\frac{1}{4}$. We then sum these two probabilities to get a probability of $\frac{1}{2}$ that the half-sibs have a maternal allele in common at this locus, whether a_1 or a_2.

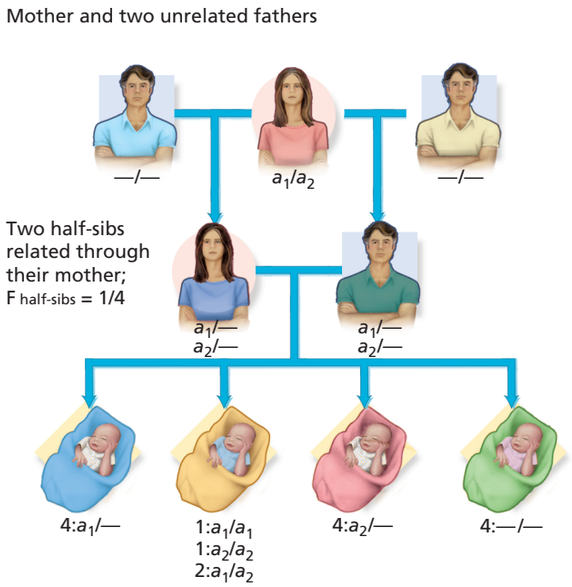


FIGURE 3.16A **Genetic effects of Incest**

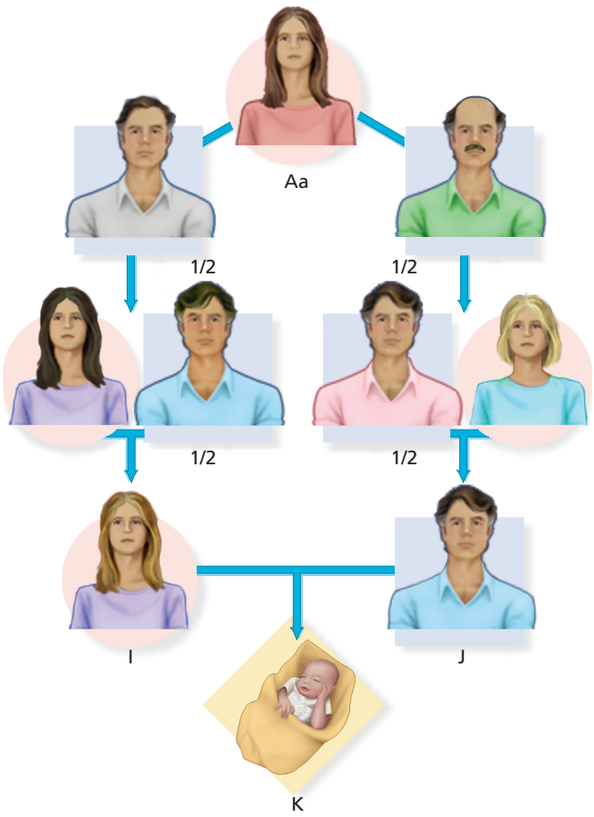


FIGURE 3.16B **A Simple Rule for Estimating the Inbreeding Coefficient, F.**

In the case that the half-sibs have an allele in common, there is a probability of $\frac{1}{4}$ that a child of theirs would be homozygous for that locus. (From the rules of Mendelian segregation, two heterozygotes have a chance of $\frac{1}{4}$ of producing an offspring that is homozygous for a particular allele when there are only two alleles.) Multiplying this probability by the chance that the half-sibs do have an allele in common, 1/2, we find that the probability of homozygosity arising from half-sib inbreeding is 1/8.

If we choose an allele at random from one of two half-sibs just described, there is a 50 percent chance that this allele is a maternal allele. As we have shown earlier, the chance that the two half-sibs share a maternal allele in common is also 50 percent. Thus the chance that a randomly chosen allele in two half-sibs is identical by common ancestry is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. This probability is sometimes called the **inbreeding coefficient** or ***F*** value. In large populations that are mating at random, the average inbreeding coefficient will be close to zero. In organisms that have been inbreeding for a number of generations, *F* may approach 1, its maximum value. In a sense, the inbreeding coefficient is a measure of deviation from random mating.

Inbreeding coefficients can be relatively easy to calculate. You take the two individuals whose inbreeding coefficient you wish to calculate, and you follow the pedigree connecting them. For each connection in the pedigree, you multiply by $\frac{1}{2}$. Thus two individuals related by just one common parent—such as the half-sibs we have just discussed—have an *F* value of $\frac{1}{4}$ (Figure 3.16A). When they have just one common grandparent, they would have an *F* value of 1/16 (Figure 3.16B). With multiple parents and grandparents in common, these calculations have to be repeated for each independent pathway through the pedigree. This can get complicated. However, most types of inbreeding involve the same patterns.

3.17 When relatives mate frequently, homozygotes increase in frequency while heterozygotes decrease in frequency

Let's consider cases in which inbreeding is sustained from generation to generation. In these situations, groups of inbreeding organisms are called **inbred lines**.

Although human inbreeding causes genetic disease, there are some good reasons for inbreeding in agriculture. We regularly eat parts of inbred organisms, both plant and animal. When inbreeding is sustained, the genetic polymorphism of randomly mating, or **outbred**, populations is lost (Figure 3.17A). Inbred lines tend to *fix* single alleles at each locus. (The term **fixation** in genetics refers to 100 percent frequency.) Which allele will be fixed is not usually predictable. But if several different inbred lines are produced, plant or animal breeders can choose the line with the best attributes for their purposes. The advantage of working with inbred lines is that inbred lines remain "true to breed." Their lack of genetic variation ensures that deviant offspring will be rare.

Figure 3.17B presents an example of how inbreeding can reduce the heterozygote frequency. An outbred population of plants has six genotypes at locus *A*. All these genotypes are represented in the original population. In three inbred lines derived from the outbred population, only two genotypes are represented, one genotype by two different inbred lines. Both of these genotypes specify short plants, though this is only coincidental. In this way, by accident, inbreeding has produced a directional change in plant height.

Some plants can **self-fertilize**; that is, they can be fertilized by pollen from their own flowers. **Selfing** is a form of inbreeding. To breed with another plant, plants are often dependent on **pollinators** like bees and birds, which carry pollen from one plant to another. When both selfing and external pollination are possible, all that is required for inbreeding is the loss of pollinators. When all the pollinators are lost, reproduction must occur by selfing, if it is to occur at all. Self-fertilization is not as common in animals, but it does occur in some snails and worms. In some species of nematodes—very simple worms—reproduction is normally by self-fertilization. This makes inbreeding, along with homozygosity, very common in these species; and *F* (the inbreeding coefficient) approaches 1.

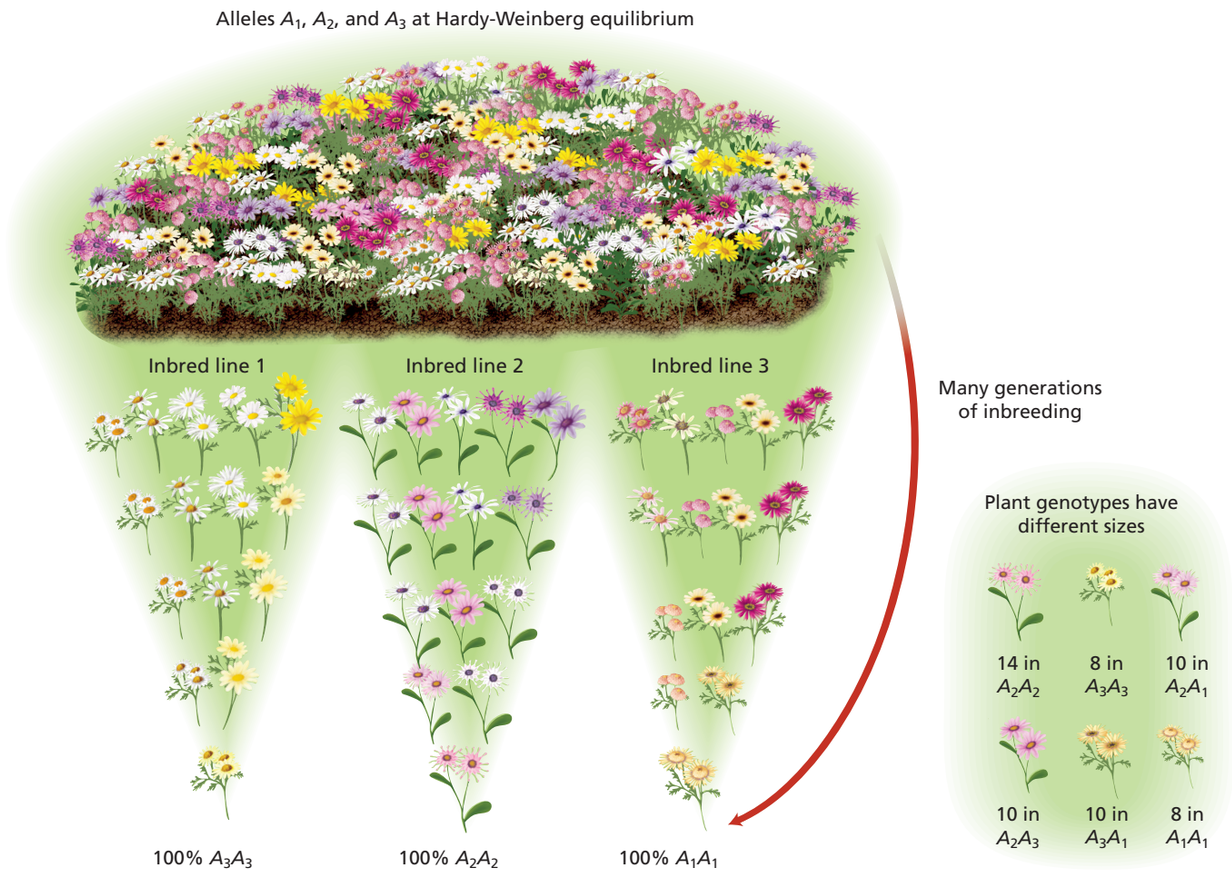


FIGURE 3.17B Differentiation of Inbred Lines from Each Other Again we suppose that several inbred lines are derived from an outbred population. When different genotypes are associated with different phenotypes, such as height, then inbreeding will produce inbred groups with consistent differences in such characters.

FIGURE 3.17A Creating Inbred Lines Imagine a plant that has one generation per year. If we start with a greenhouse population of 38, two families might buy plants for their gardens. The Gardenias might buy 5 plants, while the Mudges buy 3. If these plants mate only within their home gardens, and some plants are lost each year because of bad gardening, then the two families may end up with genetically distinct plants in their gardens.

3.18 Inbreeding tends to reduce the variance of quantitative characters within inbred lines

It is fairly easy to understand that inbreeding will tend to fix particular alleles at individual loci, increasing the level of genetic homozygosity in a population. It may be more difficult to understand that similar effects occur where the inheritance of quantitative characters is concerned; but they do.

Recall that the variation of a quantitative character has two basic components: genetic variation and environmental variation. For now, we will assume that changes in the mating system will not affect the environmental component of variation. It is the effect on genetic variation that is the basis for predicting that inbreeding will tend to reduce the variance of quantitative characters.

To understand the effect of inbreeding, it is important to distinguish its effect on a group of inbred lines from its effect on a single line. An entire collection of different inbred lines may possess a great deal of genetic variation. Just think of the enormous variation among all the different breeds of dog (see Figure 3.18A). Domestic dogs are descendants of one, or a few, populations of dog that humans domesticated 50 to 150 thousand years ago, when these dogs were essentially wolves. Yet an individual breed of dog does not contain all this variation. The effect of inbreeding on the individual breed is very different from its effect on an ensemble of inbred lines.

Let us focus on a single inbred line. As inbreeding proceeds, each locus has an increased likelihood of becoming homozygous, as we saw in the preceding module. Over all loci, many will become homozygous as a result of inbreeding. When loci that affect quantitative characters become homozygous, there will be less quantitative genetic variance (V_G) affecting those quantitative characters. This occurs because quantitative genetic variance requires genetic polymorphism. As Figure 3.18B shows, sustained inbreeding reduces the genetic component of the phenotypic variance (V_P) during sustained inbreeding, making the phenotypic variance approach the value of the environmental variance (V_E). This reduction in variance increases the predictability of the characters of the inbred line.

If these inbred lines are horses, dogs, cows, or tomatoes, this increased predictability may increase the value of the inbred line. Indeed, dog shows and similar competitions for agricultural animals often focus on specific standards that have been established for breeds. Deviations from those standards result in lost points during competition. Such deviations can be prevented best by maintaining the purity of the breed, avoiding any crosses with animals from other breeds or crosses with mongrels. Long-standing human practices have thereby fostered the continued inbreeding of dogs as well as other animals and plants. ❖

FIGURE 3.18A Three Inbred Dog Breeds Proceeding from top to bottom: American beagle, Belgian sheepdog, Pekingese.

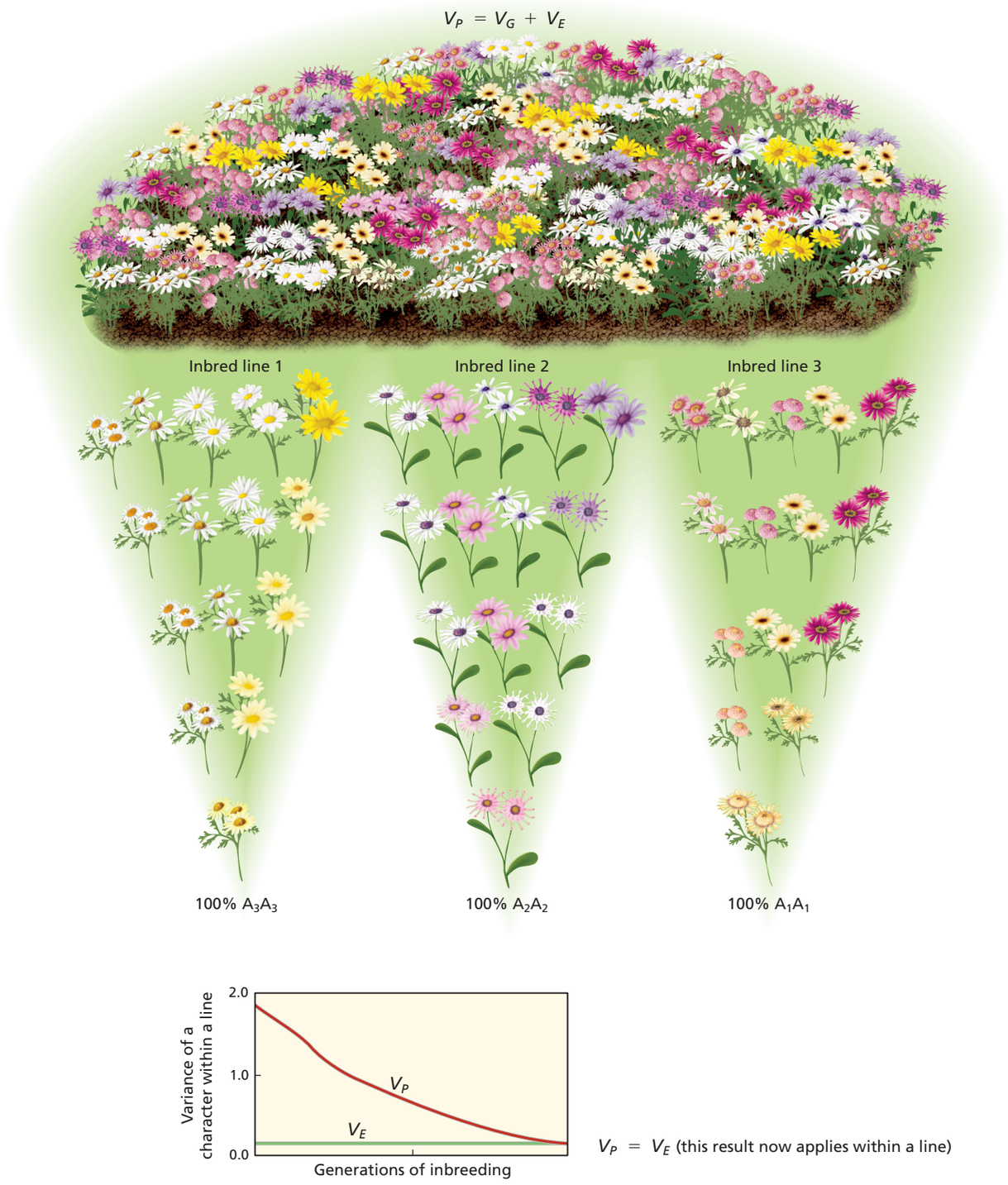


FIGURE 3.18B Uniformity within Inbred Lines Inbreeding bleeds the genetic variation out of individual inbred lines, so that the phenotypic variance (V_P) eventually equals the environmental variance (V_E). This is how breeders create lines that breed pure; examples of such inbred lines range from dog breeds to flower varieties.

3.19 Inbreeding tends to reduce the average value of beneficial characters

Inbreeding does more than reduce genetic variances. It also reduces the value of characters that are related to fitness or function, when populations begin as outbred. This reduction is known as **inbreeding depression**.

Figure 3.19A shows the impact of inbreeding on the dog pelvis. In highly inbred large breeds, like German shepherds and Saint Bernards, the joint connecting the femur to the pelvis is undermined. The head of the femur tends to become detached—a condition known as *hip dysplasia*. High levels of inbreeding in dogs also produce infertility, inappropriate aggression, blindness, and so on, as Figure 3.19B shows. The phenomenon is well known in a variety of agricultural animals and plants, and it can be produced in laboratory breeding experiments at will.

Inbreeding depression has an opposite. When unrelated inbred lines are crossed, their hybrids usually show considerable superiority. This is called **hybrid vigor**. For example, mongrels usually have fewer health problems than purebred dogs do. Most grains used in agriculture are hybrids of inbred lines. These grains produce higher yields, and they are more resistant to disease. Inbreeding depression and hybrid vigor are thus fundamental for agricultural breeding programs.

The puzzle is why inbreeding depression should be so common. One clue comes from organisms that normally inbreed in nature, such as self-fertilizing worms. These species do *not* show inbreeding depression or hybrid vigor. Inbred lines of such organisms show no decline in function or fertility, nor do crosses of their inbred lines always give hybrid vigor. Therefore, mere homozygosity does not cause inbreeding depression.

Wild populations of outbreeding organisms are fairly heterozygous. Inbred lines, on the other hand, are highly homozygous. Therefore, relative to inbreeding, outbreeding will tend to select for alleles that produce higher fitness when heterozygous. Long-standing inbreeding, by contrast, will select strictly on an

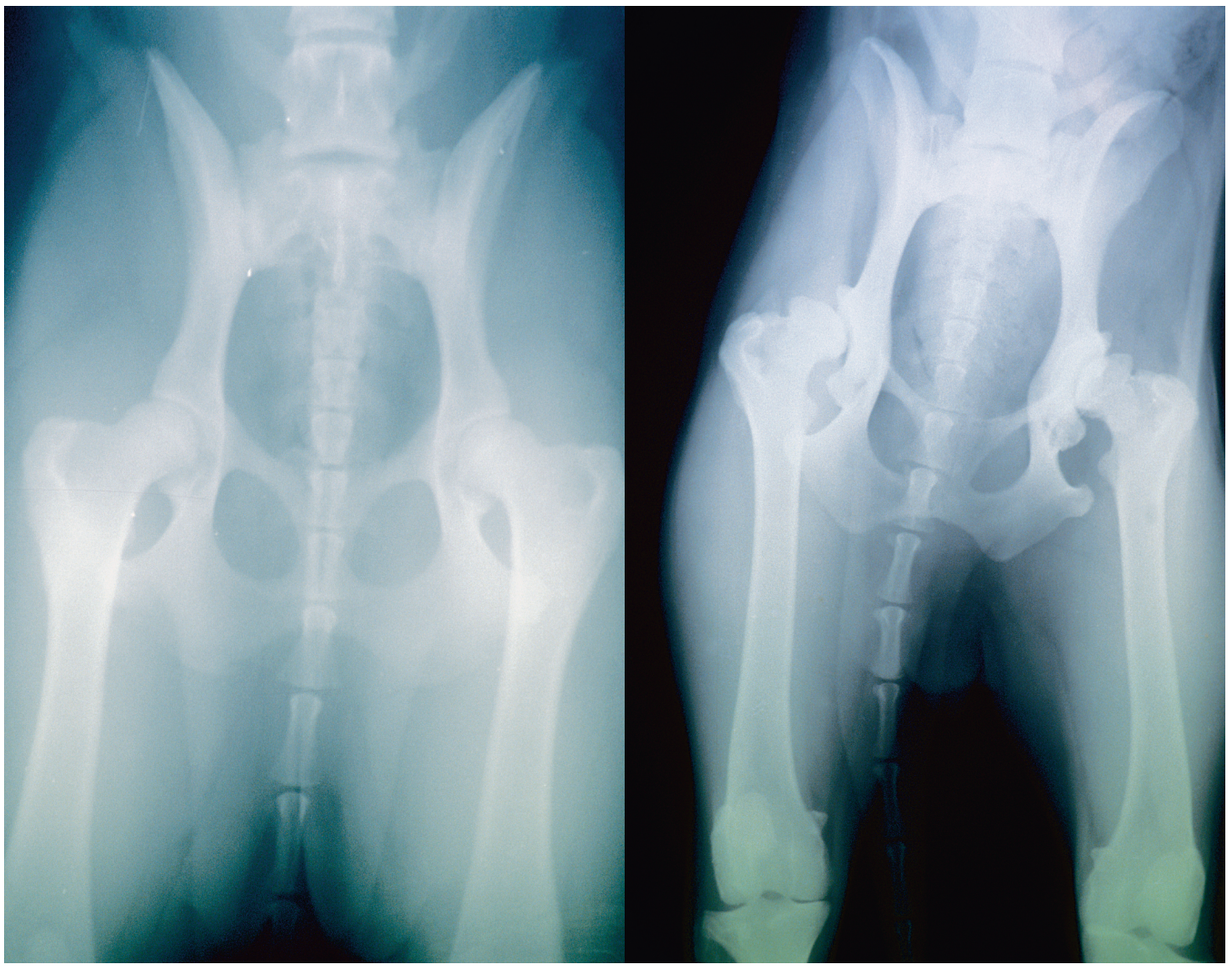


FIGURE 3.19A Hip Dysplasia in a Dog The x-ray of a normal dog is shown on the left. The dog with hip dysplasia is shown on the right.

allele's effect when homozygous, because that is the normal genetic situation with sustained inbreeding.

Consider a rare allele that is fairly benign when present in heterozygous form, but a disaster when there are two copies of the allele in the genome. Some human **genetic diseases** are thought to fit this pattern, one example being cystic fibrosis. Many people carry one copy of the gene for this genetic disease, but they show few, if any, bad effects. They are only carriers. Those with two copies of the gene suffer a debilitating and life-shortening disease, a disease that renders males almost totally infertile. And for evolution, infertility is a major problem. (Cystic fibrosis, and other examples of genetic disease, are discussed further in Chapter 4.) Inbreeding among carriers of the genetic disease cystic fibrosis would produce many more afflicted individuals than normally occur in the general population. But if humans were always inbred, this situation would not occur—because selection would then virtually eliminate the gene for cystic fibrosis.

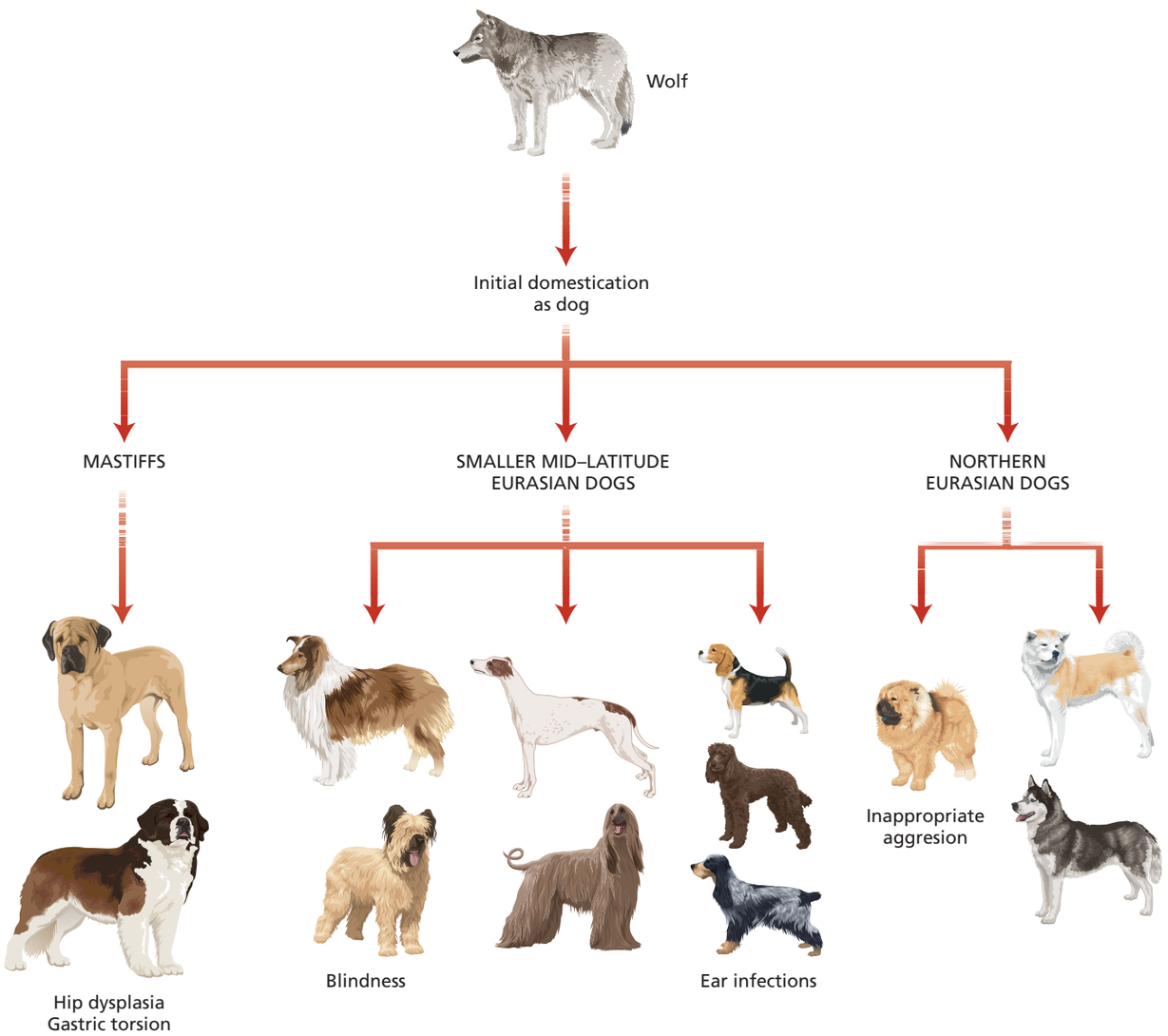


FIGURE 3.19B The History of Dog Breeding, with Annotations for Health Problems in Purebreds The history of dog breeds is in flux as of this writing, but that does not affect the health problems of inbred breeds. Mongrels are usually free of the health problems of purebred dogs.

3.20 Inbreeding can arise from the subdivision of populations

Inbreeding can be more subtle than the mating of relatives or the systematic inbreeding of agricultural animals. It can arise from the geographical distribution of a species. That is, who mates with whom can be made nonrandom by the mere accident of location.

Figure 3.20A illustrates this principle of **subdivision of populations**. In our hypothetical example, plants are living on the side of a large mountain. Soil suitable for the growth of this plant is present only at a few locations—one pasture high on the mountain, and one on each of the east and west mountain sides. If these plants have pollinators that do not disperse very far—such as small beetles—then the three plant populations may be largely isolated from each other. When this occurs, chance alone will cause genetic differentiation of these small populations, as we will discuss in the next module. There are four possible patterns of flowering—no flowers, pink flowers, yellow flowers, and blue flowers—determined genetically in part. The three populations differ in the genes that affect flowering, such that each population is a bit differentiated from the other. This differentiation gives rise to more variation in the species as a whole than would be expected from study of just one of the isolated populations. Conversely, there is less variation within each of these isolated populations than there is at the level of the species as a whole.

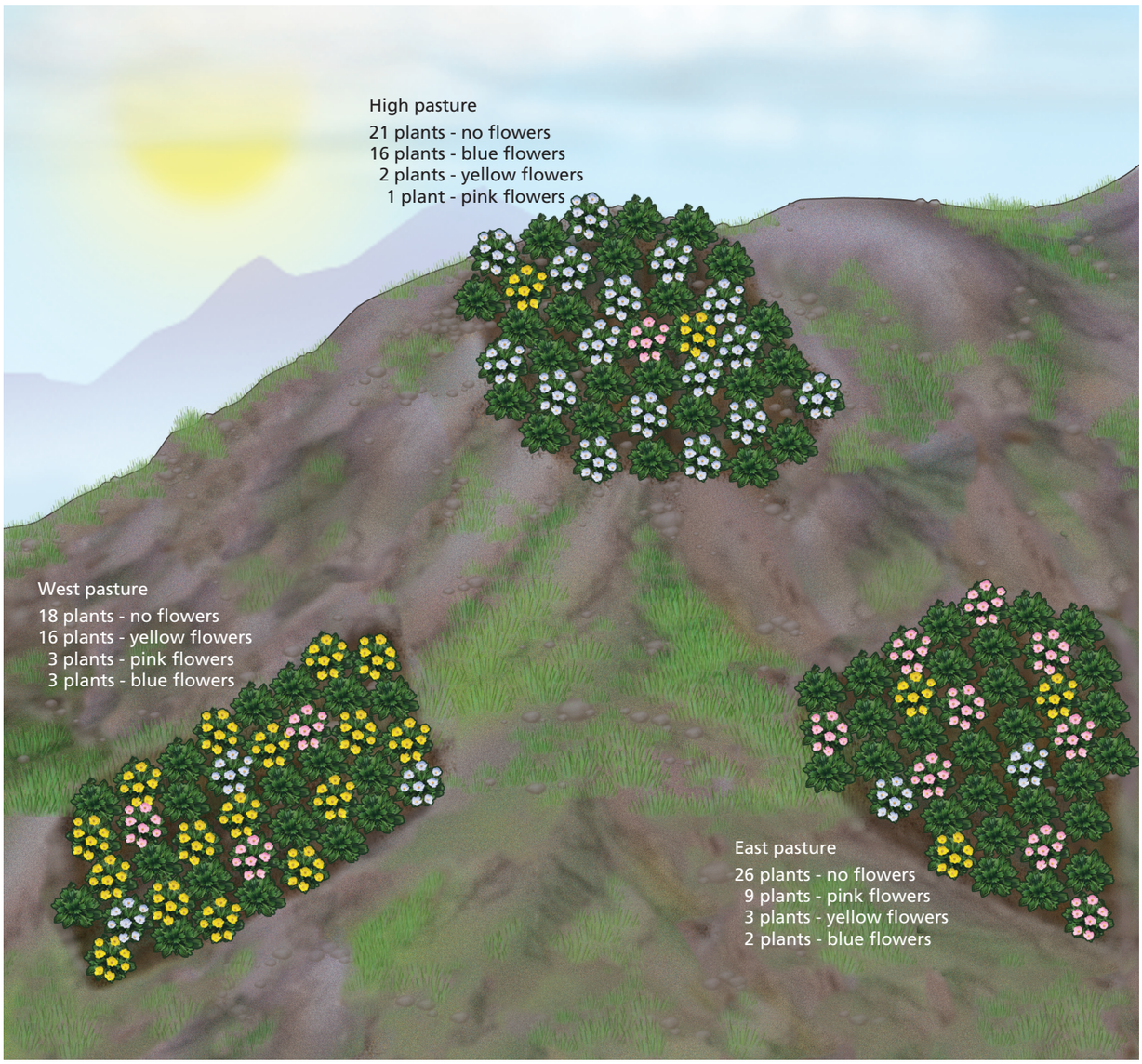


FIGURE 3.20A Some populations in nature often contain more of one genetic variant than do other populations of the species; genetic variation is in this case described as subdivided.

Species with very large populations that frequently intermate do not have this type of population structure. Seagulls and pine trees, for example, usually mate widely, because both marine birds and conifer pollen can travel great distances, breaking down local subdivision and differentiation of populations.

The impact of population subdivision is known as the **Wahlund effect**. Figure 3.20B gives a calculation showing the effect on heterozygosity of variance in allele frequency between subpopulations. The more variance in allele frequency, the greater the depression in the frequency of heterozygotes in the population as a whole. Conversely, the isolation of subpopulations increases the frequency of homozygous individuals.

$$H = 2pq - 2V_q$$

The frequency of heterozygotes

The normal frequency of heterozygotes, with no population subdivision

Twice the variance in the allele frequency over the subpopulations

FIGURE 3.20B **Wahlund Effect** The more subpopulations vary in allele frequency, the fewer heterozygotes will be present in the population as a whole.

GENETIC DRIFT

3.21 Genetics is like card games, and genetic drift is like a trip to Las Vegas

Think of **genetic drift** as the working out of chance on the next level up from inheritance in individual families. Mendelian genetics considers the inheritance of particular characters with known parents. It is very much like a game of chance, such as coin tossing or cards (Figure 3.21A). If we mate two heterozygotes, *Aa* and *Aa*, what are the chances that we will get an *AA* genotype with just one child? We know that the chances are $\frac{1}{4}$, because the chance that the gamete from one parent is *A* is $\frac{1}{2}$, and the same for the other parent. With independent production of gametes by the two parents, the probability of this happening twice is just the product $\frac{1}{2} \times \frac{1}{2}$, which equals $\frac{1}{4}$. This problem is just the same as the chance of getting two heads in a row when we toss a coin.

Just as the amount of money in your wallet or purse will rise and fall as a result of the many individual games that you play in Las Vegas, so does the frequency of individual alleles in the population rise or fall.

But at the next level up, in the genetics of whole populations, the situation is different. In the genetics of populations, the effects of the individual genetic processes combine. And there are as many of these processes as there are individuals producing gametes from which the next generation is made. Population genetics without selection is like going to Las Vegas for a weekend's gambling. (Figure 3.21B). You will be playing many rounds of poker or blackjack, putting quarters in slot machines, each play like the production of gametes by a single mated couple. The financial outcome for you, whether you will be richer or poorer, is a higher-level chance process. In this sense, we can describe genetic drift as a population's random production of gametes and zygotes to create the next generation, a higher-level process laid on top of the lower-level process of genetics itself.

Just as the amount of money in your wallet or purse will rise and fall as a result of the many individual games that you play in Las Vegas, so does the frequency of individual alleles in the population rise or fall. Both are essentially determined by "luck," which is to say by nothing in particular.

In small populations, allele frequencies change in discrete steps. To understand why, consider a very small population of just two individuals (Figure 3.21C). If we consider an autosomal locus with two alleles, *A* and *a*, there are five possible values for the frequency of the *A* allele: 0, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1. Two configurations will produce an allele frequency of $\frac{1}{2}$: The population may be composed of two heterozygotes, or it may have one *AA* homozygote and one *aa* homozygote. When these two individuals reproduce to create the next generation of two individuals, the frequency of the *A* allele may change to any

FIGURE 3.21A A Game of Cards in the Eighteenth Century

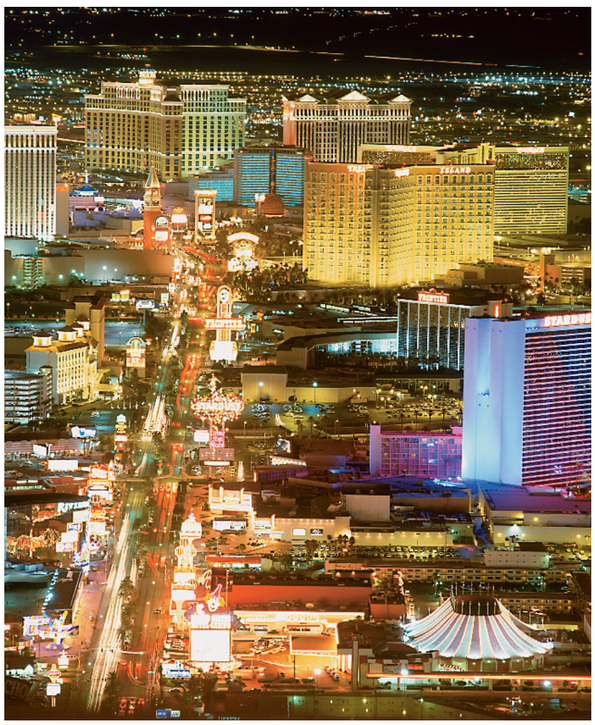


FIGURE 3.21B Las Vegas around the End of the Twentieth Century

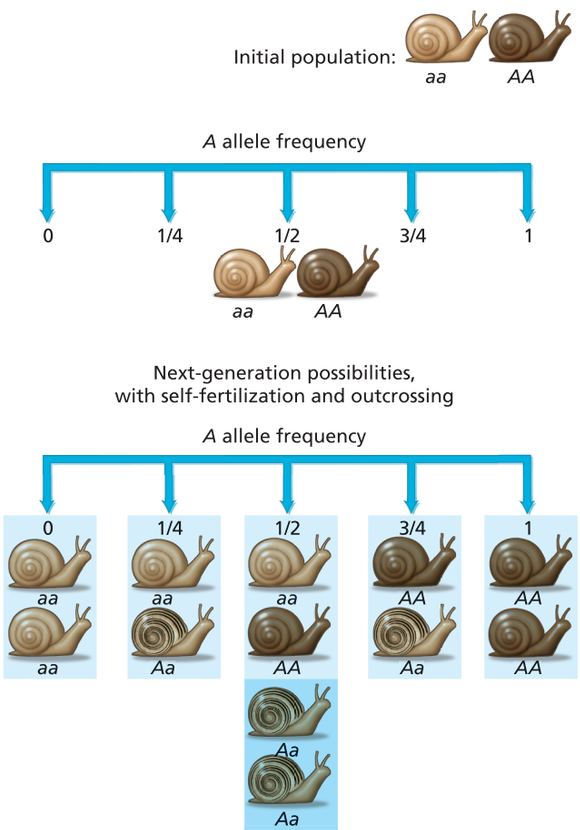


FIGURE 3.21C Genetic Drift with a Population Size of Two Six different population types can arise, including one that is the same as the initial population type: *aa, AA*.

one of these five values. Those of you paying close attention may not understand how a sexually reproducing species with one *AA* individual and one *aa* individual could produce two *AA* offspring, for instance. The answer is that our simple model permits self-fertilization, although it can be easily modified to account for sexual reproduction without self-fertilization.

The important point is that with random genetic drift, as long as both alleles are present, allele frequencies can change from their current values to any of the other possible values. As it turns out, the probabilities of changing to any one of these other values are not all equal, and they depend on the population size. In a population of size two at an allele frequency of 0.5, the chance of one allele frequency going to zero in the next generation is 1/16. However, in a population of 100 at the same allele frequency, the chance of going to zero in one generation is about 10^{-61}. ❖

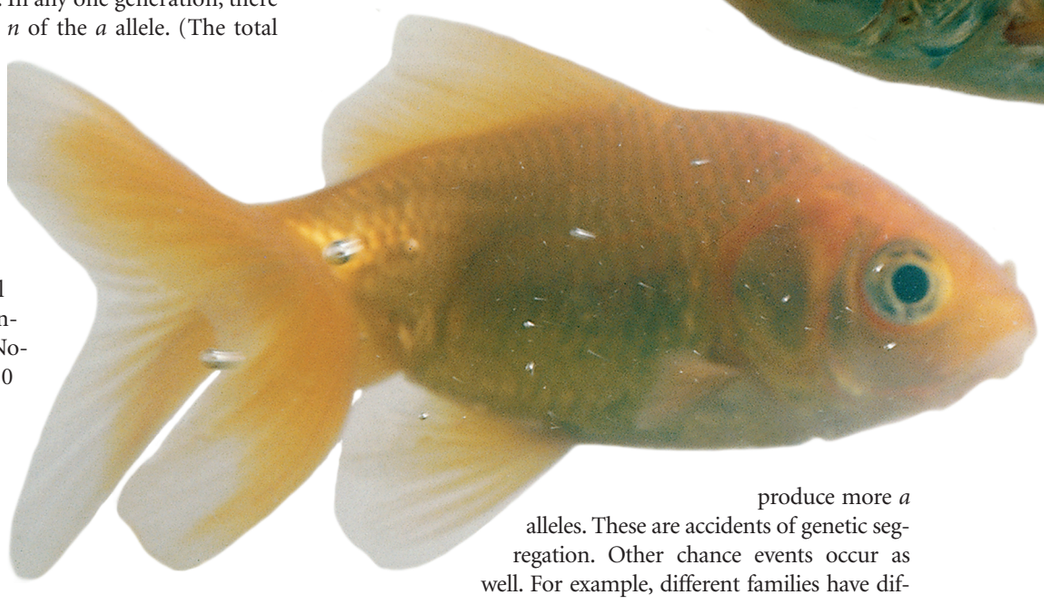
3.22 Populations can undergo evolutionary change from genetic drift alone

Let's take a closer look at genetic drift. In any one generation, some alleles will be lost due to accidents of Mendelian segregation during gamete production. A parent with both *a* and *A* alleles, a heterozygote, may have offspring that receive only the *a* allele. On average, this particular accident will tend to be canceled out by the opposite happening to another heterozygous parent, which has offspring receiving only *A* alleles.

Think of a population as a column of alleles, as shown in the first column in Figure 3.22A. In any one generation, there might be *N* of the *A* allele and *n* of the *a* allele. (The total number of individuals in a diploid population will be one-half of $N + n$.) In this case, we have 10 *A* alleles and six *a* alleles, and eight individuals. Figure 3.22B shows in detail what happens as the frequency of the *A* alleles increases and decreases, the total number of alleles holding constant, over two generations. (Notice that both figures start with 10 *A* alleles and six *a* alleles.)

Figure 3.22A provides a way of visualizing changes in allele frequency over several generations. Note that we group all of the *A* alleles with each other, and all of the *a* alleles with each other. This grouping enables us to represent the process of accidental sampling as an expansion or contraction of the *A* and *a* parts of the column.

Figure 3.22A follows the genetic drift of the population as a whole for generations 1 to 7. Various chance processes affect allele frequencies. Some heterozygotes produce more of allele *A*, and some produce more *a* alleles. These are accidents of genetic segregation. Other chance events occur as well. For example, different families have different levels of fertility. Some offspring do not survive to adulthood. All these chance events make allele frequencies fluctuate.



Intuitively, we can see that there will be more fluctuation in gene frequencies when populations are smaller. With many individuals producing gametes and many families, accidental biases in favor of one allele over the other will average out. For this reason, we expect genetic drift to be more rapid with inbreeding, and slow when population size is very large. Again, this is somewhat like what happens on a trip to Las Vegas. If you start with a large amount of money, you are not as likely end up broke. Your large cash reserve should allow you to survive strings of bad luck without going broke. (Please note, however, that we are not in the business of advising you how to gamble. We just offer an analogy.) Think of genetic drift in the same way you would think of being "ahead" or "behind" in a gambling situation, and you have its essential features. ❖

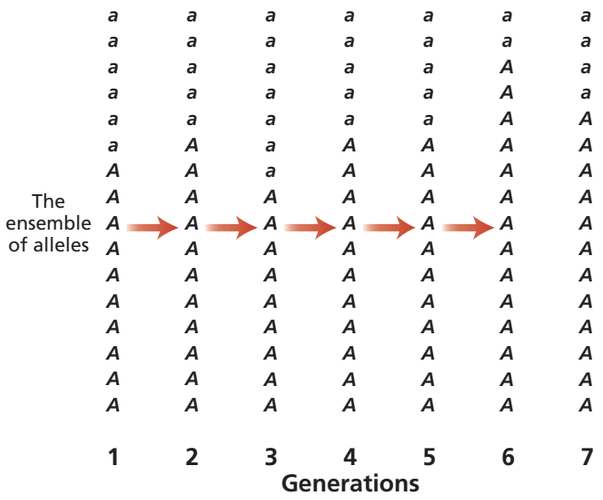


FIGURE 3.22A Genetic Drift in a Population of Eight Organisms with One Diploid Locus and Two Alleles, *a* and *A*

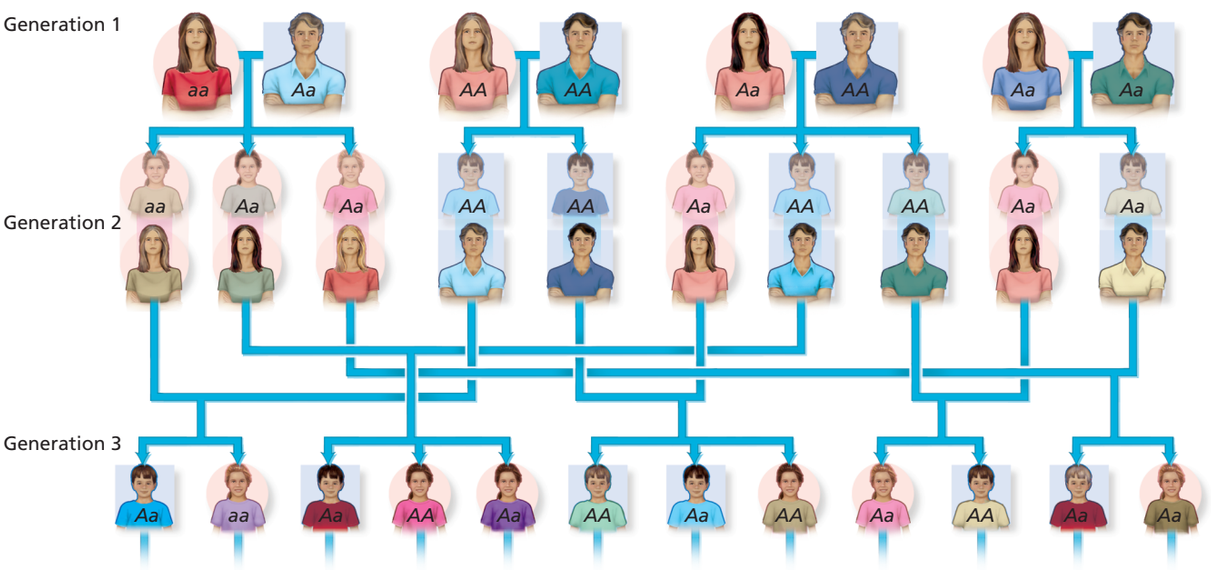


FIGURE 3.22B **An Example of How Accidents of Genetics and Reproduction Can Produce Genetic Drift**

3.23 Genetic drift can lead to the loss or fixation of alleles

Because genetic drift makes alleles bounce up and down in frequency, accidental loss of alleles is possible. A key factor is population size. Large populations—for practical purposes, those of more than 1000 breeding individuals—make the loss of alleles unlikely, for two reasons. First, genetic drift is very "slow" in such large populations, as mentioned earlier. Chance accidents will average out, so allele frequencies will not fluctuate much. Second, large populations receive more genetic mutations per generation, because they have large numbers of individuals having new offspring, which could carry new mutations. With frequent mutation, lost alleles will be regenerated by the mutational process alone. Thus genetic drift is expected to lead to rapid loss or fixation of alleles only when populations are relatively small—for practical purposes, fewer than 100 individuals.

When populations are small, genetic drift can cause the frequencies of alleles to wander erratically. Thus it is not surprising that some alleles might accidentally fall to very low frequencies, or even disappear altogether from the population. We can think of this process as analogous to the path of a drunk wandering down an unfenced pier in the dark. The drunk is highly likely to fall off one side of the pier or the other. In our analogy, when the drunk falls off one side, one allele is lost (for example, *A*), and when the drunk falls off the other side, the other allele is lost (for example, *a*). When one of two alleles is lost, the other becomes fixed, as shown in Figure 3.23A, where the loss of *a* would be equivalent to the fixation of *A*. (If there are more than two alleles, the geometry of the pier has to be more complicated, but the basic evolutionary event is still analogous to falling off an edge.) In very large populations, mutation acts as fencing along the sides of the pier; but small populations have too few mutations and so lack fencing.

What is the probability that an allele fixes? More common alleles fix more often. The probability that an allele fixes is simply equal to its initial frequency. For example, new mutations necessarily have a frequency of $1/2N$ in diploid populations of size N, because at first they are present in just one copy out of all $2N$ alleles at a locus. Therefore, they have a chance of fixing of only $1/2N$, compared to all other alleles in the population. This is a kind of "fairness" because, without selection, every allele in the population is equal. Which allele "wins" and becomes fixed by drift is then an accident, like picking a card from a shuffled deck of $2N$ cards. Because there are a total of $2N$ alleles in a population, they each have a chance of $1/2N$ of fixing, just like one card out of $2N$ cards has a chance of $1/2N$ of being picked at random. Figure 3.23B shows some possible gene-frequency trajectories with genetic drift.

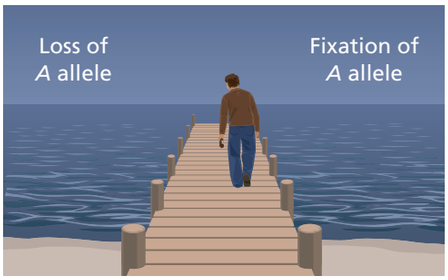


FIGURE 3.23A Genetic Drift Leads to Loss or Fixation of an Allele This process can be compared to the path of a drunk wandering down a pier at night. The probability that allele *A* will be fixed = p_i, its initial frequency. Initially common alleles are more likely to be fixed accidentally.

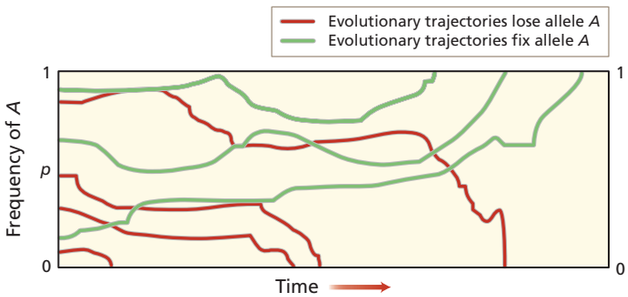


FIGURE 3.23B Multiple Examples of Genetic Drift through Time

SUMMARY

1. Genetics is perhaps the greatest achievement of twentieth-century biology. At its most fundamental level, the genetic engine transfers hereditary information from one generation to the next. That hereditary information is packaged into chromosomes, either prokaryotic or eukaryotic in configuration. Some gametes, cells, or organisms have one complete set of chromosomes (a haploid genome), while others have two sets (a diploid genome) or more.
2. The genetics of populations are determined by the frequencies of genotypes, and thus the frequencies of genes. With random mating for one generation, and no selection, the frequencies of the genotypes at a single locus can be calculated from the product of gene frequencies. The genotype frequencies are stable at this Hardy-Weinberg equilibrium. When the frequencies of gametes are the product of the frequencies of the alleles at multiple loci, there is linkage equilibrium.
3. Quantitative characters are affected by multiple genetic loci and multiple environmental influences. Sometimes these effects are additive; sometimes they are not. When inheritance is additive, the resemblance between relatives is determined by the genetic variance.
4. With random mating, the frequencies of genes on sex chromosomes evolve toward equality. In the absence of selection, random mating and genetic recombination lead to the evolution of linkage equilibrium.
5. Inbreeding occurs when relatives mate. If inbreeding occurs more than expected by chance, the level of homozygosity rises in the population, while the level of heterozygosity falls. Inbreeding leads to increased genetic disease. It also depresses functional characters, such as survival and fertility. Inbreeding increases when populations are subdivided.
6. Genetic drift results from the combination of individual genetic accidents. Genetic drift leads to the accidental fixation or loss of particular alleles by chance alone.

REVIEW QUESTIONS

1. The person who discovered genetics was . . . ?
2. Gametes are produced using what cellular process?
3. Hardy-Weinberg equilibrium arises after how many generations, with random mating and no selection?
4. Quantitative characters are influenced by which two major factors?
5. Evolution proceeds toward linkage equilibrium because of what process?
6. What are the effects of inbreeding on the genotypic composition of a population?
7. What does genetic drift do to gene frequencies?
8. Heritability measures what biological tendency?
9. Why do mutts live longer than purebred dogs?

KEY TERMS

addition rule
allele
arithmetic mean
blood group
chromosome
coupled phase
dominance
diploid
endosperm
epistastis
Fisher, R. A.
fixation
Galton, Francis
gamete
gametogenesis
gene expression
genetic disease
genetic drift
genotype
haploid
Hardy-Weinberg equilibrium
heritability
heterozygosity
homozygosity
hybrid vigor
inbred line
inbreeding
inbreeding coefficient, *F*
inbreeding depression
independent assortment
linkage disequilibrium
linkage equilibrium
loci
mean
median
meiosis
Mendel, Gregor
meroploidy
Morgan, Thomas Hunt
multiplication rule
outbred
Pearson, Karl
phenotype
pollinator
polyploid
polyteny
population
quantitative character
recessive
recessive genetic disease
recombination
repulsion phase
resemblance between relatives
self-fertilization, selfing
subdivision of populations
variance
variance, environmental
variance, genetic, additive genetic
variance, phenotypic
Wahlund effect
X chromosome
Y chromosome
zygote

FURTHER READINGS

Ayala, Francisco J., and John A. Kiger. 1980. *Modern Genetics*. Menlo Park, Calif.: Benjamin Cummings.

Falconer, D. S., and T.F.C. Mackay. 1996. *Introduction to Quantitative Genetics*, 4th ed. Harlow, Essex: Longman.

Hartl, Daniel L. 2000. *A Primer of Population Genetics*, 3rd ed. Sunderland, MA: Sinauer.

Hartl, Daniel L., and Elizabeth W. Jones. 1999. *Essential Genetics*, 2nd ed. Sudbury, MA: Jones and Bartlett.

Hedrick, Philip W. 2000. *Genetics of Populations*, 2nd ed. Boston: Jones and Bartlett.

Whitehouse, H.L.K. 1969. *Towards an Understanding of the Mechanism of Heredity*, 2nd ed. London: Edward Arnold.