13

Genetics

Introduction

In this chapter, we will study the ways that genetic information is passed between generations and how it is expressed. Cells can make exact copies of themselves through asexual reproduction. The genes such cells carry can be turned off and on to vary the cells' behaviors, but the basic information they contain can be changed only by mutation, a process that is somewhat rare to begin with and usually kills the cell anyway.

Genetic material is mixed in sexual reproduction, but the result of such mixing is seldom expressed as a "blend" of the properties' expressions. Rather, the rules for the combination of genetic information are somewhat complex. Sexual reproduction thus results in offspring that are different from the parents. Much research shows that the ultimate genetic source of this variation is mutation, but the most immediate source is the scrambling of preexisting mutations.

The variations produced by sexual reproduction serve as a basis for evolutionary selection, preserving the most desirable properties in a particular environmental context.

13.1 Asexual Cell Reproduction: Mitosis

Asexual reproduction of a cell results from the copying and equal distribution of the genetic material of a single cell. Each resultant daughter cell then possesses the same genes as the parent cell. If we are considering a single-celled organism, an environment for which the parent cell is suited should therefore also be suitable for the daughter cells. If we are considering a multicellular organism, the daughter cell may take on functions different from those of the parent cell by selectively turning genes off. This process creates the various tissues of a typical multicellular organism.

In this section, we will take a brief look at the mitosis cell division cycle, the process by which a cell reproduces an exact copy of itself. This is complementary to

R.W. Shonkwiler and J. Herod, *Mathematical Biology: An Introduction with Maple and Matlab*, Undergraduate Texts in Mathematics, DOI: 10.1007/978-0-387-70984-0_13,
© Springer Science + Business Media, LLC 2009

the more detailed scrutiny of the cell cycle undertaken in Chapter 12 *needed for the discussion of cancer.*

Eukaryotic mitosis gives each of two daughter cells the same genes that the parent cell had.

The actual process of eukaryotic mitosis is comparable to a movie, with sometimes complex actions flowing smoothly into one another, without breaks. For reference, however, mitosis is usually described in terms of five specific stages, named interphase, prophase, metaphase, anaphase, and telophase. It is important to remember, however, that a cell does not jump from one stage to the next. Rather, these stages are like "freeze-frames," or preserved instants; they are guideposts taken from the continuous action of mitosis (see [1] for further discussion).

Most of the time a cell's nucleus appears not to be active; this period is called *interphase*. If one adds to an interphase cell a stain that is preferentially taken up by nuclei and then examines the cell through a microscope, the nucleus appears to have no internal structure over long periods of time. This appearance is actually quite misleading, because, in fact, the nucleus is very active at this time. Its activity, however, is not reflected in changes in its outward appearance. For example, the addition of radioactive thymine to an interphase cell often leads to the formation of radioactive DNA. Clearly DNA synthesis takes place in interphase, but it does not change the appearance of the nucleus.

Biologists further subdivide interphase into three periods: G_1, during which preparations for DNA synthesis are made; S, during which DNA is synthesized; and G_2, during which preparations are made for actual cell division. (The "G" stands for "gap.") If we could see the DNA of a human skin cell during G_1 we would find 46 molecules. Each molecule, as usual, consists of two covalent polynucleotides, the two polymers being hydrogen-bonded to one another in a double helix. Genetic information is linearly encoded into the base sequence of these polynucleotides.

When we discussed DNA structure in Chapter 8, we associated a gene with the nucleic acid information necessary to code for one polypeptide. Thus a gene would be a string, not necessarily contiguous, of perhaps a few hundred to a few thousand bases within a DNA molecule. It is convenient to define a gene in another way, as a *functional unit of heredity*, a definition that has the virtue of generality. It can therefore include the DNA that codes for transfer RNA or ribosomal RNA, or it can just be a section of DNA that determines a particular observable property, such as wing shape or flower color. In this general definition, each DNA molecule is called a *chromosome*, where each genetic region, or *gene locus*, on the chromosome, determines a particular observable property.

In Figure 13.1.1, one chromosome is illustrated for a cell progressing through mitosis. (A human skin cell, for example, has 46 chromosomes, and each one behaves like the one in the figure.) The structure of the chromosome at G_2 cannot be seen in a microscope, so we must surmise its structure by its appearance in the next stage (prophase).

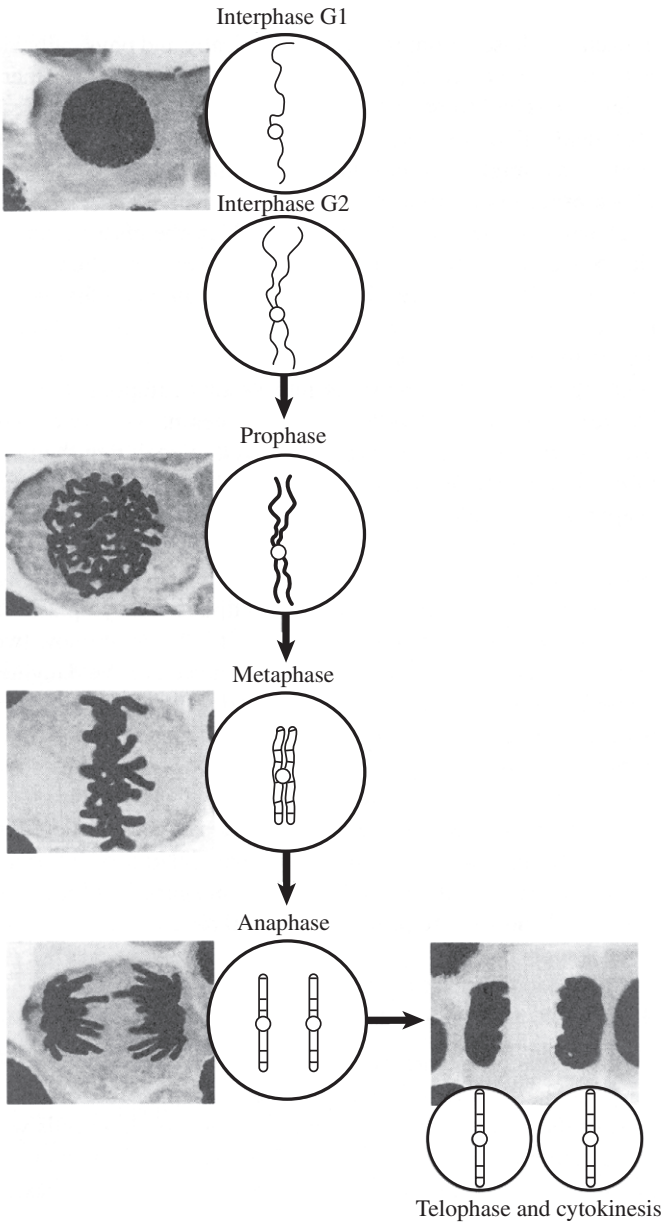


Fig. 13.1.1. The stages of mitosis. The figure shows actual photographs of a dividing cell's chromosomes. The line drawings show how the individual chromosomes are behaving during that stage of division. During mitosis, each chromosome replicates lengthwise and the two copies go to different daughter cells. Thus each daughter cell ends up with exactly the same genetic complement as the parent cell. (Photos of mitosis taken from *Radiation and Chromosomes Biokit*, item F6-17-1148, Carolina Biological Supply Company, Burlington, NC. Used with permission.)

At *prophase*, the nuclear membrane disappears and the chromosomes become visible for the first time, resembling a ball of spaghetti. If we could grab a loose end of a chromosome and separate it from the others, we would see that it looks like that shown in the figure beside the prophase cell. It consists of two halves, called sister *chromatids*, lying side by side and joined at a *centromere*. The two chromatids of each prophase chromosome are chemically and physically identical to each other because one of each pair was manufactured from the other in the preceding S phase. Each chromatid therefore contains a double-stranded DNA molecule that is identical to the DNA of its sister chromatid. The two chromatids are still referred to as a single chromosome at this point.

As prophase progresses, the chromosome becomes shorter and fatter, and it moves to the center of the cell. The stage at which the chromosomes reach maximum thickness and are all oriented at the cell's center is called *metaphase*. Chromosomes at metaphase have reached their maximum visibility, and a view through a microscope often shows them all arranged neatly in the cell's equatorial plane, as shown in the photo in Figure 13.1.1.

At *anaphase*, each chromosome splits into its two component chromatids, which are now referred to as individual chromosomes in their own right, and one copy moves toward each end, or *pole*, of the cell. Recall that the two sister chromatids of each chromosome are identical to each other. In summary, what happens in anaphase is that identical double-stranded DNA is delivered to each pole.

At *telophase*, the chromosomes collect together at each pole and a new nuclear membrane forms around them. The cell then divides its cytoplasm in such a way that one new nucleus is contained in each half.[1] There are now two cells where there was only one, but the crucial point is that each of the daughter cells now has the same DNA code that the original cell had. Put another way, two cells have been formed, each having the same genes as the parent cell.

One way to look at asexual reproduction is to think of each chromosome as a piece of paper, with information written on it. A human skin cell has 46 pages, labeled 1–46. At S phase, an exact copy is made of each page, and during mitosis each daughter cell gets one copy of each page. No new information is created, nor is any lost. Each daughter cell gets the same genetic information, i.e., each daughter cell ends up with 46 pages, labeled 1–46.

A karyotype is a picture of a cell's chromosomes.

It is not difficult to obtain a picture of most organisms' chromosomes. For example, it is a routine laboratory procedure to take a sample of a person's blood and isolate some of their white blood cells. (Mammalian red blood cells won't do because they lose their nuclei as they mature.) These white cells are then cultured in a test tube and their nuclear material is stained as they enter metaphase, which is when chromosomes are most easily visualized. The cell, with its chromosomes, is photographed through a microscope. The chromosomes are then cut out of the photograph and arranged

[1] The actual splitting of the cell is called *cytokinesis*.

in a row, according to size. This picture is a *karyotype*. An example is shown in Figure 13.1.2.

There are several interesting features of the illustrated karyotype:

1. These are metaphase chromosomes and therefore are lengthwise doubled, joined at a centromere. Each chromosome consists of two chromatids, a feature that sometimes confuses students. The problem is that the chromosomes must be photographed at metaphase because that is when they are most easily visible and distinguishable from one another. This is also the point at which they are in a duplex form. You may want to refer back to the discussion of Figure 13.1.1 to clarify the distinction between chromosome and chromatid.

2. There are 46 chromosomes in this cell. This is the number found in most of the cells of the human body, the exceptions being mature red blood cells, which lack nuclei, and certain cells of the reproductive system, called *germinal cells*, to be discussed later in this chapter. Any cells of our body that are not germinal are said to be *somatic* cells, a category that therefore includes virtually the entire bulk of our body: skin, blood, nervous system, muscles, the structural part of the reproductive system, etc. Our somatic cell chromosome number is thus 46.

3. The chromosomes in the karyotype seem to occur in identical-appearing pairs, called *homologous pairs*. Evidently, our human chromosomal complement is actually two sets of 23 chromosomes. It is very important to understand the difference between a homologous pair of chromosomes and the two chromatids of a single metaphase chromosome. The karyotype shows 23 homologous pairs; each member of each pair consists of two chromatids. Each chromatid contains a double-helical DNA molecule that is identical to the DNA of its sister chromatid, but which is different from the DNA of any other chromatid.

Asexual reproduction can generate daughter cells that differ from each other.

We could imagine an amoeba, a common single-celled eukaryote, dividing by mitosis to yield two identical amoebas. We could just as easily imagine a skin cell of a human, a multicellular eukaryote, dividing by mitosis to give two identical human skin cells. Indeed, this is the way that our skin normally replaces those cells that die or are rubbed off. In both cases, the daughter cells have the same DNA base sequence that the parent cell had, and that is reflected in the identical physiology and appearance of the daughter cells.

There is another possibility: consider a single fertilized human egg. It divides by mitosis repeatedly to form a multicellular human, but the cells of a developed human are of many sizes, shapes, and physiological behaviors. Liver cells look and behave one way, nerve cells another, and muscle cells still another. Mitosis seems not to have been conserved. How could cells that have exactly replicated their DNA in mitosis and then partitioned it out equally have yielded different progeny cells?

One possibility is that cells in each unique kind of tissue of a multicellular organism have lost all their genes except those essential to the proper functioning of that particular tissue. Thus liver cells would have retained only those genes needed for

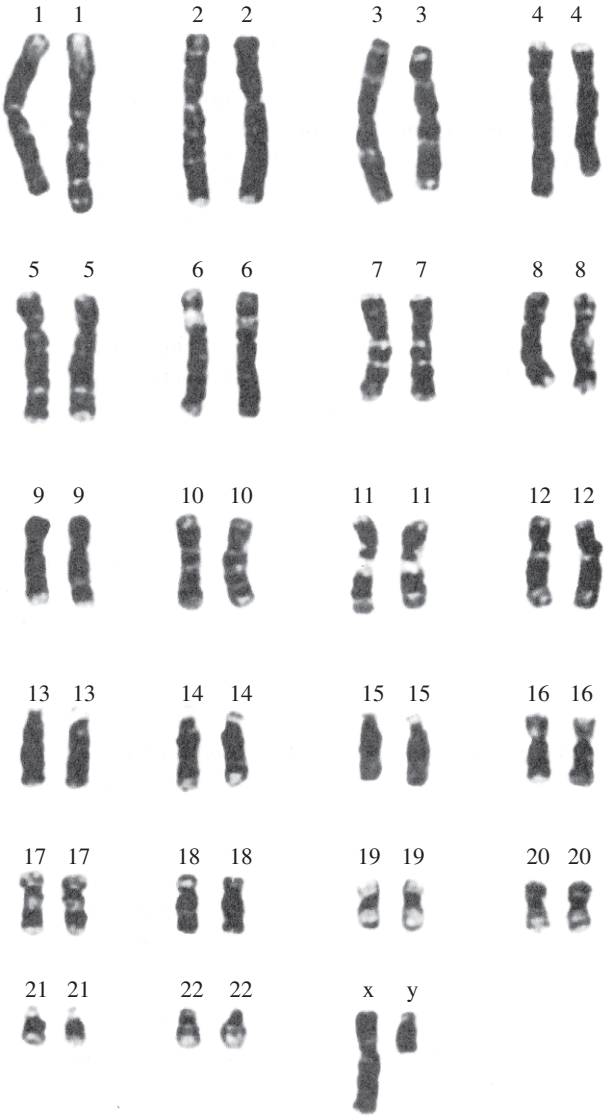


Fig. 13.1.2. A karyotype of a normal human male. The chromosomes were photographed at metaphase and images of the individual chromosomes were then cut out and arranged by size. The result is a group of 22 chromosome pairs, called homologs, each pair of which is matched by length, centromere location, and staining pattern. Because this is a male's karyotype, the 23rd pair of chromosomes (sex chromosomes; X and Y) do not match each other. Each of the chromosomes shown in the figure consists of two identical daughter chromatids, but they are so closely associated that they are often indistinguishable at metaphase. However, note the right-hand homolog of number 18; the two chromatids can be distinguished. (Photo of karyotype arranged from *Human Karyotypes, Normal Male*, item F6-17-3832, Carolina Biological Supply Company; Burlington, NC. Used with permission.)

liver functioning and muscle cells would have retained only those genes needed for muscle functioning. This possibility is easy to reject by a simple experiment: In the cells of a plant stem the genes necessary for stem growth and function are obviously active, and there is no evidence of genes involving root formation. If the stem is broken off and the broken end inserted into soil, within a few weeks the plant will often start to grow roots at the broken stem end. Clearly the genes for root growth and function were in the cells of the stem all along, but were reversibly turned off. A similar experiment has been done on a vertebrate, in which a nucleus from a specialized somatic tissue, the intestinal lining of a tadpole, has been used to grow a whole tadpole and the subsequent toad. We can conclude that mitosis generates different tissues of multicellular organisms when selected genes are turned off or on in the course of, or in spite of, asexual cell division.

The process by which unspecialized cells of a multicellular organism take up specialized roles—liver, nerve, skin, etc.—is called *differentiation*. Differentiation is not restricted to embryos, but can occur all our lives, e.g., in bone marrow, where unspecialized stem cells can become specialized blood cells. Differentiation is only one part of *development*, which includes all the changes in an organism in its life, from conception to death. Other aspects of development include tissue growth and deterioration, as described in Section 9.2.

Some cell types rarely divide.

Certain cells of multicellular organisms seem to have a very limited, even nonexistent, capacity for division. For example, muscle cells don't divide; the muscle enlargement associated with exercise comes from cellular enlargement. Fat cells get larger or smaller, but their numbers stay the same (which is why cosmetic liposuction works—the lost fat cells can't be replaced). Cells of the central nervous system don't divide, which explains the seriousness of spinal injuries. Liver cells rarely divide unless part of the liver is cut away—in which case the liver cells undergo division to replace those removed. Note the implication here: Genes controlling liver cell division haven't been lost. They were shut off, and can be reactivated.

13.2 Sexual Reproduction: Meiosis and Fertilization

Sexual reproduction involves the creation of an offspring that contains genetic contributions from two parents. A type of cell division called meiosis halves the chromosome number of germinal cells to produce sperms or eggs. A sperm and an egg then combine in fertilization to restore the double chromosome number. The new offspring now has genetic information from two sources for every characteristic. The ways that these two sets of information combine to produce a single property are complex, and this is the subject of the study of classical genetics.

Sexual reproduction provides variation upon which evolutionary selection can act.

Recall the Darwinian model: More organisms are born than can survive, and they

exhibit variability. Those with favored characteristics survive and may pass the favored properties to their offspring. It is tempting to credit genetic mutation with this variability and let it go at that. The fact is that all of the ten (nontwin) children in a hypothetical large family look different and virtually *none* of the variations among them are the result of mutations in their, or their parents', generation. This fact, surprising at first, seems more reasonable when we consider the accuracy of DNA base pairing, the "proofreading" capability of some kinds of DNA polymerase and the existence of repair mechanisms to correct DNA damaged by such mutagens as radiation. Thus DNA sequences tend to be conserved over many generations. We can therefore conclude that most of the variations among the ten children of the same family are the result of scrambling of existing genes, not the result of recent mutation. The cause of this shuffling of the genetic cards is sexual reproduction. Of course, the variant genes *originated* through mutation, but virtually all of them originated many generations earlier (see [2] for further discussion).

Sexual reproduction involves the combination of genetic material from two parents into one offspring.

Refer back to the karyotype in Figure 13.1.2. The human chromosome complement consists of 23 homologous pairs or, put another way, of two sets of 23 each. The sources of the two sets of 23 can be stated simply: we get one set from each of our parents when a sperm fertilizes an egg. What is not so simple is how the genetic material in those 46 chromosomes combines to make each of us what we are. The rules for combination will be the subject of Section 13.3. Our more immediate concern, however, is the means by which we generate cells with 23 chromosomes from cells having 46.

Meiosis halves the chromosome number of cells.

A special kind of reductional cell division, called *meiosis*, creates *gametes* having half the number of chromosomes found in somatic cells.[2]

The chromosomes are not partitioned at random however; rather, every gamete winds up *with exactly one random representative of each homologous pair*, giving it one basic set of 23 chromosomes. Such a cell is said to be *haploid*. A cell that has two basic sets of chromosomes is said to be *diploid*. We see that somatic cells are diploid and germinal cells are haploid. Thus meiosis in humans converts diploid cells, with a chromosome number of 46, to haploid cells with a chromosome number of 23.

Meiosis is diagramed in Figure 13.2.1 for a hypothetical organism having two homologous pairs; its diploid number is 4. Each chromosome is replicated in interphase and thus contains two identical chromatids joined at a centromere. In a departure from meiosis, homologs bind together, side by side, in a process called *synapsis*, to form *tetrads* consisting of two chromosomes (four chromatids). The homologs then separate to end the first meiotic division. Next, the chromatids separate to complete

[2] Gametes are often called *germinal cells* to distinguish them from somatic cells.

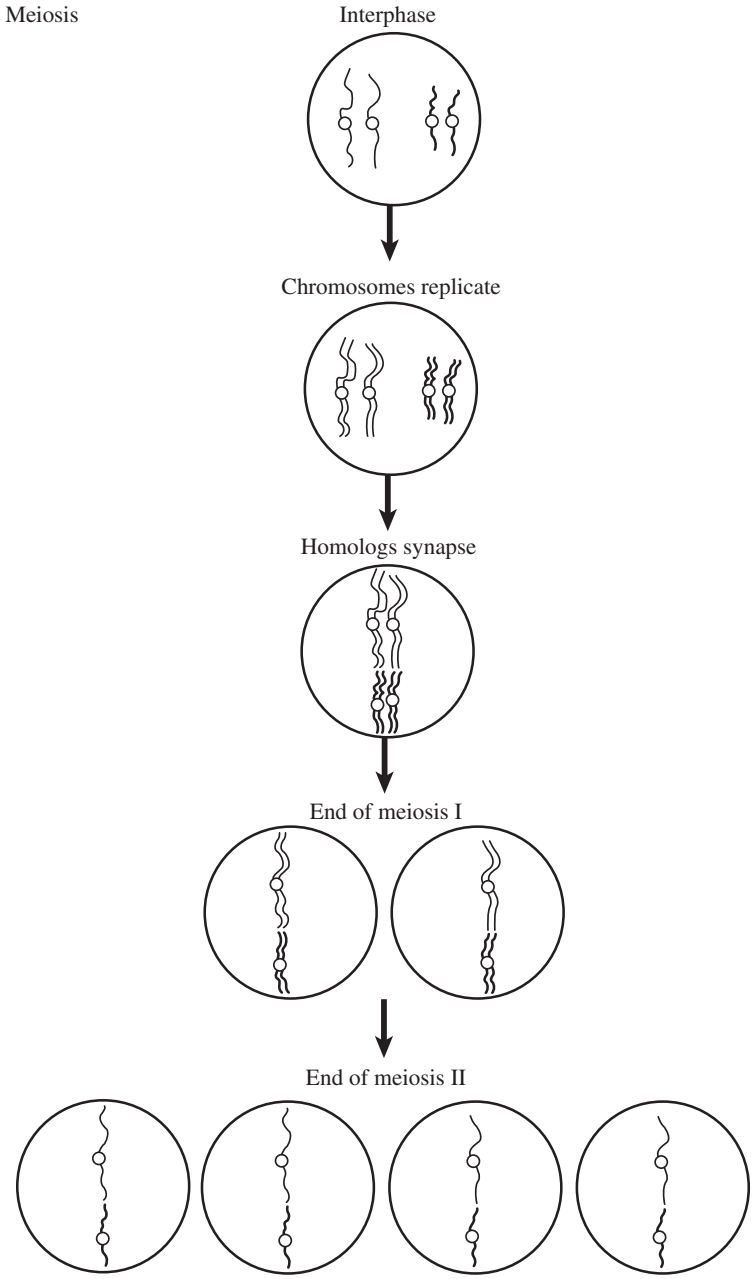


Fig. 13.2.1. The stages of meiosis. The cell shown has two homologous pairs. Each chromosome replicates lengthwise to form two chromatids, synapses to its homolog, and then two cell divisions ensue. The daughter cells each end up with exactly one representative of each homologous pair. Thus a diploid cell at the start of meiosis results in four haploid cells at the end of meiosis.

the second meiotic division. The result is four cells, each containing two chromosomes, the haploid number for this hypothetical organism. Note that the *gametes' chromosomes include exactly one representative of each homologous pair*.

The process of meiosis (perhaps followed by developmental maturation of the haploid cell) is called *gametogenesis*. Specifically in animals, the formation of male gametes is called *spermatogenesis*, and it yields four sperms, all similar in appearance. The formation of female gametes is called *oogenesis* and yields four cells, but three of them contain almost no cytoplasm. The latter three are called *polar bodies*, and they die. Thus oogenesis actually produces only one living egg, and that one contains all the cytoplasm of the diploid precursor. The reason for this asymmetry is that once the egg is fertilized, the first several cell divisions of the fertilized egg (called a *zygote*) remain under the control of cytoplasmic factors from the mother. Evidently, all the cytoplasm from the egg precursor is needed in a single egg for this process.

The concept of sexual reproduction can be incorporated into the alternation of generations.

We can diagram the alternation of the diploid and haploid generations:

$$\cdots \longrightarrow \text{diploid} \overset{\text{meiosis}}{\longrightarrow} \text{haploid} \overset{\text{fertilization}}{\longrightarrow} \text{diploid} \longrightarrow \cdots .$$

Note that the diploid and haploid generations are equally important because they form a continuous string of generations. On the other hand, the two generations are not equally *conspicuous*. In humans, for instance, the haploid generation (egg or sperm) is microscopic and has a lifetime of hours to days. In other organisms, mainly primitive ones like mushrooms and certain algae, the haploid generation is the conspicuous one, and the diploid generation is very tiny and short lived.

Another way to show the alternation of generations is in the diagram in Figure 13.2.2.

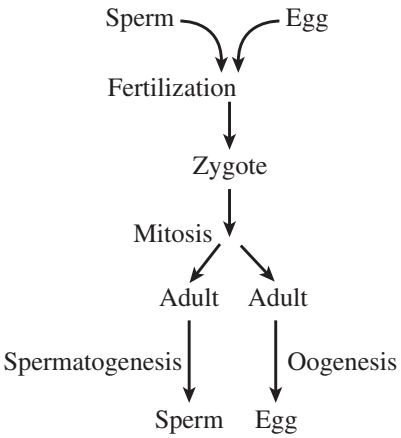


Fig. 13.2.2.

13.3 Classical Genetics

Classical genetics describes the many ways that the genetic material of two parents combines to produce a single observable property. For instance, a red-flowered plant and a white-flowered plant usually produce an offspring with a single color in its flower. What that color will be is not predictable unless a geneticist has already studied flower colors in that plant—because there are about a dozen ways that parental genes can combine. We describe many of those ways in this section.

Classical genetics describes the result of interactions in genetic information.

A diploid human cell carries 23 homologous pairs of chromosomes: One member of each pair comes from a sperm cell of the male parent, and the other member comes from an egg cell of the female parent. Other diploid organisms may have chromosome numbers ranging from a few up to hundreds, but the same principle about the origin of homologous pairs holds. What we will consider now is how the genetic information from the two parents combines to produce the characteristics that appear in the offspring and why the latter are so variable. Let us first examine a chromosome at G_1 phase, because that is the usual condition in a cell.

Genes, defined generally as functional units of heredity, are arranged linearly along the chromosome (Figure 13.3.1). Each gene locus affects some property, say, flower color or leaf shape in a plant. The order in which these loci appear is the

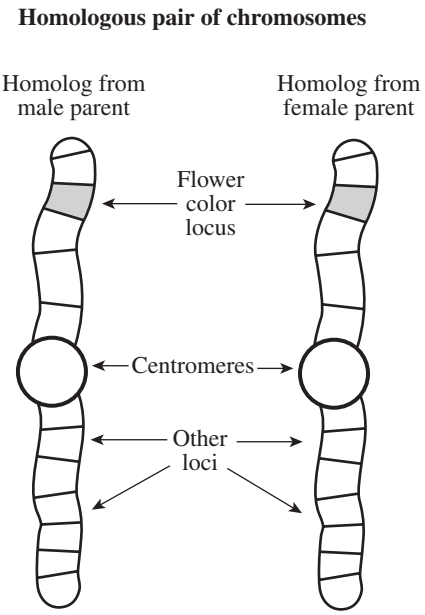


Fig. 13.3.1. This shows a simple model of the chromosome. The genes are lined up along the length of the chromosome, like beads on a string. A hypothetical flower color locus is labeled.

same on each member of the homologous pair. Thus it is common to refer to the "flower color" locus, meaning the section of either member of a homologous pair that is the gene that determines flower color. Clearly, each property is determined by two such sections, one on each homolog. *Each parent, then, contributes to each genetic property in the offspring*.

The behavior of chromosomes provides a basis for the study of genetics.

The pioneering geneticist was Gregor Mendel, who studied the genetics of peas, a common flowering plant. Peas, like many flowering plants, have male and female reproductive structures in the same flower. The male part makes pollen that is carried to the female part of that or another plant; the pollen then produces a sperm cell and fertilizes an egg. It a straightforward matter to dissect out the male part of a flower to prevent the plant from self-pollinating. Further, it is simple to use pollen from the male part of one plant to fertilize an egg of another plant and thus to make controlled matings. The seed that results from fertilizing an egg can be planted and the appearance of the offspring studied. The principles of chromosomal behavior and gene interaction in peas are the same as for humans.

Mendel had two groups, or populations, of plants that were *true breeding*. A population is true breeding if its freely interbreeding members always give rise to progeny that are identical to the parents, generation after generation. Members of a population of true-breeding red-flowered peas fertilize themselves or other members of the population for many generations, but only red-flowered plants ever appear. Mendel made a cross between a plant from a true-breeding red-flowered population and one from a true-breeding white-flowered population.

Mendel did not know about chromosomes, but we do and we will make use of that knowledge, which will simplify our learning task in the discussion to follow. We will therefore represent the cross in the following way: The gene for flower color is indicated by the labeled arrow in Figure 13.3.1. Note that each of the two homologs has such a gene locus.[3] The genetic information for red flower color is symbolized by the letter R, and the plant has two copies, one from each parent. (The reason for the copies being alike will become clear shortly.) Using the same convention, the genetic information at the flower color locus of the two homologs in the other (white-flowered) parent is symbolized by w.

Meiosis produces gametes containing one, and only one, representative of each homologous pair, as shown in Figure 13.3.2. A gamete from each parent combines at fertilization to reestablish the diploid condition. The offspring has flower color genetic information Rw. It turns out that this pea plant produces only red flowers, indistinguishable from the red parent. Evidently red somehow masks white; we say that red information is *dominant* to white, and white is *recessive* to red.

At this point, we need to define several terms. The variant forms of information for one property, symbolized by R and w, are *alleles*, in this case flower color alleles. The allelic composition is the organism's genotype; RR and ww are *homozygous*

[3] For learning purposes, we will ignore all other chromosomes, as if they do not have loci that affect flower color. In actual fact, this may not be true.

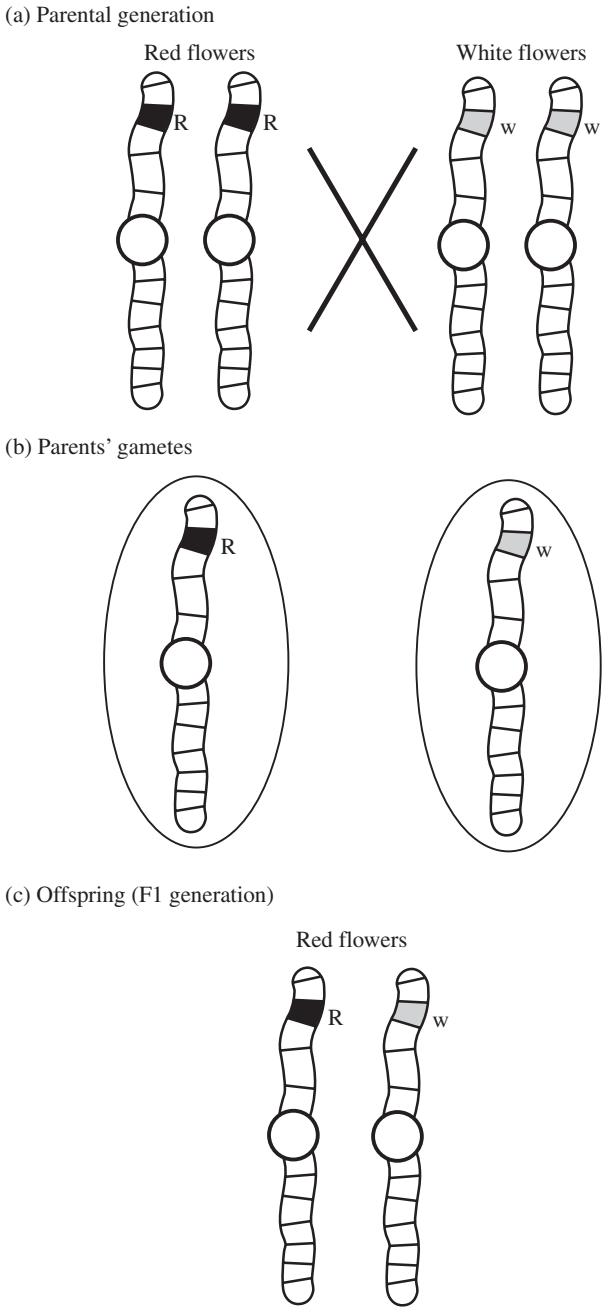


Fig. 13.3.2. The behavior of chromosomes and their individual loci during a cross between two homozygous parents. The parents (RR and ww) each contribute one chromosome from the homologous pair to form gametes. The gametes combine in fertilization to restore the diploid number of two. The offspring's flowers will be red.

genotypes and Rw is the *heterozygous* genotype. What the organism actually looks like, red or white, is its *phenotype*. Thus the initial, or parental, cross, was between a homozygous red-flowered plant and a homozygous white-flowered plant. The result in the first *filial*, or F1, generation was all heterozygous, red-flowered plants.

To obtain the F2 generation, we self-cross the F1, which is equivalent to crossing it with one just like itself. Figure 13.3.3 shows the gametes obtained from each parent in the F1 generation. They combine in all possible ways at fertilization. The result is a ratio of 1 RR, 2 Rw, and 1 ww, which gives a 3:1 ratio of red-to-white phenotypes.

An experiment of the sort just described, involving a single property like flower color, is called a *monohybrid cross*. We used the chromosome model, whereas Mendel actually ran the experiment; satisfyingly, both give the same results. Let us now make a *dihybrid cross*, involving the two properties of flower color and stem length, which we specify to be *unlinked*, which means that their genetic loci are on different homologous pairs. The cross is diagramed in Figure 13.3.4. Note that we have quit drawing in the chromosomes—we understand that the genes are on chromosomes and that drawing the latter is redundant. The F1 self-cross now can be represented as RwLs $\times$ RwLs. Note the phenomenon of *independent assortment*: Each gamete gets one and only one representative of each homologous pair, and the behavior of one pair in meiosis is independent of the behavior of the other pair. Thus meiosis in the F1 generation results in equal numbers of gametes containing RL, Rs, wL, and ws. The outcome of the cross is shown in the array, called a *Punnett square*, at the bottom of the figure.

The dihybrid cross yields a 9:3:3:1 phenotypic ratio of offspring. We should ask whether the inclusion of stem length in any way interferes with the 3:1 ratio of flower color. Among the 16 offspring in the Punnett square, we see 12 red and 3 white, which gives the 3:1 ratio. We might have anticipated this—that the two properties would not affect their separate ratios—after all, they are unlinked and the two homologous pairs assort independently.

We must obtain large numbers of progeny in order to get the expected ratios of offspring.

Suppose we make a cross like Rw $\times$ Rw in peas (red $\times$ red) and get only four progeny. We should not expect an exact 3:1 ratio of phenotypes in this experiment. After all, if we flipped a coin two times, we would not be certain to get one head and one tail. Rather, we expect to get approximately the 1:1 ratio only if we flip the coin many times, say 2000. The same reasoning holds in genetics—we must make enough Rw $\times$ Rw crosses to get many offspring, say 4000, and *then* we would obtain very close to 3000 red and 1000 white offspring.

The ratios 3:1 and 9:3:3:1 are often called *Mendelian ratios*, because they are what Mendel reported. There is a bit of a problem here: Statisticians have examined Mendel's data and some have concluded that the experimental data are too good, i.e., consistently too close to the 3:1 and 9:3:3:1 expected ratios. For the sample sizes Mendel reported, it would be expected that he would have gotten somewhat larger deviations from "Mendelian" ratios.

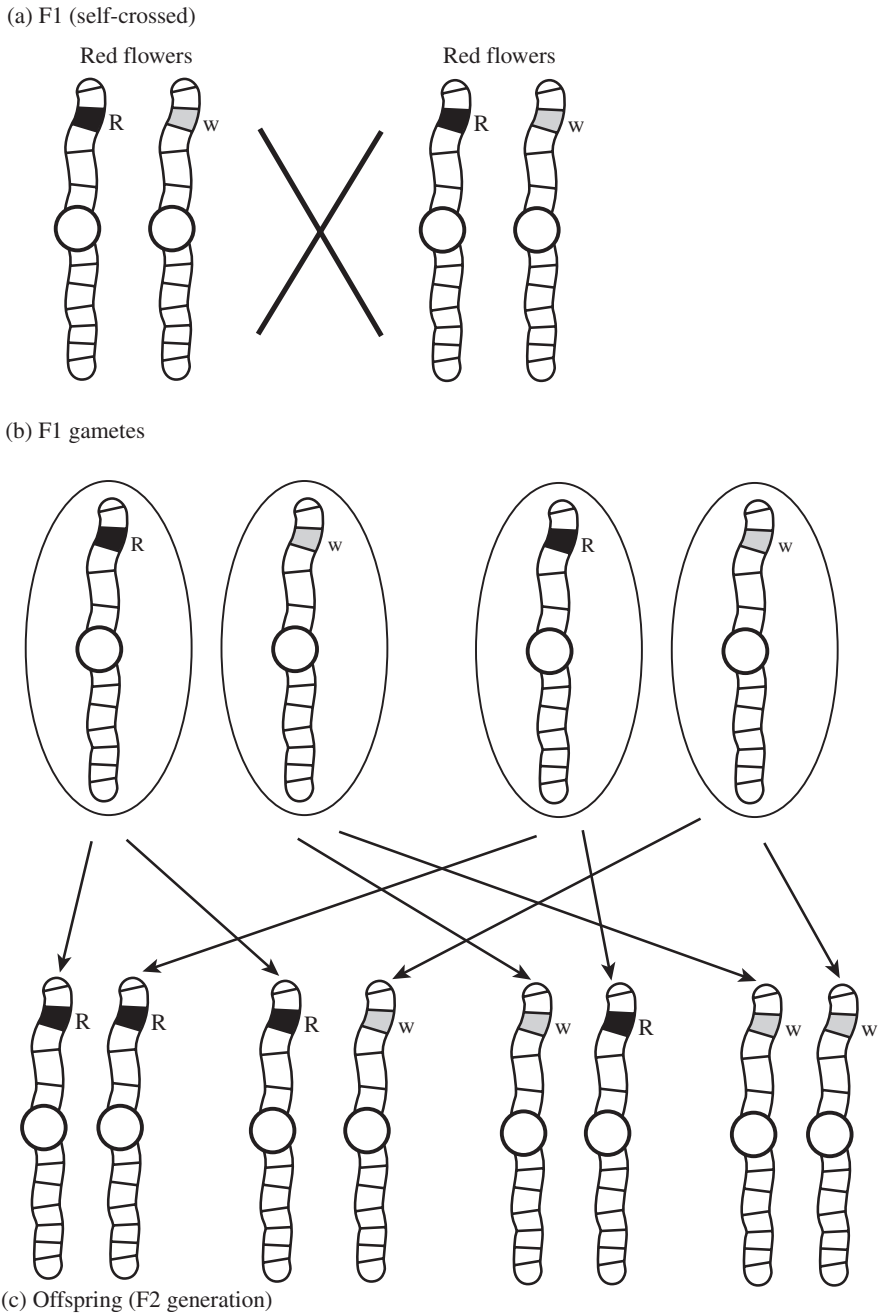


Fig. 13.3.3. A cross between two heterozygotes. Each F1 from Figure 13.3.2 makes gametes having the genes R and w with equal probability. When the gametes combine to make the F2 generation, the result is offspring of genotypes RR, Rw, and ww in the ratio 1:2:1.

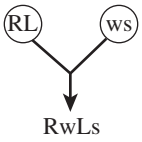
R = red flowers
w = white flowers
L = long stems
s = short stems

(a) Parental generation

RRLL × wwss

(b) Parental gametes

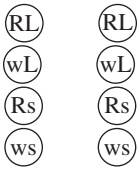
(c) F1 generation



(d) Self-cross F1

RwLs × RwLs

(e) F1 gametes



(f) Punnett square to give F2 generation

	RL	wL	Rs	ws
RL	RRLL	RwLL	RRLs	RwLs
wL	RwLL	wwLL	RwLs	wwLs
Rs	RRLs	RwLs	RRss	Rwss
ws	RwLs	wwLs	Rwss	wwss

9:3:3:1 ratio of phenotypes

Fig. 13.3.4. A complete dihybrid cross between plants whose flower color locus and stem length locus are on different homologous pairs, i.e., the two properties are not linked. The result is a 9:3:3:1 ratio of phenotypes in the F2. In this figure, only the allelic symbols are shown; the chromosomes are not drawn.

Sexual reproduction leads to variation in several ways.

We shall concern ourselves with organisms in which the diploid generation is the most conspicuous, e.g., humans, and we will examine the variations introduced into the diploid organism by sexual reproduction. It should always be borne in mind, however, that haploid organisms are under genetic control also.

Earlier it was pointed out that while mutation is the ultimate cause of genetic variation, there is only a very small chance that a given locus will mutate between two generations, will be unrepaired, and will not kill the cell. In spite of this, there are great variations among even the offspring of a single mating pair. We are now in a position to understand the sources of this immediate variation. First, look at the Punnett square of the dihybrid cross in Figure 13.3.4. Note that the F1 (RwLs) yields the gametes RL, Rs, wL, and ws, and yet the gametes of the parental generation were RL and ws. Thus two new combinations have turned up in the gametes of the F1. The reason is that the flower color locus and the stem length locus are unlinked—they are on different homologous pairs—and every homologous pair assorts independently of every other pair. Thus in the gametes of the F1, R paired up with L as often as R paired up with s. There were therefore $2^2 = 4$ combinations of chromosomes in the gamete. A human has 23 homologous pairs, all of which assort independently; thus a person can produce 2^{23} different combinations of chromosomes in their gametes, using independent assortment alone!

Second, when homologous chromosomes synapse they can exchange pieces in a process called *crossing over*. Let us cross two true-breeding parents, AABB $\times$ aabb, as shown in Figure 13.3.5. Notice that the two gene loci are *linked*, i.e., on the same chromosome. The F1 genotype is AaBb, and we *test-cross* it.[4] Some of the gametes of the F1 are the expected ones, AB and ab, but as the figure shows, crossing over, in which the homologs break and rejoin in a new way, produces gametes with two new allelic combinations, Ab and aB. These two new kinds of gametes, called *recombinant* gametes, are different from the gametes of either members of the parental generation. When the various gametes are paired up with the ab gametes in the test-cross, the following *phenotypes* appear in the F2 generation: Ab, aB, AB, and ab. The last two of these are the same phenotypes as the parental generation, and the first two are recombinant offspring, having phenotypes not seen in the previous crosses. We see that crossing over rearranges genetic material and presents novel phenotypes upon which selection can act.

How often does such crossing over occur? Actually, it is not unusual to find at least one example in every tetrad. Furthermore, crossing over is predictable: The farther apart two loci are, the more likely crossing over is to occur between them. The frequency of crossing over, measured by the frequency of recombinant offspring, is used by geneticists as a measure of the distance between two loci.

Note that we can account for an immense number of allelic combinations just using independent assortment and crossing over, without a mention of mutation. Independent assortment and crossing over account for virtually all the phenotypic variation seen in members of a single family generation. This variation, in the main, is what Darwinian selection works on.

A final point is worth mentioning here: Self-fertilization might be considered to be a limiting form of sexual reproduction.[5] Suppose that allele A is completely

[4] A test-cross is a cross with a homozygous recessive individual.

[5] Think of it this way: A self-cross is just like a cross between two separate, but genetically identical, parents.

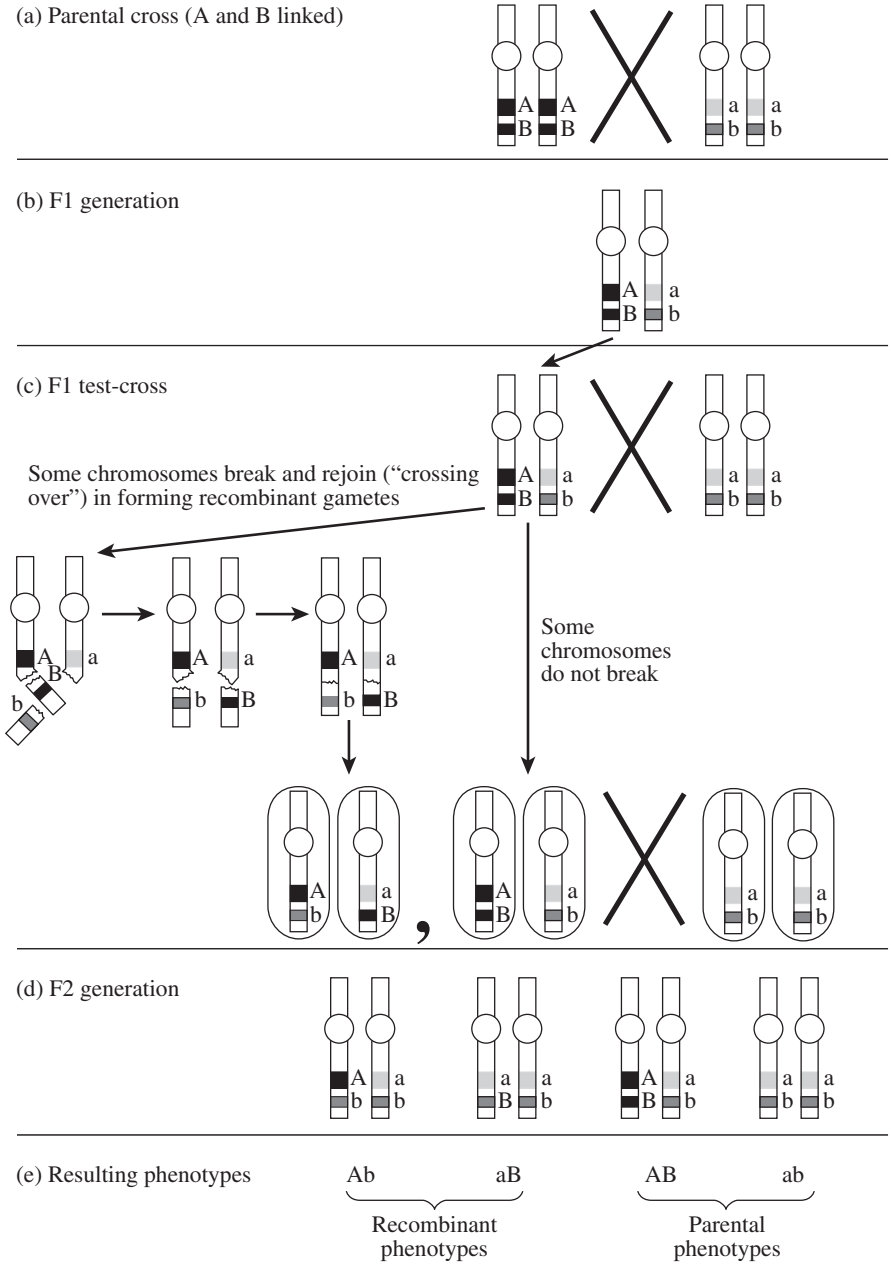


Fig. 13.3.5. A complete dihybrid cross, in which loci A and B are on the same chromosome, i.e., the two properties are linked. The results are predictable until the F1 test-cross at (c), when the chromosomes may break, yielding new combinations of the two loci. Notice that the resulting phenotypes at (e) include two (Ab and aB) that are unlike either of the two original parents.

dominant to allele a: If we self-cross an individual of genotype Aa, variant offspring appear in the ratio of 3:1, a mark of sexual reproduction. Asexual reproduction in the same organism yields only one kind of offspring—Aa. Where self-fertilizing organisms might run into evolutionary problems is in *continued* self-fertilization, which minimizes variation. This is shown by the following example: Take a population that is 100% heterozygotes (Aa) and self-cross all individuals. Note that the result is 50% heterozygotes and 50% homozygotes. Now self-cross all of that generation and note that 75% of the next generation will be homozygotes. After a few more generations of self-fertilization, virtually the entire population will be homozygous, either AA or aa. This can create problems for the population in two ways: First, suppose that the recessive allele is an unfavorable one that is usually masked by the dominant allele. As shown above, self-fertilization increases homozygosity, and homozygous recessive individuals would be selected out. Second, when homozygotes fertilize themselves, independent assortment and crossing over can occur, but they cannot generate variation. (You should verify this statement by schematically working out the cross.)

Here is an idea to think about: We sometimes hear about the "rescue" of a species that is near extinction. The last few members of the species are brought together to be bred in a controlled environment, free from whatever forces were causing the extinction in the first place. Suppose now that a particular species has been depleted until only one male and one female are left. This mating pair must serve to reestablish the species. It is to be expected that each member of this pair would be heterozygous for at least a few unpleasant recessive genes. In light of the information in the preceding paragraph, what unique problems will the reconstituted species face?

A group of questions for practice and for extending Mendelian genetics.

1. Refer to the definition of "true breeding" two sections back. In the discussion of the monohybrid cross and Figures 13.3.2 and 13.3.3, "true breeding" was asserted to mean "homozygous." Suppose for a moment that a member of a supposedly true-breeding population were a heterozygote. Show that being heterozygous is inconsistent with the definition of true breeding.

2. Suppose you are given a red-flowered pea. A *test-cross* will enable you to determine whether this dominant phenotype is a heterozygote (Rw) or a homozygote (RR). Cross it with a homozygous recessive individual (ww); the cross is therefore either RR $\times$ ww or Rw $\times$ ww. Note the different results obtained, depending on the genotype of the dominant phenotype. How do we know that a white-flowered plant is homozygous?

3. The red-flower allele in peas completely masks the white-flower allele, i.e., red is *completely dominant* to red. If we cross a true-breeding red-flowered snapdragon with a true-breeding white-flowered one, the F1 offspring are all pink. We say that dominance is *incomplete*, or *partial*, for snapdragon flower color; partial dominance is a very common phenomenon. Cross two pink snapdragons to get offspring with a phenotypic ratio of 1 red:2 pink:1 white.

4. Foxes with platinum fur have the genotype Pp and silver foxes are pp. The genotype PP kills the fetus right after conception, i.e., it is *lethal*. Evidently, the gene locus for fur color controls other properties as well, among them at least one very basic metabolic process. Show that a cross of two platinum foxes gives a 2:1 phenotypic ratio of offspring.

5. There is a notable exception to the statement that every chromosome in a mammalian diploid cell has an exact homolog. Mammalian males have one chromosome called an X chromosome and one called a Y chromosome. Females have two Xs and no Ys. These *sex chromosomes* carry a number of genes having to do with gender and many others that do not. Despite the fact that they are not homologous, the X and Y chromosomes in a male can synapse over a portion of their length to facilitate meiosis. A well-known recessive gene on the X chromosome is for hemophilia, a blood-clotting disorder. Let us represent a heterozygous ("carrier") female as $X^h\ X^+$, where "X" indicates X-linkage, "h" indicates the hemophilia allele, and "+" represents the normal allele. Note that a male of genotype X^hY will show the disorder because there is no possibility of a dominant allele on his Y chromosome to mask the hemophilia allele on his X chromosome. Cross a carrier female with a hemophilic male to show that a female can get hemophilia. Cross a carrier female with a normal male to show that no daughters and half the sons would be affected.

6. Often there are more than two choices for the alleles for a property, a phenomenon called *multiple alleles*. The presence of certain molecules on red blood cells is determined by the alleles A, B, and O. For example, the genotypes AA and AO yield the A molecule, the genotypes BB and BO yield the B molecule, the genotype OO yields neither molecule, and the genotype AB yields both the A and B molecules. The latter case, expression of both alleles, is called *codominance*. Cross an AB parent with an O parent; what ratio of offspring is obtained? Could an O-type man be the parent of an AB child? Is it possible that a particular A-type man is the father of an A-type child by an A-type mother?

7. The expression of some genes is determined by the environment. The gene for dark pigmentation in Siamese cats is expressed only in cool parts of the cat's body—nose, ears, and tail tip. The expression of the gene for diabetes mellitus, a deficiency in sugar metabolism, is affected by diet and the person's age. As an example, environmental effects might cause a dominant allele not to be expressed under certain conditions, and an individual with genotype AA or Aa might show the recessive phenotype. How might you determine that such an individual is actually of the dominant genotype?

13.4 Genetic Drift

Natural selection is not the only mechanism at work in evolution. The other important mechanism is *genetic drift*. Genetic drift is a random or stochastic process in which

the allelic fractions of a population change from generation to generation due to chance alone.

At one level chance works on mating itself. In a pair of diploid sexually reproducing parents (such as humans), not all of the parent's alleles will be passed on to their progeny due to chance assortment of chromosomes at meiosis. This is called *sampling error*. Extended over an entire population, sampling error will be mitigated by the law of large numbers but not completely eliminated. Thus population-wide, frequencies of alleles change from generation to generation.

In addition to that, not all offspring of a new generation survive and reproduce. Of course, natural selection works on this principle when the underlying cause is differential fitness. But fitness is not always the issue; random events can intervene with the result that even the most fit individuals fail to reproduce.

Genetic drift is not self-correcting. The population does not have a genetic memory and is not urged back to some previous genetic state. The changes that arise in the allelic frequencies of the previous generation become the basis of the new gene pool. As more and more generations pass, allelic frequencies can range far from where they started. Surprising as it may seem, over time genetic drift moves the fraction of every allele (not subject to natural selection) to either 1 or 0. This will even happen due to sampling error alone.

Genetic drift is an example of the mathematical process known as a *random walk*. In a simplified version, suppose that the frequency of some allele, say A, can change by the amount s, the step size, in each generation; assume $s = \frac{1}{10}$. Right now A comprises the fraction f of the gene pool for its trait; suppose $f = 60\%$. From generation to generation, f will move up by 10% or down by 10% with some probability $p > 0$. The walk will go back and forth along the points $0\%, 10\%, \ldots, 90\%, 100\%$, but eventually f will become 0% or 100% and be trapped there. In random walk terminology, the walk has been *absorbed*. The larger the value of s or p, the fewer expected number of generations until absorption.

The effects of genetic drift are accentuated on small populations. Loss of alleles due to limited mating outcomes and random environmental occurrences represent a larger fractional allelic change to the population when it is small. So the step size s of the previous paragraph is larger. There are two extreme situations in which genetic drift is of primary importance due to small population size. The first is when some catastrophic event occurs to a population and its numbers fall to a very low level; this is called the *bottleneck effect*. The second is when a subpopulation of the whole becomes reproductively isolated; this is the *founders effect*.

An example of the bottleneck effect occurred to the northern elephant seal. This animal was hunted almost to extinction. By 1890, there were fewer than 20 animals remaining. Although it now numbers around 30,000, there is very little genetic variation in this population. As a result the population is highly vulnerable to extinction, for example from disease. In any case, the present elephant seal population is sure to have large differences in allelic frequencies from its pre-1890 counterpart.

Founder effect occurs when a small subpopulation of a species becomes reproductively isolated. These are the founders of the isolated population and their allelic frequencies will be its norm. But these frequencies will most certainly be much dif-

ferent from those of the parent population for many traits. Native American Indians constitute an example in that their ancestors crossed the Bering Strait in small numbers to found societies in the Western Hemisphere. Unlike other races, American Indians lack the blood group B, in all probability due to the absence of its allele among the founders.

The founders effect often results in the high prevalence of normally rare diseases. The Amish people of Pennsylvania constitute a closed population stemming from a small number of original German immigrants, about 200. But the Amish carry unusual concentrations of gene mutations that cause a number of otherwise rare inherited disorders. One of these, *Ellis–van Creveld syndrome*, involves dwarfism, polydactyl abnormalities (extra fingers or toes), and, in about half of the afflicted, a hole between the two upper chambers of the heart.

Since founders and their progeny are genetically isolated and must interbreed, their recessive genes will pair up much more frequently than occurs in the parent population. But recessive genes are often defective, hence the increased incidence of these kinds of genetic disorders. In the Amish, Ellis–van Creveld syndrome has been traced back to a single couple who came to the area in 1744.

Obviously, genetic drift is highly important as a mechanism of evolution. Arguments over how important as compared to natural selection is an unsettled issue among geneticists. It is known that genetic drift can overcome natural selection if the selection pressure is weak. The fixation of less fit alleles is an integral feature of the evolutionary process.

13.5 A Final Look at Darwinian Evolution

We close out our discussion of biology with a last look at the Darwinian model of evolution, which we introduced in Section 3.1. *Fitness is measured by the persistence of a property in subsequent generations. If a property cannot be inherited, it cannot be selected. Thus acquired properties like facelifts cannot be selected, nor can genetic properties of sterile individuals, like a mule's hardiness.*

Populations evolve; individuals do not. An individual is born with a fixed set of genes; mutations in somatic cells are not transmitted to offspring, and mutations in germinal cells can be seen only in the offspring.

Some organisms do not exhibit sexual reproduction but rather reproduce only asexually. Their only source of variation is therefore mutation. Nevertheless, such organisms have long evolutionary histories.

Fitness is measured by the ability to project genes into subsequent generations.

Common phrases like "struggle for survival" and "survival of the fittest" can be very misleading because they bring to mind vicious battles to the death between two contestants. The fact is that, except arguably among humans, violence is rarely the route by which Darwinian fitness is achieved in the biological world. Even the noisy, aggressive encounters between male animals seen on television nature programs seldom result in serious injury to participants. We must look to much more subtle interactions as a source of fitness.

One group of organisms may be slightly more able than another to tolerate heat, to thrive on available food, or to elude predators. Subtle pressure is the norm in evolution; it works slowly, but there is no hurry. *Drosophila*, a common fruit fly, is used in many genetic experiments because it is easy to raise, has a short life span, and has many simple physical properties, such as eye color, whose modes of genetic transmission are easy to follow. If a large number of red-eyed and white-eyed *Drosophila* are put together in an enclosure and left to their own devices, the fraction of flies with white eyes will decrease steadily for several tens of generations and finally reach zero. Close observation reveals the reason: A female *Drosophila*, either red-eyed or white-eyed, will generally choose not to mate with a white-eyed male if a red-eyed male is available. Thus there is a definite selection for the red-eye genetic trait.

Humans are not excluded from such subtle pressures: "Personals" ads in newspapers contain wish lists of traits people prefer in a mate. Height and affluence (control of territory?) are prized male traits, and hourglass figures and youth (ability to bear children?) are valued female traits.

Regardless of the strength of the selective pressure or the nature of the properties being selected, there is really only one way to measure the evolutionary value of a trait, and that is the degree to which it is propagated into future generations. A shy, ugly person who has lots of fertile children has a high degree of fitness. We see that one generation of propagation is not enough; the trait must be persistent. For example, mules are known for their hardiness, but they are sterile offspring of horses and donkeys. As a result, the hardiness of a mule cannot confer any evolutionary advantage.[6]

Populations evolve; individuals do not.

A *population* is a group of organisms of the same species, living in the same area. As before, we will restrict our discussion here to populations of organisms for which the diploid generation is most conspicuous, e.g., humans.

If we observe a population over many generations, the "average" phenotypic property will change, in keeping with our earlier discussion of species formation and genetic drift. Thus the average height may increase, or the typical eye color may darken. We now ask and answer two questions: At what points in the alternation of generations do the changes occur, and what kinds of changes are relevant to evolution?

The Darwinian model stipulates that favored properties may be transmitted to offspring; in any case, they certainly must be *capable* of transmission for the model to apply. A diploid individual is conceived with a set of genes that are relatively fixed for that individual's lifetime. Exceptions to this statement might involve mutations in somatic cells and infection by lysogenic viruses (see Chapter 10). As long as these changes do not occur in germinal cells or germinal cell precursors, they cannot be transmitted to the next generation and thus have no evolutionary effect. In addition,

[6] There is a peculiar example of a noninheritable trait—a desire for a large family—that might be passed from one generation to another by teaching and which could have a strong positive selective value. This was discussed in Section 4.1.

there are many phenotypic properties that favor reproduction but that, because they are not of a genetic nature, cannot be transmitted to offspring. Examples are suntans, exercise-strengthened bodies, and straightened teeth.

Genetically transmissible variations must originate via one of at least three routes, all of which require sexual reproduction (in other words, an *intervening haploid generation*) for their expression:

1. independent assortment;
2. crossing over;
3. mutation in a sperm, or an egg, or in their precursors in a parent prior to conception, or in a zygote at a very early stage of development—the altered genetic material in any one of these cases should turn up in those cells of the reproductive system that undergo meiosis to form the next generation of gametes.

We can conclude that because the Darwinian model requires changes that are inheritable, and because the observation of inheritable changes requires the observation of more than one generation, *it is the population that evolves*. Changes restricted to the somatic cells of individuals are not genetically transmitted to offspring; thus in terms of evolution, an individual is fixed. Over a period of time, however, the average, or typical, characteristics of the population evolve.

Some organisms do not exhibit sexual reproduction.

Sexual reproduction is unknown (and probably nonexistent) in several kinds of organisms, for example, most bacteria, blue-green algae, and some fungi. In those cases, all reproduction is asexual, which would seem to limit severely the possibilities of variation. Nonetheless, these organisms seem to have gotten along fine over long periods of history. We must conclude that some combination of three things applies: Either these organisms have not been exposed to large fluctuations in their environments, or they possess an innate physiological flexibility that permits them to get along in different environments, or their spontaneous mutation rates are sufficiently high to generate the variation necessary for adapting to new environmental situations.[7]

13.6 The Hardy–Weinberg Principle

Diploidism and sexual reproduction complicate the calculation of inheritance probabilities. But remarkably, the results are the same as if alleles were balls selected for combination from an urn. This is the Hardy–Weinberg principle. Although its veracity depends on random mating, among other properties, it continues to provide good approximations in many other situations as well.

Mendelian inheritance follows the laws of probability.

We will be concerned with probabilities associated with Mendelian inheritance for a

[7] There is now good evidence that bacteria, including asexual ones, can pass small pieces of DNA, called plasmids, to other bacteria.

diploid organism. As explained in Section 13.2, meiosis produces four haploid cells of two different kinds, each equally likely to participate in fertilization. Then the probability is $\frac{1}{2}$ that a given kind of gamete will do so.

Consider first a single locus for which there are only two alleles, say A and a. Hence there are three distinct genotypes, the homozygotes AA and aa, and the heterozygote Aa (or aA). If one parent is AA and the other Aa, then the possible zygote genotypes resulting from a mating may be represented by an *event tree* as follows.

Let the first branch point in Figure 13.6.1 correspond to the allele donated by the first parent, AA. There are two possible alleles, and so here the diagram will have two branches. But for the parent AA, both branches lead to the same result, namely, the contribution of allele *A* to the offspring. Let the second branch point correspond to the allele donated by the second parent. Again there are two possibilities, but this time the outcomes are different as indicated.

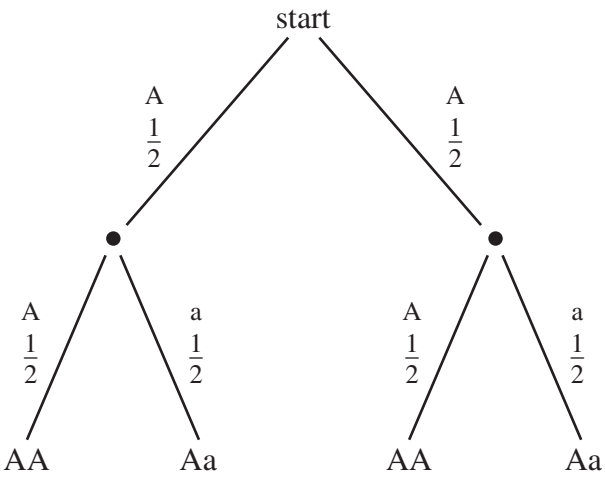


Fig. 13.6.1. Probabilities for the offspring of an AA with Aa mating.

Nhe resulting probabilities may be calculated in several ways. Since all the legs, or *edges*, of the diagram are equally likely, so are the resulting outcomes, each having probability $\frac{1}{4}$. Hence

$$\Pr(\mathrm{AA}) = \frac{1}{2} \quad \text{and} \quad \Pr(\mathrm{Aa}) = \frac{1}{2}.$$

Alternatively, starting at the top node, the *root node*, and traversing the two edges to the left leading to AA gives a probability of $\frac{1}{4}$ for this outcome by multiplying the probabilities along each edge of the path ($\frac{1}{2} \cdot \frac{1}{2}$). This way of calculating the probabilities is the method of *conditional probabilities*, since the probabilities along the branches leading away from any node are conditioned on the sequence of events

leading to the node. Altogether, the probability of an AA zygote by this method is $\frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} = \frac{1}{2}$, since AA can occur in two different ways according to the tree.

Finally, the probabilities can be calculated by the principle of independence (see Section 2.8). The selection of a gamete from the AA parent will result in an A with probability 1. The selection of a gamete from the Aa parent is independent and will result in an A with probability $\frac{1}{2}$. Therefore, the probability of an AA zygote is $1 \cdot \frac{1}{2}$.

The complete list of Mendelian inheritance probabilities is given in Table 13.6.1.

Table 13.6.1. Mendelian inheritance probabilities.

Parent	**Zygote genotypes**		
genotypes	**AA**	**Aa**	**aa**
AA × AA	1		
AA × Aa	$\frac{1}{2}$	$\frac{1}{2}$	
Aa × Aa	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Aa × aa	$\frac{1}{2}$	$\frac{1}{2}$	
aa × aa	1		

Random allelic combination preserves allelic fractions.

Let n_{AA} denote the number of AA genotypes in a population, and likewise let n_{aa} denote the number of aa genotypes. For reasons that will shortly become clear, let n_{Aa} denote one-half the number of Aa genotypes. Then the size of the entire population N is the sum $N = n_{\mathrm{AA}} + 2n_{\mathrm{Aa}} + n_{\mathrm{aa}}$. Let n_{A} and n_{a} denote the number of A alleles and a alleles, respectively, carried by the population. Thus $n_{\mathrm{A}} + n_{\mathrm{a}} = 2N$, since the population is diploid.

Similarly, let p_{AA}, p_{Aa}, p_{aa}, p_{A}, and p_{a} denote their corresponding fractions of the population. Then $p_{\mathrm{A}} + p_{\mathrm{a}} = 1$ and $p_{\mathrm{AA}} + 2p_{\mathrm{Aa}} + p_{\mathrm{aa}} = 1$. Moreover,

$$p_{\mathrm{A}} = \frac{n_{\mathrm{A}}}{2N} = \frac{2n_{\mathrm{AA}} + 2n_{\mathrm{Aa}}}{2N} = p_{\mathrm{AA}} + p_{\mathrm{Aa}},$$

and similarly

$$p_{\mathrm{a}} = p_{\mathrm{Aa}} + p_{\mathrm{aa}}.$$

Now imagine that all the *alleles* of the population are pooled and two are selected at random from the pool to form a pair. The selection of an A happens with probability p_{A}, while the selection of an a happens with probability p_{a}. (We assume that the pool is so large that the removal of any one allele does not appreciably change the subsequent selection probability.) Then, for example, the probability of forming an AA pair is p_{A}^2, since we assume that the selections are made independently. In the same way, the other possible pair selections are calculated, with the results shown in Table 13.6.2. As always, these probabilities are also the (approximate) fractions of the various outcomes in a large number of such pairings.

Table 13.6.2. Mendelian inheritance probabilities.

Female gametes	**Male gametes (frequencies)**	
(frequencies)	**A (p_A)**	**a (p_a)**
A (p_A)	AA (p_A^2)	Aa ($p_A p_a$)
a (p_a)	aA ($p_A p_a$)	aa (p_a^2)

From the table we can calculate the fraction, p'_A, of A alleles among the resultant pairs. Each pair of type AA contributes two A alleles, and while each Aa pair contributes only one, there are twice as many such pairs. Hence

$$p'_A = \frac{2p_A^2 + 2p_A p_a}{2} = p_A(p_A + p_a) = p_A.$$

In this it is necessary to divide by 2 because each pair has two alleles. Thus the fraction of A alleles among a large number of pairings is the same as their fraction in the original gene pool, p_A. The same is (consequently) true for the a allele, $p'_a = p_a$.

Of course, the process of gene maintenance for bisexual diploid organisms is much more complicated than the simple random pairing of alleles selected from a common pool that we have explored here (see Section 13.3). Nevertheless, we will see in the next subsection that the results are the same if mating is random.

Random mating preserves allelic fractions.

Again consider a one-locus, two-allele system and suppose mating is completely random. Then the probability of an AA × Aa mating, for example, is $2p_{AA}(2p_{Aa})$, since the first parent could be AA and the second Aa or the other way around. Altogether, there are six different kinds of matings; their probabilities are listed in Table 13.6.3.

Table 13.6.3. Mendelian inheritance probabilities.

Genotype mating	**Probability**
AA × AA	$(p_{AA})^2$
AA × Aa	$2p_{AA}(2p_{Aa})$
AA × aa	$2p_{AA}p_{aa}$
Aa × Aa	$(2p_{Aa})^2$
Aa × aa	$2(2p_{Aa})p_{aa}$
aa × aa	$(p_{aa})^2$

Now apply the Mendelian inheritance laws to calculate the probability of the various possible zygotes, for example, an AA zygote. First, an AA results from an AA × AA parentage with probability 1. Next, an AA results from an AA × Aa parentage with probability $\frac{1}{2}$ (see Figure 13.6.1), and finally an AA results from an Aa × Aa cross with probability $\frac{1}{4}$. Now, by the method of conditional probabilities

as discussed at the beginning of this section, we have

$$\begin{aligned}\Pr(\mathrm{AA}) &= p_{\mathrm{AA}}^2 \cdot 1 + 2p_{\mathrm{AA}}(2p_{\mathrm{Aa}}) \cdot \frac{1}{2} + (2p_{\mathrm{Aa}})^2 \cdot \frac{1}{4} \\ &= p_{\mathrm{AA}}^2 + 2p_{\mathrm{AA}}p_{\mathrm{Aa}} + p_{\mathrm{Aa}}^2 \\ &= (p_{\mathrm{AA}} + p_{\mathrm{Aa}})^2 = p_{\mathrm{A}}^2.\end{aligned}$$

Similarly, we leave it to the reader to show that

$$\Pr(\mathrm{aa}) = (p_{\mathrm{aa}} + p_{\mathrm{Aa}})^2 = p_{\mathrm{a}}^2$$

and

$$\Pr(\mathrm{Aa}) = 2(p_{\mathrm{AA}} + p_{\mathrm{Aa}})(p_{\mathrm{Aa}} + p_{\mathrm{aa}}) = 2p_{\mathrm{A}}p_{\mathrm{a}}.$$

But this shows that the fractions of alleles A and a are again p_{A} and p_{a}, respectively, among the offspring just as among their parents, assuming that the various genotypes are equally likely to survive. This is the same result we calculated in the last section. In other words, the effect of random genotype mating is indistinguishable from that of random gamete recombination. This is the *Hardy–Weinberg principle*.

Hardy–Weinberg principle. *Under the condition that mating is random and all genotypes are equally fit, the fractions of alleles will stay the same from generation to generation.*

A consequence of the Hardy–Weinberg principle is that after at most one generation, the fractions of genotypes also stabilize and at the values

$$\begin{aligned}p'_{\mathrm{AA}} &= p_{\mathrm{A}}^2, \\ 2p'_{\mathrm{Aa}} &= 2p_{\mathrm{A}}p_{\mathrm{a}} \\ p'_{\mathrm{aa}} &= p_{\mathrm{a}}^2.\end{aligned}$$

For example, suppose that initially 70% of a population is AA and the remaining 30% is aa. Then the fractions of alleles in subsequent generations are also 70% and 30% for A and a, respectively. Therefore, after one generation, the fractions of genotypes will be

$$\begin{aligned}\mathrm{AA}: &\quad (0.7)^2 = 0.49, \\ \mathrm{Aa}: &\quad 2(0.7)(0.3) = 0.42, \\ \mathrm{aa}: &\quad (0.3)^2 = 0.09.\end{aligned}$$

In some cases the Hardy–Weinberg principle is applicable even when mating is not random. Mating would fail to be random for example if the homozygote for a recessive gene is impaired or unviable. But in fact, the homozygotes in these cases are so rare that the induced error is very small. Keep in mind that for a recessive gene a, the homozygote AA and heterozygote Aa are indistinguishable, so that random mating among them is a reasonable assumption.

The Hardy–Weinberg principle breaks down when there is migration, inbreeding, or nonrandom mating, that is, phenotypes are selected for some attribute.

Sex-linked loci give rise to different rates of expression between males and females.

In the event that males (or females) have one or more nonhomologous chromosomes, the foregoing derivations must be modified. One consequence of nonhomologous chromosomes is that there can be a large difference in expression of a sex-linked character between males and females. For definiteness, suppose the male has the nonhomologous pair XY, while the female has the homologous pair XX.[8] For this case, fractions of alleles for genes on either the X or the Y chromosome are identical to genotype fractions for the male. For example, suppose a recessive sex-linked allele occurs with frequency p among a population. Then p is also the rate at which the allele will occur in males. However, the rate at which the homozygous condition will occur in females is p^2.

An example of such an allele is color blindness in humans. Through various studies, it is believed that the frequency of the recessive allele is 8% as derived from the incidence rate in males. Therefore, the incidence rate in females ought to be $(0.08)^2 = 0.0064$ or 0.6%. Actually, the female incidence of the disease is about 0.4%. The discrepancy is an interesting story in its own right and stems from the fact that there are four different kinds of color blindness, two of which are red blindness and the other two green blindness. The bearer of defective genes for different types, such as these two, can still see normally.

Another possibility that can arise relative to sex-linked genes is that the allelic fractions are different between males and females. This can happen, for instance, when males and females of different geographical backgrounds are brought together. Let F be the fraction of allele A in the females and let M be its fraction in males. Then $f = 1 - F$ is the fraction of a in females and $m = 1 - M$ is its fraction in males. Assuming an equal number of males and females, the population frequencies work out to be

$$p_{\mathrm{A}} = \frac{M + F}{2} \quad \text{and} \quad p_{\mathrm{a}} = \frac{m + f}{2} = 1 - p_{\mathrm{A}},$$

and these will remain constant by the Hardy–Weinberg principle. However, the values of M and F will change from generation to generation.

To follow these fractions through several generations, we need only keep track of F and M, since f and m can always be found from them. Let F_n and M_n refer to generation n with $n = 0$ corresponding to the initial fractions.

Since a male gets his X chromosome from his mother, the allelic frequencies in males will always be what it was in females a generation earlier; thus

$$M_{n+1} = F_n.$$

On the other hand, the frequency in females will be the average of the two sexes in the preceding generation, since each sex contributes one X chromosome; hence

$$F_{n+1} = \frac{1}{2}M_n + \frac{1}{2}F_n.$$

[8] This is a mammalian property. In birds, the situation is reversed.

In matrix form, this can be written

$$\begin{bmatrix} M_{n+1} \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} M_n \\ F_n \end{bmatrix}.$$

In the exercises, we will investigate where this leads.

Before we leave this example, there is another observation to be made. We used matrix T above,

$$T = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

in conjunction with multiplication on its right to update the column of male/female fractions M_n and F_n. But in this example there is a biological meaning to left multiplication on the matrix T. In each generation, there will be a certain fraction of the alleles on the X chromosome in males that originally came from the females. It is possible to track that distribution.

To fix ideas, suppose that a ship of males of European origin runs aground on a South Sea island of Polynesian females. Further, suppose (hypothetically) that the alleles for a gene on the X chromosomes of the Europeans, the E-variant, are slightly different from those of the Polynesians, the P-variant, in, say, two base pairs. So the distribution of E-variant and P-variant chromosomal alleles of the emigrating males can be described by the (row) pair

$$\begin{pmatrix} 1 & 0 \end{pmatrix},$$

where the first element is the fraction originating with the males and the second is the fraction originating with the females. The distribution of these fractions can be traced through the generations by a matrix calculation similar to that above, only this time using matrix multiplication on the left. In the first generation, we have

$$\begin{pmatrix} 1 & 0 \end{pmatrix} \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix},$$

showing that all the alleles in the males in this generation come from the females. In the second generation, the fraction works out to

$$\begin{pmatrix} 0 & 1 \end{pmatrix} \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix},$$

or 50–50. Of course, the calculation can be continued to obtain the fractions for any generation.

The same calculation can give the female ratios, by starting with the initial female ratio of $(0 \ 1)$.

13.7 The Fixation of a Beneficial Mutation

A beneficial mutation does not necessarily become a permanent change in the gene pool of its host species. Its original host individual may die before leaving progeny,

for example. Under the assumption that such a mutation is dominant (rather than recessive) and that individuals with the mutation behave independently, it is possible to derive the governing equations for calculating the fixation probability. One way of measuring the value of a vital factor is the expected number of surviving offspring, beyond self replacement, an adult will leave. For an r-strategist, the chance that a beneficial mutation will become permanent is about twice the overreplacement value of the mutation to its holder.

Probability of fixation of a beneficial mutation is the complement of the fixed point of its probability-generating function.

Let p_k be the probability that a chance mutation appearing in a zygote will subsequently be passed on to k of its offspring. A convenient method of organizing a sequence of probabilities, such as $p_k, k = 0, 1, \ldots$, is by means of the polynomial

$$f(x) = p_0 + p_1 x + p_2 x^2 + \cdots,$$

in which the coefficient of x^k is the kth probability. This polynomial is called the *probability-generating function* for the sequence p_k. The probability-generating function is purely formal, that is, it implies nothing more than a bookkeeping device for keeping track of its coefficients. Note that $f(1) = 1$. And $f(0) = p_0$ is the probability that the mutation disappears in one generation. Also note that the expected number of offspring to have the mutation is given (formally) by

$$\sum_{k=1}^{\infty} k p_k = f'(1)$$

(see Section 2.8). To say that the mutation is beneficial is to say that this expectation is greater than 1, that is,

$$\sum_{k=1}^{\infty} k p_k = 1 + a > 1$$

for some value a, which is a measure of the benefit in terms of overreplacement in fecundity.

Now if two such individuals with this mutation live and reproduce independently of each other (as in a large population), then the probability-generating function for their combined offspring having the mutation is

$$p_0^2 + 2p_0p_1x + (2p_0p_2 + p_1^2)x^2 + (2p_0p_3 + 2p_1p_2)x^3 + \cdots, \qquad (13.7.1)$$

which is proved by considering each possibility in turn. There will be no mutant offspring only if both parents leave none; this happens with probability p_0^2 by independence. There will be one mutant offspring between the two parents if one leaves none and the other leaves exactly one; this can happen in two ways. There will be two mutant offspring if the first leaves none while the second leaves two, or they both leave one, or the first leaves two while the second leaves none; this is $(2p_0p_2 + p_1^2)$. The other terms of (13.7.1) may be checked similarly.

But note that (13.7.1) is exactly the polynomial product $f^2(x)$,

$$(p_0 + p_1x + p_2x^2 + \cdots)(p_0 + p_1x + p_2x^2 + \cdots)$$
$$= p_0^2 + 2p_0p_1x + (2p_0p_2 + p_1^2)x^2 + \cdots .$$

More generally, m independent individuals with the mutation as zygotes will pass on the mutation to their combined offspring with probability-generating function given by the mth power $f^m(x)$.

Now start again with one mutant zygote and consider the probability-generating function f_2 for generation 2. Of course, the outcome of generation 2 depends on the outcome of generation 1. If there are no mutants in generation 1, and this occurs with probability p_0, then there are none for certain in generation 2. Hence this possibility contributes

$$p_0 \cdot 1$$

to f_2. On the other hand, if the outcome of generation 1 is one, then the probability-generating function for generation 2 is $f(x)$; so this possibility contributes

$$p_1 f(x).$$

If the outcome of generation 1 is two mutant individuals (and they behave independently), then the probability-generating function for generation 2 is, from above, $f^2(x)$; so this possibility contributes

$$p_2 f^2(x).$$

Continuing this line of reasoning yields the result that the probability-generating function for generation 2 is the composition of the function f with itself, $f_2(x) = f(f(x))$,

$$f_2(x) = p_0 + p_1 f(x) + p_2 f^2(x) + p_3 f^3(x) + \cdots = f(f(x)).$$

More generally, the probability-generating function for generation n, $f_n(x)$, is given as the composition $f \circ f \circ \cdots \circ f$ of f with itself n times, or

$$f_n(x) = \underbrace{f(f(\cdots f(x)\cdots))}_{n \text{ times}}.$$

Now the probability that the mutation dies out by the nth generation is the constant term of $f_n(x)$ or $f_n(0)$. Hence the probability that the mutation dies out or vanishes some time is the limit

$$V = \lim_{n\to\infty} \underbrace{f(f(\cdots f(0)\cdots))}_{n \text{ times}}.$$

Applying f to both sides of this equality shows that V is a fixed point of f,

$$f(V) = V.$$

The fixed point of $f(x)$ is where the graphs $y = f(x)$ and $y = x$ intersect (see Figure 13.7.1). Since $f'(x)$ and $f''(x)$ are nonnegative for $x > 0$ (having all positive or zero coefficients), and since $f(1) = 1$, we see that there can be either zero or one fixed point less than $x = 1$. If there is a fixed point less than one, then V is that value; otherwise, $V = 1$.

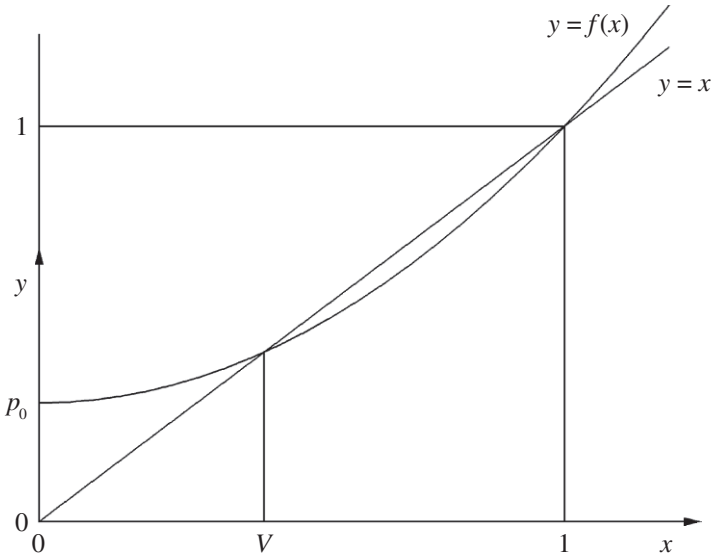


Fig. 13.7.1. V is the fixed point of $f(x)$.

For example, suppose that a mutation arose on the X-chromosome of a human female about the time that "Lucy" walked the earth (2 million years ago). Further suppose that the following probabilities of producing surviving (female) offspring pertained to the holder of such a mutation:

- probability of leaving no female offspring, $p_0 = 0.35$;
- probability of leaving one female offspring, $p_1 = 0.25$;
- probability of leaving two female offspring, $p_2 = 0.20$;
- probability of leaving three female offspring, $p_3 = 0.1$;
- probability of leaving four female offspring, $p_4 = 0.1$,
- and zero probability of leaving more than four female offspring.

Then the probability-generating function is

$$f(x) = 0.35 + 0.25x + 0.2x^2 + 0.1x^3 + 0.1x^4.$$

Its fixed points can be found by solving the roots of the fourth-degree polynomial

$$0.1x^4 + 0.1x^3 + 0.2x^2 + (0.25 - 1)x + 0.35 = 0.$$

With the following code, the appropriate root is found to be 0.62:

```
MAPLE
> f:=.1*x^4+.1*x^3+.2*x^2+(.25-1)*x+.35;
> fsolve(f,x,0..1);

MATLAB
> p=[.1 .1 .2 (.25-1) .35];
> min(roots(p))
```

Hence the probability of fixation is the complementary probability 0.38.

The chance that a mutation will become permanent for an r-strategist is about twice its overreplacement benefit.

Under certain conditions, the probability that an individual will have k offspring over its life is $b^k e^{-b}/k!$ for some constant b.[9] Once again we encounter the ubiquitous *Poisson distribution*. The conditions are approximately satisfied by many r-strategists. In this case, the probability-generating function is

$$f(x) = e^{-b}\left(1 + \frac{b}{1!}x + \cdots\right) = e^{-b}e^{bx} = e^{b(x-1)}.$$

Let the benefit of the mutation be a; then

$$1 + a = f'(1) = be^{b(1-1)} = b,$$

so $b = 1 + a$. Now let F be the fixation probability of the beneficial mutation, that is, the probability that the mutation will become permanent; then $F = 1 - V$. Since $V = f(V)$ (from the previous section), we have

$$1 - F = e^{-(1+a)F}.$$

Taking logarithms,

$$(1 + a)F = -\ln(1 - F) = F + \frac{F^2}{2} + \frac{F^3}{3} + \cdots.$$

The infinite series is the Taylor series for the middle term. Divide by F and subtract 1 to get

$$a = \frac{F}{2} + \frac{F^2}{3} + \cdots.$$

If a is small, then approximately

$$a \approx \frac{F}{2},$$

so the fixation probability is about $2a$.

[9] The conditions are (a) the probability of an offspring over a short period of time Δt is proportional to Δt; (b) the probability of two or more offspring over a short period of time is essentially zero; and (c) offspring occur independently of one another. The distribution would also apply if offspring occurred in batches; the k counts batches.

Exercises/Experiments

1. In this problem, assume a diploid organism having three loci per homologous chromosomal pair and two alleles per locus.

(a) If the organism has only one such chromosomal pair, how many different genotypes are possible?

(b) Same question if there are two chromosomal pairs.

(c) Suppose there are two chromosomal pairs with genes α, β, and γ on one of them, while genes δ, ϵ, and ϕ lie on the other. How many different haploid forms are there?

(d) For a given genotype as in (c), how many different gametes are possible? That is, suppose that a particular individual has the homologous chromosomes (1) (A, b, C) and (A, B, C) and (2) (d, e, F) and (D, e, F). How many haploid forms are there?

(e) What is the maximum number of different offspring possible from a mating pair of organisms as in (d)? What is the minimum number? How could the minimum number be achieved?

(f) Work out a graph showing how the number of haploid forms varies with (i) number of chromosomal pairs or (ii) number of genes per chromosomal pairs. Which effect leads to more possibilities?

2. For a given diploid two-allele locus, the initial fractions of genotypes are AA : p, Aa : q, and aa : r (hence $p + q + r = 1$). Recall that the frequencies in the next generation will be $p_{\mathrm{AA}} = x$, $p_{\mathrm{Aa}} = y$, and $p_{\mathrm{aa}} = z$, where

$$x = \left(p + \frac{1}{2}q\right)^2, \qquad y = 2\left(p + \frac{1}{2}q\right)\left(r + \frac{1}{2}q\right), \qquad z = \left(r + \frac{1}{2}q\right)^2.$$

Under the assumption that the various genotypes are selected neither for nor against, show that these ratios will be maintained in all future generations; i.e., show that

$$x = \left(x + \frac{1}{2}y\right)^2,$$
$$y = 2\left(x + \frac{1}{2}y\right)\left(z + \frac{1}{2}y\right),$$
$$z = \left(z + \frac{1}{2}y\right)^2.$$

Hence when the Hardy–Weinberg principle holds, genotype frequencies stabilize in one generation,

MAPLE

```
> x:=(p+q/2)^2;
> y:=2*(p+q/2)*((1-p-q)+q/2);
> z:=1-x-y;
> X:=(x+y/2)^2;
```

```
> simplify(X-x);
```

MATLAB

```
> p=.7; q=0; r=.3;
> x=(p+q/2)^2
> y=2*(p+q/2)*(r+q/2)
> z=(r+q/2)^2
> X=(x+y/2)^2
> Y=2*(x+y/2)*(z+y/2)
> Z=(z+y/2)^2
```

etc.

3. In this problem, we want to see how many homozygous recessives for a trait result from homozygous parents and how many result from heterozygous parents (see Figure 13.7.2). The question is, given an aa progeny, what is the probability the parents were aa × aa?

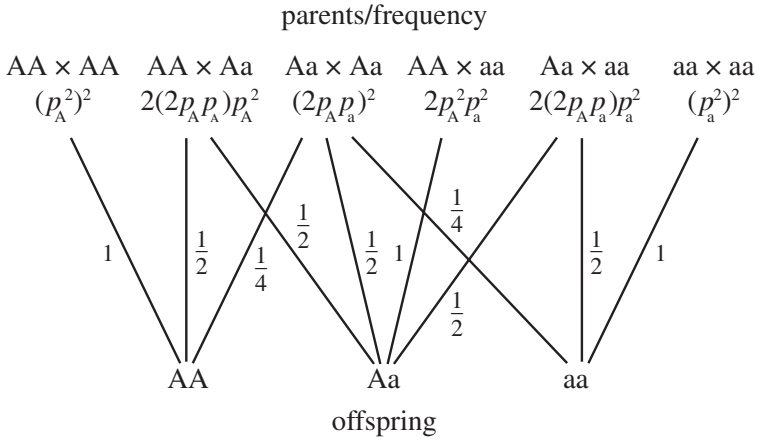


Fig. 13.7.2.

Since the progeny is known to be aa, the universe for this problem is the paths of the tree leading to aa; its frequency is given by

$$u = (2p_Ap_a)^2 \cdot \frac{1}{4} + 2(2p_Ap_a)p_a^2 \cdot \frac{1}{2} + (p_a^2)^2 \cdot 1.$$

So the relative frequency in which this occurs via aa × aa parents is

$$\frac{(p_a^2)^2 \cdot 1}{u}.$$

(a) Calculate the probable parentage of an aa progeny via Aa × Aa genotypes and Aa × aa genotypes.

(b) Make three graphs of these probable parentages over the range of frequencies of allele a from 0.25 to 0.001, say.

(c) If $p_a = 0.01$, then what is the chance that an aa individual had heterozygous parents? Same question for at least one heterozygous parent.

For part (a):

MAPLE
```
> u:=(2*pA*pa)^2*(1/4)+2*(2*pA*pa)*pa^2*(1/2)+(pa^2)^2;
> pA:=1-pa;
> aaxaa:= pa->(pa^2)^2/u;
> #similarly for AaxAa and Aaxaa
```

For part (b):

MAPLE
```
> plot(aaxaa(pa),pa=0.001..0.25);
```

MATLAB
```
> pa=(.001:.001:.25);
> pA=1-pa;
> u=(2*pA.*pa).^2*(1/4)+2*(2*pA.*pa).*pa.^2*(1/2)+(pa.^2).^2;
> aaxaa=pa.^4./u;
> plot(pa,aaxaa)
```

4. This problem refers to the sex-linked loci subsection of Section 13.6.

(a) Let the starting fraction of allele a in males be $M_0 = 0.1$ and in females be $F_0 = 0.3$. By performing the matrix calculation

$$\begin{bmatrix} M_{t+1} \\ F_{t+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} M_t \\ F_t \end{bmatrix}$$

repeatedly, find the limiting fractions M_∞ and F_∞. What is the ratio $\frac{M_\infty}{F_\infty}$?

(b) Do the same for the starting ratios $M_0 = 1$ and $F_0 = 0$. What is the limiting ratio $\frac{M_\infty}{F_\infty}$?

(c) Let T be the matrix in part (a):

$$T = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Show that T satisfies

$$T \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

We say that this column vector, with both components 1, is a right eigenvector for T with eigenvalue 1.

(d) As in part (a), iterate the calculation

$$(0\ 1) = (1\ 0) \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

this time multiplying the matrix on the left by the vector, to obtain the limit. This will represent the ultimate distribution of the original male vs. female alleles. Show that

$$\left(\tfrac{1}{3}\ \tfrac{2}{3}\right)$$

is a left eigenvector for T. What is the eigenvalue?

MAPLE

```
> with(LinearAlgebra):
> T:=Matrix([[0,1],[1/2,1/2]]);
> v:=Vector([1/10,3/10]);
> #the next to see the trend
> for n from 0 to 10 do
    evalf((T^n).v);
  od;
> #Now get the eigenvalue and eigenvector
> Eigenvectors(T);
```

MATLAB

```
> T=[0 1; .5 .5]
> x=[1 0]
> for k=1:20
> x=x*T
> end
> y=x % print eigenvector
> lambda = y(1)/x(1) % and evalue
```

In the output, the first item is the eigenvalue (as above), the second is its multiplicity (how many times repeated, should be 1 here), and the third is the eigenvector. Eigenvectors may be multiplied by any constant, so if $\binom{1}{1}$ is an eigenvector, so is $\binom{3}{3}$.

5. Two hypotheses that explain the greater incidence of early baldness in males than in females are (1) an autosomal dominance that is normally expressed only in males and (2) an X-linked recessive. If the first is correct and Q is the frequency of the gene for baldness, what proportion of the sons of bald fathers are expected to be bald? What proportion are from nonbald fathers? What are the corresponding expectations for the X-linked recessive hypothesis.

Data gathered by Harris[10] found that 13.3% of males in the sample were prematurely bald. Of 100 bald men, 56 had bald fathers. Show that this is consistent with the sex-limited dominance hypothesis but not the sex-linked recessive. (Note that it is easier to get data about the fathers of bald sons than it is to wait for the sons of bald fathers to grow up to get data about bald sons.)

6. Suppose an organism that is capable of both sexual and asexual reproduction reproduces c fraction of the time asexually (by cloning) and $1 - c$ fraction of the time sexually with random mating. Let P_t be the fraction of the genotype AA in generation t and let p be the frequency of allele A. Assumethat c is independent of genotype, and consequently p will remain constant from generation to generation. However, the frequency of genotype AA can change. Using the Hardy–Weinberg principle, show that the change in this fraction is given by

$$P_{t+1} = cP_t + (1 - c)p^2.$$

Find the limiting fraction P_∞.

[10] H. Harris, The inheritance of premature baldness in men, *Ann. Eugenics*, **13** (1946), 172–181.

```
  MAPLE
> restart;
> # first try
> F:=x–>c*x+(1-c)*p^2;
> x:=0; y:=F(x); w:=F(y);
> simplify(%);
> restart;
> F:=x–>c*x+(1-c)*p^2;
> y[0]:=0;
> for n from 1 to 10 do
    y[n]:=F(y[n-1]):
  od:
> simplify(y[10]);

  MATLAB
> c=1/3; P(1)=.7;
> p=.8
> for k=1:30
>    P(k+1)=c*P(k)+(1-c)*p^2;
> end
> P
> c=.9;
> for k=1:30
>    P(k+1)=c*P(k)+(1-c)*p^2;
> end
> P % how does P(infinity) depend on c?
> p=.3
> for k=1:50
>    P(k+1)=c*P(k)+(1-c)*p^2;
> end
> P % compare P(infinity) with p^2
```

7. Suppose the frequency of a recessive allele is p (equal to $\frac{1}{1000}$, say); therefore, the frequency of homozygotes under the hypothesis of random mating will be p^2. But what if mating is not random? In this problem we want to investigate this somewhat.

First, suppose the species is capable of self-fertilization. Then clearly the offspring of a homozygous adult will again be homozygous. On the other hand, the heterozygous Aa will produce A and a haploid cells in 50–50 mix as before. Hence as before, an offspring will be AA with $\frac{1}{4}$ chance, aa with $\frac{1}{4}$ chance, and Aa with $\frac{1}{2}$ chance. We record these observations in the following 3×3 matrix:

$$T = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{bmatrix}.$$

In this, the rows correspond to the genotypes AA, Aa, and aa in that order and so do the columns.

Next, suppose we start out with a mix of genotypes, say, their fractions are p, q, and r, respectively, $p + q + r = 1$. Then after one generation, the new fractions p', q', and r' will be given by the matrix product

$$(p'\ q'\ r') = (p\ q\ r)T.$$

(a) Using specific values for the starting fractions, find the limiting fractions after many generations.

Next, consider parent/child matings and calculate the probability that a homozygous recessive aa will be the result. First, condition on the parent (sketch a tree diagram) that from the root node, there will be three edges corresponding to the possibilities that the parent is AA, Aa, or aa. The AA branch cannot lead to an aa grandoffspring, so there is no need to follow that edge further. The Aa parent occurs with frequency $2p(1-p)$, as we have seen, and the aa parent with frequency p^2.

Next, condition on the genotype of the child. Use the Hardy–Weinberg principle for probabilities of alleles A and a. Starting from the Aa node, the possibilities are AA with probability $\frac{1}{2}(1-p)$, Aa with probability $\frac{1}{2}p + \frac{1}{2}(1-p) = \frac{1}{2}$, and finally aa with probability $\frac{1}{2}p$. You do the possibilities from the aa node.

Now assign the offspring probabilities using Mendelian genetics. From the Aa node along the path from root through the Aa parent, the probability of an aa offspring is $\frac{1}{4}$. From the aa node through the Aa parent, the probability is $\frac{1}{2}$, and so on.

(b) Altogether, the result should be

$$P(\text{aa offspring}) = \frac{1}{2}p\left(\frac{3}{4} + \frac{3}{2}p - p^2\right).$$

Finally, consider sibling matings. As in part (a) above, we want to investigate the trend of the population toward homozygosity. Starting with the parents, there are six possible matings by genotype, AA × AA, AA × Aa, and so on through aa × aa. Consider the AA × Aa parents. Their offspring are AA and Aa both with frequency $\frac{1}{2}$. Therefore, the sibling mating possibilities are AA × AA with frequency $\frac{1}{4}$, AA × Aa with frequency $\frac{1}{2}$, and Aa × Aa with frequency $\frac{1}{4}$.

Justify the rest of Table 13.7.1. The corresponding transition matrix T is

Table 13.7.1.

Parent	**Sibling mating frequencies**					
genotypes	**AA × AA**	**AA × Aa**	**Aa × Aa**	**AA × aa**	**Aa × aa**	**aa × aa**
AA × AA	1	0	0	0	0	0
AA × Aa	$\frac{1}{4}$	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0
Aa × Aa	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{16}$
AA × aa	0	0	1	0	0	0
Aa × aa	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
aa × aa	0	0	0	0	0	1

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{16} & \frac{1}{4} & \frac{1}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{16} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(c) Make up an initial distribution of genotypes ($p\ q\ r\ s\ t\ u$), track the change in distribution over a few generations, and find the limiting distribution.

Questions for Thought and Discussion

1. Discuss the concept of fitness as it is used in the Darwinian model. What kinds of selection factors might be involved in the case of humans?
2. A woman with type A blood has a child with type O blood. The woman alleges that a certain man with type B blood is the father. Discuss her allegation and reach a conclusion, if possible.
3. In *Drosophila*, females are XX and males are XY. On the X chromosome, there is an eye-color gene such that red is dominant to eosin and to white, and eosin is dominant to white. What is the result of crossing an eosin-eyed male with a red-eyed female whose mother had white eyes?
4. Mitosis is a conservative form of cell replication, because each daughter cell gets an exact copy of the genetic material that the parent cell had. How can we explain the fact that most of our tissues were formed by mitosis and yet are different?
5. Suppose there is an organism that reproduces only by self-fertilization, which is the highest degree of inbreeding. Start with a heterozygote for a single property and let it and its descendants reproduce by self-fertilization for three generations. Note how the fraction of homozygotes increases with each generation. What implication does this have if the recessive allele is harmful? Or suppose it is not harmful?
6. Combining the concepts of the central dogma of genetics with that of meiosis, trace the path of hereditary control of cellular chemistry from one generation to another.
7. In a hypothetical laboratory animal, a solid-color allele is dominant to striped, and long hair is dominant to short hair. What is the maximum number of phenotypes that could result from the mating of a long, solid animal with a short, striped animal?

References and Suggested Further Reading

[1] Cell division and reproduction:
W. S. Beck, K. F. Liem, and G. G. Simpson, *Life: An Introduction to Biology*, 3rd ed., Harper–Collins, New York, 1991.

[2] GENETICS:
D. T. Suzuki, A. J. F. Griffiths, J. H. Miller, and R. C. Lewontin, *An Introduction to Genetic Analysis*, W. H. Freeman, 3rd ed., New York, 1986.
[3] SEXUAL SELECTION IN HUMANS:
D. M. Buss, The strategies of human mating, *Amer. Sci.*, **82** (1994), 238–249.
[4] MATHEMATICAL GENETICS:
J. F. Crow and M. Kimura, *An Introduction to Population Genetics Theory*, Harper and Row, New York, 1970.
[5] MATHEMATICAL GENETICS:
J. B. S. Haldane, A mathematical theory of natural and artificial selection V: Selection and mutation, *Cambridge Philos. Soc. Proc.*, **23**-VII (1927), 838–844.