3. The offspring of phenotypically different parents can be reared together in a uniform environment, often referred to as a **common garden**. Differences among offspring from different parents that persist in such circumstances are likely to have a genetic basis. Propagation in the common garden for at least two generations is advisable to distinguish genetic from maternal effects.

Fundamental Principles of Genetic Variation in Populations

We embark now on our study of genetic variation and the factors that influence it—that is, the factors that cause evolution within species. *The definitions, concepts, and principles introduced here are absolutely essential for understanding evolutionary theory.* We begin with a short verbal description of these ideas, followed by an explanation of a very important formal model.

At any given gene locus, a population may contain two or more alleles that have arisen over time by mutation. Sometimes one allele is by far the most common (and may be called the WILD TYPE), and the others are very rare; sometimes two or more of the alleles are each quite common. The relative commonness or rarity of an allele—its proportion of all gene copies in the population—is called the **allele frequency** (sometimes imprecisely referred to as the "gene frequency"). In sexually reproducing populations, the alleles, carried in eggs and sperm, become combined into homozygous (two copies of the same allele) and heterozygous (one copy each of two different alleles) genotypes. The **genotype frequency** is the proportion of a population that has a certain genotype (Figure 9.3). The proportions of the different genotypes are related to the allele frequencies in simple but important ways, as we will soon see. If the genotypes differ in a phenotypic character, the amount of variation in that character will depend not only on how different the genotypes are from each other, but also on the relative abundance (the frequencies) of the genotypes—which in turn depends on the allele frequencies.

Any alteration of the genotype frequencies in one generation will alter the frequencies of the alleles carried by the population's gametes when reproduction occurs, so the genotype frequencies of the following generation will be altered in turn. *Such alteration, from*

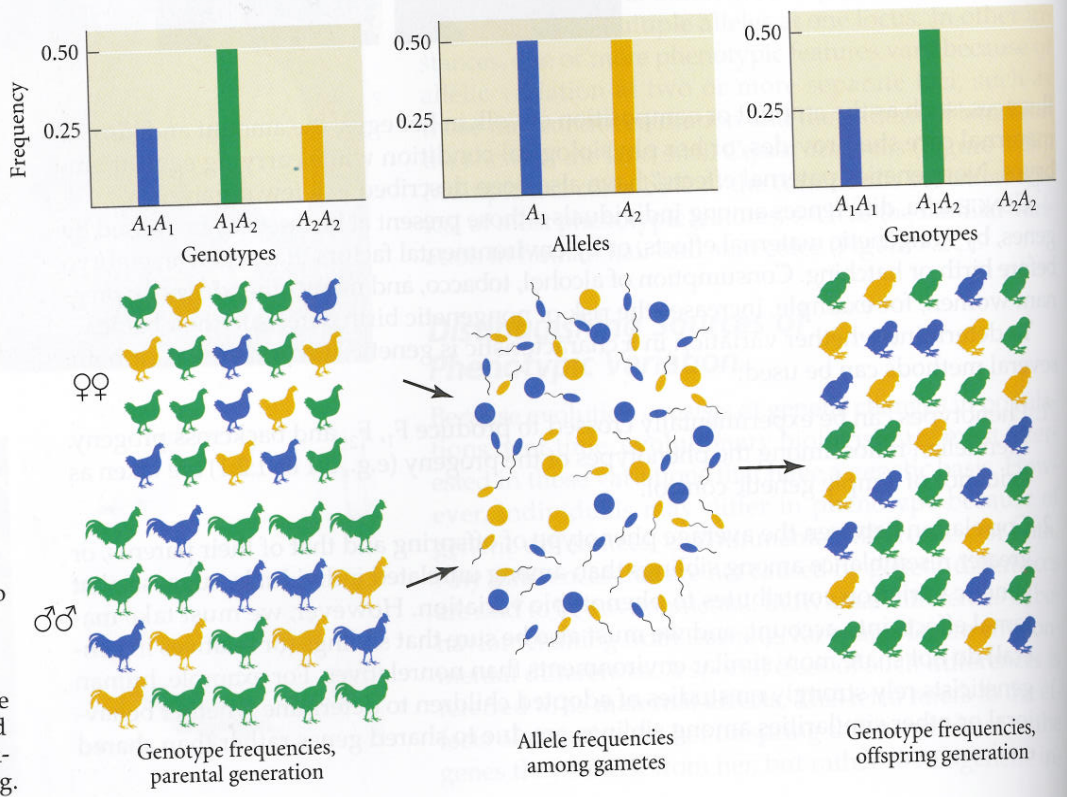


Figure 9.3 For a locus with two alleles (A_1, A_2), this diagram shows the frequency of three genotypes among females and males in one generation, the allele frequencies among their eggs and sperm, and the genotype frequencies among the resulting offspring.

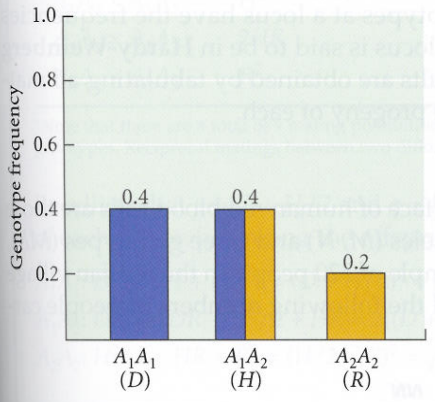
generation to generation, is the central process of evolutionary change. However, the genotype and allele frequencies do not change on their own; something has to change them. *The factors that can cause the frequencies to change are the causes of evolution.*

Frequencies of alleles and genotypes: The Hardy-Weinberg principle

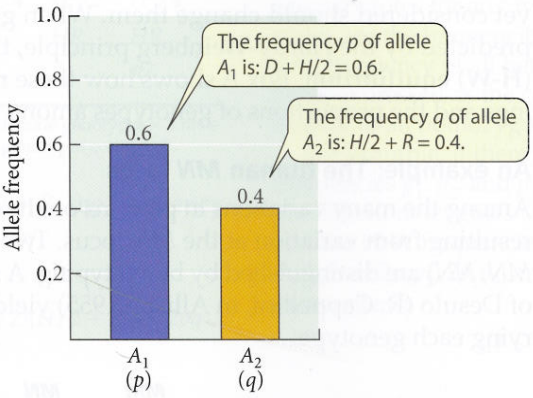
Imagine that a diploid population has 1000 individuals, and that for a particular gene locus in that population, there exist only two alleles, A_1 and A_2. Thus there are three possible genotypes for this locus: A_1A_1, A_2A_2 (both homozygous), and A_1A_2 (heterozygous).

Let us say 400 individuals have the genotype A_1A_1, 400 are A_1A_2, and 200 are A_2A_2. Let the frequencies of genotypes A_1A_1, A_1A_2, and A_2A_2 be represented by D, H, and R respectively, and let the frequencies of the alleles A_1 and A_2 be represented by p and q. Thus the genotype frequencies are $D = 0.4$, $H = 0.4$, and $R = 0.2$, and the allele frequencies are $p = 0.6$ and $q = 0.4$ (Figure 9.4A,B). The frequency of allele A_1 is the sum of all the alleles carried by A_1A_1 homozygotes, plus one-half of those carried by heterozygotes. Hence $p = D + H/2 = 0.6$. Likewise, the frequency of allele A_2 is calculated as $q = R + H/2 = 0.6$.

(A) Parental genotype frequencies (not in equilibrium)



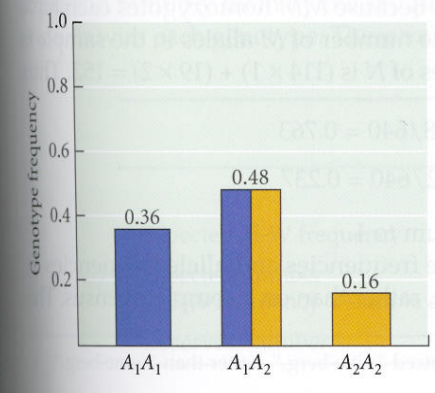
(B) Parental allele frequencies



(C)

Offspring	Probability of a given mating producing the genotype		
A_1A_1	$\Pr[A_1 \text{ egg}] \times \Pr[A_1 \text{ sperm}] = p \times p = p^2$	0.6^2	$= 0.36$
A_1A_2	$\Pr[A_1 \text{ egg}] \times \Pr[A_2 \text{ sperm}] = p \times q = pq$	$0.6 \times 0.4 = 0.24$	$= 0.48$
	$\Pr[A_2 \text{ egg}] \times \Pr[A_1 \text{ sperm}] = q \times p = pq$	$0.4 \times 0.6 = 0.24$	
A_2A_2	$\Pr[A_2 \text{ egg}] \times \Pr[A_2 \text{ sperm}] = q \times q = q^2$	0.4^2	$= 0.16$

(D) Offspring genotype frequencies



(E) Offspring allele frequencies

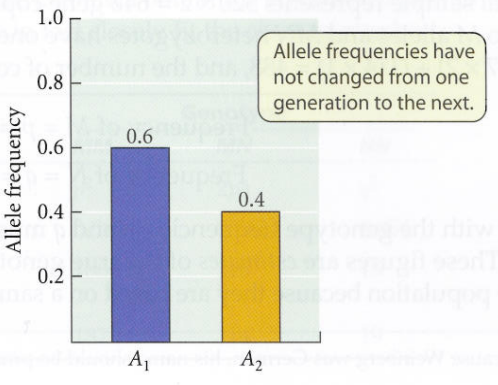


Figure 9.4 A hypothetical example illustrating attainment of Hardy-Weinberg genotype frequencies after one generation of random mating. (A) Genotype frequencies in the parental population. (B) Allele frequencies in the parental population. (C) Calculation of genotype frequencies among the offspring. (D) Genotype frequencies among the offspring, if all assumptions of the Hardy-Weinberg principle hold. (E) Allele frequencies among the offspring.

Now let us assume that each genotype is equally represented in females and males, and that *individuals mate entirely at random*. (This is an entirely different situation from the crosses encountered in elementary genetics exercises, in which familiar Mendelian ratios are produced by nonrandomly crossing females of *one* genotype with males of *one* genotype.) Random mating is conceptually the same as mixing eggs and sperm at random, as might occur, for instance, in those aquatic animals that release gametes into the water. A proportion p of eggs carry allele A_1, q carry A_2, and likewise for sperm. The probabilities (Pr) for all possible allele combinations are shown in Figure 9.4C, and the resulting offspring genotype and allele frequencies are shown in Figure 9.4D and E.

Notice that in the new generation, the allele frequencies are still $p = D + H/2 = 0.6$ and $q = H/2 + R = 0.4$. *The allele frequencies have not changed from one generation to the next*, although the alleles have become distributed among the three genotypes in new proportions. These proportions, denoted here as p^2, $2pq$, and q^2, constitute the HARDY-WEINBERG DISTRIBUTION of genotype frequencies.

The principle that genotypes will be formed in these frequencies as a result of random mating is named after G. H. Hardy and W. Weinberg,* who independently calculated these results in 1908. The Hardy-Weinberg principle, broadly stated, holds that whatever the initial genotype frequencies for two alleles may be, *after one generation of random mating, the genotype frequencies will be* $p^2{:}2pq{:}q^2$. Moreover, both these genotype frequencies and the allele frequencies *will remain constant in succeeding generations* unless factors not yet considered should change them. When genotypes at a locus have the frequencies predicted by the Hardy-Weinberg principle, the locus is said to be in **Hardy-Weinberg (H-W) equilibrium**. Box A shows how these results are obtained by tabulating all matings and the proportions of genotypes among the progeny of each.

An example: The human *MN* locus

Among the many variations in proteins on the surface of human red blood cells are those resulting from variation at the *MN* locus. Two alleles (*M*, *N*) and three genotypes (*MM*, *MN*, *NN*) are distinguished by blood typing. A sample of 320 people in the Sicilian village of Desulo (R. Ceppellini, in Allison 1955) yielded the following numbers of people carrying each genotype:

MM	*MN*	*NN*
187	114	19

We can now estimate the *frequency* of each genotype as the proportion of the total sample (320) that carries that genotype. Thus,

$$\text{Frequency of } MM = D = 187/320 = 0.584$$

$$\text{Frequency of } MN = H = 114/320 = 0.356$$

$$\text{Frequency of } NN = R = 19/320 = 0.059$$

Note that these frequencies, or proportions, must sum to 1.

Now we can calculate the allele frequencies. Each person carries two gene copies, so the total sample represents $320 \times 2 = 640$ gene copies. Because *MM* homozygotes each have two *M* alleles and *MN* heterozygotes have one, the number of *M* alleles in the sample is $(187 \times 2) + (114 \times 1) = 488$, and the number of copies of *N* is $(114 \times 1) + (19 \times 2) = 152$. Thus

$$\text{Frequency of } M = p = 488/640 = 0.763$$

$$\text{Frequency of } N = q = 152/640 = 0.237$$

As with the genotype frequencies, p and q must sum to 1.

These figures are *estimates* of the true genotype frequencies and allele frequencies in the population because they are based on a sample, rather than on a complete census. The

*Because Weinberg was German, his name should be pronounced "vine-berg," rather than "wine-berg."

BOX 9A Derivation of the Hardy-Weinberg Distribution

Let the frequencies of genotypes A_1A_1, A_1A_2, and A_2A_2 be D, H, and R, respectively, and let the frequencies of alleles A_1 and A_2 be p and q. The genotype frequencies sum to 1, as do the allele frequencies. The probability of a mating between a female of any one genotype and a male of any one genotype equals the product of the genotype frequencies. For reciprocal crosses between two different genotypes, we take the sum of the probabilities. The mating frequencies and offspring produced are:

Mating	Probability of mating*	Offspring genotype		
		A_1A_1	A_1A_2	A_2A_2
$A_1A_1 \times A_1A_1$	D^2	D^2		
$A_1A_1 \times A_1A_2$	$2DH$	DH	DH	
$A_1A_1 \times A_2A_2$	$2DR$		$2DR$	
$A_1A_2 \times A_1A_2$	H^2	$H^2/4$	$H^2/2$	$H^2/4$
$A_1A_2 \times A_2A_2$	$2HR$		HR	HR
$A_2A_2 \times A_2A_2$	R^2			R^2

*Note that there are a total of 9 mating possibilities: 3 female genotypes × 3 male genotypes. Reciprocal matings between two different genotypes count as 2.

Recalling that $p = D + H/2$ and $q = H/2 + R$, the frequency of each genotype among the offspring is:

A_1A_1: $D^2 + DH + H^2/4 = (D + H/2)^2 = p^2$

A_1A_2: $DH + 2DR + H^2/2 + HR = 2[(D + H/2)(H/2 + R)] = 2pq$

A_2A_2: $H^2/4 + HR + R^2 = (H/2 + R)^2 = q^2$

This result may also be obtained by recognizing that if genotypes mate at random, gametes, and therefore genes, also unite at random to form zygotes. Because the probability that an egg carries allele A_1 is p, and the same is true for sperm, the probability of an A_1A_1 offspring is p^2. A Punnett square shows the probability of each gametic union:

		Sperm	
		A_1 (p)	A_2 (q)
Eggs	A_1 (p)	A_1A_1 (p^2)	A_1A_2 (pq)
	A_2 (q)	A_1A_2 (pq)	A_2A_2 (q^2)

The Hardy-Weinberg principle can be extended to more complicated patterns of inheritance, such as multiple alleles. For example, if there are k alleles (A_1, A_2, ... A_k), the H-W frequency of homozygotes for any allele A_i is p_i^2, and that of heterozygotes for any two alleles A_i and A_j is $2p_ip_j$, where p_i and p_j are the frequencies of the two alleles in question. The total frequency of all heterozygotes combined (H) is sometimes expressed as the complement of the summed frequency of all homozygous genotypes, or $H = 1 - \sum p_i^2$.

For example, if there are three alleles A_1, A_2, and A_3 with frequencies p_1, p_2, and p_3, the H-W frequencies of the three possible homozygotes (A_1A_1, A_2A_2, A_3A_3) are p_1^2, p_2^2, and p_3^2; and the H-W frequencies of the heterozygotes (A_1A_2, A_1A_3, A_2A_3) are $2p_1p_2$, $2p_1p_3$, and $2p_2p_3$. The total frequency of heterozygotes is

$$H = 1 - (p_1^2 + p_2^2 + p_3^2)$$

larger the sample, the more confident we can be that we are obtaining accurate estimates of the true values. (This assumes that the sample is a *random* one, i.e., that our likelihood of collecting a particular type is equal to the true frequency of that type in the population.)

Our hypothetical example (see Figure 9.4) showed that for a given set of allele frequencies, a population might or might not have H-W genotype frequencies. (The frequencies 0.36, 0.48, and 0.16 in the offspring generation were in H-W equilibrium, but the frequencies 0.40, 0.40, and 0.20 in the parental generation were not.) Our real example shows a close fit to the H-W frequency distribution. The allele frequencies of *M* and *N* were $p = 0.763$ and $q = 0.237$. If we calculate the *expected genotype frequencies* under the H-W principle and multiply them by the sample size (320), we obtain the *expected* number of individuals with each genotype. These values in fact closely fit the *observed* numbers:

	Genotype		
	MM	*MN*	*NN*
	p^2	$2pq$	q^2
Expected H-W frequency	0.582	0.362	0.056
Expected number (H-W frequency × sample size)	186	116	18
Observed number	187	114	19

The significance of the Hardy-Weinberg principle: Factors in evolution

The Hardy-Weinberg principle is the foundation on which almost all of the theory of population genetics of sexually reproducing organisms—which is to say, most of the genetic theory of evolution—rests. Its importance cannot be overemphasized. We will encounter it in the theory of natural selection and other causes of evolution. It has two important implications: First, genotype frequencies attain their H-W values after a single generation of random mating. If some factor in the past had caused genotype frequencies to deviate from H-W values, a single generation of random mating would erase the imprint of that history. Second, according to the H-W principle, not only genotype frequencies, but also allele frequencies, remain unchanged from generation to generation. A new mutation, for example, will remain at its initial very low allele frequency indefinitely.

Like any mathematical formulation, the Hardy-Weinberg principle holds only under certain assumptions. Since allele frequencies and genotype frequencies often do change (i.e., evolution occurs), the assumptions of the Hardy-Weinberg formulation must not always hold. Therefore *the study of genetic evolution consists of asking what happens when one or more of the assumptions are relaxed.*

The most important assumptions of the Hardy-Weinberg principle are the following:

1. *Mating is random.* If a population is not **panmictic**—that is, if members of the population do not mate at random—the genotype frequencies may depart from the ratios $p^2{:}2pq{:}q^2$.
2. *The population is infinitely large* (or so large that it can be treated as if it were infinite). The calculations are made in terms of probabilities. If the number of events is finite, the actual outcome is likely to deviate, purely by chance, from the predicted outcome. If we toss an infinite number of unbiased coins, probability theory says that half will come up heads, but if we toss only 100 coins, we are likely not to obtain exactly 50 heads, purely by chance. Similarly, among a finite number of offspring, both the genotype frequencies and the allele frequencies may differ from those in the previous generation, *purely by chance.* Such random changes are called **random genetic drift**.
3. *Genes are not added from outside the population.* Immigrants from other populations may carry different frequencies of A_1 and A_2; if they interbreed with residents, this will alter allele frequencies and, consequently, genotype frequencies. Mating among individuals from different populations is termed **gene flow** or **migration**. We may restate this assumption as: There is no gene flow.
4. *Genes do not mutate from one allelic state to another.* Mutation, as we have seen (in Chapter 8), can change allele frequencies, although usually very slowly. The Hardy-Weinberg principle assumes no mutation.
5. *All individuals have equal probabilities of survival and of reproduction.* If these probabilities differ among genotypes (i.e., if there is a consistent difference in the genotypes' rates of survival or reproduction), then the frequencies of alleles and/or genotypes may be altered from one generation to the next. Thus the Hardy-Weinberg principle assumes that there is *no natural selection* affecting the locus.

Inasmuch as nonrandom mating, chance, gene flow, mutation, and selection can alter the frequencies of alleles and genotypes, these are the major factors that cause evolutionary change within populations.

Certain subsidiary assumptions of the Hardy-Weinberg principle can be important in some contexts. First, the principle, as presented, applies to autosomal loci; it can also be modified for sex-linked loci (which have two copies in one sex and one in the other). Second, the principle assumes that alleles segregate into a heterozygote's gametes in a 1:1 ratio. Deviations from this ratio are known and are called SEGREGATION DISTORTION or MEIOTIC DRIVE.

If the assumptions we have listed hold true for a particular locus, that locus will display Hardy-Weinberg genotype frequencies. But if we observe that a locus fits the Hardy-Weinberg frequency distribution, we cannot conclude that the assumptions hold true! For example, mutation or selection may be occurring, but at such a low rate that we cannot

detect a deviation of the genotype frequencies from the expected values. Or, under some forms of natural selection, we might observe deviations from Hardy-Weinberg equilibrium if we measure genotype frequencies at one stage in the life history, but not at other stages.

Frequencies of alleles, genotypes, and phenotypes

At Hardy-Weinberg equilibrium, the frequency of heterozygotes is greatest when the alleles have equal frequencies (Figure 9.5). When an allele is very rare, almost all its carriers are heterozygous. If one allele (say, A_1) is dominant, masking a recessive allele (A_2) in heterozygotes, the dominant phenotype (genotypes A_1A_1 and A_1A_2) makes up $p^2 + 2pq$ of the population. Because almost all copies of a rare recessive allele are carried by heterozygotes, the allele may not be detected readily. If, for example, the frequency of a recessive allele is $q = 0.01$, only $(0.01)^2 = 0.0001$ of the population displays this recessive allele in its phenotype. Thus populations can carry **concealed genetic variation**. However, *a dominant allele may well be less common than a recessive allele.* "Dominance" refers to an allele's phenotypic effect in the heterozygous condition, not to its numerical prevalence. In the British moth *Cleora repandata*, for example, black coloration is inherited as a dominant allele, and "normal" gray coloration is recessive. In a certain forest, 10 percent of the moths were found to be black (Ford 1971).

Inbreeding

The Hardy-Weinberg principle assumes that a population is panmictic. What if mating does not occur at random? One form of nonrandom mating is **inbreeding**, which occurs when individuals are more likely to mate with relatives than with nonrelatives, or, more generally, when the gene copies in uniting gametes are more likely to be identical by descent than if they joined at random. Gene copies are said to be **identical by descent** if they have descended, by replication, from a common ancestor, relative to other gene copies in the population.

Box B shows that as inbreeding proceeds, the frequency of each homozygous genotype increases and the frequency of heterozygotes decreases by the same amount. The frequency of heterozygotes is $H = H_0(1 - F)$, where H_0 is the heterozygote frequency expected if the locus were in H-W equilibrium, and F is the **inbreeding coefficient**.

The ways in which genotype frequencies change are easily seen if we envision mating of individuals only with their closest relatives—namely, themselves—by self-fertilization. (You may have trouble imagining this if you think about people, but think about plants and you will see that it is not only plausible, but quite common.) The homozygous genotype A_1A_1 can produce only A_1 eggs and A_1 sperm, and thus only A_1A_1 offspring; likewise, A_2A_2 individuals produce only A_2A_2 offspring. Heterozygotes produce A_1 and A_2 eggs in equal proportions, and likewise for sperm. When these eggs and sperm join at random, ¼ of the offspring are A_1A_1, ½ are A_1A_2, and ¼ are A_2A_2. (The two allele copies carried by the homozygous offspring are identical by descent.) Thus the frequency of heterozygotes is halved in each generation, and eventually reaches zero. Conversely, F increases as inbreeding continues; in fact, F can be estimated by the deficiency of heterozygotes relative to the H-W equilibrium value (see Box B). If CONSANGUINEOUS mating (mating among relatives) is a consistent feature of a population, F will increase over generations at a rate that depends on how closely related the average pair of mates is.

The most extreme form of inbreeding, **self-fertilization** or **selfing**, occurs in many species of plants and a few animals. The wild oat *Avena fatua*, for example, reproduces mostly by selfing, and it has a low frequency of heterozygotes at the sev-

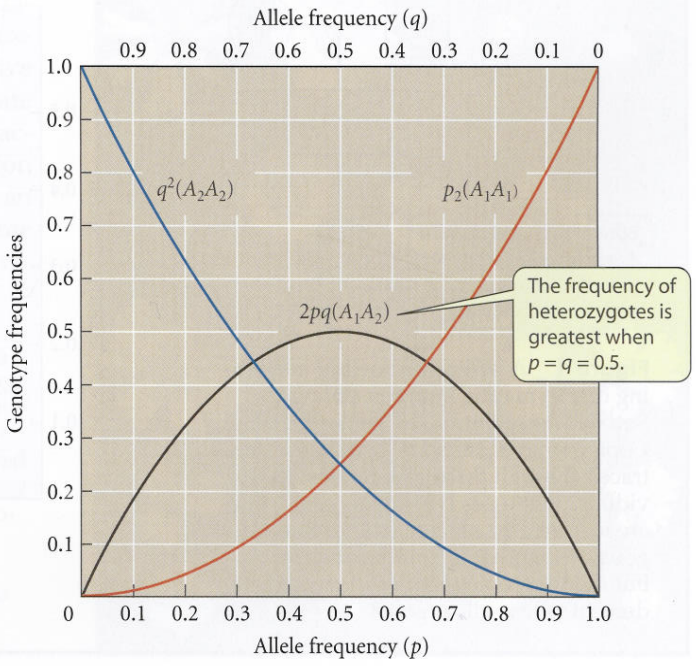


Figure 9.5 Hardy-Weinberg genotype frequencies as a function of allele frequencies at a locus with two alleles. Heterozygotes are the most common genotype in the population if the allele frequencies are between ⅓ and ⅔. (After Hartl and Clark 1989.)