

I. Brief Overview

There are five major forces (and some others of interest only to specialists) that can change allele and/or genotype frequencies in populations. These forces are the causes of evolution. In their absence, the Hardy-Weinberg law applies, allele and genotype frequencies will not change, and there will be no evolution. The five major forces are listed in the table below, along with columns telling whether the force is a strong or a weak force for causing evolution, or is dependent on certain conditions. If the force is very dependent on conditions, the conditions under which it would cause more evolutionary change are listed in the third column.

<u>Force</u>	<u>Strong, Weak or Dependent?</u>	<u>If dependent, condition(s) under which it would be stronger</u>
Selection	Strong	-----
Migration	Dependent	1. more migration 2. greater allele frequency difference among the populations
Mutation	Weak	-----
Non-Random Mating	Strong or Dependent	the mating is very non-random
Genetic Drift	Dependent	a very small population

We will now consider each of the forces in more detail.

II. Selection

Natural selection is perhaps the most important, by no means the only, force causing evolutionary change. Populations of many organisms are rather small, and genetic drift and non-random mating usually have significant effects in small populations. Migration also plays a large role in many organisms.

Previously, we considered the three types of selection (stabilizing, directional, disruptive) when they acted on a polygenic character, that is, a character trait whose genetic component is made up of many alleles. Examples are height, weight, and top running speed — characters for which there is usually a bell-shaped distribution in the population.

Below, we consider models of selection acting on one locus with two alleles. We start with a Hardy-Weinberg population, let selection act, then randomly mate the survivors to produce the next generation. We then compare the allele frequencies in the starting generation with those in the succeeding generation to look at the amount of genetic change produced by selection. We can always develop an equation relating gene frequency in one generation to gene frequency in the succeeding generation; such an equation is called a recurrence equation, because it **recurs** each generation. That is, one can find the allele frequency three generations into the future by using the equation 3 times, each time using the results of the previous calculation as the starting point for the succeeding calculation. This is illustrated below. On occasion, we can **solve** such a recurrence equation — such a solution expresses gene frequency at any number 'n' generations in the future without having to use the recurrence equation n times. This is also illustrated below.

A. The general case. In the following example, we take a plant that has two alleles at a flower color locus, without dominance. Thus, we have three phenotypes, each corresponding to a genotype. This information may be gotten by using the techniques of Mendel. We start out with a population in the Hardy-Weinberg allele frequencies, and assign relative fitnesses to each genotype. Traditionally, these fitnesses are denoted W_1 , W_2 , and W_3^* .

				SUM
Phenotypes	red	pink	white	----
Hardy-Weinberg Genotypic frequencies (note that they sum to 1)	p^2	$2pq$	q^2	1
Relative Fitnesses	W_1	W_2	W_3	-----
Genotypic frequencies after selection (note that they no longer sum to 1, but that they still represent the whole population. Thus we must make them sum to 1 by dividing each by the total of all of them)	p^2W_1	$2pqW_2$	q^2W_3	less than 1 unless all fitnesses are 1

We now need to calculate the new frequencies of the A_1 and A_2 alleles, which are denoted p_1 and q_1 , indicating that these are the frequencies 1 generation later (p_2 and q_2 would represent the frequencies two generations later, etc. Sometimes p_0 and q_0 are used to denote the allele frequencies in the initial or "0th" generation; sometimes the initial allele frequencies are not subscripted at all. I have not subscripted the initial p and q in this example).

In the next generation, p_1+q_1 must also =1, since they represent all the allele frequencies. We calculate p_1 by taking the frequency of the A_1A_1 after selection, adding $\frac{1}{2}$ the frequency of the heterozygotes after selection, and dividing by the sum of all the post-selection genotype frequencies. Dividing by the sum of the post-selection genotype frequencies **normalizes** the allele frequencies so they will sum to 1.

Recurrence Equations for Selection:

$$p_1 = \frac{p^2W_1 + pqW_2}{p^2W_1 + 2pqW_2 + q^2W_3} \qquad q_1 = 1 - p_1 \equiv \frac{pqW_2 + q^2W_3}{p^2W_1 + 2pqW_2 + q^2W_3}$$

The above equations apply in all cases, no matter what the fitness values or the allele frequencies.

*Such relative fitnesses may be calculated if one knows the average number of offspring produced by each of the three genotypes. If each red plant produces 35 [surviving] offspring on average, each pink produces 53, and each white produces 22, these numbers are the absolute fitnesses of the genotypes. We divide each of these three absolute fitnesses by the largest of the three to find relative fitnesses. Thus, the genotype with the highest absolute fitness will have a relative fitness of 1.0, and all the other relative fitnesses will be ≤ 1.0 .

B. Solution of a special case. In the case of selection against a recessive lethal, where the fitnesses of the genotypes are as shown below, the recurrence equation can be solved to. Using the solution, one can then calculate wither how much gene frequency change will take plane in 'n' generations, or, solving for 'n', the number of generations it would take to effect a certain amount of gene frequency change. This is illustrated below:

Derive and solve the recurrence equation for selection against a recessive lethal. Using the solution, tell how many generations it would take to reduce q from 0.5 to 0.01.

A_1A_1	A_1A_2	A_2A_2	genotypes
p^2	$2pq$	q^2	H-W proportions
1	1	0	fitnesses

solve for n:

$$q_1 = \frac{pq}{p^2+2pq} = \frac{pq}{p(p+2q)} = \frac{q}{p+q+q} = \frac{q}{1+q}$$

$$nq_0 = \frac{q_0}{q_n} - 1$$

$$q_2 = \frac{q_1}{1+q_1} = \frac{\frac{q_0}{1+q_0}}{1+\frac{q_0}{1+q_0}} = \frac{\frac{q_0}{1+q_0}}{\frac{1+q_0+q_0}{1+q_0}} = \frac{\frac{q_0}{1+q_0}}{\frac{1+2q_0}{1+q_0}} = \frac{q_0}{1+2q_0}$$

$$q_3 = \frac{q_2}{1+q_2} = \frac{\frac{q_0}{1+2q_0}}{1+\frac{q_0}{1+2q_0}} = \frac{\frac{q_0}{1+2q_0}}{\frac{1+2q_0+q_0}{1+2q_0}} = \frac{\frac{q_0}{1+2q_0}}{\frac{1+3q_0}{1+2q_0}} = \frac{q_0}{1+3q_0}$$

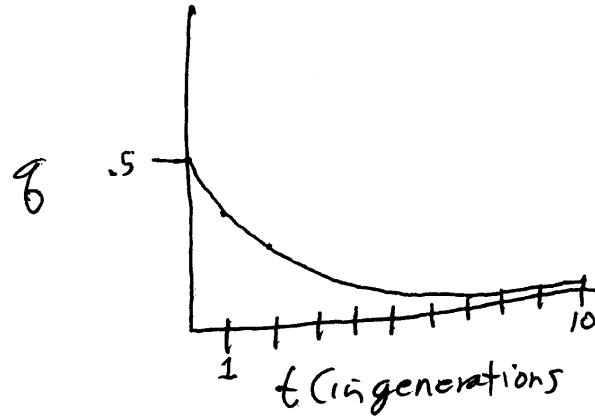
use above equation,
 where $q_0 = 0.5$ and $q_n = 0.01$:

therefore, by mathematical induction,

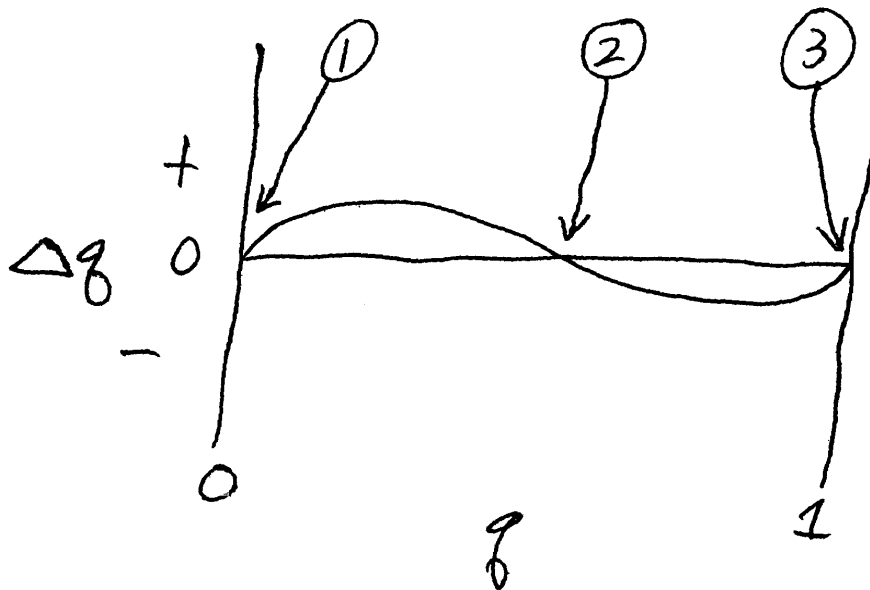
$$q_n = \frac{q_0}{1+nq_0}$$

Answer is 98 generations.

C. Graphs of gene frequency change. Graphs of the change in gene frequency, which are graphs of the course of evolutionary change, can be shown in two ways. The first is to plot q on the Y-axis and time on the x-axis. For selection on a recessive lethal, the graph is one of exponential decay, as sketched below:



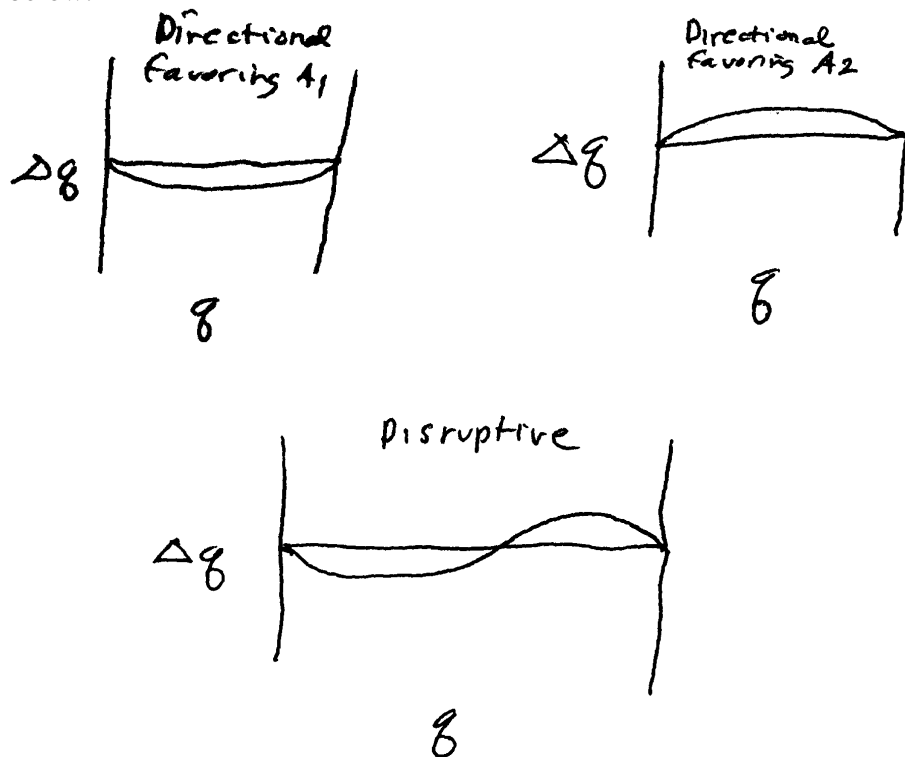
D. Δq graphs. Delta- q (Δq) graphs are the other way to show the course of evolutionary genetic change, and are more useful for analyzing equilibria, which are allele frequencies at which there will be no more gene frequency change (no change=equilibrium). A Δq graph plots gene frequency change (Δq) against gene frequency (q). A Δq graph for stabilizing selection (heterozygote has the highest fitness) at a single locus with two alleles would look as follows:



Wherever the curve crosses or touches the $\Delta q = 0$ line, there is an equilibrium. In the above sketch there are three equilibria, at points indicated by the circles numbers 1, 2, and 3. Points 1 and 3 are called "trivial" equilibria, since at these points the population is homozygous and selection cannot act. Point 2 is an "interesting" equilibrium, since both

alleles are present in the population at this equilibrium. In addition to trivial/interesting, equilibria can also be classified as to their stability. If an equilibrium is locally stable, gene frequency will tend to return to the equilibrium if disturbed (moved away from the equilibrium a little bit) in either direction. An equilibrium is unstable if gene frequency will move farther from the equilibrium if disturbed. An equilibrium is metastable if gene frequency will move toward the equilibrium if disturbed in one direction, but move away from the equilibrium if disturbed in the other direction. In the above sketch, point 1 is a trivial unstable equilibrium, point 2 is an interesting stable equilibrium, and point 3 is another trivial unstable equilibrium. Note that the Δq graph for stabilizing selection has a single internal stable equilibrium. Such an equilibrium is said to be 'globally' stable, since no matter where gene frequency starts (except the trivial equilibria), selection will force it to the single stable equilibrium. If there is more than one stable equilibrium, each must be 'locally stable' rather than 'globally stable', since each would only 'attract' gene frequencies from a part of the 'globe' of gene frequency possibilities.

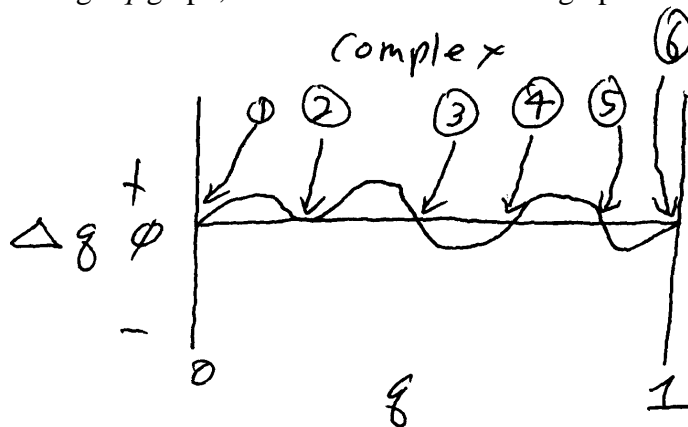
Directional selection produces a Δq graph which always has positive values of Δq , or always has negative values of Δq . Simple disruptive selection (heterozygote with the lowest fitness) produces a Δq graph with an internal UNstable equilibrium. Sketches are shown below.



E. Complex Δq graphs. If selection isn't a simple case of stabilizing, directional or disruptive, but is frequency-dependent (the fitnesses change as the allele frequency changes) or otherwise complex, the Δq graph may be similarly complex and may have several equilibria. Each of these equilibria may be classified as trivial or interesting, and stable, unstable or metastable. If an equilibrium is stable, one also wishes to know whether it is globally or only locally stable.

In the real world, where there is a certain amount of chaos (due to random events; one may also say that in the real world, evolution is at least a partly stochastic process). This means that real-world populations will never stay at metastable or unstable equilibria, because stochastic (random) events will move them off the equilibrium some time or other. In a real population, gene frequencies will be found in transition, or very near some stable equilibrium — if evolution is occurring, gene frequencies will be changing; once evolution is completed, the gene frequencies will be very near only a stable equilibrium. So, if a population starts with a given gene frequency 'q', that frequency will tend to move to the next stable equilibrium it encounters, and stay there forever, as long as the set of fitnesses that produced the Δq graph we are working with do not change. An example of the interpretation of a complex Δq graph follows.

For the following Δq graph, fill in the table below the graph.

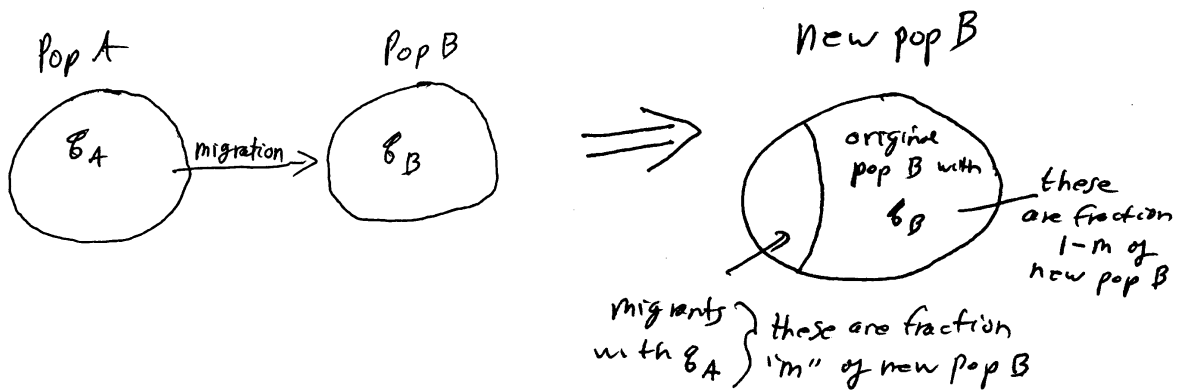


Point on graph	Is point an equilibrium? (Y/N)	<u>If equilibrium,</u> is point trivial or interesting? (T/I)	<u>If equilibrium,</u> is point stable, metastable, or unstable? (S/M/U)	<u>ONLY if STABLE equilibrium,</u> it is globally or locally stable? (G/L)
1	Y	T	U	-----
2	Y	I	M	-----
3	Y	I	S	L
4	Y	I	U	-----
5	Y	I	S	L
6	Y	T	U	-----

Note the following: if there is more than one stable equilibrium, they must be local — only if there is one stable equilibrium can it be global (but is not necessarily global). If the tangent to the curve has a negative slope at the equilibrium, the equilibrium is stable; if the slope is 0, the equilibrium is metastable, and if the slope is positive, the equilibrium is unstable (you may remember using first and second derivatives in your calculus class to do similar curve analyses).

III. Migration. Migration is the movement of individuals from one population to another. As individuals move, of course, they carry their genes with them from one population to the other. When the migrating individuals arrive in the new population, they interbreed with those individuals already there, and thus contribute their genes to the 'gene pool' (the total of all alleles in the population) of the new population.

We model the 1-locus, 2 allele case of migration of individuals from a population 'A' to a population 'B' as follows: q_A is the frequency of the A_2 allele in population A; q_B is the frequency of the A_2 allele in population B. After some individuals from population A migrate into population B, we have a new population B. This new population B is composed partly of the original individuals in population B, and partly of migrants from population A. A sketch follows:



The new frequency of the A_2 allele in population B is denoted by q'_B . It is simply the weighted mean of the allele frequencies in populations A and B, weighted (multiplied) by their proportions in the new population B. That is: $q'_B = mq_A + (1 - m)q_B$. We can also get the Δq equation, which is always the new allele frequency minus the old:

$$\Delta q_B = q'_B - q_B \equiv mq_A + (1 - m)q_B - q_B \equiv mq_A + q_B - mq_B - q_B \equiv mq_A - mq_B \equiv m(q_A - q_B).$$

Since $\Delta q = m(q_A - q_B)$, we know that the amount of gene frequency change (evolution) due to migration depends on two factors: m , the amount of migration, and $q_A - q_B$, the difference in allele frequency of the two original populations. More migration, and/or greater allele frequency differences in the original populations will cause more evolution.

III. Mutation. Mutation is important to evolution because mutation is the original source of all genetic variation. However, because mutation rates are so low (averaging 1 mutation at a given locus in 50,000 individuals [100,000 alleles] per generation [rates in bacteria are much lower, about 1 in a billion per locus per cell division]), mutation is very weak as a force changing allele frequencies. Only in the virtual absence of other forces will mutation play a role in determining allele frequencies at a locus. However, there probably such cases. If alleles are selectively neutral, such as perhaps the third base of some codons (where a change in the third base does not affect the amino acid), mutation may be a determinant of allele frequencies. For such genes, mutation may produce a "molecular clock" in which mutations occur at a fairly constant rate per million years, and the number of differences between organisms is then a measure of how many years it has been since they diverged from a common ancestor.

In addition to the "normal" kinds of mutations, two additional types deserve consideration in evolutionary thought. The first is the class of transposable elements (also called transposons or jumping genes), which are segments of DNA which code for a protein that copies the DNA and inserts a copy somewhere else along the chromosome. The only function of such DNA seems to be replication, but such replication can result in thousands of copies of such a transposable element scattered more or less at random throughout an organism's chromosomes (this occurs in eukaryotes only). If a copy of a transposable element is inserted into a functional gene, it may destroy or alter the function of the gene. Transposable elements also can move among species.

A very evolutionarily important mutation type is gene duplication. Such duplication may affect only one gene, many genes, whole chromosomes, or the entire genome. Duplication of the entire genome (polyploidy) is common in flowering plants: the majority of flowering plants in fact appear to have undergone one or more genome duplications. In animals, there appear to have been several genome duplications, one of which may have given rise to the vertebrates. Gene duplication, whether singly or 'en masse', give evolutionary flexibility to the organisms, because (to be teleological about it) one copy of a gene is free to mutate while the other copy keeps the organisms alive. In humans, for example, the various hemoglobin genes (α , β , γ , δ , ζ) are all obvious duplications arising from an original myoglobin gene in lower animals.

IV. Genetic Drift. Genetic drift is a change in allele frequencies due to small population size.

While the actual mathematics of genetic drift are beyond the scope of this class, drift can be illustrated by a thought experiment. Think of an average class of students. If the whole class of individuals were put on some island, and mated at random, what would the offspring generation be like? Well, they would be pretty much like the class put on the island — a variety of skin colors, hair colors, body sizes and shapes and so forth. But now take only two people at random from the class (say, the female with the lowest last four social security number digits and the male with the lowest last 4 social security number digits), put them on the island, let them mate and produce offspring. By chance, they may both have dark hair. If so, there is a good chance that all the offspring will also have dark hair. In other words, if the number of individuals starting a new population is small

enough, they will not carry all the genetic variation present in the larger population from which they came. This is the phenomenon of genetic drift — a change in allele frequencies due to small population size. Genetic drift is only important in very small populations, however — by the time you get to 20 or so randomly chosen breeding individuals, there is little effect on allele frequencies.

Another consequence of small population size is inbreeding. Inbreeding increases homozygosity (decreases heterozygosity), and homozygosity is generally (though not always) bad. If a population is very small in size, after a few generations, all the organisms are likely to be fairly closely related, since there were only relatives available with which to breed.

Genetic drift in nature can occur in two ecological ways. One is called the bottleneck effect, in which a large population is reduced in numbers to a very small population, due to some catastrophe. After the catastrophe, the population rebounds in numbers. However, the population will have lost some genetic variation when it was very small. The population of cheetahs is commonly considered to be an example. Cheetahs have very little genetic variation; not very many thousand years ago there were probably only 1-3 female cheetahs breeding in the world. Male cheetahs have a high proportion of abnormal sperm, and cheetahs in general do not reproduce as well as the other cats. California sea otters may have suffered a similar fate when their numbers were reduced by hunting.

The other way genetic drift can occur is called the founder effect. In this case, a few individuals leave a large population, go to another geographic area, and found a new population. The effects are the same: the small number of individuals cannot represent all the genetic variation found in the population from which they come, and inbreeding will usually cause loss of heterozygosity in subsequent generations.

V. Non-random mating. There are several types of non-random mating. Many types of non-random mating are formally equivalent to selection, and can be treated as such. For instance, if female mice prefer to mate to male mice with long tails, tail length would (if tail length were heritable and there no force countering the mating preference) increase in subsequent generations — this would then be directional selection. If females preferred to mate with mice with average size noses, the situation would be formally equivalent to stabilizing selection.

Another type of non-random mating is assortative mating — either positive assortative mating in which mating individuals are more similar than an average pair of individuals chosen at random, or negative assortative mating in which mating individuals are less similar than the average pair chosen at random. Positive assortative mating in general increases homozygosity (remember that even if the like types are heterozygotes [in the single gene case], the cross of two heterozygotes produces offspring of which 50% are homozygotes); negative assortative mating in general increases heterozygosity.

Inbreeding is another major type of departure from random mating. The definition of an inbred individual is one whose parents are more closely related than the average pair of parents in the population. In practice, we mean that an inbred individual has parents

who are fairly closely related (at, say, what would be, for humans, the second-cousin level or closer). The effect of inbreeding is on genotypic frequencies, not on allele frequencies.

Inbreeding decreases the proportion of heterozygotes in the population (and, of course, therefore simultaneously increases the proportion of homozygotes in the population). For reasons not fully understood, homozygosity is generally bad for individuals and heterozygosity is generally good, though there are plenty of exceptions. In the example below, a population initially in the H-W genotypic frequencies is selfed (individuals mating to themselves, the closest possible inbreeding and not uncommon in many organisms, such as plants). Each generation, the proportion of homozygotes in the population is reduced 50%, with the eventual frequencies of the homozygotes becoming the allele frequencies p and q . However, one generation of random mating will restore the population to the H-W genotypic frequencies. Note that when A_1A_1 mates to A_1A_1 , all the offspring will be A_1A_1 , but when A_1A_2 mates to A_1A_2 , only $\frac{1}{2}$ the offspring will be heterozygotes.

H-W Genotypic freqs:	p^2	$2pq$	q^2
after 1 gen. of selfing:	$p^2 + \frac{1}{2}q$	pq	$q^2 + \frac{1}{2}pq$
after 2 gens. of selfing	$p^2 + \frac{3}{4}pq$	$\frac{1}{2}pq$	$q^2 + \frac{3}{4}pq$
after 3 gens. of selfing	$p^2 + \frac{7}{8}pq$	$\frac{1}{4}pq$	$q^2 + \frac{7}{8}pq$
after 4 gens. of selfing	$p^2 + \frac{15}{16}pq$	$\frac{1}{8}pq$	$q^2 + \frac{15}{16}pq$
by induction
after infinite time	$p^2 + pq = p$	0	$q^2 + pq = q$

If inbreeding is between less-closely-related individuals, the decrease in heterozygosity is not as rapid as with selfing.

An inbreeding coefficient, F (which is a measure of homozygosity produced by inbreeding), can be calculated from a pedigree diagram. Such F values are useful for animal and plant breeding, but we will not go through the calculations here. Generally, in breeding programs, breeders try to get as much heterozygosity as possible, since that generally increases yields. To do this, breeders often inbreed strains of, say, corn (maize). As the strains are inbred through successive generations, they generally have smaller yields. Each strain will, by the inbreeding, be made homozygous for many of its genes; however, the different inbred lines will be homozygous for at least partly different alleles. Once a breeder has many inbred strains, each strain is crossed to each other strain. Some of the hybrids will be heterozygous for a great many alleles, because the parent strains were homozygotes of opposite types for these alleles. Thus, some of the crosses produce plants which are very heterozygous, and thus, in general, will have very high yields. There is a whole mathematics of breeding genetics, which is beyond the scope of this class.