Evolutionary Biology
(Unit-7)
Sem 6 Hons (CC-6-14-TH)

Population Genetics

Population genetics is a field of biology that studies the genetic composition of biological populations, and the changes in genetic composition that result from the operation of various factors, including natural selection. Population geneticists pursue their goals by developing abstract mathematical models of gene frequency dynamics, trying to extract conclusions from those models about the likely patterns of genetic variation in actual populations, and testing the conclusions against empirical data. A number of the more robust generalizations to emerge from population-genetic analysis are discussed below.

Population genetics is intimately bound up with the study of evolution and natural selection, and is often regarded as the theoretical cornerstone of modern Darwinism. This is because natural selection is one of the most important factors that can affect a population's genetic composition. Natural selection occurs when some variants in a population out-reproduce other variants as a result of being better adapted to the environment, or ‘fitter’. Presuming the fitness differences are at least partly due to genetic differences, this will cause the population's genetic makeup to be altered over time. By studying formal models of gene frequency change, population geneticists therefore hope to shed light on the evolutionary process, and to permit the consequences of different evolutionary hypotheses to be explored in a quantitatively precise way.

The field of population genetics came into being in the 1920s and 1930s, thanks to the work of R.A. Fisher, J.B.S. Haldane and Sewall Wright. Their achievement was to integrate the principles of Mendelian genetics, which had been rediscovered at the turn of century, with Darwinian

natural selection. Though the compatibility of Darwinism with Mendelian genetics is today taken for granted, in the early years of the twentieth century it was not. Many of the early Mendelians did not accept Darwin's 'gradualist' account of evolution, believing instead that novel adaptations must arise in a single mutational step; conversely, many of the early Darwinians did not believe in Mendelian inheritance, often because of the erroneous belief that it was incompatible with the process of evolutionary modification as described by Darwin. By working out mathematically the consequences of selection acting on a population obeying the Mendelian rules of inheritance, Fisher, Haldane and Wright showed that Darwinism and Mendelism were not just compatible but excellent bed fellows; this played a key part in the formation of the 'neo-Darwinian synthesis', and explains why population genetics came to occupy so pivotal a role in evolutionary theory.

The Hardy-Weinberg Principle

The Hardy-Weinberg principle, discovered independently by G.H. Hardy and W. Weinberg in 1908, is one of the simplest and most important principles in population genetics. To illustrate the principle, large population is considered of sexually reproducing organisms. The organisms are assumed to be *diploids*, meaning that they contain two copies of each chromosome, one received from each parent. The gametes they produce are *haploid*, meaning that they contain only one of each chromosome pair. During sexual fusion, two haploid gametes fuse to form a diploid zygote, which then grows and develops into an adult organism. Most multi-celled animals, and many plants, have a lifecycle of this sort.

Suppose that at a given locus, or chromosomal 'slot', there are two possible alleles, A_1 and A_2; the locus is assumed to be on an autosome,

not a sex chromosome. With respect to the locus in question, there are three possible genotypes in the population, A_1A_1, A_1A_2 and A_2A_2 (just as in Mendel's pea plant example above). Organisms with the A_1A_1 and A_2A_2 genotypes are called *homozygotes*; those with the A_1A_2 genotype are *heterozygotes*. The proportions, or relative frequencies, of the three genotypes in the overall population may be denoted $f(A_1A_1)$, $f(A_1A_2)$ and $f(A_2A_2)$ respectively, where $f(A_1A_1) + f(A_1A_2) + f(A_2A_2) = 1$. It is assumed that these genotypic frequencies are the same for both males and females. The relative frequencies of the *A* and *B* alleles in the population may be denoted p and q, where $p + q = 1$.

The Hardy-Weinberg principle is about the relation between the allelic and the genotypic frequencies. It states that if mating is random in the population, and if the evolutionary forces of natural selection, mutation, migration and drift are absent, then in the offspring generation the genotypic and allelic frequencies will be related by the following simple equations:

$$f(A_1A_1) = p^2, \quad f(A_1A_2) = 2pq, \quad f(A_2A_2) = q^2$$

Random mating means the absence of a genotypic correlation between mating partners, i.e. the probability that a given organism mates with an A_1A_1 partner, for example, does not depend on the organism's own genotype, and similarly for the probability of mating with a partner of one of the other two types.

That random mating will lead the genotypes to be in the above proportions (so-called *Hardy-Weinberg* proportions) is a consequence of Mendel's law of segregation. To see this, note that random mating is in effect equivalent to offspring being formed by randomly picking pairs of gametes from a large ‘gamete pool’ and fusing them into a zygote. The gamete pool contains all the successful gametes of the parent organisms. Since we are assuming the absence of selection, all parents contribute equal numbers of gametes to the pool. By the law of segregation, an A_1A_2 heterozygote produces gametes bearing the A_1 and A_2 alleles in equal proportion. Therefore, the relative frequencies of

the *A* and *B* alleles in the gamete pool will be the same as in the parental population, namely *p* and *q* respectively. Given that the gamete pool is very large, when we pick pairs of gametes from the pool at random, we will get the ordered genotypic pairs $\{A_1A_1\}$, $\{A_1A_2\}$, $\{A_2A_1\}$, $\{A_2A_2\}$ in the proportions $p^2:pq:qp:q^2$. But order does not matter, so we can regard the $\{A_1A_2\}$ and $\{A_2A_1\}$ pairs as equivalent, giving the Hardy-Weinberg proportions for the unordered offspring genotypes.

This simple derivation of the Hardy-Weinberg principle deals with two alleles at a single locus, but can easily be extended to multiple alleles. (Extension to more than one locus is trickier; see section 3.6, 'Two-Locus Models and Linkage', below.) For the multi-allelic case, suppose there are n alleles at the locus, $A_1 \ldots A_n$, with relative frequencies of $p_1 \ldots p_n$ respectively, where $p_1 + p_2 + \ldots + p_n = 1$. Assuming again that the population is large, mating is random, evolutionary forces are absent, and Mendel's law of segregation holds, then in the offspring generation the frequency of the A_iA_i genotype will be p_i^2, and the frequency of the (unordered) A_iA_j genotype ($i \neq j$) will be $2p_ip_j$. It is easy to see that the two allele case above is a special case of this generalized principle.

Importantly, whatever the initial genotypic proportions, random mating will automatically produce offspring in Hardy-Weinberg proportions (for one-locus genotypes). So if generations are non-overlapping, i.e. parents die as soon as they have reproduced, just one round of random mating is needed to bring about Hardy-Weinberg proportions in the whole population; if generations overlap, more than one round of random mating is needed. Once Hardy-Weinberg proportions have been achieved, they will be maintained in subsequent generations so long as the population continues to mate at random and is unaffected by evolutionary forces such as selection, mutation etc. The population is then said to be in *Hardy-Weinberg equilibrium*—meaning that the genotypic proportions are constant from generation to generation.

The importance of the Hardy-Weinberg principle lies in the fact that it contains the solution to the problem of blending that troubled Darwin.

As we saw, Jenkins argued that with sexual reproduction, the variation in the population would be exhausted very rapidly. But the Hardy-Weinberg principle teaches us that this is not so. Sexual reproduction has no inherent tendency to destroy the genotypic variation present in the population, for the genotypic proportions remain constant over generations, given the assumptions noted above. It is true that *natural selection* often tends to destroy variation, and is thus a homogenizing force; but this is a quite different matter. The 'blending' objection was that sexual mixing *itself* would produce homogeneity, even in the absence of selection, and the Hardy-Weinberg principle shows that this is untrue.

Another benefit of the Hardy-Weinberg principle is that it greatly simplifies the task of modelling evolutionary change. When a population is in Hardy-Weinberg equilibrium, it is possible to track the genotypic composition of the population by directly tracking the allelic frequencies (or gametic frequencies). That this is so is clear—for if we know the relative frequencies of all the alleles (at a single locus), and know that the population is in Hardy-Weinberg equilibrium, the entire genotype frequency distribution can be easily computed. Were the population not in Hardy-Weinberg equilibrium, we would need to explicitly track the genotype frequencies themselves, which is more complicated.

Primarily for this reason, many population-genetic models assume that Hardy-Weinberg equilibrium obtains; as we have seen, this is tantamount to assuming that mating is random with respect to genotype. But is this assumption empirically plausible? The answer is sometimes but not always. In the human population, for example, mating is close to random with respect to ABO blood group, so the genotype that determines blood group is found in approximately Hardy-Weinberg proportions in many populations (Hartl 1980). But mating is not random with respect to height; on the contrary, people tend to choose mates similar in height to themselves. So if we consider a genotype that influences height, mating will not be random with respect to this genotype (see 'Non-Random Mating').

The geneticist W.J. Ewens has written of the Hardy-Weinberg principle, ‘it does not often happen that the most important theorem in any subject is the easiest and most readily derived theorem for that subject’. The main importance of the principle, as Ewens stresses, is not the gain in mathematical simplicity that it permits, which is simply a beneficial side effect, but rather what it teaches us about the preservation of genetic variation in a sexually reproducing population.

Factors disrupting Hardy-Weinberg equilibrium

The term ‘evolution’ is generally defined as any change in a population's genetic composition over time. The four factors that can bring about such a change are: natural selection, mutation, random genetic drift, and migration into or out of the population. A brief introduction to the standard population-genetic treatment of each of these factors is given below.

Selection-Mutation Balance

Mutation is the ultimate source of genetic variation, preventing populations from becoming genetically homogeneous in situations where they otherwise would. Once mutation is taken into account, the conclusions drawn in the previous section need to be modified. Even if one allele is selectively superior to all others at a given locus, it will not become fixed in the population; recurrent mutation will ensure that other alleles are present at low frequency, thus maintaining a degree of polymorphism. Population geneticists have long been interested in exploring what happens when selection and mutation act simultaneously.

Continuing with our one-locus, two allele model, let us suppose that the A_1 allele is selectively superior to A_2, but recurrent mutation from A_1 to A_2 prevents A_1 from spreading to fixation. The rate of

mutation from A_1 to A_2 per generation, i.e. the proportion of A_1 alleles that mutate every generation, is denoted u. (Empirical estimates of mutation rates are typically in the region of 10^{-6}.) Back mutation from A_2 to A_1 can be ignored, because we are assuming that the A_2 allele is at a very low frequency in the population, thanks to natural selection. What happens to the gene frequency dynamics under these assumptions? Recall equation (1) above, which expresses the frequency of the A_1 allele in terms of its frequency in the previous generation. Since a certain fraction (u) of the A_1 alleles will have mutated to A_2, this recurrence equation must be modified to:

$$p' = (p^2 w_{11} + pq w_{12}) (1 - u) / w$$

to take account of mutation. As before, equilibrium is reached when $p' = p$, i.e. $\Delta p = 0$. The condition for equilibrium is therefore:

$$p = p^* = (p^2 w_{11} + pq w_{12}) (1 - u) / w \qquad (3)$$

A useful simplification of equation (3) can be achieved by making some assumptions about the genotype fitnesses, and adopting a new notation. Let us suppose that the A_2 allele is completely recessive (as is often the case for deleterious mutants). This means that the A_1A_1 and A_1A_2 genotypes have identical fitness. Therefore, genotypic fitnesses can be written $w_{11} = 1$, $w_{12} = 1$, $w_{22} = 1 - s$, where s denotes the difference in fitness of the A_2A_2 homozygote from that of the other two genotypes. (s is known as the *selection co-efficient* against A_2A_2). Since we are assuming that the A_2 allele is deleterious, it follows that $s > 0$. Substituting these genotype fitnesses in equation (3) yields:

$$p^* = p (1 - u) / p^2 + 2pq + q^2(1 - s)$$

which reduces to:

$$p^* = 1 - (u/s)^{½}$$

or equivalently (since $p + q = 1$):

$$q^* = (u/s)^{½} \qquad (4)$$

Equation (4) gives the equilibrium frequency of the A_2 allele, under the assumption that it is completely recessive. Note that as u increases, q^* increases too. This is highly intuitive: the greater the mutation rate from A_1 to A_2, the greater the frequency of A_2 that can be maintained at equilibrium, for a given value of s. Conversely, as s increases, q^* decreases. This is also intuitive: the stronger the selection against the A_2A_2 homozygote, the lower the equilibrium frequency of A_2, for a given value of u.

It is easy to see why equation (4) is said to describe *selection-mutation* balance—natural selection is continually removing A_2 alleles from the population, while mutation is continually re-creating them. Equation (4) tells us the equilibrium frequency of A_2 that will be maintained, as a function of the rate of mutation from A_1 to A_2 and the magnitude of the selective disadvantage suffered by the A_2A_2 homozygote. Importantly, equation (4) was derived under the assumption that the A_2 allele is completely recessive, i.e. that the A_1A_2 heterozygote is phenotypically identical to the A_1A_1 homozygote. However, it is straightforward to derive similar equations for the cases where the A_2 allele is dominant, or partially dominant. If A_2 is dominant, or partially dominant, its equilibrium frequency will be lower than if it is completely recessive; for selection is more efficient at removing it from the population. A deleterious allele that is recessive can 'hide' in heterozygotes, and thus escape the purging power of selection, but a dominant allele cannot.

Before leaving this topic, one final point should be noted. Our discussion has focused exclusively on deleterious mutations, i.e. ones which reduce the fitness of their host organism. This may seem odd, given that beneficial mutations play so crucial a role in the evolutionary process. The reason is that in population genetics, a major concern is to understand the causes of the genetic variability found in biological populations. If a gene is beneficial, natural selection is likely to be the major determinant of its equilibrium frequency; the rate of sporadic mutation to that gene will play at most a minor role. It is only where a

gene is deleterious that mutation plays a major role in maintaining it in a population.

Random Genetic Drift

Random genetic drift refers to the chance fluctuations in gene frequency that arise in finite populations; it can be thought of as a type of ‘sampling error’. In many evolutionary models, the population is assumed to be infinite, or very large, precisely in order to abstract away from chance fluctuations. But though mathematically convenient, this assumption is often unrealistic. In real life, chance factors will invariably play a role, particularly in small populations. The term 'random drift' is sometimes used in broad sense, to refer to any stochastic factors that affect gene frequencies in a population, including for example chance fluctuations in survival and mating success; and sometimes in a narrower sense, to refer to the random sampling of gametes to form the offspring generation (which arises because organisms produce many more gametes than will ever make it into a fertilized zygote). The broader sense is used here.

To understand the nature of random drift, consider a simple example. A population contains just ten organisms, five of type *A* and five of type *B*; the organisms reproduce asexually and beget offspring of the same type. Suppose that neither type is selectively superior to the other—both are equally well-adapted to the environment. However, this does not imply that the two types will produce identical numbers of offspring, for chance factors may play a role. For example, it is possible that all the type *B*s might die by accident before reproducing; in which case the frequency of *B* in the second generation will fall to zero. If so, then the decline of the *B* type (and thus the spread of the *A* type) is the result of random drift. Evolutionists are often interested in knowing whether a given gene frequency change is the result of drift, selection, or some combination of the two.

The label ‘random drift’ is slightly misleading. In saying that the spread of the *A* type is due to random drift, or chance, we do not mean that no

cause can be found of its spread. In theory, we could presumably discover the complete causal story about why each organism in the population left exactly the number of offspring that it did. In ascribing the evolutionary change to random drift, we are not denying that there is such a causal story to be told. Rather, we mean that the spread of the *A* type was not due to its adaptive superiority over the *B* type. Put differently, the *A* and the *B* types had the same *expected* number of offspring, so were equally fit; but the *A* types had a greater *actual* number of offspring. In a finite population, actual reproductive output will almost always deviate from expectation, leading to evolutionary change.

An analogy with coin tossing can illuminate random drift. Suppose a fair coin is tossed ten times. The probability of heads on any one toss is ½, and so the *expected* frequency of heads in the sequence of ten is 50%. But the probability of *actually* getting half heads and half tails is only 242/1024, or approximately 23.6%. So even though the coin is fair, we are unlikely to get equal proportions of heads and tails in a sequence of ten tosses; some deviation from expectation is more probable than not. In just the same way, even though the *A* and *B* types are equally fit in the example above, it is likely that some evolutionary change will occur. This analogy also illustrates the role of population size. If we tossed the coin a hundred times rather than ten, the proportion of heads would probably be very close to ½. In just the same way, the larger the population, the less important the effect of random drift on gene frequencies; in the infinite limit, drift has no effect.

Drift greatly complicates the task facing the population geneticist. In the example above, it is obviously impossible to *deduce* the composition of the population in the second generation from its composition in the first generation; at most, we can hope to deduce the probability distribution over all the possible compositions. So once drift is taken into account, no simple recurrence relation for gene frequencies, of the sort expressed in equation (1) above, can be derived. One important effect of random drift is to decrease the degree of heterozygosity in a population over time. This happens because, given enough time, any finite population

will eventually become homozygous through drift (though if the population is large, the approach to homozygosity will be slow.) It is easy to see why this is—for gene frequencies of 0 and 1 are ‘absorbing boundaries’, meaning that once the boundary is reached, there is no way back from it (apart from mutation). So eventually, a given allele will eventually become fixed in a population, or go extinct, the latter being the more likely fate. Indeed mathematical models show that a neutral allele arising by mutation has a very low probability of becoming fixed in a population; the larger the population, the lower the probability of fixation.

Another important effect of random drift is to cause the different subpopulations of a species to diverge genetically from each other, as the chance accumulation of alleles will probably proceed differently in each, particularly if the alleles confer little selective advantage or disadvantage. By chance, one population may become fixed for allele A_1, while a second population becomes fixed for another allele A_2. This possibility is an important one, for if we ignore it, we may mistakenly conclude that the A_1 allele must have been advantageous in the environment of the first population, the A_2 allele in the environment of the second, i.e. that selection was responsible for the genetic differentiation. Such an explanation *might* be right, but it is not the only one—random drift provides an alternative.

The question of whether drift or selection plays a more important role in molecular evolution was much debated in the 1960s and 1970s; it lay at the heart of the heated controversy between ‘selectionists’ and ‘neutralists’. The neutralist camp, headed by M. Kimura, argued that most molecular variants had no effect on phenotype, so were not subject to natural selection; random drift was the major determinant of their fate. Kimura argued that the apparently constant rate at which the amino acid sequences of proteins evolved, and the extent of genetic polymorphism observed in natural populations, could best be explained by the neutralist hypothesis (Kimura 1977, 1994). Selectionists countered that natural selection was also capable of explaining the molecular data. In recent years, the controversy has subsided somewhat, without a clear victory

for either side. Most biologists believe that some molecular variants are indeed neutral, though fewer than were claimed by the original neutralists.

Migration

Migration into or out of a population is the fourth and final factor that can affect its genetic composition. Obviously, if immigrants are genetically different from the population they are entering, this will cause the population's genetic composition to be altered. The evolutionary importance of migration stems from the fact that many species are composed of a number of distinct subpopulations, largely isolated from each other but connected by occasional migration. (For an extreme example of population subdivision, think of ant colonies.) Migration between subpopulations gives rise to gene flow, which acts as a sort of ‘glue’, limiting the extent to which subpopulations can diverge from each other genetically.

The simplest model for analysing migration assumes that a given population receives a number of migrants each generation, but sends out no emigrants. Suppose the frequency of the A_1 allele in the resident population is p, and the frequency of the A_1 allele among the migrants arriving in the population is p_m. The proportion of migrants coming into the population each generation is m (i.e. as a proportion of the resident population.) So post-migration, the frequency of the A_1 allele in the population is:

$$p' = (1 - m)\, p + m p_m$$

The change in gene frequency across generations is therefore:

$$\Delta p = p' - p$$

$$= - m\, (p - p_m)$$

Therefore, migration will increase the frequency of the A_1 allele if $p_m > p$, decrease its frequency if $p > p_m$, and leave its frequency

unchanged if $p = p_m$. It is then a straightforward matter to derive an equation giving the gene frequency in generation t as a function of its initial frequency and the rate of migration. The equation is:

$$p_t = p_m + (p_0 - p_m)(1 - m)^t$$

where p_0 is the initial frequency of the A_1 allele in the population, i.e. before any migration has taken place. Since the expression $(1 - m)^t$ tends towards zero as t grows large, it is easy to see that equilibrium is reached when $p_t = p_m$, i.e. when the gene frequency of the migrants equals the gene frequency of the resident population.

This simple model assumes that migration is the only factor affecting gene frequency at the locus, but this is unlikely to be the case. So it is necessary to consider how migration will interact with selection, drift and mutation. A balance between migration and selection can lead to the maintenance of a deleterious allele in a population, in a manner closely analogous to mutation-selection balance, discussed above. The interaction between migration and drift is especially interesting. We have seen that drift can lead the separate subpopulations of a species to diverge genetically. Migration opposes this trend—it is a homogenising force that tends to make subpopulations more alike. Mathematical models suggest that that even a fairly small rate of migration will be sufficient to prevent the subpopulations of a species from diverging genetically. Some theorists have used this to argue against the evolutionary importance of group selection, on the grounds that genetic differences between groups, which are essential for group selection to operate, are unlikely to persist in the face of migration.

Non-Random Mating

Recall that the Hardy-Weinberg law, the starting point for most population-genetic analysis, was derived under the assumption of random mating. But departures from random mating are actually quite common. Organisms may tend to choose mates who are similar to them phenotypically or genotypically—a mating system known as ‘positive

assortment'. Alternatively, organisms may choose mates dissimilar to them—'negative assortment'. Another type of departure from random mating is inbreeding, or preferentially mating with relatives.

Analysing the consequences of non-random mating is quite complicated, but some conclusions are fairly easily seen. Firstly and most importantly, non-random mating does not in itself affect gene frequencies (so is not an evolutionary 'force' on a par with selection, mutation, migration and drift); rather, it affects genotype frequencies. To appreciate this point, note that the gene frequency of a population, at the zygotic stage, is equal to the gene frequency in the pool of successful gametes from which the zygotes are formed. The pattern of mating simply determines the way in which haploid gametes are 'packaged' into diploid zygotes. Thus if a random mating population suddenly starts to mate non-randomly, this will have no effect on gene frequencies.

Secondly, positive assortative mating will tend to decrease the proportion of heterozygotes in the population, thus increasing the genotypic variance. To see this, consider again a single locus with two alleles, A_1 and A_2, with frequencies p and q in a given population. Initially the population is at Hardy-Weinberg equilibrium, so the proportion of A_1A_2 heterozygotes is $2pq$. If the population then starts to mate completely assortatively, i.e. mating only occurs between organisms of identical genotype, it is obvious that the proportion of heterozygotes must decline. For $A_1A_1 \times A_1A_1$ and $A_2A_2 \times A_2A_2$ matings will produce no heterozygotes; and only half the progeny of $A_1A_2 \times A_1A_2$ matings will be heterozygotic. So the proportion of heterozygotes in the second generation must be less than $2pq$. Conversely, negative assortment will tend to increase the proportion of heterozygotes from what it would be under Hardy-Weinberg equilibrium.

What about inbreeding? In general, inbreeding will tend to increase the homozygosity of a population, like positive assortment. The reason for this is obvious—relatives tend to be more genotypically similar than randomly chosen members of the population. In the majority of species,

including the human species, inbreeding has negative effects on organismic fitness—a phenomenon known as ‘inbreeding depression’. The explanation for this is that deleterious alleles often tend to be recessive, so have no phenotypic effect when found in heterozygotes. Inbreeding reduces the proportion of heterozygotes, making recessive alleles more likely to be found in homozygotes where their negative phenotypic effects become apparent. The converse phenomenon—‘hybrid vigour’ resulting from outbreeding—is widely utilised by animal and plant breeders.

Solving of Simple Problems related to estimation of allelic and genetic frequencies

Allele Frequency Definition

The allele frequency is the number of individual alleles of a certain type, divided by the total number of alleles of all types in a population. In simple terms, the allele frequency describes how common an allele is within a population.

Allele Frequency Overview

The allele frequency is different from the *phenotypic ratio* in that it accounts for all alleles, even if they are recessive and are “hidden” within carrier organisms. The phenotypic ratio only describes the phenotypes, or actual physical features that are present within a population. **To find the allele frequency, scientists must consider heterozygous individuals, which may be hiding a recessive allele.**

Allele frequency is most commonly calculated using the **Hardy-Weinberg equation**, which describes the relationship between two alleles within a population. When more than two alleles are present, scientists must use more complex methods to determine the actual allele frequency. Allele frequency can change over time as evolution acts upon a population and the population adapts by increasing or decreasing the frequency of certain alleles.

Calculating allele frequencies is a complex topic, which combines aspects of math and genetics. In general, all of the alleles in a population add up to 100%. So, we can use mathematical formulas to predict and determine the allele frequency of an allele in a population.

How to Calculate Allele Frequency

To find the number of alleles in a given population, you must look at all the phenotypes present. The phenotypes that represent the allele are often masked by dominant and recessive alleles working in conjunction. **To analyze the allele frequency in a population, scientists use the Hardy-Weinberg (HW) equation.** The Hardy-Weinberg equation is written as follows:

$$\mathbf{1 = p^2 + 2pq + q^2}$$

P and q each represent the allele frequency of different alleles. The term p^2 represents the frequency of the homozygous dominant genotype. The other term, q^2, represents the frequency of the homozygous recessive genotype.

While it would be impossible to count all of the hidden alleles, it is easy to count the number of recessive phenotypes in a population. Recessive phenotypes are caused by two recessive alleles. Therefore, q^2 can be easily observed by dividing the total number of recessive phenotypes by the total number of individuals. Let's look at an example of how we can use this information to calculate the allele frequency of any given allele.

Allele Frequency Example

In a simplified scenario, p and q are the only alleles in the population, and the population is not developing any mutations. **If this is the case, the sum of the allele frequencies of p and q must equal 1 because with only two alleles the combined frequency must equal 100%.**

Finding q

In this example, consider a hypothetical population of rabbits. A certain recessive allele within rabbits causes the rabbits to be white, while all of the other rabbits are black. Only a rabbit with two recessive alleles for a particular gene will be white. When we observe the population, we find that there are 16 white rabbits and 84 black rabbits.

Since we already know what q^2 is simply by observing the population, we can take the square root of q^2 to find q. In this case, the white rabbits contain two recessive alleles. The white rabbits account for 16 of the 100 total rabbits. In a percentage, this is exactly 16%, or

0.16. This number is equivalent to q^2. Taking the square root, we find that the allele frequency of q (white) is 0.4, or 40%.

Finding p

Once we know q, we can simply subtract q from 1 to find the frequency of p. This works only in a simplified scenario, where p and q are the only alleles and account for 100% of the total alleles. In this case, p will be equal to 60% of the alleles, or 0.6.

Common Mistakes to Avoid

Trying to Find p First

One mistake that students commonly make is trying to calculate p by observing the population, then taking the square root. This does not work in typical recessive/dominant allele relationships, simply because a dominant allele can hide a recessive allele. For instance, if we were to calculate the square root of .84 (proportion of black rabbits), we would get nearly 92%. **This overestimates the p allele frequency because of the fact that heterozygous phenotypes are actually hiding a recessive allele and should not be counted towards p.**

Relating Allele Frequency to Fitness

A common misconception of allele frequency is that it is directly related to the evolutionary fitness of a particular allele. **Just because an allele is frequent or infrequent has no bearing on the fitness of that allele.** For example, many recessive traits that are deleterious “hide” in a

population. This can mean that while it appears to exist at really low levels, it is in fact just hiding in the hybrids of the population.

Other times, a new beneficial mutation will have a very low allele frequency. A new allele must establish itself in a population by out competing other alleles. To do this is must be continuously replicated across many generations. In this way, many beneficial alleles are still highly underrepresented in the population because the population has not had time to evolve.

THE INTUITIVE APPROACH

The Hardy-Weinberg law can be used under some circumstances to calculate genotype frequencies from allele frequences. Let A1 and A2 be two alleles at the same locus,

p is the frequency of allele A1 $0 \leq p \leq 1$

q is the frequency of allele A2 $0 \leq q \leq 1$ and $p + q = 1$

where the distribution of allele frequencies is the same in men and women, i.e.:
hommes (p,q) femmes (p,q)

if they procreate : $(p + q)^2 = p^2 + 2pq + q^2 = 1$

where:

p^2 = frequency of the A1 A1 genotype <-- HOMOZYGOTE

2pq = frequency of the A1 A2 genotype <-- HETEROZYGOTE

q^2 = frequency of the A2 A2 genotyp <-- HOMOZYGOTE

these frequencies remain constant in successive generations.

Example : autosomal recessive inheritance with alleles A and a, and allele frequencies p and q:

--> frequency of the AA = p^2 and the phenotypes [[A] = p^2 +

genotypes: :]: 2pq

Aa = 2pq

[a] = q^2

aa = q^2

Example : phenylketonuria (recessive autosomal), of which the deleterious gene has a frequency of 1/100:
--> q = 1/100
therefore, the frequency of this disease is q^2 = 1/10 000, and the frequency of heterozygotes is 2pq = 2 x 99/100 x 1/100 = 2/100;

Note that there are a lot of heterozygotes: 1/50, two hundred times more than there are individuals suffering from the condition. .

For a rare disease, p is very little different from 1, and the frequency of the heterozygotes = 2q.

We use these equations implicitly, in formal genetics and in the genetics of pooled populations, usually without considering whether, and under what conditions, they are applicable.

THE HARDY-WEINBERG EQUILIBRIUM

The Hardy-Weinberg equilibium, which is also known as the panmictic equilibrium, was discovered at the beginning of the 20th century by several researchers, notably by Hardy, a mathematician and Weinberg, and physician.
The Hardy-Weinberg equilibrium is the central theoretical model in population genetics. The concept of equilibrium in the Hardy-Weinberg model is subject to the following hypotheses/conditions:

1. The population is panmictic (couples form randomly (panmixia), and their gametes encounter each other randomly (pangamy))

2. The population is "infinite" (very large: to minimize differences due to sampling).
3. There must be no selection, mutation, migration (no allele loss /gain).
4. Successive generations are discrete (no crosses between different generations).