Lecture 25

Population Genetics

Until now, we have been carrying out genetic analysis of individuals, for the next three lectures we will consider genetics from the point of view of groups of individuals, or populations.

We will treat this subject entirely from the perspective of human population studies where population genetics is used to get the type of information that would ordinarily be obtained by breeding experiments in experimental organisms.

At the heart of population genetics is the concept of **allele frequency**

Consider a human gene with two alleles: **A** and **a**

The frequency of **A** is *f*(**A**) ; the frequency of **a** is *f*(**a**)

Definition: $p = f(A)$ $q = f(a)$

p and ***q*** can be thought of as probabilities of selecting the given alleles by random sampling. For example, ***p*** for a given population of humans is the probability of finding allele **A** by selecting an individual from that population at random and then selecting one of their two alleles at random.

Since ***p*** and ***q*** are probabilities and in this example there are only two possible alleles;

$$p + q = 1$$

Correspondingly, there are three possible **genotype frequencies**:

$$f(A/A) + f(A/a) + f(a/a) = 1$$

We usually can't get allele frequencies directly but must derive them from the frequencies of the different genotypes that are present in a population

$$p = f(A/A) + \tfrac{1}{2} f(A/a)$$

(homozygote) (heterozygote)

$$q = f(a/a) + \tfrac{1}{2} f(A/a)$$

Example: **M** and **N** are different blood antigens specified by alleles of the same gene. The antigens are codominant so a simple blood test can distinguish the three possible genotypes.

$$f(^{M}/_{M}) = 0.83, \quad f(^{M}/_{n}) = 0.16, \quad f(^{N}/_{N}) = .01$$

$$p = f(M) = .83 + .08 = 0.91$$

$$q = f(N) = .01 + .08 = 0.09$$

Note: we can get both ***p*** and ***q*** with just two of the genotype frequencies because the three genotype frequencies must total to a frequency of 1.0:

$$f(^{M}/_{M}) + f(^{M}/_{N}) + f(^{N}/_{N}) = 1$$

Now let's think about how the inverse calculation would be performed. That is, how to derive the genotype frequencies from the allele frequencies. To do this we must make an assumption about the frequency of mating of individuals with different genotypes. If we assume that the gametes *mix at random*, we can calculate the compound probabilities of obtaining each possible combination of alleles.

		egg
	A	**a**
sperm	(***p***)	(***q***)
A (***p***)	$^{A}/_{A}$ (p^2)	$^{A}/_{a}$ (***pq***)
a (***q***)	$^{A}/_{a}$ (***pq***)	$^{a}/_{a}$ (q^2)

Thus the genotype frequencies for the next generation are:

$$f(^{A}/_{A}) = p^2, \quad f(^{A}/_{a}) = 2pq, \quad f(^{a}/_{a}) = q^2$$

We can now calculate the new p_1 for this generation using the formula for deriving allele frequencies from genotype frequencies:

$$p_1 = f(A/A) + {}^1/_2\, f(A/a)$$

$$= p^2 + pq$$

$$= p\,(p + q)$$

$$= p$$

We obtain the simple but very important result that when mixing of gametes occurs at random, the allele frequencies do not change from one generation to the next.

This is a condition known as **Hardy-Weinberg Equilibrium**

If we know the genotype frequencies and allele frequencies then we can ask whether the population is in **H-W** equilibrium for that gene by determining whether the genotype frequencies reflect random mixing of alleles. Consider two different populations that have different genotype frequencies and different allele frequencies but have different genotype frequencies.

	M/M	M/N	N/N	p	q
US Caucasians	0.29	0.5	0.21	0.54	0.46
American Inuit	0.84	0.16	0.008	0.92	0.08

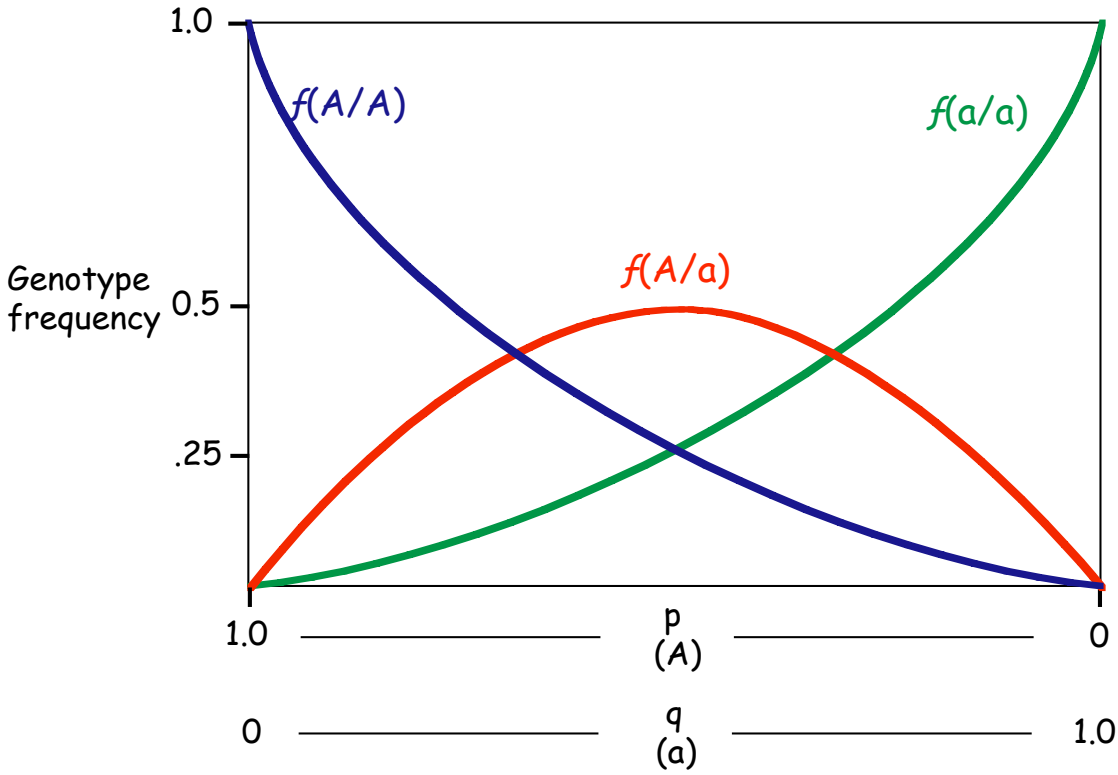
Although the allele frequencies are quite different, both populations have the genotype frequencies and allele frequencies that fit **H-W** equilibrium.

Consider the two sample populations that have the same allele frequencies but have different genotype frequencies.

	A/A	A/a	a/a	p	q
Population I:	0.20	0.20	0.60	0.3	0.7
Population II:	0.09	0.42	0.49	0.3	0.7

Only population II satisfies **H-W** criteria: $p^2 = 0.09$, $2pq = 0.42$, $q^2 = 0.49$

Here is a helpful way to look at frequencies in **H-W** equilibrium:



Before we needed at least two of the genotype frequencies to calculate allele frequency but if we know that the population is in **H-W** equilibrium we can get both allele frequencies and all genotype frequencies from just one of the genotype frequencies or one of the allele frequencies.

How good is the random mating assumption in actual human populations? The chief criteria necessary for a population to be **H-W** equilibrium is *random mating* among individuals in the population. These are some of the conditions that affect random mating assumption and therefore may affect **H-W** equilibrium:

1) **Genotypic effects on choice of partner:**

Examination of allele frequencies and genotype frequencies for most genes in the human populations reveals that they closely fit **H-W** equilibrium. The implication is that in general, humans select their mates at random with respect to individual genes and alleles. This may seem odd given that personal experience says that choosing a mate is anything but random. However the usual criteria for selecting mates such as character, appearance, and social position are largely not

determined genetically and, to the extent that they are genetically determined, these are all very complex traits that are influenced by a large number of different genes. The net result is that our decision of with whom we have children does not in general systematically favor some alleles over others.

One of the exceptional conditions that produce a population that is *not* in **H-W** equilibrium is known as **Assortative Mating.** Which means preferential mating between like individuals. For example, individuals with inherited deafness have a relatively high probability of having children together. But even this type of assortative mating will only affect the genotype frequencies related to deafness.

2) **New mutations:**

Although new mutations continually arise, mutation rates are usually sufficiently small that in any single generation their effect on allele frequencies is negligible. As will be discussed in the next lecture, the effect of mutations compounded over many generations can have a significant effect on allele frequencies.

3) **Selection (differences in survival or reproduction of different genotypes)**

Like new mutations, the effect of selection is usually small in any single generation and therefore usually does not affect **H-W** equilibrium. An exception would be a recessive lethal mutation that would render the genotype frequency of the homozygote = 0 regardless of the genotype frequency of the heterozygote. As will be discussed in the next lecture, the effect of selection can have a significant effect over many generations.

4) **Genetic drift/Founder effect:**

For small populations only a small number of individuals pass their alleles on to the next generation. Under these circumstances, chance fluctuations in the alleles that are transmitted can cause significant changes in allele frequency. These effects are usually insignificant for large populations such as in the U.S.

To see how this would happen, consider a gene in a very large population with a single major dominant allele **A** and 10 minor recessive alleles a_1, a_2, a_3 ...a_{10} with allele frequencies $f(a_1) = f(a_2) = f(a_3) \ldots = 10^{-4}$ and ($f(A) \approx 1$)

Now imagine that a group of 500 individuals from this population move to an island starting a new population. The aggregate frequency of recessive alleles (a_n) is 10^{-3}. Thus, only one of the recessive alleles will likely be in the initial 1000 alleles included in the island population. If the selected allele happens to be a_1, the new frequencies in the island population will be: $f(a_1) = 10^{-3}$, and $f(a_2) = f(a_3) = f(a_4) \ldots = 0$.

Thus in a stochastic fashion, most of the minor alleles will be lost, whereas an occasional rare allele will experience an increase in frequency. The smaller the founding population the more likely that a rare allele will be lost and the greater the increase in frequency experienced by the alleles that happen to be selected.

5) **Migration of individuals between different populations**

When individuals from populations with different allele frequencies mix, the combined population will be in **H-W** equilibrium after one generation of random mating. The combined population will be out of equilibrium to the extent that mating is assortatative.

If we are considering rare alleles we can make the following approximations allowing us to avoid a lot of messy algebra in our calculations.

For $f(\mathbf{a}) = q$, and $f(\mathbf{A}) = p$,

If $q \ll 1$ then $p \approx 1$

From **H-W**:

$$f(\mathbf{A/A}) = p^2 \approx 1,\ f(\mathbf{A/a}) = 2pq \approx 2q,\ f(\mathbf{a/a}) = q^2$$

Since most genetic diseases are rare, these approximations are valid for many of the population genetics calculations that are of medical importance.

For example, albinism occurs in $^1/_{20,000}$ individuals. Let's say that this condition is due to a recessive allele **a** of a single gene that is in **H-W** equilibrium.

$$f(^{a}/_{a}) = 5 \times 10^{-5} = q^2$$

$$q = \sqrt{5 \times 10^{-5}} \quad = 7 \times 10^{-3}$$

$$f(^{A}/_{a}) = 2pq \approx 2q = 1.4 \times 10^{-2}$$

We will now calculate the fraction of alleles for albinism that are in individuals that are homozygous for albinism.

Number of alleles in homozygotes $\approx 2 \times N (q^2)$ N = population size

Number of alleles in heterozygotes $\approx N (2q)$

The ratio is: $\dfrac{2 \times N (q^2)}{N (2q)} = q$

Thus, for albinism (since $q = 7 \times 10^{-3}$) the fraction of alleles in homozygotes is 7×10^{-3}. That is, > 99% of the alleles are in heterozygotes.