Review of Population Genetics Equations

1. Hardy-Weinberg Equation:

$$p^2 + 2pq + q^2 = 1$$

Derivation: Take a gene with two alleles; call them **A** and **a**. (Dominance doesn't matter for our purposes; this works equally well with codominance or incomplete dominance.) In a population, some members will have the **AA** genotype, some will have the **Aa** genotype, and some will have **aa**.

Now, imagine that you can somehow take all the gametes produced by the members of the population—for simplicity, we'll assume that these are eggs and sperm. Some gametes, of course, carry **A**, and some **a**.

p = freq (A)
q = freq (a)

NOTICE that: the frequency of an allele is equal to the probability that a randomly chosen gamete will be carrying that allele. Also notice that **p+q=1**.

What's the chance that an egg and sperm drawn randomly will both be carrying **A**? Obviously, it's $\mathbf{p \times p}$, or $\mathbf{p^2}$. And the chance that both gametes will both bear **a** is $\mathbf{q \times q}$, or $\mathbf{q^2}$. There are two other possibilities: sperm with **A** and egg with **a**, or sperm with **a** and egg with **A**. The chance of either one happening is $\mathbf{p \times q}$, and the total probability of producing a zygote with the **Aa** genotype is twice that: **2pq**. All of these probabilities sum to 1. So $\mathbf{p^2 + 2pq + q^2 = 1}$. [Since **p+q=1**, $\mathbf{(p+q)^2 = p^2 + 2pq + q^2 = 1^2 = 1}$.]

WHO CARES? It's like this: this equation links allele frequency to genotype frequency, assuming certain conditions are met. This means that:

1) If you know allele frequencies, you can predict the genotype frequencies, and compare them with the actual frequencies. If they don't match, then one of your assumptions is violated—maybe there is natural selection going on, or immigration, or non-random mating. . .

2) If you know genotype frequencies, you can predict allele frequencies, and compare them with the actual frequencies. Again, if they don't match, then one of your assumptions is violated.

3) If you know phenotype frequencies, then you can estimate genotype and allele frequencies—but you can't test the underlying assumptions.

2. Fitness:

$$p^2 w_{AA} + 2pq w_{Aa} + q^2 w_{aa} = \overline{w}$$

$$p^2(w_{AA} / \overline{w}) + 2pq(w_{Aa} / \overline{w}) + q^2 (w_{aa} / \overline{w}) = 1$$

Derivation: **w** in general means "fitness": a measurement of the relative ability ofindividuals with a certain genotype to reproduce successfully. w_{AA}, for instance, means the relative ability of individuals with the **AA** genotype to reproduce successfully. **w** is always a number between 0 and 1. Adding **w**s to the Hardy-Weinberg equation allows you to predict the effect of selection on gene and allele frequencies in the next generation.

Take the Hardy-Weinberg equation and multiply each term (the frequency of each genotype) by the fitness of that genotype. Add those up and you get the mean fitness, $\overline{w}$ ("w-bar"). Divide through by $\overline{w}$, and you get the second equation. Here, each term of the equation is multiplied by the fitness of a genotype divided by the mean fitness. If a genotype is fitter than average, this quotient is greater than 1, and that genotype will increase in frequency in the next generation. If a genotype is less fit than average, the quotient is less than 1, and that genotype will decrease in frequency in the next generation.

3. Mutation:

$$q_t = q_{t-1} + up_0 - vq_0$$
$$\Delta q = up - vq$$

Derivation: Imagine that in each generation, allele **A** mutates to allele **a** with a frequency of **u**, and that allele **a** "back-mutates" to **A** with a frequency of **v**. Then in each generation, **q**, the frequency of the **a** allele, increases by a factor of **up** (the rate of mutation of **A** to **a** times the frequency of **A**) and decreases by a factor of **vq**. These will eventually balance each other out, so that $\Delta q = 0$ (i.e. allele frequencies don't change any further). When $\Delta q = 0$, it must be true that $up = vq$. From this, with a little algebraic jugglery, you can derive the formula

$$\hat{q} = u / (u + v)$$

where $\hat{q}$ ("q-hat") is the equilibrium frequency. Similar equations let you derive $\hat{p}$.

This isn't all that useful an equation, however. In real life, mutation rates are usually on the order of 10^{-5} per locus per generation. For example, in humans, the Huntington's chorea mutation spontaneously appears about once in every 200,000 gametes produced. This means that mutation, by itself, has very little effect on allele frequencies.

4. Inbreeding:

freq (AA) = p^2 + pqF
freq (Aa) = 2pq - 2pqF
freq (aa) = q^2 + pqF

Derivation: **F** is the **inbreeding coefficient**, and it is the probability that two alleles in a diploid zygote are **identical by descent**—in other words, that they are both descended from the same recent ancestor within the population. The effect of inbreeding is to increase the frequency of homozygotes and decrease the frequency of heterozygotes. These equations show how this effect is quantified.

Given that **F** is the probability that two alleles are identical by descent, what is the probability that a given genotype will be **AA**? There are two ways in which a genotype can be **AA**. First, one of the parental gametes could be carrying **A** and the other could be identical by descent; the probability of this happening is **p × F**. The other possibility is that one gamete could have the **A** allele, the other gamete could have the **A** allele, *and* the two alleles are not identical by descent. The probability of that being the case is **p × p × (1-F)**, since if **F** is the probability of two alleles being identical by descent, **1-F** is the probability of two alleles not being identical by descent. So the total probability of a given genotype being **AA** is the sum of these two: **pF + p^2(1-F)**. We can expand that to **pF + p^2 – p^2F**, rearrange to **p^2 + pF – p^2F**, and regroup to **p^2 + pF(1-p)**. **1-p**, of course, is **q**, so the formula for the frequency of the **AA** genotype is **p^2 + pqF**. You can work out the other two formulas in much the same way.

One way to use this formula is to calculate **F**; if you know **p** and **q** for a population, and you know the actual frequency of, say, the **AA** genotype, you can plug those numbers in and calculate **F.** If there is no inbreeding, **F = 0**, and you have the basic Hardy-Weinberg equilibrium; if **F = 1**, you have a completely inbred population with no heterozygotes at all. To give a concrete example: if you had a population in which only full siblings mated with each other, after ten generations F would be about 0.85.

5. Effective Population Size:

Here's the general idea: Factors like genetic drift (see below) depend on population size. But the census size of a population—the raw number of members—may not tell you what's really going on. In a population of living things, some members may not be reproducing (because they're too young or too old, because there's an excess of males or females, because a few "privileged" males mate with multiple females. . . etc. etc.) The **effective population size** is the size of a hypothetical ideal population, all of whose members have an equal probability of breeding, that has as much inbreeding and genetic drift as the real population.

Exactly how you calculate this depends a lot on the situation and can be a little tricky, but here's a sample: In a population of diploid individuals with separate sexes, $\mathbf{N_e}$, the

effective population size, is equal to $\mathbf{(4N_fN_m) / (N_f + N_m)}$, where $\mathbf{N_f}$ is the number of females and $\mathbf{N_m}$ is the number of males. If $\mathbf{N_f}$ and $\mathbf{N_m}$ are equal, then $\mathbf{N_e}$ reduces to $\mathbf{N_f + N_m}$, which is just **N** (the actual population size). But suppose that this population is dominated by a single alpha male who does all the breeding (which is close to what happens in some animals, such as elephant seals). Then $\mathbf{N_m} = 1$, and the formula becomes $\mathbf{N_e = 4N_f / (N_f + 1)}$.

So if you have a population of 100 elephant seals, 50 males and 50 females, **N** = 100, but $\mathbf{N_e} = 200/51 = 3.9$. In other words, a population of 100 elephant seals in which only one male mates will have as much genetic drift and inbreeding as a population of *four* elephant seals in which all members could mate! This has a lot of implications for conservation and breeding programs.

6. Genetic Drift

$$\mathbf{V_q = p_0q_0/2N}$$

where V_q is the **variance** in the allele frequencies after one generation, N_e is the effective population size, and p_0 and q_0 are the allele frequencies that you started with. The variance of any set of numbers is a measure of how "spread out" the numbers are; to be more exact, it's the sum of the differences between each number and the mean of the set. The square root of the variance is the **standard deviation**, which you might be more familiar with.

7. Narrow-Sense Heritability:

h^2 = slope of least-squares regression of mean offspring phenotype on mean parental phenotype

Derivation: Use this when you're dealing with a continuous trait—height, weight, color, number of parts, etc., as long as it can be quantified by integers or real numbers—rather than a simple Mendelian dominant/recessive trait. For each pair of parents, find the mean of the trait; for each set of offspring, find the mean value of the trait. Then graph parental means vs. offspring means on a Cartesian graph, and take the slope of the regression line through the set of points. If the slope is nearly 0, then parental traits have nothing to do with what's seen in the offspring; if the slope is close to 1, then parental traits and offspring traits are tightly correlated.

BIG DISCLAIMER: The trait doesn't have to be genetically controlled AT ALL for this to work! Poverty, for example, would show a strong narrow-sense heritability (children from rich families are usually rich themselves, children of poor families often stay poor). So would religious affiliation (with some exceptions, children usually adopt the religion of theirparents). But there are no "poverty genes", or "Baptist genes".

8. Response to Selection for a Continuous Trait:

$R = h^2S$

Derivation: For a continuous trait, calculate the mean phenotype of the first generation (**t**) and the mean phenotype of the second generation after a round of selection (**t***). The difference between the two is **S**, the **selection differential** for a continuous trait.

$S = t^* - t$

S multiplied by the narrow-sense heritability, $\mathbf{h^2}$, is the response to selection, **R**. This is a prediction of how a population will respond to directional selection on a continuous trait.