Genetics 453 J. Felsenstein
Evolutionary Genetics Winter, 2002

Outline of lectures 3 - 6

Population genetics

1. Reproduction of genotypes in a population. Hardy-Weinberg proportions.
2. Assumptions
   - Random mating
   - No mutation
   - No migration
   - No differential viability
   - No differential fertility
   - Infinitely large population
3. Genotypes with current *genotype frequencies* P, Q, and R of AA, Aa, and aa have a fraction $p = P + \frac{1}{2}Q$ of their genes being A rather than a. The is the *gene frequency* (note the difference between this and a genotype frequency). The gene frequency of the a allele is, for the same reasons, $q = \frac{1}{2}Q + R$. These can also be computed by counting the fractions of A and a among the individuals.
4. Random mating is equivalent to random union of gametes. Imagine making a pot of female gametes, a pot of male gametes, and drawing a pair, one from each.
5. The result is that AA, Aa, and aa have expected genotype frequencies p^2, $2pq$ and q^2.
6. The gene frequency in this offspring population is again p.
7. If we again mate these individuals randomly, the gene frequencies in the second generation are again p and q.
8. Thus the genotype frequencies become these "Hardy-Weinberg" proportions, and stay that way forever. The gene frequencies remain forever p.
9. Mendelian genetic systems thus do not tend to lose genetic variability just because of random mating. Blending inheritance would lose it. The fundamental reason is that segregation in a heterozygote yields gametes that are $\frac{1}{2}A$ and $\frac{1}{2}a$, whereas in blending inheritance it is as if they were all medium-sized A's.
10. The gene frequency of a sample of individuals can be computed in two different ways, both of which get the same answer.

(a) We can take the numbers of the three genotypes (n_{AA}, n_{Aa}, and n_{aa}, and compute the genotype frequencies P, Q, and R from them by dividing by their sum. The we get the frequency of A among the gametes those individuals produce, which is $p = P + (1/2)Q$.

(b) We can simply count the numbers of A and a genes among the individuals in our sample, and get the gene frequency from that:

Genotype	AA	Aa	aa
Number	n_{AA}	n_{Aa}	n_{aa}
A copies:	$2n_{AA}$	n_{Aa}	-
a copies:	-	n_{Aa}	$2n_{aa}$

The result is that the frequency of A is $p = (2n_{AA} + n_{Aa})/(2N)$, where $N = n_{AA} + n_{Aa} + n_{aa}$ is the total number of individuals in our sample.

11. When we relax the assumption of no differential viability and no differential fertility, we now have natural selection going on.

12. The *absolute fitness* of each genotype is the expected contribution a newborn individual of that genotype makes to the next generation. This is the product of $\frac{1}{2}$(viability)(fertility). The one-half is because each offspring it has only gets one-half of its genes from that parent.

13. Many populations (all?) are subject to density-dependent population size regulation. If we can assume that this falls "fairly" on all genotypes, then it simply multiplies all viabilities by the same number, and/or multiplies all fertilities by the same number. It will do this if the density-dependent population size regulation acts at a different life stage, in a way unrelated to whatever causes the other fitness differences.

14. If this is true, then the ratios of the absolute fitnesses do not change as the population changes density, only the multiplier that makes them into abolute fitnesses.

15. Then we can define *relative fitness* of a genotype as the ratio between its absolute fitness and the absolute fitness of some reference genotype. Thus relative fitneses might be $1 : 0.8 : 0.7$ for the three genotypes, for example, when AA is the reference genotype.

16. With this one-locus case, one can compute the gene frequency after natural selection. The genotype frequencies at the beginning of the generation are of course $p^2 : 2pq : q^2$. When we count them by their contributions to the next generation (as a result of differential survival and fertility) they are in the proportions $p^2 w_{AA} : 2pqw_{Aa} : q^2 w_{aa}$.

17. These three numbers don't add to one, usually. So we can make them into frequencies by dividing by their sum. The sum is the *mean fitness* $\bar{w} = p^2 w_{AA} + 2pqw_{Aa} + q^2 w_{aa}$, which is also the average value of the relative fitness of a randomly chosen newborn.

18. Dividing the sum by $\bar{w}$, we get the three frequencies:

$$\frac{p^2 w_{AA}}{\bar{w}} \ : \ \frac{2pq w_{Aa}}{\bar{w}} \ : \ \frac{q^2 w_{aa}}{\bar{w}}$$

Taking the frequencies of these three after selection (i.e. according to their contributions to the next generation) the gene frequency of A in that next generation will be the frequency of AA plus half that of Aa or

$$p' = \frac{p^2 w_{AA} + \frac{1}{2} \times 2pq w_{Aa}}{\bar{w}}$$

or

$$p' = \frac{p^2 w_{AA} + pq\ w_{Aa}}{p^2 w_{AA} + 2pq\ w_{Aa} + q^2 w_{aa}} = p\frac{(p^2 w_{Aa} + pq\ w_{Aa})}{\bar{w}} = p'\frac{\bar{w}_A}{\bar{w}}$$

19. Note the rightmost expression: it says simply that the new gene frequency is the old one (p) times the mean fitness of the genotypes that a randomly-chosen A allele happens to find itself in ($\bar{w}_A$), divided by the mean fitness of everybody. In short, the gene frequency will increase if the mean fitness of A's is bigger than the mean fitness of random individuals.

20. *Rate of change of gene frequencies as a result of natural selection.* Suppose that the fitnesses of AA, Aa, and aa are $(1+s)^2 \ : \ 1+s \ : \ 1$. s is called a *selection coefficient.* The curve of gene frequency change is a logistic curve (see overheads). The time taken to change between any two gene frequencies is (approximately) inversely proportional to s.

21. *Overdominance.* When the fitness of the heterozygote is higher than that of either homozygote, natural selection will bring the gene frequency toward an interior equilibrium, retaining both alleles. This is a *polymorphism.* The exact equilibrium gene frequency depends on the fitnesses (in fact, if fitnesses are written as $1-s \ : \ 1 \ : \ 1-t$ the equilibrium frequency of A is $t/(s+t)$. If we plot fitness against gene frequency we get a quadratic curve, with a peak precisely at this equilibrium gene frequency.

22. This movement of gene frequencies can be rationalized in terms of the mean fitness of A compared to the mean fitness of everybody. In an overdominant case, when A is rare it is present mostly in heterozygotes. In that case A copies have a higher mean fitness than a (which, being common, are mostly located in aa individuals. So A increases when rare. When A is common and a is rare, the argument is reversed, with a being mostly in heterozygotes and having the advantage.

23. *Underdominance.* When s and t are both negative, so that the heterozygote is the worst genotype, the gene frequency will move continually away from the interior equilibrium, which is now an unstable equilibrium. The gene frequency tends toward 1 or 0. Which one it goes to depends on which side of the unstable equilibrium it started from. Note that the outcome depends on the exact starting point. The plot of fitness against gene frequency is again a quadratic curve, but now it has a minimum at the unstable equilibrium. The stable equilibria are now 0 and 1.

24. *Selection and fitness.* In all of these cases the gene frequency changes so that the mean fitness either improves or remains the same, it never declines. In each case the population "climbs" the *adaptive surface* or *fitness surface* until it comes to rest at the top. Incidentally, this is true for constant relative fitnesses, and for any number of alleles. It is not perfectly true when fitness is controlled by multiple loci. But in a lot of cases it is true that there is a net gain of mean relative fitness from the beginning of the evolution of the gene frequencies to the end.

25. Is "all for the best in this best of all possible worlds?" (At least in terms of evolution resulting in optimal organisms). The underdominance case shows that while evolution at a single locus (with constant relative fitnesses) results in improvement of the mean fitness, the population can sometimes come to rest on an equilibrium which is not the highest possible one. It depends on the starting point. A gene is evaluated by natural selection against the backgrounds in which it occurs, and that decides whether it will increase. If the fitnesses of AA, Aa, and aa are $1.2 : 0.7 : 1$, then when A is rare it is mostly occurring with a's in heterozygotes, which have fitness 0.7. By comparison, the a's are occurring in homozygotes which have fitness 1. So a seems to be better and copies of it survive and reproduce better. But in fact, AA would be the best genotype. However natural selection is not making a global assessment of the effects of combining alleles, so it misses this and we end up with aa.

26. Thus the opportunistic nature of natural selection causes us to climb the nearest peak on the adaptive surface, not the highest one. If one could always do the latter, would we be able to fly (unaided) at 500 miles per hour, swim to the depths of the ocean, while composing brilliant sonatas all the time? There does not seem to be any way to know, without a comprehensive understanding of organisms.