# Outline of lectures 6 - 9

1. **Genetic Drift**. Our discussions of gene frequency change have assumed infinitely large populations so far. This ensures that when a gene frequency of (say) 0.333 is expected, it is precisely achieved. But in small populations (and even in medium-sized populations) the expected gene frequency is not precisely achieved because of random sampling (i.e., the random happenstance of births, deaths, and Mendelian segregations).

2. These random fluctuations do not, as you might think, average out over time. In fact they are cumulative. When a population of (say) 1000 individuals which is expected to have a gene frequency of 0.4 turns out to have a gene frequency of 0.42, then the next generation is produced at that new gene frequency. Then the second generation is expected to have a gene frequency of 0.42, but it might turn out to have 0.41. Now the process starts from 0.41. The process of gene frequency change by these random fluctuations is called *random genetic drift* or just *genetic drift*. The process has no memory of where it started: it is always starting anew at the new frequency.

3. Genetic drift leads ultimately to the population "fixing" for one allele or another. The $A$ allele is either fixed or lost in each drifting population. Interestingly enough, the probability of $A$ fixing is the same as its initial gene frequency. This is simply because one of the gene copies in the original population wins the Mendelian lottery and is the progenitor of all the copies in the population, and the chance that it is a $A$ is the same as the fraction of $A$'s in the original population.

4. Mutation and migration can counteract this fixation of alleles by re-introducing the missing allele. When a series of interconnected populations exchange migrants, they all tend to drift together, though more slowly, as now it is like having one big population.

5. **Migration.** Migration moves genes from one population to another. It is important not to get too concerned with genotypes migrating, because reproduction in the new population will separate the copies of the genes that migrated together and combine them with the other copies into a *gene pool*. We can just think of migration as averaging the gene frequencies.

6. The averaging is a weighted average. So a population that consists of 80% individuals from population I and 20% from population II will have a gene frequency which is $0.8p_1 + 0.2p_2$, where $p_1$ and $p_2$ are the gene frequencies in populations I and II. Thus if they have gene frequencies 0.9 and 0.4 for allele $A$, the resulting gene frequency in the admixed population will be: $0.8(0.9) + 0.2(0.4) = 0.80$.

7. Continued migration between populations will bring their frequencies closer and closer to each other, at a rate that is controlled by the migration rates. In the end their gene frequencies will be expected to be the same as if all the gene copies at that locus had been dumped into one big gene pool.

8. Natural selection and genetic drift can cause gene frequencies in different populations to differ. When there is a geographic continuum, such as north-south, if natural selection differs in these geographic regions a pattern of smooth change of gene frequency (a *cline*) can be set up. It is not easy to tell from a cline's shape whether the selection regime changes smoothly as one goes from south to north, or whether it changes abruptly at an environmental boundary. The smoothing effect of migration will create a smooth pattern of gene frequency in either case.

9. (Example). Clausen, Keck, and Hiesey used a "common garden experiment" at Stanford University to test morphology of the yarrow *Achillea lanulosa* collected from different altitudes in a transect across the Sierra Nevada. The higher the plants came from, the shorter they were in the common garden.

10. (Example). Antonovics, Bradshaw, and Turner (*Advances in Ecological Research* 7: 1-85 1971) examined heavy metal tolerance in plants growing on Welsh mines. Plants from within the mine area showed higher tolerance to the metal produced by that mine (and spread around in its soil due to the dumping of tailings). The tolerance declines smoothly as one reaches the boundary of the mine. This was assessed by testing the descendants of the plants in the lab, so these are almost certainly genetic clines.

11. (Example). Johnston and Selander (*Science* 144: 548-550 1964) studied geographical variation in house sparrows (*Passer domesticus*) in North America. They have differentiated quite a bit geographically since their introduction by a lunatic in the mid-1800's. They follow the well-known "laws" of geographic differentiation in northern climes: they are larger in northern areas (Bergmann's Rule), have relatively shorter appendages in northern areas (Allen's Rule), and are darker in northern areas (Glogler's Rule).

12. **Mutation.** Mutation rates are low (underlying point mutation rates are about $10^{-8}$ per base per year). Some kinds of DNA changes such as trinuclotide repeat expansions and contractions can go faster, up to one a mutation rate of about 0.001 per locus.

13. Under mutation, we expect an equilibrium frequency to be reached between forward and backward mutation. But mutation is a very weak and slow force, and this mutational equilibrium will not be reached for a number of generations equal to the the reciprocal of the sum of the mutation rates, so this can be millions of generations. Natural selection can easily overwhelm this weak force which pushes the gene frequencies toward the mutational equilibrium.

14. Forward mutation rates are much higher than backward ones. This is because there are a lot of points in a gene that can mutate to make it inactive, but once one of those has happened, there is only one place that it can mutate to restore the original

sequences (there are so called "second-site revertants" possible too, but not too many of them).

15. It is like typographical errors. There may be many errors that can arise in a sentence, but once we get to the sentence: "To bf or not to be, that is the question", there is only one mutant that will revert it to be a sensible English sentence. So back mutations from the nonfunctional class to the functional class are rarer than forward mutations.

16. The objection has been raised that natural selection cannot work because only an extremely tiny fraction of sequences at a locus will even be marginally functional. So mutants will almost always be totally nonfunctional. This objection, a favorite of mathematicians and physicists, is invalid because mutants do not reach random sequences, but neighbors of functional ones. These are much more likely to be functional.

17. There is an analogy in that argument to the word game where one gets from one word to another by changing single letters. One can go from WORD to GENE by doing: WORD → WORE → GORE → GONE → GENE. There are in all $26^4 = 456,976$ 4-letter words possible, but all but a small fraction are "nonfunctional" words like BGHJ. Nevertheless we can move to nearby words that are functional, if we start at a functional word. Like functional genes, meaningful words are "clustered" in clumps that are close to each other in sequence space.

18. The sequence space is very big: with 1000 bases, there are $4^{1000} \simeq 10^{602}$ sequences in all. Each sequence has 3,000 neighbors one mutation away. Like the famous "six degrees of separation" of humans, no two of these genes are more than 1000 steps apart (in fact, an average pair of unrelated sequences of length 1000 is only 750 steps apart).

19. When mutants are deleterious, natural selection will oppose mutation and will hold the deleterious mutants to low frequencies. For example, with a recessive deleterious mutant whose fitnesses are $1 : 1 : 1 - s$, when the frequency of the deleterious allele $a$ is $q$, in a (large) population of $N$ individuals we expect $Nq^2$ homozygotes, a fraction $s$ of whom die (we are assuming the selection is all by inviability just to make the argument easier to follow). Each death eliminates two copies of $a$. So the total loss of $a$ copies is $2Nq^2s$ per generation. $2Nu$ new copies of $a$ arise by mutation if the mutation rate is $u$. Equating these, we get $q^2s = u$ so that the equilibrium frequency $q = \sqrt{u/s}$. This can be low. If $u = 10^{-8}$ and $s = 0.01$, for example, $q = 0.001$.

20. When the mutant is selected against also in heterozygotes, the result is even lower equilibrium frequencies of the deleterious mutation. If fitnesses are $1 : 1 - hs : 1 - s$, most selection is in the heterozygote (as $aa$ homozygotes are rare). The same sort of argument gives $q = u/(hs)$. This is typically lower: when $u = 10^{-8}$ and $s = 0.01$, then when $h = 0.01$ (so that the effect of the mutant allele in one copy is only 1% of its effect in the homozygote, we get $q = 0.0001$.