GENOME 453 J. Felsenstein
Evolutionary Genetics Winter, 2015

Outline of lectures 7 - 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 the randomness of 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 when that generation reproduces, the fact that the gene frequency was originally 0.4 is forgotten, and the next generation is then expected to have a gene frequency of 0.42. Their offspring then might turn out to actually have 0.41. Now the process starts from 0.41, and forgets both the 0.42 and the earlier 0.4.

3. 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 originally started: it is always starting anew at the new frequency. There is no tendency to have successive changes average out and push the gene frequency back to the original gene frequency – the process has lost any memory of that gene frequency.

4. 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.

5. The analogy has been made to a "drunkard's walk" in which the drunkard steps inaccurately and forgets where he has been. This will be demonstrated in class, without actual intoxication.

6. (For a demonstration *with* actual intoxication, see the YouTube video of a lecture by Andrew Hendry of McGill University. He does the drunkard's walk with a bottle of Dalmore scotch. See here. He credits this to me – he was a Fisheries student at UW in the late 1990s and took this course.)

7. The changes of gene frequency come because individual copies are accidentally having more or fewer descendants than expected. If the A alleles as a whole happen to have many more descendants, over a long period, than the a alleles, this also means that within them, some copies of the A allele have many more descendants than other copies of A.

8. Interestingly enough, the probability of A fixing is the same as its initial gene frequency. This is simply because one copy of the gene in the original population wins the Mendelian and population lottery and is accidentally the progenitor of all the copies in the population. In pure genetic drift, with no fitness differences between the alleles, the chance that a copy wins this lottery does not depend on which allele it is. The chance that it is one which turns out to be A is thus the same as the fraction of As in the original population.

9. On average, genetic drift does not favor one allele over another. If we have a large number of populations drifting independently, some will fix for A, some for a. The overall gene frequency of A among all the populations will not change, but the gene frequency in any one population will change dramatically. (In effect, what has happened is that genetic drift converts variability within populations into differences between populations).

10. Thus Mendelian genetics does not dramatically reduce variation in the population (as blending inheritance does), but genetic drift reduces variation, by having alleles drift to fixation or loss. This can be quite a slow effect if population size is large. Thus a population of size 1,000,000 will take about 4,000,000 generations to lose much of its variability. Smaller populations lose variability proportionately faster.

11. 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.

12. Let us now discuss mutation and migration as they would be in the absence of genetic drift, in a very big population. But keep in mind that these "deterministic" forces (selection, migration, and mutation) actually interact with genetic drift. You will see that in the computer simulation exercise where it will be *very* dramatic.

13. **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 arrived together in a single individual. Random mating will combine them with the other copies at random. In effect there is a *gene pool*. We can just think of migration as averaging the gene frequencies.

14. 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 $0.8 \times p_1 + 0.2 \times p_2$, where p_1 and p_2 are the gene frequencies in populations I and II.

15. That is true because, if we choose a gene copy at random in the resulting population, the chance that it is from population I is 80%, and a fraction p_1 of that it time it is A. It can also be A because the gene is from population II 20% of the time, and a fraction p_2 of those are A. The sum is the weighted average formula above.

16. Thus if they have gene frequencies 0.7 and 0.4 for allele A, the resulting gene frequency in the admixed population will be: $0.8 \times (0.7) + 0.2 \times (0.4) = 0.64$.

17. 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 become equal, and to be the same as if all the gene copies at that locus had been dumped into one big gene pool.

18. The rate at which gene frequencies are equalized depends on the exact geography, but it is true that it has something to do with the rate of migration. Thus if populations exchange 1% of their genes each generation, they will take about $1/0.01 = 100$ generations to move substantially towards equal gene frequencies.

19. Natural selection and genetic drift can cause gene frequencies in different populations to differ. When there is a geographic continuum, such as one that runs 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.

20. (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 when grown in the common garden. These differences must be genetic, as in the garden at Stanford all the plants are raised in a common environment.

21. (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.

22. (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. All these populations are descended from the same flock of birds.

23. Nevertheless 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). Their geographic pattern has thus settled into a cline, differentiated by selection but smoothed by migration. The similarity of the cline to similar patterns of differentiation in Europe is strong evidence for these differences being due to natural selection.

24. **Mutation.** Mutation rates are low (underlying point mutation rates are about 10^{-8} per base per year). In organisms as large as mice the mutation rates can be directly

measured by screening large numbers of offspring for mutants. Some kinds of DNA changes such as trinucleotide repeat expansions and contractions can go faster, up to a mutation rate of about 0.001 per generation.

25. Under mutation, we expect an equilibrium frequency to be reached between forward and backward mutation. Neither allele gets lost, because mutation always reintroduces the other. The gene frequencies slowly move to an equilibrium frequency at which as many As are being converted to a each generation as as are being converted to A. 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. We will see this below when we calculate the equilibrium between selection against deleterious mutants and mutation which causes them to arise.

26. Forward mutation rates are much higher than backward ones. This is not a mysterious pattern – it 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 back, to restore the original sequence (this is not always true – there are so called "second-site revertants" possible too, but not too many of them).

27. When we talk about "forward" mutations to a state of nonfunctionality, and "backward" mutation to being functional, we are grouping DNA sequences into classes, functional and nonfunctional. Underlying these groupings are many possible alleles whose DNA sequences differ. And of course the model is oversimplified: in real life, alleles can occur that are more functional than the original state, and alleles that are partially functional are found too.

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

29. 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, if we think of mutants as drawing new sequences from a pool of totally random sequences, these mutants will almost always be totally nonfunctional. sequences, these mutants would almost always be totally nonfunctional. Similarly with sentences: if we draw a random sentence of the same word lengths as the Hamlet sentence above, it will be something like "`Fi th om qjn yx qs, dpag du ekd czfnvphg`". It is hopeless to try to understand this sentence in any language. (I generated that sentence using a random number generator making 30 numbers from 1 to 26).

30. Some critics feel that this invalidates evolution. This objection is a favorite of mathematicians and physicists who haven't thought enough about how genotypes work or how evolution actually works (it is also important to some objections by creationists and Intelligent Design advocates). It is invalid because mutants do not reach randomly chosen sequences, but neighbors of functional ones. These are much more likely to be functional.

31. In that argument there is an analogy to the word game where one gets from one word to another by changing single letters. One can go from WORD to GENE by making these one-letter changes: WORD → WOR**E** → **G**ORE → GO**N**E → G**E**NE. 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.

32. The sequence space is very big: with 1000 bases, there are $4^{1000} \simeq 10^{602}$ sequences in all. By comparison, there are about 10^{80} electrons in the universe. Each sequence has 3,000 neighbors one mutation away. Like the famous "six degrees of separation" of people in the U.S., 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). So although the space of possible sequences is enormous beyond comprehension, the number of related sequences only a few steps away is quite large.

33. For example, there are 3,000 sequences one step away from a 1000-base sequence, 4.9995×10^6 two steps away, and 4.48650900×10^8 sequences three steps away. That is a very big number but still a tiny fraction of all possible sequences.

34. 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 $2N$ to cancel out and find $q^2s = u$ so that we can solve for the equilibrium frequency at which these two processes balance each other, $q = \sqrt{u/s}$. This is an equilibrium between selection removing deleterious mutations and mutation introducing them. The equilibrium frequency of the deleterious mutations can be low. If $u = 10^{-8}$ and $s = 0.01$, for example, $q = 0.001$.

35. Although the gene frequency of allele a is only 0.001 in this case, that is still 100,000 times higher than the mutation rate. So the copies of the a allele in the population represent a great many generations of past mutations that have accumulated.

36. When the mutant is not purely recessive, but is also selected against in heterozygotes, the result is even lower equilibrium frequencies of the deleterious mutation. If fitnesses are $1 : 1-hs : 1-s$, so that most selection is in the heterozygote (as aa homozygotes

are rare). For example, if $h = 0.01$, the reduction in fitness in the heterozygote is quite small. If $s = 0.01$, then the fitnesses of AA, Aa, and aa are $1 : 0.9999 : 0.99$. This would seem to be an insignificant reduction of fitness in the heterozygotes, and it is unlikely that we would notice this decrease in survival or fertility if we were breeding these organisms in the lab. But there are so many more heterozygotes around that this small reduction does most of the eliminating of a alleles.

37. The same sort of argument as before, equating the numbers of a copies introduced with the numbers eliminated, gives $q = u/(hs)$. This is typically lower than if a were purely recessive: when $u = 10^{-8}$ and $s = 0.01$, then when $h = 0.01$ we get a much lower equilibrium frequency of the mutant: $q = 0.0001$.

38. An interesting case to use to test your understanding of the processes of population genetics would be where two loci each have alleles that are advantageous, but where both together are not advantageous. For example if a crocodile's upper jaw was longer than its lower jaw we could imagine a locus where a rare allele occurs that makes the upper jaw shorter, and another that has a rare allele that makes the lower jaw longer. Either of these would be helpful, but both together would not be. What will happen? The table of fitnesses might look like this (OK, I admit, this is a very naive model of gene action on jaw length):

	BB	*Bb*	*bb*
AA	0.8	0.9	1.0
Aa	0.9	1.0	0.9
aa	1.0	0.9	0.8

Assume for the moment that the genotypes at the two loci are independent, in that a *BB* genotype at locus B is just as likely to have an AA genotype at locus A as is a *bb* individual – this state, called "linkage equilibrium" is expected when there is a reasonable amount of recombination between these two loci.

See if you can figure out the answers to these questions:

- What combinations of gene frequencies at the two loci have the highest fitness?
- If selection "climbs" the adaptive surface, and starts out with low frequencies of A and of B, what is expected to happen?
- What is the effect of genetic drift early in the process?
- Is it possible for both A and B to be lost?
- If different outcomes occur in different populations, what will the effect of migration be?
- If a population reaches, say, a state of being all *AA bb*, what will the effect of mutation be?