Outline of lectures 13-15

Genetic Variation and Neutrality

1. Until 1966, evolutionary geneticists had a limited range of genetic variation they could examine, and nothing like an unbiased sample of variation at the gene level:
   - They could look at morphological variation, but could not know how many loci were varying.
   - They could look at fitness variation uncovered by making whole chromosomes homozygous (for example, in *Drosophila*) but they did not know exactly how many loci were contributing to the effects. They were by definition concentrating on variation chosen because of its large fitness effects, so they couldn’t know whether these effects were at typical loci.

2. Previous to that time there were two major theories about variation at the genetic level. H. J. Muller’s *classical* view was that most loci would have a very common “wild-type” allele, and mutation would maintain, in a mutation/selection balance, low frequencies of mutant alleles that were deleterious. Theodosius Dobzhansky’s *balance* view was that most loci would have multiple alleles maintained by strong overdominance or frequency-dependent selection.

3. In the early 1960’s people started using the technique of protein electrophoresis to study variation at individual enzyme loci.

4. Gel electrophoresis was invented by Oliver Smithies in 1957 (the same guy who won the Nobel Prize 50 years later for unrelated work). It was related to earlier paper electrophoresis methods. A sample of blood is put in a gel made of potato starch or acrilamide, and subject to an electric current for a few hours. This is done under pH and temperature conditions that do not denature the proteins. The gel is then stained for the product (or the substrate) of one particular protein, and bands are seen, where the active enzyme protein is on the gel.

5. These bands show how far the protein of that enzyme has migrated through the gel. It is affected by both charge and conformation of its protein molecule. The method can detect a single amino acid substitution (though some are not detected) and does so in a way that has nothing to do with the fitness effect of the substitution.

6. The single-locus studies in the early 1960’s often found variation at such loci, but there was no overall survey to see how typical this was. People tended to publish “lo and behold” papers showing that they had found that their one locus showed variation. It
wasn’t clear how many other people found no variation at the locus they studied and decided not to publish.

7. In 1966, Lewontin and Hubby and (independently) Harry Harris surveyed populations (respectively *Drosophila pseudoobscura* and humans) at multiple loci. Both projects found a lot of variation, which was a bit of a surprise.

8. The amount of variation is usually summarized in one of two kinds of statistic: heterozygosity or polymorphism:

- **Polymorphism** is the fraction of loci at which the commonest allele is less than 95% in frequency (i.e. all the rarer alleles add up to more than 5%).
- **Heterozygosity** is measured by taking the gene frequencies at each locus, and computing

\[
1 - (p_1^2 + p_2^2 + p_3^2 + \ldots + p_{10}^2)
\]

which is the predicted heterozygosity, as it is 1-(the homozygosity). This is then averaged across loci.

9. Typical values of heterozygosity seen by protein electrophoresis would be about 15% for invertebrates, about 7% for vertebrates. The variation of these typical values is big. For example, in amphibians one can find groups with heterozygosity values around 15%. These are for a variety of enzyme loci, measured by electrophoresis.

10. One issue is whether these loci may be regarded as typical loci.

11. Polymorphism and Heterozygosity, although both are measures of genetic variability, do not necessarily show concordant patterns when we compare natural populations.

12. Enzyme polymorphism reveals only small fraction of the variability on the DNA level. This began to be examined once DNA sequencing was available. The alcohol dehydrogenase (ADH) locus in *Drosophila melanogaster* has two electrophoretic alleles (*S* and *F*). Marty Kreitman (1984) sequenced 11 different copies of the ADH gene (at the time, before PCR was invented, about all that could be done with the time and resources he had) with a total of 43 mutations. One single site was responsible for the difference between the two electrophoretic alleles and was the only site that changed the protein sequence. All the rest were synonymous substitutions.

13. The pattern of variation at the DNA level is similar to what is seen at the protein level – a lot of variation. If we take two copies of the same chromosome (say the two copies that an individual has) we will see one difference about every 1000 bases. This varies a bit (from about 1 in 500 to 1 in 1500) depending on the species. It is a slightly different figure from how often a SNP (Single Nucleotide Polymorphism) is found, since that requires that more than two copies be looked at and that the variation is not rare among them.
14. The observations immediately contradicted Muller’s view, as they projected that hundreds to thousands of loci would be heterozygous in a typical individual. Later it was also realized that Dobzhansky’s mechanism of strong overdominant selection would not fit the observations either. With hundreds of loci becoming homozygous when a whole chromosome was made homozygous, it would predict far stronger inbreeding depression than was actually observed.

15. Many surveys of patterns of variation find suggestive patterns, such as higher heterozygosity in invertebrates than in vertebrates, but do not settle the issue of whether the variation is maintained by selection. They also find that some categories of loci, such as enzymes in the glycolytic pathways, are less variable.

16. Some interesting cases:

- Horshoe crabs (*Limulus polyphemus*) have shown a fair amount of stasis in phenotype over hundreds of millions of years. Do they lack variability at the genetic level? No, showed Robert Selander using electrophoresis in the 1970s. They have normal levels of variability.

- The Northern Elephant Seal (*Mirounga angustirostris*) was reduced to just a few individuals about 1900, and has now recovered to about 150,000 individuals (and still increasing). The Southern Elephant Seal, a very similar species, was not reduced to a small number of individuals, though it was hunted. Sure enough, the Northern Elephant Seal shows very little genetic variability – most loci are fixed for one allele.

17. Lewontin and Hubby had suggested that the data could be explained by either balancing selection or by neutral mutation. Motoo Kimura, the greatest population geneticist of his generation, advocated and greatly developed the latter position, using his formidable theoretical powers to greatly advance understanding of neutrality. His colleague Tomoko Ohta has argued for the importance of nearly-neutral mutations.

18. If neutral mutations are occurring at a locus at a rate $\mu$ per copy per generation, and each one is to a new allele (the “infinite isoalleles” model) and the effective population size is $N_e$, Crow and Kimura showed in 1964 that the expected amount of heterozygosity at the locus is $4N_e\mu/(4N_e\mu + 1)$. The alleles continually turn over, with no equilibrium gene frequencies of any allele, but the level of variation is roughly predictable. (The effective population size is the population size corrected for other details of the life cycle that affect the rate of genetic drift).

19. Low selection coefficients can maintain alleles segregating in populations. All that is required is that $4N_e s > 1$ which means that for $N_e = 10^6$ $s$ can be as low as 0.00000025 and still maintain the alleles.

20. Recall that, for protein electrophoretic variation, Lewontin and Hubby observed about 15% heterozygosity in *Drosophila*, Harris observed about 7% in humans. A $4N_e\mu$ value
of about 0.18 which would be obtained by having $N_e = 10^6$ and $\mu = 1.8 \times 10^{-7}$ will do this.

21. Less variability at some loci or in some parts of the genome is compatible with both theories, as the neutral mutation theory says that the variation is not maintained by selection, but it does not rule out there being selection against deleterious mutants. It is thus not a statement that all mutations are neutral – “purifying selection” can remove deleterious mutants without invalidating the neutral theory.

22. Laboratory experiments such as “population cages” with Drosophila can rule out large selection coefficients above 0.01, thus rejecting Dobzhansky’s view, but they are totally incapable of detecting whether selection is 0 or 0.0001, as lab experiments involve smaller populations and much shorter time spans than apply in natural populations. For example a population of 10,000 individuals (about the limit for a population cage), in a period of 100 generations (about the limit too, that’s 8 years or so), could not detect any natural selection below about 0.0002 even if it were there.

23. A selection coefficient as small as $1/(4N_e)$ can be effective in nature, and that can be far smaller than anything we can detect in the lab. Nature can run an “experiment” millions of generations long with hundreds of millions of flies.

24. The controversy remains largely unresolved after 35 years, although it is most likely that much of our “junk” DNA accumulates mostly neutral mutations. For amino acid variations at protein loci the controversy is still unresolved.

25. Studies on pairs of whole human genomes find hardly any regions where there is evidence of strong balancing selection.

26. At the same time, genome comparisons of closely related organisms show signs of non-neutral patterns of substitution at protein-coding loci. and this is argued to come down against the neutral mutation theory.

27. Whatever its ultimate fate, the neutral theory has played a major role as a “null hypothesis” against which comparisons could be made.