Genetics 453 J. Felsenstein
Evolutionary Genetics Winter, 2002

Outline of lectures 15-16

Electrophoretic 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, and they were then by definition concentrating on variation chosen because of its large fitness effects.

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 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 (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. 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.

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 the protein molecules. 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 examples in the early 1960's often found variation at such loci, but there was no overall survey to see how typical this was.

7. In 1966, Lewontin and Hubby and (independently) Harry Harris surveyed populations (respectively *Drosophila pseudoobscura* and humans) at multiple loci. Both groups 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 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, they 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. The alcohol dehydrogenase (ADH) locus in *Drosphila melanogaster* has two electrophoretic alleles (S and F), Kreitman (1984) sequenced 11 different variants with a total of 43 mutations. One single site was responsible for the two electrophoretic alleles and was the only site that changed the protein sequence. All the rest were synonymous substitutions.

13. The observations immediately contradicted Muller's view. Later it was also realized that Dobzhansky's mechanism of strong selection would not fit the observations either. They would predict far stronger inbreeding depression than was actually observed.

14. 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.

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

16. If neutral mutations are occurring at a locus at a rate μ 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.

17. 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.

18. 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.

19. 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.

20. Laboratory experiments such as "population cages" with *Drosophila* can rule out large selection coefficients above 0.01, thus rejecting Dobzhansky's view, but are totally incapable of detecting whether selection is 0 or 0.001, as lab experiments involve smaller populations and much shorter time spans than apply in natural populations.

21. The controversy remains unresolved after 30 years, although it is most likely that much of our "junk" DNA accumulates mostly neutral mutations. For the functional loci the controversy is still unresolved.