Biology 494: Handout 15: 12 pp. *9 Sept 1994*

Some Notes on Genetics

Genetics nowadays consists of several different fields, all of which are important to evolutionary biologists. There is classical" or "Mendelian" or transmission genetics, which concerns itself with what happens when certain kind of individuals are crossed. Molecular genetics examines the molecular nature of the genetic materials, mapping genotypes, and probing into gene functions and therefore into the initial stages of ontogenies, including gene interactions. Population genetics is concerned with what happens from generation to generation, not in test crosses, but in whole populations of organisms; with changes in gene frequencies and how genic interactions are translated into changes in kinds and proportions of phenotypes; with the details of microevolution, the relative importance of each component of microevolution, and with what kinds of selection exist and how they work. Population geneticists wish to explain the presence of alternative alleles (or heterozygosity or genetic variation) in populations (why didn't one allele replace all others?); and they wish to explain how or why changes in gene frequencies take place. Population geneticists are sometimes called mathematical population geneticists, ecological geneticists, or evolutionary geneticists. Currently, there is a group of geneticists who call themselves evolutionary geneticists and refer to their field as "the new evolutionary genetics," by which they mean that their emphasis is, to a much greater degree, on natural selection -- on conflicts and confluences of interest among units at different levels in the hierarchy of organization of life. They might examine, for example, intragenomic conflict: how nuclear genes evolve, as compared to heritable elements that lie outside the nucleus, and the consequences for the makeup and operation of the cell and the organism.

As I see it, Sir Ronald A. Fisher more or less started the field of population genetics in 1930 with a poorly read and poorly understood (but remarkably important) book called *The Genetical Theory of Natural Selection* (his almost equally famous contemporaries were J. B. S. Haldane, a British mathematician, and Sewall Wright, an American who lived into his nineties, had his life chronicled by a Cornell historian named William Provine, and so got in the last "licks" in his arguments with the other two). Early in his book Fisher compared alleles, in a way that I call "either-or" population genetics (will selection go this way or that, favor this trait or that, therefore whatever alleles are responsible), as opposed to quantitative genetics, which attempts to determine not only which way selection will go but how fast? Even if we behaviorists and adaptationists tend to begin at the other *end* of ontogeny, so to speak, population

genetics is important to us for a lot of reasons, not least because we need to be sure that we know enough about its topics not to err as we go about the business of analyzing phenotypic traits with a behaviorist's or an adaptationist's approach.

Incidentally, the basis of Fisher's population genetics and that of his contemporaries - and of population genetics in general -- is not a settled issue yet. Thus, see S. A. Frank and Montgomery Slatkin. 1992. Fisher's Fundamental Theorem of Natural Selection. *TREE* 7(3):92-95, a discourse at least in part trying to deal with the problem that mathematical population genetics has never escaped dependency on something called the "average fitness" of populations -- in this sense actually interpreting its results under a group selection concept. Fisher tried to escape this trap after he saw what Sewall Wright was doing with it in his notion of an "adaptive landscape" and of progressive adaptation of the group to its environment, saying (Fisher) that . . . the principle of Natural Selection . . . refers only to the variation among individuals (or co-operative communities), and to the progressive modification of structure or function only in so far as variations in these are of advantage to the individual . . . and affords no corresponding explanation for any properties of animals or plants which, without being individually advantageous, are supposed to be of service to the species to which they belong (p. 49,1958 edition). His own writings, however, as well as what the above two recent authors have written (the senior author, Steve Frank, began his evolutionary biology as an undergraduate in this course), indicate that, despite this disclaimer and others, he never quite did escape.

Early in his book Fisher discussed the "traits" of alleles. He said they can be:

1. new mutants or alleles of long-standing (producers of what is generally called "wild type" phenotypes -- the presumably long-existent phenotype in the "natural" or "outdoor" population)

2. dominant or recessive

3. rare or common

4. advantageous or disadvantageous

Next he noted that associations among these categories are not random. Thus, those alleles which produce phenotypes that we call"wild type" are usually common (abundant) alleles and usually dominant. Mutants are usually disadvantageous and rare (and recessive). He asked why this should be so and discussed the question in the following terms:

1. Suppose an allele is dominant and deleterious. We probably will not even detect its presence because it will be present so briefly and in so few individuals. It will immediately be exposed to selection and eliminated.

2. Suppose it is recessive and deleterious. If it is completely recessive we will not be able to detect its presence, and neither will selection, until it becomes frequent enough that heterozygotes mate so that some homozygous recessives are produced. If it happens to be lethal in the homozygous condition we might never see its effects because it might simply cause mated heterozygotes to have fewer offspring than other kinds of pairs. It will increase in frequency according to the mutation rate until homozygotes are produced and then change in frequency until the net rate of mutation toward it is balanced by its selective disadvantage.

3. Suppose it is recessive and beneficial. Again, we won't see its effects (and neither will selection) until homozygotes are produced. If you cannot figure out why it is said that selection works on the phenotype and not on the genotype directly, just consider a recessive allele that is still too rare to produce homozygotes. Until electrophoresis we could not even detect the presence of an allele except by its effects on one or more parts of the phenotype (traits), usually developmentally (ontogenetically) quite remote from the genotype.

4. Now suppose it is dominant and beneficial. It will spread.

So the answer to Fisher's question about nonrandom allelic associations involves natural selection, and in more ways than are yet apparent. Thus, dominance and recessiveness are not "traits" of alleles *per se,* but refer to the effects of alleles on the phenotype *when particular other alleles are present.* Further, they refer to effects of alleles *in relation to* the effects of particular other alleles; the effect of an allele may be dominant in relation to an effect of a second allele and recessive in relation to the same effect of still another allele. Moreover, said Fisher, mutation itself -- the event of mutation -- does not cause dominance or recessiveness because back mutation to a dominant wild type is known. Fisher concluded that alleles must *acquire* dominance by -- guess what -- natural selection! Recessive mutants that become advantageous acquire dominance as they spread, and wild type dominants that become disadvantageous become recessive as they are displaced. The rest of the genome is part of an allele's environment of selection, as is the organism itself, its morphology, physiology, and behavior, and the genetic materials in the somatic cells.

Suppose that dominance is incomplete in an important effect of a new allele. If the effect is deleterious, not only will the allele tend to become less prevalent via selection, but anything that renders the effect more recessive will be saved, and will cause that deleterious allele to persist a little longer in the population. If the effect is beneficial, not only will the allele increase in frequency but anything that renders the effect more dominant will be saved. This is what I mean by "either-or" population genetics, and of course it is a method employed not only by R. A. Fisher but by adaptationists everywhere (so to speak).

What kinds of genetic changes could do this? Fisher reasoned that if other genes at other loci modified the effects of an allele beneficially toward dominance or recessiveness, they could be favored on that account. This notion has been through so much argument that a population geneticist at Cornell (Bruce Wallace) called it one of the great debates of population genetics. But, like most of Fisher's ideas, it has weathered the storms well. It is supported by the fact that dominance in "supergenes" (tightly linked sets of genes that affect the same trait) that produce particular patterns of mimicry or cryptic coloration in butterflies and moths appears to have become altered (in both directions at different times and in different places), and also by the fact that if one takes a dominant allele or supergene out of its usual genetic environment by making a cross between individuals from populations distant from one another, dominance is usually incomplete or imperfect: intermediate phenotypes are obtained. Moreover, dominance can be modified in either direction by selection in the laboratory.

One requirement for Fisher's hypothesis seemed to be that alleles that modify the effects of particular alleles in the population ought to have a way to remain with the alleles they modify during recombination. Such "linkage" could be assisted by loci moving closer together on the chromosome (through saving the appropriate chance translocations) and, once they are close together, by (saving) inversions and any other chance events that reduce the likelihood of translocation of these genes between chromosomes (after an inversion, which refers to a segment of the chromosome reversing itself, crossing-over of the involved genes is less likely because during meiosis in genomes heterozygous for the inversion -- see below -- the genes involved in the inversion may not be able to line up opposite their counterparts, and they are unlikely to cross-over to a different locus).

Probably the greatest challenge to Fisher's particular hypothesis for dominance modification by selection of modifying genes was the suggestion that if dominance is only changed by modifying alleles, most of the genome might become tied up in modifying dominance. A British ecological geneticist named P. M. Sheppard, however, noted that all one

has to imagine is that there are genes that *canalize* development in a generalized way, preventing "breakouts" from a particular pattern of development (that is, maintaining the developing embryo or individual within a certain range of variation at every stage). In this hypothesis, for any allele that led to an unacceptable "exaggeration" of the developing phenotype at any stage, the action of a canalizing gene would represent a dominance modifier, and, presumably, one such gene could canalize a very significant portion of development (for example, by becoming active in response to morphological, physiological, or behavioral traits of the developing organism, that are controlled by many different genes, whenever they exceeded certain parameters), thus modifying dominance in quite a large array of genes.

Also, selection of dominance is only effective in heterozygotes. So a condition for Fisher's hypothesis is that there have to be "enough" heterozygotes in populations, although how many that means is moot. Notice that all of those relationships between alleles discussed by Fisher will be changing in heterozygotes.

Gel electrophoresis and other methods of looking more closely at genes or their (immediate) products has coincidentally revealed far more allelic polymorphisms (therefore more heterozygosity) than anyone ever suspected. The very well-known Harvard population geneticist, Richard C. Lewontin, once said that the amount of polymorphism revealed was "embarrassingly large to be explained by selection alone." Not surprisingly, from such an argument, many biologists immediately supposed that most alleles must be neutral with respect to one another, and that this explains why so many remain in gene pools. Even this argument -- which is certainly not without controversy "backhandedly" gives credit to natural selection as the principal guiding force of evolution: it implies (oversimplifying slightly) that only advantageous alleles or alleles without selective effects can remain in populations.

Polymorphism, or genetic variation within populations, is what population geneticists want to explain. So let's try here to make an exhaustive list of possible explanations for genetic variability (or "polymorphisms") in populations:

1. recurrent mutations

2. accidental fluctuations (drift)

3. frequency dependent selection or advantages in rareness (e.g., sex ratios)

4. temporal and spatial variations in selection

a. geographic (clinal) (e.g., field and ground crickets vary in size and life cycle with latitude)

b. geographic (mosaic) (disruptive selection leading to speciation or balanced polymorphism (e.g., Guy Bush's studies of trypetid flies)

c. temporal (directional selection -- "transitional" polymorphisms)

d. temporal (cyclic)

i. selection causing different life cycles (e.g., field crickets with two generations)

ii. selection affecting different life cycle stages (e.g., t-alleles, caterpillars and moths darkened by different genes)

e. different sexes ("division of labor" within species)

5. Heterozygote advantages

a. intermediate is best (more likely to be a transient polymorphism)

b. the heterozygote has all the best features

Many population geneticists believe (or believed) that heterozygote advantages were the principal reason for maintenance of genetic polymorphisms. Suppose an allele A1 is dominant and beneficial for effect X and allele A2 is dominant and beneficial for effect Y. A homozygote A1A1 has the poorer effect for trait X and the better one for trait Y, and a homozygote A2A2 vice versa. The heterozygote has the best of possible worlds. Of course, if a new allele providing superiority in both X and Y arises, it will replace A1 and A2.

Developmental polymorphisms deserve separate discussion, because they imply that the developing organism has a recurrent possibility of encountering two or more distinctly different environments, and has somehow evolved a genetic background that allows it to detect environmental conditions that precede and predict which of these conditions will occur, and then respond by producing the appropriate phenotype (notice that this appears to be a phenotypic or ontogenetic "system" superior to one that involves genetic polymorphisms, which are never advantageous in themselves, or as polymorphisms per se -- because it will always be better to be able to produce the particular phenotype

appropriate to a particular environment through developmental switches, learning, or whatever mechanism of "plasticity" -- so long as too much possibility of error is not introduced). My favorite example (of course) involves crickets and other insects that produce a long-winged flying, migrating, non-aggressive, non-territorial form or a short-winged, non-flying, nonmigrating, aggressive, territorial form. Sometimes, as in migratory locusts, the different forms are so distinct in so many ways that they were for long times believed to represent different species.

Of course, "polymorphisms" do not fall simply into these two categories. Instead, their genetic correlates are often confusing, equivocal, and resistant to analysis (as with left- and right-handedness and homosexuality and heterosexuality), and they often do not eventuate as two or a few quite distinct phenotypes but as a large number, or even an essential continuum, as with behaviors that are regularly learned. There are many phenotypic variations that I would think should be considered along with polymorphisms which most people would never think of in that context. But the ontogenetic problems are all one big difficult package.

In introducing the "how fast" (do gene frequency changes occur) or quantitative aspects of their field, population geneticists tend to begin with something called the Hardy-Weinberg Law. It was dreamed up by a British mathematician named Hardy and a German physician named Weinberg in 1908, not long after the terms gene and genotype were invented by a Dane named Johannsen. Punnett, for whom the Punnett Square is named, was a Britisher enthusiastic about Mendelism, and a friend of Hardy's. The story is that one night at the dinner table at Trinity College in Cambridge, Punnett asked Hardy to figure out whether or not recessive genes would be eliminated in a population in the absence of selection -- in other words, would they disappear just because they were recessive. Apparently your intuition is supposed to tell you they would: presumably, if your intuition doesn't work that way you might find little (well, less) usefulness in the Hardy-Weinberg Equilibrium.

A Punnett Square is a diagram in which you can determine the frequency of different kinds of genotypes if you know the kinds of alleles available; all of you have encountered it in introductory biology courses. One can set up a similar diagram and plot in it not the gametes from a cross between two individuals but the frequencies of the alleles in a population (as indicated by the frequencies of the phenotypes to which they give rise), thereby calculating the frequencies of the genotypes that will occur (or vice versa) and transforming himself (well --partway . . .) from a Mendelian (or "transmission") geneticist to a population geneticist.

Consider two alleles A and a, with frequencies of p and q, respectively.

	p	q
	90%A	10%a
p 90%A	0.81 AA	0.09 Aa
q 10%A	0.09 Aa	0.01 aa

If we wish, we can make it into a binomial equation and say that:

$(A)2 + 2Aa + (a) = 1$

$p2 + 2pq + q2 = 1$

$0.81 + 0.09 + 0.09 + 0.01 = 1$

If conditions are met, genotype frequencies and gene frequencies go to equilibrium (do not change further) in one generation. If you know q2 only, for example, you can predict all the rest. The Hardy-Weinberg equilibrium describes conditions under which no evolution occurs. It shows that in the absence of selection and with mating random with no immigration or emigration, no sampling error (drift), and no mutation (or with equal back mutations) the frequency of the genes won't change. In other words, Hardy's question was answered: recessiveness has no direct effect on frequency. In a back-handed way the Hardy-Weinberg equilibrium represents a kind of proof of evolution. It tells us that evolution is inevitable. Consider those conditions:

1. absence of selection: alleles have to be neutral
2. random mating: we can't have incest avoidance or viscosity or selective mating of any kind
3. no mutations: there are a few loci for which no mutations have yet been detected, but . . .
4. no sampling error: any time a new population is founded there's likely to be drift.

Under Hardy-Weinberg conditions, then, the gene frequencies of a particular generation depend upon the gene frequencies of the previous generation and not upon the genotype frequencies. The frequencies of different genotypes produced through random mating depend only upon the gene frequencies. After one generation, genotype frequencies will remain stable.

Suppose you were considering two alleles in a population, and you saw that in a population of 140 individuals there were 3 homozygous A1A1 and 3 homozygous A2 A2 and the rest were heterozygotes. How would you know if this meant there is heterozygote advantage (heterosis, overdominance, hybrid vigor)? You could calculate what the genotype frequencies should be from the Hardy-Weinberg equilibrium and run a chi-square or some statistical test to see if this particular deviation is significant.

Did you ever consider that a population might be composed of all heterozygotes for a locus with only two alleles? Both homozygotes could be lethal.

You can prove that under Hardy-Weinberg conditions (above) the equilibrium genotype frequencies will be reached in one generation. If you begin with AA = 80; Aa - 20 (90/10) or with AA =. 90, Aa = 0, aa =10 (90/10); or with AA =170, Aa = 20, aa =10 (180/20). All of these combinations involve the same gene frequencies as in my example (above), and therefore the same equilibrium genotype frequencies after one generation.

So, if anyone asks: what makes evolution go? You can tell them it's because the Hardy-Weinberg equilibrium rests on assumptions never realized in nature.

I wouldn't want you to believe that whenever no measurable gene frequency changes are occurring it means evolution is not occurring. Selection does two things: It spreads new, better genes through populations and it hangs on to genes that keep mutating only to inferior alleles. If someone tells you, for example, that some protein is the same, or very nearly so, in organisms that seem by the fossil record to have diverged several million years ago, than I would say you have some evidence for a hypothesis that a better allele than the one making that protein has not been produced by mutation during all that time. Eliminating all those inferior alleles for so long is powerful selection. Evolution is involved in the *maintenance* of adaptation, and some people think that's a better definition of it than anything else.

Now, in fact, you can use this same method as above to calculate the effects of certain amounts or intensities of selection -- to determine rates of gene frequency change or rates of evolutionary change. In other words, we can state predictive generalizations about directions of selection (will

it go this way or that way -- either-or population genetics) and often understand them without calculations. But discussing *rates of* change requires mathematical calculations -- and sometimes those mathematical calculations can also surprise our intuitions about directions of selection. An example I remember from some recent reading is that Sir Ronald Fisher was able to show that matings of double cousins have a greater inbreeding effect than do uncle-niece or aunt-nephew matings, even though in Britain the law allows double cousin marriages and doesn't allow the other (but I don't remember why it should be so!). This is not really socially relevant, for we can also note that the law sometimes forbids marriages between stepchildren and step parents when this would cause no inbreeding at all. But it might have surprised your intuition, and I couldn't think of a better example.

Calculating gene frequency changes under selection, the population geneticists operate something like this:

Let genes A and a have frequencies p and q, with A completely dominant to a. Strength or intensity of selection is measured as the proportionate reduction in gametic contribution of a particular genotype, compared with the favored genotype. We can call this reduction S (selection coefficient) and say that the favored genotype has a fitness of 1, the disfavored one a fitness of 1-S. If the selective coefficient is 0.1, for example, then the disfavored genotype produces 90 zygotes for every 100 zygotes produced by the favored genotype.

In this fashion you can use a binomial equation to calculate the frequencies of genotypes after one generation of selection. And you can calculate the "average fitness" of the population by multiplying the value given for the less reproductive allele by, in this case, 0.9.

	AA	Aa	aa
fitness of different genotypes	1	1	1-s

So we can calculate genotype frequencies after one generation in this fashion:

$$p2(1.0) + 2\ pq(1.0) + q2(1.0 - s) = 1\text{-}sq2 \text{ or } W$$

From this, you can figure out how to calculate the gene frequencies, the changes in gene frequencies per generation, and lots of other things that I won't even try.

W-bar, used to symbolize the "average fitness of the genotypes in the population," is lowered (above) because the proportion of the less

reproductive genotype, aa, will be lowered. But what is this figure, W-bar? It is the "average" fitness of genotypes in the population. It is a figure that has little meaning except when evolution is a process of group selection -- except when gene frequencies are determined by groups going extinct rather than individuals failing to reproduce. It's a figure that actually can be *reduced* by selection at the individual level. Now, usually, one supposes that mathematics is so nice and quantitative and internally consistent that you just cannot make mistakes with it. Let me tell a story that will also relate to criticisms of group selection conclusions by ecologists and behaviorists.

Suppose you obtain, as you do, a figure that can be described as the "average fitness of the genotypes in the population." But then you shorten it to "average fitness of the population" and then to "average population fitness" and then to "population fitness." After all, you're aware that the population is the unit of evolutionary changes (individuals and genes don't evolve; populations do), and you're a *population* geneticist -- why not?

In 1950 H.J. Muller referred to the deleterious recessive alleles in a population as that population's "genetic load." He was talking about humans, and the term caught on: All of our bad genes are our "genetic load." He could even talk about how many genetic deaths would be necessary to get rid of this "genetic load."

The idea of genetic load fit in nicely with the notion of population fitness. If a population had a deleterious allele, it had a genetic load and a lowered W-bar or average fitness.

But here the population geneticists fell into a trap, which I have discussed in a part of the article excerpted in Handout 14. The trap existed because they always set the fitness of the best genotype at 1.0 to keep the population's fitness from climbing (numerically) in their calculations. Because of this mathematical convenience, every time any population had more than one allele at a locus, it was seen as having a genetic load and a lowered fitness (do you see why?). Even if the new allele was better, the population still had a load because now the old allele would be poorer. In fact, it had an enormous load the minute the beneficial allele appeared because at first the beneficial allele would be rare (and the amount of the load would depend on the relative frequencies of the two alleles, the load being large when the deleterious -- here previously existing -- allele was common). As J. B. S. Haldane put it, the population would have to undergo many "genetic deaths" to get rid of the old allele.

Now I'm oversimplifying a bit, but there are two monstrous mistakes here that we can identify. First, population geneticists goof in assigning the notion of fitness to the population. The only fitness that

counts very importantly or consistently in evolution is the relative fitness of the individuals within a population. As Williams put it in 1966, natural selection is better versus worse (usually among individuals) in the immediate situation. If a new beneficial allele appears in a population, that population's size and density can both increase as the new allele spreads; they increase because of the appearance of the new allele, which is being interpreted as causing a lowered population fitness! Furthermore, a monomorphic population can keep its fitness at 1.0 and go steadily to extinction because no new beneficial alleles appear. An anthropologist named Alice Brues pointed out that the real cost is in not evolving! In this particular sense genetic load theory is silly.

It is curious that population geneticists made the mistake of thinking of the population as the unit of selection as a result of the mathematics of genetic load theory, the ecologists did it as a result of trying to understand population regulation, and the behaviorists did it because they didn't understand altruism.

What was the other mistake some population geneticists made? How complex is selection across time and space in a species' geographic range? No one has the foggiest notion. Can variations and oscillations in selection all by themselves maintain genetic variations in a population? Everybody admits that clinal variation can; but what about temporally oscillatory and geographically mosaic variability in selection? Population geneticists have concluded that it would be unlikely, at least beyond certain points, because the standard calculations indicate that the size of the effect of selection on gene frequencies, measured by proportionate changes in gene frequencies, depends not only on the reduction of fitness in the genotypes carrying the deleterious gene but also on its frequency in the population: selection is less effective on rare alleles. Nevertheless, every calculation or assignment of a selective coefficient to a whole population is some kind of average across both space and time. There is not necessarily any effort to look within the population to see how selection may vary across the geographic range of the population or from time to time in the same places. There is no particular upper limit with respect to either of those things, and because the environment of an allele is not only the external environment but also the phenotype (body, soma: physiology, morphology, and behavior) of its bearer, as well as its genotype, it is at least theoretically possible to think of every copy of every allele as being in a different environment. This suggests (to me) that maybe there is no particular amount of genetic variation in a population that can legitimately be called "embarrassingly large" to be explained by selection alone.