Natural Selection, Variation, and the Hardy-Weinberg Law

I. Definitions:

- A. Natural Selection: the process by which better adapted forms in a population increase in frequency in the population (at the expense of less well-adapted forms)
- B. Variation: the **heritable** differences among individuals in a population
- C. Adaptation: a feature of an organism that allows it to reproduce better than if it lacked the feature.
- D. fitness: the number of offspring an individual produces. Relative fitness is the number of offspring an individual produces divided by the number of offspring produced by the most fecund individual)

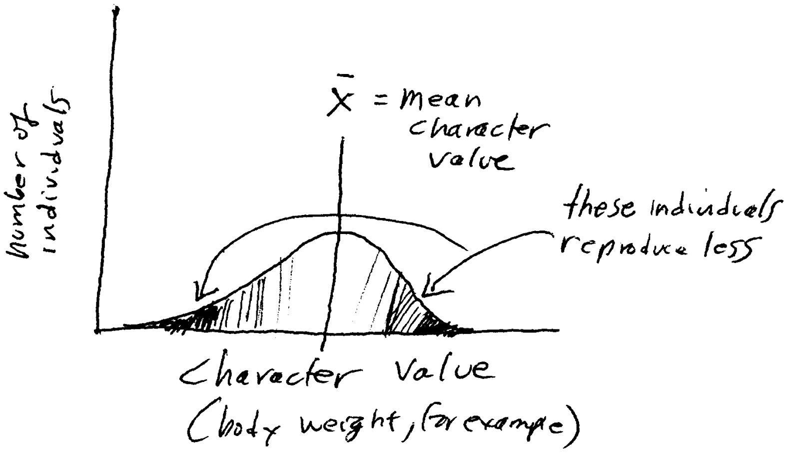
II. Conditions. If the following conditions exist, there WILL be natural selection.

- A. **excess fecundity**: more organisms produced than can survive to reproduce: this produces automatic competition for resources important to survival
- B. **heritable variation in fitness**: i. e., variation in traits important to survival and/or reproduction

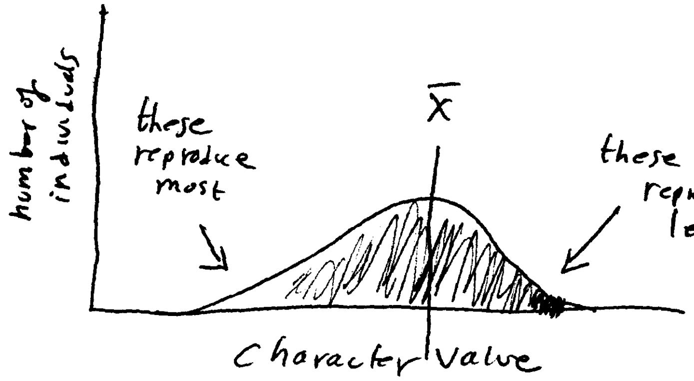
(Ridley says 4 conditions: reproduction, heredity, variation, and variation in "fitness"; I cover these with the two conditions above.)

III. Major modes (types) of natural selection

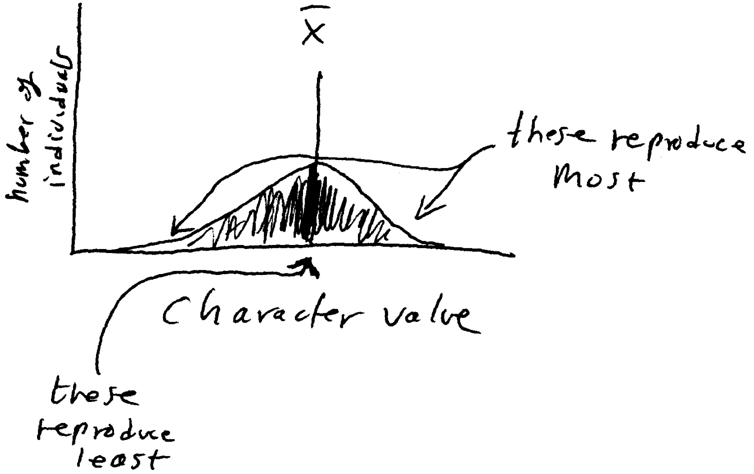
- A. stabilizing — individuals nearest the mean of the character distribution reproduce best; those farthest from the mean reproduce least. This is probably the most common mode of selection, as most populations are pretty well adapted to their environments for most traits. Birth weight in humans is the classic example: babies with very low or even very high birth weights don't survive infancy as well as those between 7 and 9 lbs. birth weight (data from 1935-1946 in England). Results in no change in mean character value (unless selection is at least partly directional), may result in a decreased range of character values (decreased variance).



B. directional — individuals at one end or the other of the character distribution reproduce best. This is probably the next most common mode of selection (after stabilizing selection). Many traits in nature are directionally selected, such as the speed at which rabbits can run — the faster a rabbit can run, other things being equal, the more likely it is to escape from a coyote. However, even if there is selection for faster running speed, there may be counter-selection. For instance, if selection for faster speed tends to select for rabbits with longer leg bones (to cover more distance in one jump), those longer bones may be more brittle and tend to break, in which case the rabbit still becomes coyote food. Thus, selection may not be able to improve running speed. Intense directional selection often has unintended consequences, as plant and animal breeders continually find out. This is because selection on one trait is most often really selection on the **many** genes contributing to that trait, as well as those genes next to (linked to) the selected genes on the same chromosomes. Many breeds of dogs, for instance, are prone to hip dysplasia: selection for other traits has also increased the frequency of the genes causing hip dysplasia in these breeds.



C. dispersive/disruptive — individuals at both ends of the character distribution reproduce better than those near the center of the distribution. Apparently much less common in nature than either stabilizing or directional selection. Does **not** result in new species if there continues to be random mating among individuals. May increase the range of character values (increase variance); does not change mean unless there is a selection is at least partly directional (that is, unless the disruptive selection has a directional component).



IV. Variation in natural populations: Mutation

A. allelic: generally, 1/4 - 1/3 of loci polymorphic (have at least two alleles) in a population, 10% of loci in an individual are heterozygous (cheetahs are notable for being low in variation). Morphological loci often lower in polymorphism than biochemical loci, probably because there is more intense selection on morphological traits than the average protein. DNA variation does not necessarily translate into protein variation (many 3rd base changes in codons do not change the amino acid translated); protein variation does not necessarily translate into fitness variation (because some alleles may not have an effect on fitness: these are called neutral alleles).

B. transitions (purine-purine or pyrimidine-pyrimidine), transversions (purine-pyrmidine), frame-shift, nonsense (stop), synonymous are all types of biochemical mutations

C. gross (large) or chromosomal: some species highly polymorphic (Drosophila, some grasshoppers), some not (primates). Inversion polymorphisms best studied; also some translocation polymorphisms. These both have effects on recombination. Human chromosome structure differs from the rest of the great apes in having one translocation (Robertsonian fusion): our #2 chromosome is composed of the same genetic material as found in two of the chromosomes of the other great apes. We also differ by some inversions in some chromosomes.

D. rates: in eukaryotes, averages about 1 in 100,000 (10^{-5}) per locus per generation (1 in 10,000 [10^{-4}] to 1 in 1,000,000 [10^{-6}] is most of the range). In bacteria, generally 10^{-8} to 10^{-9}. This difference may be partly due to number of cell generations per reproductive generation. In bacteria, it is 1, in eukaryotes, it is unknown, but only 42 doublings make the whole three trillion cells in a human form just one cell. As far as the gametes go, it is probably near 20 (10-35) cell generations per reproductive generation.

E. how does mutation happen? at random with respect to fitness, except some new evidence, still tentative, of at least partly directed mutation in bacteria. Insertions of transposable elements may "direct" mutations.

V. Population Genetics — **"The theory of population genetics is the most important, most fundamental body of theory in evolutionary biology."** -- Ridley

A. Introduction to population genetics: allele (gene) and genotype frequencies.

1. In simple Mendelian genetics, we are only concerned with the mating between two individuals, and thus only consider, for one locus, the two alleles each of those individuals carries. When working with a whole population, instead of doing matings between single individuals, we mate all the females at (usually) random with all the males to produce the next generation of the population. In the simplest case, we would study one locus with 2 alleles, without dominance. We would therefore have three possible genotypes: A_1A_1, A_1A_2, and A_2A_2. The frequency of the A_1 allele is the proportion of the A_1 allele among all the alleles in the population; the frequency of the A_2 allele is the proportion of the A_2 allele among all the alleles in the population. To find the frequency of the A_1 allele, we simply count the number of A_1 alleles in the population and divide by the total number of alleles (at that locus) in the population. We commonly denote the frequency of the A_1 allele as: $f(A_1)$,

and often shorten this even further to the single letter p. The frequency of the A_2 allele is denoted $f(A_2)$ or simply q. The following is an example of how to determine these frequencies in a sample of a population:

Let us sample a population of plants in which flower color is determined by a single locus with two alleles and no dominance. Let red flower color be one homozygous type, pink be the heterozygote, and white be the other homozygote. We thus have:

Phenotypes:	Red	Pink	White
Genotypes:	A_1A_1	A_1A_2	A_2A_2
Numbers in our sample:	50	45	20

To calculate p (the frequency of the A_1 allele), we note that there are 50+45+20=115 individuals in the sample. Each of them has two alleles at the locus, so the sample has a total of 230 alleles at this locus. How many of these 230 alleles are the A_1 type? Each A_1A_1 (Red) individual has TWO A_1 alleles, each A_1A_2 (pink) individual has only one A_1 allele, and the A_2A_2 (white) individuals have no A_1 allele, because both of their alleles are of the A_2 type.

Thus, there are 50+50 A_1 alleles carried by the A_1A_1 individuals **PLUS** 45 A_1 alleles carried by the A_1A_2 individuals in this sample. This is a total of 145 A_1 alleles out of a total of 230 alleles. Therefore $p = \frac{145}{230} = 0.63$(to two decimal places). Note that all the rest of the alleles must be A_2 alleles, since there are <u>ONLY</u> A_1 and A_2 alleles at this locus. Therefore, the proportion of A_2 alleles must be: 1 minus the proportion of A_1 alleles. That is, q = 1-p, or p + q = 1. In the above sample, q = 1-0.63 = 0.37.

So, p = 0.63 and q = 0.37 are the ALLELE FREQUENCIES (sometimes called gene frequencies) in this sample at this locus.

It is not possible to actually count alleles to determine allele frequencies if there is dominance; however, if certain assumptions are made, allele frequencies may be estimated. One can also calculate allele frequencies for loci with more than two alleles; however, that exercise is beyond the scope of this class.

B. Introduction to The Hardy-Weinberg Law

The Hardy-Weinberg Law (H-W) is the evolutionary genetics equivalent of the law of inertia in physics. It specifies what happens to allele frequencies if there is no evolutionary force acting to change them. The H-W law states the following: **<u>allele and genotype frequencies will remain unchanged through generations unless there is some force acting to change them</u>**. Commonly, we list five forces that can change allele and/or genotypic frequencies. Since these are the forces that change allele frequencies, ***THESE ARE THE CAUSES OF EVOLUTION***.

The forces are: selection, migration, genetic drift, non-random mating, and mutation. (There are also other, generally less important forces, but they are of interest only to specialists.)

It is important to note that evolution does not just occur: it takes one of the above forces to make it happen.

Below, these forces will be gone over in some detail.

C. Allele frequencies, genotypic frequencies and the H-W law.

Allele frequencies and their calculation were covered above. Genotypic frequencies are simply the frequencies of each genotype in the population sample. In the previous example, we had 50 A_1A_1, 45 A_1A_2 and 20 A_2A_2 for a total of 115 individuals. The frequency of the A_1A_1 genotype is simply $\frac{50}{115} = 0.43$ (to two decimal places). The frequency of the A_1A_2 genotype is $\frac{45}{115} = 0.39$, and the frequency of the A_2A_2 genotype is $\frac{20}{115} = 0.17$. Note that **due to rounding error only**, these add to only 0.99. If all decimal places were kept, they would sum to 1.000 exactly. We commonly denote the frequency of the A_1A_1 genotype as $f(A_1A_1)$, the frequency of the A_1A_2 genotype as $f(A_1A_2)$, and so forth.

Note that we can also calculate p as the frequency of the homozygote for A_1 plus half the frequency of the heterozygote:

$p = .43 + \frac{.39}{2} = 0.63$

q, of course, is 1-p = 1-0.63=0.47

THIS IS THE USUAL WAY OF CALCULATING ALLELE FREQUENCIES: WE CALCULATE GENOTYPIC FREQUENCIES FIRST, THEN USE THESE TO CALCULATE ALLELE FREQUENCIES. THIS IS EXACTLY THE SAME AS ACTUALLY COUNTING THE ALLELES IN THE SAMPLE.

In terms of p and q, the H-W law states that for any allele frequencies p and q, the genotypic frequencies of the next generation will be in the proportions p2 = $f(A_1A_1)$, 2pq = $f(A_1A_2)$, and q2 = $f(A_2A_2)$. This can be easily seen in the following table, where we cross all the males in a population to all the females in a population. Among the males, p = $f(A_1)$ and q = $f(A_2)$; among the females the same proportions of alleles are present. [This is equivalent to saying that if 30% of people in a room are blond, we expect that 30% of both the males and 30% of the females should be blond. That is, there are no differences in allele frequencies between the sexes.] The table looks like a simple Mendelian cross between two

heterozygotes, but there is a crucial difference: we are crossing a population of males with a frequency of the A_1 allele which is probably NOT equal to ½ to a population of females with the same frequency of the A_1 allele. Thus, the proportions of the genotypes in the offspring do NOT always equal ¼, ½, ¼, but they always do equal p^2, 2pq and q^2 for whatever values p and q have.

		allele frequencies among males	
		$p = f(A_1)$	$q = f(A_2)$
allele frequencies among females	$p = f(A_1)$	$p \text{ x } p = p^2 = f(A_1A_1)$	$p \text{ x } q = pq = f(A_1A_2)$
	$q = f(A_2)$	$q \text{ x } p = pq = f(A_1A_2)$	$q \text{ x } q = q^2 = f(A_2A_2)$

Note that the frequency of the heterozygotes among all the offspring equals pq + pq or 2pq.

Thus, we have the H-W genotypic frequencies:
p^2 A_1A_1, 2pq A_1A_2, and q^2 A_2A_2.

Now calculate the allele frequencies in the offspring, whose genotypic frequencies are p^2, 2pq and q^2. The f(A1) will be $p^2 + \frac{2pq}{2} = p^2 + pq = p^2 + p(1-p) = p^2 + p - p^2 = p$

Thus, we started out with $f(A_1)=p$, and we let the males and females randomly mate with no selection, migration, genetic drift or mutation, and we end up in the next generation with $f(A_1)=p$. This shows that allele frequencies do not change under the H-W conditions. This is an example, not the actual proof, of the H-W law.

It can be shown in the actual proof of the H-W law that each generation the allele frequencies will be p and q, and each generation the genotypic frequencies will be p^2, 2pq and q^2, unless there is some force acting to change them. This actual proof of the H-W law is beyond the scope of this class.

D. Testing a sample of a population for H-W

The first thing generally that an evolutionary geneticist would want to know about a trait in population is whether the genotypes are in the H-W proportions of p^2, 2pq and q^2. If so, there is NO EVIDENCE that evolutionary change is going on in the population; if the genotypes are NOT i the H-W frequencies, one or more of the forces must be acting to cause evolutionary genetic change. In this class, we will only test a population sample in which we have one locus with two alleles and no dominance. If there are only two alleles and dominance, we cannot test a population for H-W, for the same reason that we can't count the alleles. It is possible to test for H-W if there are more than 2 alleles at a locus, but such is beyond the scope of this course.

The following is the procedure for testing a population with one locus, two alleles, no dominance, for agreement of the observed genotypic frequencies with those predicted by the H-W law. The procedure was gone over in class; this is just a reminder of how to proceed.

A population consists of 45 A_1A_1, 110 A_1A_2 and 233 A_2A_2 individuals. Test this population for Hardy-Weinberg genotype frequencies. Use 3.84 as the critical value for your test.

$$45+110+233=388=N \qquad p=\frac{45+\frac{110}{2}}{388}=0.2577 \qquad q=1-p=.7423$$

$$p^2N=(.2577)^2\text{x}388=25.8 \quad 2pqN=2\text{x } 0.2577\text{x } 0.7423\text{x}388=148.4 \quad q^2N=.7423^2\text{x}388=213.8$$

	A_1A_1	A_1A_2	A_2A_2
observed	45	110	233
expected	25.8	148.4	213.8
$\frac{(o-e)^2}{e}$	14.3	9.9	1.7

$$\chi^2=14.3+9.9+1.7=25.9$$

(with 3-2=1 d. f.) Critical value = 3.84. $\chi^2 >$critical value (25.9>>3.84), so NOT in H-W

One caution with the H-W test: if you are actually sampling a MIXTURE of individuals from two separate random mating populations, you may get results not in agreement with H-W due to this mixture. If this is the case, you will have MORE observed homozygotes than expected homozygotes, and FEWER observed heterozygotes than expected heterozygotes. This phenomenon is called the Wahlund effect. In the above example, note that there ARE more observed homozygotes than expected (45>25.8; 233>213.8) and fewer heterozygotes than expected (110<148.4). Without further knowledge, it is at least possible that the sample above was in fact a mixture of individuals from two separate populations, each in the H-W frequencies.