

B.D. Mishler, Dept. of Integrative Biology, 2-6810, bmishler@berkeley.edu

Evolution lecture #11 -- **Hardy Weinberg departures: genetic variation & drift** -- Oct. 25<sup>th</sup>, 2006  
459-464 (ch. 23) in 7th ed.  
450-456 (ch. 23) in 6th ed.

- **Genetic variation, genetic drift (summary of topics)**

- **Extent of genetic variation in natural populations**

- Examine the extent of genetic variation in natural populations, and understand the concepts of the **neutral, balancing selection** and **evolutionary lag schools** to explain this variation

- **Deviations from Hardy Weinberg (HW)**

- Explain the consequences of violating each of the assumptions of the HW law: **non-random mating, mutation, migration**, and **genetic drift** (**selection** is for next lecture)

- **Genetic drift**

- Understand the **short and long term effects of genetic drift** on the genetic structure of populations, and the consequences of **founder effects** and **bottlenecks**

- **Sexual versus asexual reproduction**

- Contrast **sexual** and **asexual** reproduction in terms of the generation of genetic variability

- **Extent of genetic variation in natural populations**

**How much genetic variation is there in natural populations?** Before 1966 there were two disparate views on the extent of overall genetic variation in natural populations: classical and balance.

The **classical** view assumes that at nearly every locus every individual is homozygous for a wild-type allele. In addition, each individual is heterozygous for rare deleterious alleles, and occasionally heterozygous for a selected allele maintained in the population by balancing selection.

The **balance** view in its extreme form on the other hand assumed that there was a lot of genetic variation in populations so that most individuals will be heterozygous for alternative alleles at very many of their loci. This genetic variation was believed to be maintained by some form of balancing selection.

The year 1966 is important in population genetics, as it marks the use of an objective test to measure the extent of genetic variation in populations—gel electrophoresis. The initial, and later, studies showed that more than approximately 30% of loci (and this is an underestimate) exhibit variation in natural populations.

So, we now know, and more recent DNA based technologies have confirmed this, that a great deal of variation does exist in natural populations. In humans approximately 1/1,000 DNA base pairs is polymorphic (referred to as a SNP—single nucleotide polymorphism). In contrast, humans differ from chimpanzees approximately every 1/100 base pairs.

From these observations, it would seem that the balance school wins out. However, the classical theory has been retained in terms of the so-called **neutral** (or neo-classical) **theory**.

Also to consider is that some, or much, of the variation in natural populations may represent a transient polymorphism—the **evolutionary lag school**. The argument is that there will ultimately be changes in a species ecosystem (via environmental changes or evolutionary advances by other species) and consequently if a species is to survive it must evolve continually and rapidly to catch up to the latest changes in its ecosystem.

**neutral school:** much of the genetic variation in populations is evolutionary noise, and the allelic variants are selectively equivalent.

**balance school:** most variation has adaptive significance and is maintained by some form of balancing selection.

**evolutionary lag school:** much of the variation in a population is transient variation, as advantageous alleles replace other alleles. Even if an allele is selected it will take a long time to become established in the population unless the selection is extremely strong (for example, with selection of 1% it takes 2,000 generations to fix an allele in a population, which equates to about 45,000 years for humans).

**Which school is right?** There is controversy as to which is the predominant factor creating the high level of genetic variation seen in most natural populations. While selection certainly operates, nevertheless, much genetic variation is probably neutral. All three factors probably play an important role.

**Apportionment of genetic variation:** initial studies of the degree of genetic differentiation of human populations and ethnic groups using allozyme data from gel electrophoresis studies showed that **most genetic variation in humans is found *within* populations** (85%), with the remaining variation equally divided (7.5% each) *between* populations within ethnic groups, and between ethnic groups. Similar results have been found with RFLPs, microsatellites, and HLA data.

## • Deviations from Hardy Weinberg (HW)

**Deviations from Hardy Weinberg assumptions:** the strength of the Hardy Weinberg (HW) law is that one can deviate from the assumptions quite a bit and the data will still approximate Hardy Weinberg proportions (HWP).

The weakness of the HW test is that the deviation from the assumptions has to be very strong in order to detect the effect of this evolutionary force, e.g., selection. Deviations from HW assumptions involve:

- (1) **Non-random mating**, e.g., inbreeding, mate-choice.
- (2) **Mutation**. The effects of mutation in populations are usually negligible as mutation rates are low—but mutation is an important force in creating new variation.
- (3) **Migration** is important if the migration rate is high and the two population are very distinct genetically.
- (4) **Genetic drift** due to small population size (chance effects)—genetic drift effects are important in both small and large (but finite) populations in terms of short and long term effects of changes in allele frequencies over generations due solely to drift effects (note that the finite size of a sample taken from a population is taken into account in the statistical tests for HWP and finite population size itself does not cause significantly detectable deviations from HWP).
- (5) **Selection** has to be strong to cause deviations from HWP, e.g., it can be detected with sickle cell anemia (selection is the topic for the next lecture).

### More details about the first four of these:

**1. Non-random mating:** individuals with certain genotypes sometimes mate with one another more commonly than would be expected on a random basis.

When like mates more often with like we term this positive assortative mating, e.g., height, IQ. Positive assortative mating increases the proportion of homozygous individuals but does not alter the allele frequencies. Negative assortative (or "disassortative") mating is preference for *different* genotypes. For example, there is evidence that a person is attracted to potential mates by pheromones indicating that the other person has different alleles in the immune system than he/she has.

With self-fertilizing plants the level of heterozygosity is reduced by 1/2 each generation (with the remaining 1/2 divided equally between the two homozygous classes. Self-fertilizing plants have more homozygotes than expected under Hardy-Weinberg and often show significant deviations from HWP.

Inbreeding (mating with close relatives) is another form of non-random mating. Relatives are more likely to carry the same recessive allele for a rare recessive trait—inbreeding increases the number of affected individuals with homozygous rare recessive traits. Marriages between first cousins have about twice the rate of birth defects as random matings.

- 2. Mutation:** in and of itself does not change allele frequencies to a noticeable extent as mutation rates are low. However, mutations are the **raw material** of evolution, the ultimate source of genetic variation. Although the frequencies of mutants are initially rare, and most are lost from the population, nevertheless some increase in frequency due to genetic drift effects and also selection (next lecture).

Mutation is any change in the DNA sequence that is transmitted to offspring. A mutation can be a change in a single nucleotide, the insertion or deletion of one or more nucleotides, the rearrangements of chromosomes or parts of chromosomes, as in the chromosomal fusion example in Fig. 23.9, the duplication of one or more chromosomes, and even the doubling of the whole genome because of an error in meiosis. Mutation does not change allele frequencies but it creates new alleles which are then affected by selection and drift.

Mutation rates of single nucleotides are very small, roughly  $2 \times 10^{-9}$  per base per generation, yet these are important mutations for evolutionary changes. Larger scale mutations may be much more frequent. Trisomy 21 is a whole-chromosome duplication in humans that causes Down Syndrome. A woman aged 39 has a 1/137 chance of having a trisomy-21 child. Mutations of this type play no evolutionary role because the extra chromosome creates sterility because of problems during meiosis.

Duplications of whole genes plays an important role in evolution. One copy can retain its function while the other can diverge to perform a similar or quite different function. Alpha and beta and other globin genes are the result of an ancient duplication.

- 3. Migration:** is the movement of individuals from one population into another, which can alter allele frequencies, and if there are large genetic differences cause a statistically significant deficiency of heterozygotes from Hardy-Weinberg expectations.

Gene flow results from the movement of gametes or individuals. A high level of gene flow prevents the divergence of different populations of a species. In the absence of gene flow, isolated populations will tend to become more different because of the combined effects of genetic drift, mutation and natural selection. A low level of gene flow moves alleles to other populations. The effect is similar to that of mutation, in creating more variability on which natural selection can act.

Gene flow is a major issue in discussion of the use of genetically engineered plants and animals. Starlink corn is an example (<http://ccr.ucdavis.edu/biot/new/StarLinkCorn.html>). Many domesticated species can interbreed with closely related wild species. When a gene is inserted into one variety of a domesticated plant, the question is whether it will spread to other varieties and to wild relative. One of many examples is Starlink corn which has a gene producing Cry9C

inserted into it. Cry9C is a protein isolated from a common soil bacteria; *Bacillus thuringiensis*. The Cry9C protein is effective against caterpillars because it binds to different sites of the insect gut and destroys the stomach cells. Starlink corn was approved for animals but not for humans. This protein has been found now in roughly 10% of corn tested by USDA inspectors. Several lawsuits have been filed as a result of adverse effects of eating taco shells and other corn products containing Cry9C. Similar issues arise with the insertion of herbicide resistance genes, but there is concern is with the escape of genes to wild relatives and the creation of super-weeds.

#### 4. **Genetic drift:** (chance effects) random change in the frequency of alleles at a locus.

**short term genetic drift effects:** cause changes in allele frequencies, both in small and large populations (Fig. 23.7 (7th) (Fig. 23.4 6th)). The change in allele frequency due to genetic drift in a small population appears larger, statistical testing can determine whether changes are larger than expected by chance.

As an example, an allele frequency change in a population of size 50 from  $p = 0.5$  to 0.56 in 1 generation is within the range expected by drift, whereas in a population of size 5,000 such a change would be much too large to be due solely to drift effects.

Think in terms of tossing a coin, if you tossed a coin 50 times you would expect 25 heads and 25 tails, but due to the finite number of tosses would not be surprised to observe 28 heads and 22 tails (the change given for the allele frequency above, 56% heads). In fact, 95% of the results of tossing a coin 50 times would fall within the range from 30 heads (20 tails) to 20 heads (30 tails).

If you toss a coin 5,000 times 95% of the results would fall within the range from 2,550 heads (2,450 tails) to 2,450 heads (2,550 tails), and an observed outcome of 2,800 heads (2,200 tails) (56% heads when 50% expected) is well outside the range expected by chance (this outcome would occur by chance less than 1 in a million times).

**Founder effect:** the change in allele frequencies when a new colony is formed by a very small number of founding individuals from a larger population.

Alleles found in the general population may be absent from the founder population, e.g., of an island population, or a religious isolate, such as the Amish and Hutterite populations in the United States.

On the other hand, by chance rare alleles from the general population may be more frequent in the founder population, e.g., a form of dwarfism in the Amish.

A founder effect can explain why a disease allele is in much higher frequency in an isolated population than in the source. In the town of San Luis in Venezuela as many as 1 in 4 has Huntington's disease, resulting from a single mutation carried by a founding individual 5

generations ago. The rate in the US is about 1/10,000, implying about 30,000 people have the disease, which is a late onset lethal dominant disease. This town provided a large pedigree that permitted the first mapping and cloning of the causative gene.

The reduction in genetic variability and the increase in rare allele frequencies in isolated populations plays an important role in modern efforts to find disease-causing genes. There are fewer such alleles and they are in high frequency. The population of Iceland is especially well studied. A company Decode Genetics is trying to exploit that strategy. Iceland was founded by a few thousand Vikings and their Celtic slaves about 1100 years ago.

**Bottleneck:** drastic reduction in population size due, e.g., to over fishing or a natural disaster such as a hurricane, will lead to changes in allele frequencies (Fig. 23.8 (7th) (Fig. 23.5 6th)).

An extreme reduction in population size can occur because of disease, environmental change and human activities. The result is that genetic drift can be very strong, leading to substantial changes in allele frequencies and loss of genetic variation. Cheetahs are one example: Cheetahs have so little genetic variability that skin can be grafted between unrelated individuals.

**Long term genetic drift effects:** are **loss** or **'fixation'** of an allele.

In both small and large populations, genetic variation is lost due to the finite size of the population. Variation is gained from mutations. The combination of these two forces means that even in the absence of selection, molecular evolution will occur; by chance a new mutant may replace all existing alleles.

It is easy to understand how genetic drift in small populations can lead to **loss** of alleles (and hence **'fixation'** of the other allele), e.g., suppose we have a constant population size of 10 haploid individuals, with an initial allele frequency of  $p = f(A) = 0.5$ ,  $q = f(B) = 0.5$ . We assume at the moment that no new mutations arise. (We put quote marks around 'fixation' as in reality there will be mutations of the allele that has become 'fixed,' but these will be mutations of this allele.)

	alleles	A	B
	1	.5	.5
	2	.6	.4
generation	3	.8	.2
	4	.7	.3
	5	.9	.1
	6	1.0	0

After the first generation reproduces, the allele frequencies in the 2nd generation are  $p = 0.6$ ,  $q = 0.4$  (see below), and we now reweight our 'coin' so that the probability of an A in the next (3rd) generation is 0.6 and of a B is 0.4, the observed frequencies in the current (2nd) generation. Due to drift effects the observed frequencies in the 3rd generation may differ from these expected values. Continuing this process, the genetic drift (random) effects result in eventual loss of one of the alleles, and "fixation" of the other.

In this example, the B allele is **lost**, and the A allele is '**fixed**' in the population. From the starting point of equal allele frequencies it is equally likely that B would be '**fixed**' and A **lost** in a subsequent random sampling of the genetic material.

The exact same process goes on in large populations, it just **takes longer**.

<u>alleles</u>	<u>A</u>	<u>B</u>
	500	500
	495	505
	510	490
	...	...
	...	...
	...	...
	1,000	0

In this case A is '**fixed**' and B **lost** from the population, from these starting conditions (equal frequencies of A and B) it is equally likely that A would be **lost**, and B '**fixed**' in the population.

**common ancestor:** the consequence of finite population sizes and hence genetic drift effects is that we have a common ancestor for all our genes. It is a different common ancestor for each gene due to independent assortment of genes on different chromosomes (Mendel's 2nd law), and recombination between genes on the same chromosome. Very closely linked genes could share a common ancestor.

For a neutral allele at an autosomal locus in a diploid organism the time to trace back to the common ancestor is on average  $4N_e$  generations where  $N_e$  is the effective population size.

The female ancestor of our mitochondrial DNA, which is inherited maternally and without recombination, has been traced to Africa about 200,000 years ago. Note that this woman is only the common ancestor for our mitochondrial DNA, and further this observation does not tell us what the population size was at that time, it certainly does not mean it was just this woman and one man.

Our nuclear genes trace back to many other common ancestors, some presumably from this time period, some more recent, and others which are older. Because of genetic recombination of nuclear genes, it is much more difficult than with mitochondrial DNA to trace back to the common ancestor

The Y chromosome will trace back to a male common ancestor and studies on Y chromosome variation are in progress in a number of laboratories.

## • Sexual versus asexual reproduction

**sexual vs. asexual reproduction:** in sexual reproduction, a new organism is formed by the fusion of 2 gametes (egg and sperm or pollen). In contrast, asexual reproduction occurs without sexual fertilization.

There is a cost to sex (50%), since an individual is only contributing half of their genetic makeup to their offspring. So what is the selective advantage of sex?

It is believed that a population of sexually reproducing organisms can, under some conditions evolve faster than with asexual reproduction, since sexual reproduction can produce more genetic variability in offspring.

The greater genetic variability is produced via independent segregation of alleles on different chromosomes, and mixing of genetic information on homologous chromosomes due to crossing over (recombination).

## Questions relating to lecture on genetic variation and genetic drift

1. In a population of self fertilizing plants of size 100, the genotypes at a codominant locus are 20 AA, 60 AB, and 20 BB individuals. What will the genotype counts be in the next generation (assume a population size of 100 again and that there are no genetic drift effects)? What would they be if there was random mating?
2. Do self-quiz questions 5, 8, and 9 on page 471 of the 7th edition of the textbook (self-quiz questions 9, and 11-13 on page 463 of the 6th edition).

### Answer to question 1 (don't peek till you try it!)

With selfing, the frequency of heterozygotes is reduced by 1/2 each generation, with the remaining 1/2 divided equally between the two homozygous classes. The genotype counts in the next generation will be 35 AA, 30 AB, and 35 BB.

If there had been random mating, the genotype counts would be 25 AA, 50 AB, and 25 BB.