Lecture 35 Hardy-Weinberg
Reading: 759-762 (2nd edition: 677-682)

- Population genetics is concerned with the distribution of genes in populations and the ways in which that distribution changes. A **population** for this purpose is a group of individuals that can be regarded separately. Ideally a population is freely interbreeding, but in most cases a population is the group available for study. Many studies of human populations are by country, state, county, city, neighborhood etc. because those are the units for which data are available.
- The change in genetic composition is called **microevolution**, which is Darwin's **descent with modification** in genetic terms.
- The basic ideas in population genetics are all in terms of various **frequencies**: **phenotypes, genotypes, alleles**, all of which are obtained by long division. To compute allele frequencies in a diploid species, you have to remember that each individual carries two copies of gene.
- The **Hardy-Weinberg fractions** are the expected genotype frequencies of a gene after one generation of **random mating**, under the assumption that the population is very large and that there is no mutation, immigration or selection. To compute HW frequencies, which I am sure you have seen before, you use the product rule to find the probability that each pair of genotypes mates, use Mendelism to find the expected genotypic fractions in the offspring, and then use the sum rule to find the total fraction of each genotype.
- There are three very important consequences of HW. (1) **Allele frequencies do not change** under random mating (2) **HW frequencies are established in one generation** of random mating (3) Under HW, the allelic state of on one chromosome is **independent** of the allelic state on the other.
- You can use the independence of alleles to find the corresponding genotype frequencies when there are **more than 2 alleles.** If the frequencies of two alleles, A_1 and A_2 at a locus are p_1 and p_2, then the frequencies of homozygotes are p_1^2 and p_2^2 and the frequency of A_1A_2 is $2p_1p_2$, regardless of how many other alleles are at the locus and what their frequencies are.
- You can also use independence to predict genotype and phenotype frequencies for X-linked loci.
- An important practical implication of HW is that the **frequency of individuals homozygous for a low frequency allele is very low**, possibly so low that there may be few or no individuals with that genotype in a population. Another way of saying that is that most copies of a rare allele are in heterozygous carriers. For example, roughly 1/2500 newborns of European descent have CF, implying that p=0.02. Hence roughly 1/25 people of European descent are carriers of one of the CF mutants.
- In humans and in most **outbreeding species**, genotypic proportions in a single geographic area are almost always very **close to HW**. HW plays an important role in the forensic use of DNA. Currently, state and federal agencies genotype at 13 SSRs, called Combined DNA Index System (**CODIS**) loci. The 13-locus genotype is called the **DNA profile**. Each locus is highly polymorphic. One use of DNA profiles is to compare the profile of a crime-scene sample with the profile of a suspect. If the profiles differ at all, then the suspect is exonerated. If the profiles match, then the prosecution will argue that the suspect was at the crime scene. That evidence will be presented to a jury in the form of a probability, called the **random match probability** (RMP) that the crime-scene profile could have come from some person other than the suspect. The RMP is computed by applying HW to find the probability

that each locus matches, and then multiplying across loci (**the product rule)** to find the 13-locus match probability. To illustrate, suppose that a crime scene sample and a suspect are both found to have genotype 14-17 at D3S1358 and genotype 10-13 at D8S1179. The frequencies of 14 and 17 at D3S1358 in California Caucasians is 0.1404 and 0.2118, so the match probability for that locus is 2x0.1404x0.2118=0.0595. At D8S1179, the frequencies are 0.0733 and 0.3393 and the match probability for that locus is 2x0.0733x0.3393=0.0497. Using the product rule, the two locus match probability is 0.0595x0.0497=0.00296. Using 13 loci, RMPs are on the order of 10^{-12}, i. e. 1 in a trillion.

- When significant deviations from HW are found, either there is **experimental error** or **something very interesting** has happened. With SSRs, **null alleles** will produce an apparent excess of homozygotes. Another cause of excess homozygosity is the **mixture of two or more populations** in which allele frequencies differ. You can see the effect clearly by considering a mixture of two populations that are fixed for different alleles at a locus. The reduction in heterozygosity in mixed populations is called the **Wahlund** effect.
- Many plant species are predominantly self-fertilizing and in those species genotype frequencies are typically far from HW. For example, in a population of wild oats, the genotype frequencies at one locus were 0.548, 0.071, and 0.381, which obviously deviates from HWE. This species, *Avena fatua*, is self-fertile and extensive self fertilization accounts for the lower than expected frequency of heterozygotes. This observation is consistent with a Mendelian experiment. In an F_3 population created from the F_2 by selfing, the frequencies are 3/8, ¼, 3/8, which deviate from HW genotype frequencies. There are too few heterozygotes.

Budowle B, Shea B, Niezgoda S, Chakraborty R (2001) CODIS STR loci data from 41 sample populations. *Journal of Forensic Science*, 46:453-489
http://journalsip.astm.org/JOURNALS/FORENSIC/PAGES/3629.htm

Problems: I-III, 2-16 (2nd ed. Ch. 20—I-III, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14)

Allele	California	Alabama
<12	0.0033	0.0033
12	0.0033	0
13	0	0.067
14	0.1367	0.1533
15	0.2800	0.2300
15.2	0	0
16	0.2167	0.2557
17	0.1967	0.2267
17.1	0	0
18	0.1500	0.1000
19	0.0133	0.0233

33.1 The table above is part of Table 2 from Budowle et al. (2001). It gives the allele frequencies at D3S1358 in Caucasians in samples from California and Alabama.

a. Assuming HW genotype frequencies, what is the fraction of Alabama Caucasians who are heterozygous for allele 17?

Ans. 2x0.2267x(1–0.2267)=0.3506

b. Assuming HW genotype frequencies in both populations, what are the fractions of individuals homozygous for allele 15 in each?

Ans. Cal, 0.28^2=0.0784, Ala, 0.23^2=0.0529

c. If you surveyed genotypes in a group of Caucasians half from Alabama and half from California, what would you expect the frequency of individuals homozygous for allele 15?

Ans. (0.0784+0.0529)/2=0.06565

d. In this mixed population, what is the frequency of allele 15?

Ans. (0.28+0.23)/2=0.255

e. What would the frequency of individuals homozygous for allele 15 be in the mixed population if genotypes were at the HW frequencies?

Ans. 0.255^2=0.065025.

f. Why are the answers to c and e different?

Ans. Because genotype frequencies in the mixed population are not in HW. The excess of homozygotes illustrates the Wahlund effect.

g. Without doing any calculations, would you expect more or fewer 15-17 heterozygotes in the mixed population than if it were at HW?

Ans. Fewer. If there are more homozygous individuals in the mixed population there have to be fewer heterozygotes.