Bio 1B Lecture Outline (please print and bring along) Fall, 2006

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

Evolution lecture #10 **-- Mendelian genetics, Hardy-Weinberg --** Oct. 23rd, 2006

ch. 14: 251-270 and ch. 23: 454-458 in 7th ed.
ch. 14: 247-266 and ch. 23: 445-449 in 6th ed.

Summary of topics

- Use Mendel's **first law (independent segregation)** to predict genotypes and phenotypes resulting from a given cross
- Genes on different chromosomes assort independently - **Mendel's second law (independent assortment)**
- Apply Mendel's principles to examples with **incomplete dominance** and **codominance** relationships of alleles, and multiple alleles
- Use the method of **gene (allele) counting** to determine allele frequencies when heterozygotes can be distinguished from the homozygote state
- Describe the **Hardy-Weinberg law**, explain the **conditions** that must be met for it to hold true, and determine if a population is in **HW proportions** (including multiple allele cases)
- Use the **HW expectations** to calculate allele frequencies for **recessive/dominant** traits or diseases, estimate the frequency of heterozygous carriers for a recessive trait, and explain why the majority of mutant alleles are carried in heterozygous individuals for rare recessive traits.

Timeline of Mendelian genetics

(You do not need to memorize the above dates, except for three with asterisks)

1858 Darwin and **Wallace** independently propose the mechanism of evolution, natural selection.

***1859 Darwin** published *On the Origin of Species*.

1865 Gregor Mendel discovered the basic laws of genetic inheritance (Mendel's laws were unknown to Darwin).

***1900** Mendel's results are rediscovered, ABO blood group system in humans are discovered and shown to be an example of Mendelian inheritance.

1944 DNA is the genetic material.

***1953 Watson** and **Crick** discover the molecular structure of DNA.

1970 Gene mapping in humans essentially limited to the X chromosome, based on the specific pattern of inheritance.

1983 Genetic linkage of Huntington disease to a chromosomal location.

1989 Cystic fibrosis gene identified.

1990 Human Genome project initiated, a handful of human disease genes had been identified.

1993 The Huntington disease gene identified.

1994 The familial breast/ovarian cancer gene (BRCA1) was identified.

1997 The first cloning of a mammal, a sheep named Dolly.

The future Documenting genetic variation of human and other genomes at the population level, identifying the genes involved in complex diseases, understanding the genetic basis of development, and much more. Many ethical issues will arise...

Overview of Mendelian genetics

Gregor Mendel, 1865: discovered basic laws of genetics

gene: sequence of DNA coding for a protein (or in some cases, part of a protein)

allele: a variant of a single gene, inherited at a particular genetic locus (A and a)

genotype: the genetic constitution of an individual; in diploid individuals it is the set of two alleles at a locus possessed by an individual (AA, Aa, and aa)

phenotype: an observable trait in an organism; it can be determined by genotype and environment and interaction between the two (see Fig. 14.6 (7th) (Fig. 14.5 6th))

homozygote: individual having two copies of the same allele at a genetic locus (AA and aa)

heterozygote: individual having two different alleles at a genetic locus (Aa)

dominant: an allele A is dominant if the phenotype of the heterozygote Aa is the same as that of the homozygote AA, but differs from the homozygote aa

recessive: an allele a is recessive if the phenotype of the homozygote aa differs from that of the heterozygote Aa and the homozygote AA, which are the same.

Mendel's law of segregation (first law): states that a simple genetic trait is determined by a pair of separable factors (now called alleles of a gene), one inherited from each parent, and when an individual produces an egg or sperm cell, only one or the other of this gene pair is randomly transmitted (alleles segregate randomly) (Figs. 14.5, 14.9, and 14.14 (7th) (Figs. 14.4, 14.8, and 14.14 (6th))).

The important feature of Mendelian inheritance is that although a recessive allele is hidden in heterozygous individuals by the dominant allele, it is not lost to the population. The heterozygote Aa individual produces on average 1/2 A and 1/2 a gametes.

Previous to this, blending inheritance was the most commonly believed pattern of inheritance; but under blending inheritance genetic variation is halved each generation.

test cross: a testcross is designed to reveal the genotype of an organism that exhibits a dominant trait, such as purple flowers in pea plants. Such a plant could be homozygous or heterozygous for the dominant allele. To reveal the genotype of the purple flower it is crossed to a homozygous recessive white flowered plant (Fig. 14.7 (7th) (Fig. 14.6 6th)).

If the purple parent plant is homozygous, all the offspring will be heterozygous and hence purple; if the purple parent plant is heterozygous, half the offspring on average will be heterozygous and purple, and the other half will be homozygous for the recessive trait and hence white.

Mendel's law of independent assortment (second law): states that alleles of different genes assort independently. This is the case if the genes are unlinked (on different chromosomes) or separated by a recombination fraction of 50% on the same chromosome.

incomplete dominance: the heterozygote AB has a phenotype intermediate to that of the two homozygotes AA and BB. For instance, in snapdragons red flowers and white flowers represent the two homozygous phenotypes, while the heterozygote has pink flowers.

E.g., a cross of a red (RR) by white (WW) flowers will result in all the offspring being pink (RW) (Fig. 14.10 (7th) (Fig. 14.9 6th)).

codominance: the heterozygote AB has a phenotype distinguishable from both homozygotes AA and BB, and both alleles are separately manifest in the phenotype. One example is the so-called MN blood group in humans, where the homozygote MM and NN phenotypes each express one type of molecule on the cell surface, whereas the heterozygote MN individuals express both types of molecule.

multiple alleles: a group of individuals may have more than two different alleles for a given gene. (Any one individual has only two alleles, which may be the same or different, one inherited from their mother, the other from their father.)

E.g., the ABO blood group system in humans is determined by a set of 3 alleles, denoted A, B, and O. The A and B alleles are codominant, while both these are dominant to the O allele, giving rise to 4 blood group phenotypes A (genotypes AA and AO), B (genotypes BB and BO), AB (genotype AB), and O (genotype OO) (Table 14.2 (7th) (Fig. 14.10 6th)).

Some genes of the HLA (human leukocyte antigen system) which is involved in the immune response have over 200 alleles. Organ transplants have a much higher success rate when donor and recipient are matched for their HLA genes, but the high level of variation makes this difficult.

polymorphic: a genetic locus is polymorphic if it has 2 or more different allelic forms.

At the population level there can be more than 2 alleles at a gene, even though a single individual has a maximum of 2 different alleles, e.g., consider the ABO blood group system with 3 alleles A, B, and O.

• Overview of different types of genetic variation:

phenotypic polymorphisms: qualitative differences in phenotype can be used as a genetic polymorphism, e.g., shell color and pattern forms in snails. (One must demonstrate that the trait has a genetic basis, and determine the number of genes involved.)

antigenic polymorphisms: sequence differences in a molecule can be detected by antibodies specific to each molecule, e.g., the ABO blood group system, and original typing of the HLA system and major histocompatibility complexes (MHCs) in other species (these are now typed by molecular methods).

chromosome markers: heritable variations in chromosome morphology.

electrophoretic polymorphisms: many proteins can be shown to be polymorphic by agarose or starch gel electrophoresis due to differences in electric charge or molecular weight at the protein (amino acid) level. This method was used extensively in the late 1960's and the 1970's. Not all mutations can be detected with this technique.

DNA polymorphisms: RFLP's, minisatellites (VNTR's—variable number of tandem repeats), microsatellites, SNP's (single nucleotide polymorphisms), indels (insertions/deletions).

• Gene (allele) counting

gene (allele) counting with incomplete dominance or codominance: When heterozygotes can be distinguished from the homozygous states (e.g., with incomplete dominance or codominance) allele frequencies are obtained by the method of gene (allele) counting.

In a 2 allele system with alleles denoted A and B,

$f(A) = f(AA) + f(AB)/2$, $f(B) = f(AB)/2 + f(BB)$, with $f(A) + f(B) = 1$,

where f(AA) is the frequency of homozygous AA individuals, f(AB) of heterozygous AB individuals, etc.

Note that no assumption of Hardy Weinberg is required.

For the observed genotype values:

Genotypes	AA	AB	BB	
Observed	50	40	10	Total: 100
Frequencies:	0.50	0.40	0.10	

$f(A) = 0.5 + (0.4)/2 = 0.7$, $f(B) = (0.4)/2 + 0.10 = 0.3$, check $f(A) + f(B) = 1$.

These calculations can be done using observed numbers of genotypes (Fig. 23.4 (7th) (Fig. 23.3 6th)):

Genotypes	AA	AB	BB	
Observed	N_{AA}	N_{AB}	N_{BB}	Total: N_{total}

$f(A) = (2N_{AA} + N_{AB})/2N_{total}$, and $f(B) = (N_{AB} + 2N_{BB})/2N_{total}$.

• Hardy Weinberg

Observed genotype frequencies in a population: consider the following examples of genotype frequencies for the Δ32 variant of the CCR5 gene, which protects against progression to AIDS, the MN blood group system, and the gene for cystic fibrosis (cc individuals are affected, Cc individuals are carriers).

CCR5 Δ32 variant in a French population

genotypes:	A/A	A/Δ32	Δ32/Δ32	
observed numbers	795	190	15	Total: 1,000

MN blood group system in an Egyptian population

genotypes:	MM	MN	NN	
observed numbers	278	489	233	Total: 1,000

Cystic fibrosis in a European population

genotypes:	CC	Cc	cc	
observed numbers	9,596	400	4	Total: 10,000

Why do we see these particular genotype frequencies, is there some relationship between allele frequencies and the genotype frequencies, why are the heterozygote frequencies so much larger than the homozygous frequencies for Δ32/Δ32 and cc?

In fact, each of these three examples have genotype frequencies close to those expected given their allele frequencies and random mating and some other assumptions described below under Hardy Weinberg.

Hardy Weinberg assumptions: Mendelian genetics was rediscovered in the early 1900's. The principle which is the foundation of population genetics—the Hardy-Weinberg law—was derived independently in 1908 by an English mathematician Hardy and a German physician Weinberg.

The geneticist Punnett brought to the attention of Hardy a remark of Yule (also a geneticist)—Yule is reported to have suggested as a criticism of the Mendelian position that if brachydactyly (short fingeredness) is dominant "in the course of time one would expect, in the absence of counteracting factors, to get 3 brachydactylous individuals to 1 normal." Of course we now understand that Yule is confusing the Mendelian 3:1 ratio in an F1 cross (Aa x Aa gives 3A-:1aa) with population features.

The importance of the Hardy-Weinberg law is that if there are no counteracting forces, then the frequencies of alleles do not change in a population.

In other words, variation in a population, under a Mendelian system, tends to be maintained. We contrast this to the previously believed blending inheritance where genetic variation is decreased each generation.

Basically we want to consider whether or not Mendelian segregation causes changes in the genetic structure of a population.

We want to consider the effect of this factor acting in isolation from all other possible factors, so we assume:

(1) random mating

(2) mutation and migration rates are negligible

(3) no selection

(4) segregation according to Mendelian rules

(5) a very large population, in essence, an infinitely large population

Hardy Weinberg equilibrium: We consider a 2 allele system with
$p = f(A)$, $q = f(B)$ in the parental generation, so that $p + q = 1$.

Under the assumptions listed above, in the offspring:
$f(AA) = p^2$, $f(AB) = 2pq$, $f(BB) = q^2$,
and these are referred to as Hardy Weinberg proportions (HWP) or Hardy Weinberg equilibrium (see Fig. 23.4 (7th) (Fig. 23.3 6th)).

Note that in the offspring generation $f(A) = p$, $f(B) = q$. (Variation is maintained.)

Thus, allele frequencies do not change under the conditions of Hardy Weinberg, and genotype frequencies (proportions) are predicted from allele frequencies.

Hardy Weinberg proportions for a given allele frequency p of allele A (hence frequency q (= 1 - p) for allele B) -- there is only one genotype distribution in exact HWP.

Composition of an equilibrium population: for a range of allele frequencies, the HWPs are given below:

A	B	AA	AB	BB
p	q	p^2	2pq	q^2
.1	.9	.01	.18	.81
.2	.8	.04	.32	.64
.3	.7	.09	.42	.49
.4	.6	.16	.48	.36
.5	.5	.25	.50	.25

Note that for a 2-allele system, the maximum heterozygosity under HWP occurs when the 2 alleles have equal frequency of 1/2.

deviations from Hardy Weinberg proportions: Given any sample is of finite size, we do not expect the genotype frequencies to be in exact HWP.

In the example above on determining allele frequencies by gene (allele) counting where we have f(A) = p = 0.7, f(B) = q = 0.3, the expected genotype frequencies under HWP are 0.49, 0.42, and 0.09, which deviate slightly from the observed values of 0.50, 0.40, and 0.10.

Using statistical testing one can show that such a small deviation is well within the range expected with this sample size (the statistical test is termed a chi-square goodness of fit, and is not covered in this course).

Another example of an observed sample with a close fit to HWP is

genotype values:	CC	CD	DD	
	35	50	15	Total: 100
Frequencies:	0.35	0.50	0.15	f(C) = 0.6, f(D) = 0.4, and
Expected HWP:	0.36	0.48	0.16	

An example with the same allele frequencies, but large (statistically significant) deviation from HWP is:

genotype values:	EE	EF	FF	
	56	8	36	Total: 100
Frequencies:	0.56	0.08	0.36	f(E) = 0.6, f(F) = 0.4, and
Expected HWP:	0.36	0.48	0.16	

One would need to investigate the possible reasons in this case for the large deviation from HWP (deficiency of heterozygotes, excess homozygotes), e.g., selection, inbreeding (**more next lecture!**)

Hardy Weinberg equilibrium for multiple alleles: For a gene with multiple alleles, labeled A_1, A_2, ... ,A_k, with frequencies p_1, p_2, ... , p_k,

the expected HWP for the homozygote A_1A_1 is $(p_1)^2$, and, for the heterozygote A_1A_2 is $2p_1p_2$, etc.

The total number of expected homozygotes under HWP is $F = \Sigma\, p_i^2$, and heterozygotes is $H = 1 - F$, termed the gene diversity index.

• Recessive traits and the square root formula

estimation of allele frequency for recessive trait, the square-root formula: Many traits of interest are not codominant or incompletely penetrant, but we may be interested in estimating the allele frequencies in this case, and the proportion of individuals heterozygous for a recessive trait (called carriers).

To do this, we **assume** HWP in the population, and then using the fact that the expected frequency of homozygotes for the recessive trait is $f(aa) = q^2$, we estimate the allele frequency of the recessive trait by $q = f(a) = [f(aa)]^{1/2}$.

Note that we cannot test whether the population is in HWP, we have used this information to obtain our estimate of the allele frequencies.

carrier frequencies:

We can also use the assumption of HWP to estimate how many carriers for the trait (heterozygotes) there are in the population, i.e., we estimate 2pq, using our estimate of q (and hence $p = 1 - q$), above.

rare recessive traits: the majority of mutant alleles are carried in heterozygotes for rare autosomal recessive traits (frequency ~2q), and most affected individuals result from heterozygote by heterozygote matings.

The following examples of carrier frequencies in rare autosomal recessive traits are given for illustration, you do not need to memorize the details.

cystic fibrosis: Among Caucasians, one of the most frequent recessive diseases is cystic fibrosis, characterized by malfunction of the pancreas and other glands. (The gene for cystic fibrosis was identified in 1989.) The incidence of the condition is about 1/2,500 individuals, thus $q^2 = 0.0004$, and our estimate of $q = 0.02$, and the estimate of carriers is ~4%.

sickle cell anemia: Among African Americans, sickle cell anemia is the most common recessive disease, with an incidence of approximately 1/400 individuals. In this case, $q^2 = 0.0025$, the estimate of $q = 0.05$, and the estimate of carriers is ~9.5% (heterozygous carriers of the disease show no ill effects except in conditions of oxygen stress). In parts of West Africa about 1/100 individuals have sickle cell anemia, and ~18% are carriers. The high frequency of the allele is due to an advantage to heterozygous individuals in malarial environments.

Tay Sachs: Among descendants of Ashkenazi Jews who settled in Eastern and Central Europe the recessive condition Tay Sachs disease occurs with an incidence of ~ 1/4,000, thus $q^2 = 0.00025$, the estimate of $q = 0.016$, and of heterozygous carriers is ~3.2%. Individuals with Tay Sachs disease lack the enzyme hexosaminidase A. It is a tragic illness which leads to death by age 3 or 4 years. Carrier status screening is possible, and is performed each year on campus. If a couple are both carriers then prenatal diagnosis can be performed. Among the non-Jewish population of North America the incidence of Tay Sachs is about 1 in 500,000 births, with $q = 0.0014$.

Questions relating to lecture on Mendel and Hardy Weinberg

1. Do self quiz questions #'s 3, 5, 6, 9, 14, and 15 on pages 272-273 of the textbook (7th edition) (questions #'s 1, 4, 6, 7, 8, 11, 16, and 17 on pages 267-268 of the 6th edition of the textbook).

2. Do self quiz questions #'s 1-3 on page 471 of the textbook (7th edition) (questions #'s 1-3 on pages 462-463 of the 6th edition of the textbook).

2. Calculate the expected genotype frequencies under Hardy Weinberg for the three examples given above of:
   CCR5 Δ32 variant in a French population
   MN blood group system in an Egyptian population
   Cystic fibrosis in a European population.