Warmup, Part 0 - Preamble: Hardy-Weinberg equilibrium and population genetics calculations

In order to get started, we have to make sure you have the basic "Newton's First Law" of population genetics well in hand – the Hardy-Weinberg equilibrium theory. Calculation of allele and genotype frequencies is central to what follows. We review this below, along with an explicit example of H-W calculations, followed by questions that everyone should answer and turn in as part of their lab report. If you already feel comfortable with the H-W calculations, you can skip the exposition section below and proceed directly to the questions. (I know the material below is presented in a very, very elementary and expository way – please, if it seems overlong and boring to you, try the 'challenge' problems at the end to make certain you are secure in your knowledge. I have tried to review basic terminology like 'allele', 'homozygous,' etc., just to make sure we're all on the same page.)

Population geneticists study frequencies of genotypes and alleles within populations rather than the ratios of phenotypes (external forms) that Mendelian geneticists use. By comparing these frequencies with those predicted by null models that assume no evolutionary mechanisms are acting within populations, they draw conclusions regarding the evolutionary forces in operation. In a constant environment, genes will continue to sort similarly for generations upon generations. The observation of this constancy led two researchers, G. Hardy and W. Weinberg, to express an important relationship in evolution. The law that describes this relationship bears their names. The Hardy-Weinberg Equilibrium Theory serves as the basic null model for population genetics.

Every individual has alleles (=variants of the same gene) that were passed on from their parents. If we take all of the alleles of a group of individuals of the same species (that is, a population) we have what is called the *gene pool*. The frequency, or proportion, of individuals in that population that possess a certain allele is called the *allele frequency*. Populations can have allele frequencies, but individuals cannot. This makes populations a reasonable hierarchical unit, or level, to study evolution, on the assumption that evolution is basically the study of the change in allele frequencies over time. (You should be prepared to defend or attack this statement! How is it true? How is it not true?)

Allele Frequencies

Consider an individual locus (= roughly, 'gene') and a population of diploid individuals where two different alleles, *A* and *a*, can be found at that locus. If your population consists of 100 individuals, then that group possesses 200 alleles for this locus (100 individuals x 2 alleles at that locus per individual). The number of *A* alleles present in that population expressed as a fraction of all the alleles (*A* or *a*) at that locus represents the frequency of the *A* allele in the population.

1. To calculate allele frequencies for populations of diploid organisms, first multiply the number of individuals in the population by 2 to obtain the total number of alleles at that locus.

2. Select one of the alleles for your first set of calculations. Let's first choose the *A* allele from the example provided above.

a. Individuals homozygous for the *A* allele ("homozygous"= all one type, e.g., all *A's*) will each possess 2 *A* alleles. Multiply the number of *AA* homozygotes by 2 to calculate the number of *A* alleles.

b. Heterozygotes will each possess only one *A* allele.

c. The total number of *A* alleles in the population = [(the number of *Aa* heterozygotes) + (2 x the number of *AA* homozygotes)]

3. The frequency of the *A* allele = [(total number of *A* alleles in the population) / (total number of alleles in population for that locus)]

4. The frequency of the *a* allele = (1 - frequency of the *A* allele)

Genotype Frequencies

Consider the same population, locus, and alleles described above. Genotype frequencies represent the abundance of each genotype within a population as a fraction of the population size. In other words, the frequency of the *AA* genotype represents the fraction of the population homozygous for the *A* allele.

1. To calculate genotype frequencies for populations of diploid organisms, first determine the number of individuals with each genotype present in the population. In the example used above, you would count the number of individuals with the following genotypes: *AA, Aa*, and *aa*.

2. To determine the frequency of each genotype, divide the number of individuals with that genotype by the total number of individuals in the population.

a. Frequency of *AA* genotype = # *AA* individuals / population size

b. Frequency of *Aa* genotype = # *Aa* individuals / population size

c. Frequency of aa genotype = # aa individuals / population size

IMPORTANT NOTE:

Unless you *know* that a population meets Hardy-Weinberg equilibrium assumptions, you *must* use the above procedure to calculate genotype frequencies. If you know that a population meets Hardy-Weinberg expectations, then you can calculate genotype frequencies using allele frequencies and the Hardy-Weinberg equations (see below).

Assertions of the Hardy-Weinberg Equilibrium Theory

The Hardy-Weinberg Equilibrium Theory refers to loci within populations that experience no
evolutionary forces. For such populations the theory asserts that:

1. Allele and genotype frequencies should remain *constant* from one generation to the next. That is, no evolution should occur at these loci. If, at a certain gene locus, there are only two alleles, each will have a frequency such that the frequency of one allele plus the other equals 1. Remember, we are discussing the frequency in a population, not in an individual. Formally, we can state the allelic frequency in a population as follows:

$$p = \text{Frequency of allele } A = \text{freq}(A)$$

$$q = \text{Frequency of allele } a = \text{freq}(a)$$

$$\text{and } p + q = 1$$

2. Given a certain set of allele frequencies, genotype frequencies should conform to those calculated using basic probability. In a one locus/two allele system such as the one described above, the genotype frequencies should be as follows:

a. Frequency of *AA* genotype = (frequency of *A* allele)^2

b. Frequency of *aa* genotype = (frequency of *a* allele) ^{2}

c. Frequency of *Aa* genotype = 2 x (frequency of *A* allele) x (frequency of *a* allele)

Within a population, the frequency of the possible combinations of a pair of alleles at one locus is related to the expansion of the binomial $(p + q)^2$. (This is because the matings are just binomial draws.) Remember that if we square one side of the equation we must square the other side, such that $(p + q)^2 = 1^2$. The expansion is

$(p + q)$ x $(p + q) = p^2 + 2pq + q^2 = 1$, where p^2 = Frequency of genotype *A/A* ; 2pq = Frequency of genotype *A/a* ; q^2= Frequency of genotype *a/a*

3. If the genotype frequencies obtained from a real population do not agree with those predicted by the Hardy-Weinberg Theory, then population geneticists know that some evolutionary mechanism or mechanisms must operate on the locus of interest. Knowledge of the theory can help narrow down the possible mechanisms. Then they can use experiments to determine which potential mechanism or mechanisms operate on the locus. As such, the Hardy-Weinberg Equilibrium Theory serves as an important tool for population geneticists.

Assumptions

Populations will conform to the Hardy-Weinberg Theory assertions only if no evolutionary forces or mechanisms influence the loci under consideration. The assumptions that populations must meet in order for the Hardy Weinberg assertions to hold include:

1. Large population size (i.e., no genetic drift). Random chance can alter allele frequencies through mating processes and death within small populations.

2. Random mating, which means that the choice of mates by individuals in the population is determined by chance, and not influenced by the genotypes of the individuals in question.

3. No difference in the mutation rates between alleles at the same locus.

4. Reproductive isolation from other populations (i.e., no gene flow or migration).

5. No differential survival or reproduction among phenotypes (i.e., no natural selection).

Example

Consider a population of 1000 individuals and the locus and alleles described above. Assume that you have no information on the presence or absence of evolutionary mechanism in this population. You find that the population consists of:

90 individuals homozygous for the *A* allele (*AA* genotype)

490 individuals homozygous for the *a* allele (*aa* genotype)

420 heterozygotes (*Aa* genotype)

1. Calculate the genotype and allele frequencies for this locus.
2. Determine whether this population meets Hardy-Weinberg Assumptions (in other words determine if evolutionary mechanisms operate in this population).

Calculation of Allele and Genotype Frequencies

Since you do not know whether this population meets Hardy-Weinberg Assumptions, you must calculate both the allele and genotype frequencies using the raw data.

1. Allele Frequencies:
   - The frequency of the *A* allele will equal:

   (total number of A alleles in the population) / (total number of alleles in population for locus) = [(90*2) + 420] / (1000*2) = 0.30
   - The frequency of the a allele will equal: (1 - 0.30) or (total number of a alleles in the

   population) / (total number of alleles in population) = [(490*2) + 420] / (1000*2) = 0.70
2. Genotype frequencies:
   - Frequency of *AA* genotype = # *AA* individuals / population size = 90/1000 = 0.09
   - Frequency of *Aa* genotype = # *Aa* individuals / population size = 420/1000 = 0.42
   - Frequency of *aa* genotype = # aa individuals / population size = 490/1000 = 0.49

3. Hardy-Weinberg Predictions:

If no evolutionary mechanisms operate on this locus, then the Hardy-Weinberg Equilibrium
Theory predicts that the genotype frequencies should be as follows:

- Frequency of *AA* = (frequency of *A* allele)^2 = $(0.3)^2$ = 0.09
- Frequency of *Aa* = 2 x (frequency of *A* allele) x (frequency of *a* allele) = 2*0.3*0.7 = 0.42
- Frequency of *aa* = (frequency of *a* allele)^2 = $(0.7)^2$ = 0.49

Conclusion

Since the observed genotype frequencies equal those predicted by the Hardy-Weinberg Equilibrium Theory, we may (tentatively) conclude that no evolutionary mechanisms operate on this locus in this population (i.e., the population meets the assumptions of the Hardy Weinberg Theory). (Why is our conclusion 'tentative'?)

Part 0b. Hardy-Weinberg warmup questions (yes, I know they seem simple)

Question 1. What is the frequency of heterozygotes (*Aa*) in a population in which the frequency of all dominant phenotypes is 0.25 and the population is in H-W equilibrium?

Question 2. Suppose the following data were accumulated for the frequencies of each of three genotypes at 5 separate loci, A through E:

AA: 0.36 *BB*: 0.0 *CC*: 1.0 *DD*: 0.70 *EE*: 0.25

Aa: 0.48 *Bb*: 0.03 *Cc*: 0.0 *Dd*: 0.20 *Ee*: 0.50

Aa: 0.16 *bb*: 0.97 *cc*: 0.0 *dd*: 0.10 *e*e: 0.25

a. Which loci are monomorphic? Which loci are polymorphic?

b. What are the allele frequencies at each locus?

c. Is there evidence that some mechanisms of evolution are acting at some loci but not others? How can this be?

Question 3. In a population a locus with 3 alleles showed the following genotypic numbers in a sample; *aa*=0, *ab*=0, *bb*=10, *ac*=0, *bc*=35, *cc*=6. Calculate allele frequencies, and determine the expected genotype frequencies under H-W conditions. Do the genotype frequencies depart from those expected by H-W? What may be going on here if they don't?

Question 4. Can a population with more than 80% heterozygotes for a locus be in Hardy-Weinberg equilibrium? Less than 10%? Explain why in a few sentences, or a bit of math.