Hardy-Weinberg Equilibrium

# Introduction

The Hardy-Weinberg model, named after the two scientists that derived it in the early part of this century, describes and predicts genotype and allele frequencies in a non-evolving population. The model has five basic assumptions: 1) the population is large (i.e., there is no genetic drift); 2) there is no gene flow between populations, from migration or transfer of gametes; 3) mutations are negligible; 4) individuals are mating randomly; and 5) natural selection is not operating on the population. Given these assumptions, a population's genotype and allele frequencies will remain unchanged over successive generations, and the population is said to be in Hardy-Weinberg equilibrium. The Hardy-Weinberg model can also be applied to the genotype frequency of a single gene.

# Importance

The Hardy-Weinberg model enables us to compare a population's actual genetic structure over time with the genetic structure we would expect if the population were in Hardy-Weinberg equilibrium (i.e., not evolving). If genotype frequencies differ from those we would expect under equilibrium, we can assume that one or more of the model's assumptions are being violated and attempt to determine which one(s).

# Questions

How do we use the Hardy-Weinberg model to predict genotype and allele frequencies? What does the model tell us about the genetic structure of a population?

# Variables

|  |  |
| --- | --- |
| p | frequency of one of two alleles |
| q | frequency of the other of twoalleles |

# Methods

The Hardy-Weinberg model consists of two equations: one that calculates allele frequencies and one that calculates genotype frequencies. Because we are dealing with frequencies, both equations must add up to 1. The equation

|  |  |
| --- | --- |
| $$p+q=1$$ | LaTeX Code: \[ p + q = 1 \] |

describes allele frequencies for a gene with two alleles. (This is the simplest case, but the equation can also be modified and used in cases with three or more alleles.) If we know the frequency of one allele (*p*) we can easily calculate the frequency of the other allele (*q*) by 1 - *p* = *q*.

In a diploid organism with alleles *A* and *a* at a given locus, there are three possible genotypes: *AA*, *Aa*, and *aa*. If we use *p* to represent the frequency of *A* and *q* to represent the frequency of *a*, we can write the genotype frequencies as (*p*)(*p*) or *p*2 for *AA*, (*q*)(*q*) or *q*2 for *aa*, and 2(*p*)(*q*) for *Aa*. The equation for genotype frequencies is

|  |  |
| --- | --- |
| $$p^{2}+2pq+q^{2}=1$$ | LaTeX Code: \[ p^2 + 2pq + q^2 = 1 \] |

One approach to the study of genetic diversity is to look at allele and genotype frequencies of allozymes. Allozymes are enzymes that show different rates of movement in gel electrophoresis due to the presence of different alleles at a single locus; they are often denoted as F (fast-moving) and S (slow-moving) alleles. Allozyme variation is an indicator of genetic variation and can be studied to quantify genetic variation among populations.

Lidicker and McCollum (1997) examined genetic variation in two populations of sea otters (*Enhydra lutris*) in the eastern Pacific. Sea otters were distributed throughout this region before fur hunting nearly led to their local extinction. Along the central California coast only one population of 50 or fewer individuals is thought to have survived; this population was protected in 1911 and has grown to its current size of approximately 1500 otters. Because of the extreme reduction in population size (a bottleneck), the population may have lost considerable genetic variation. A population from Alaska also experienced a bottleneck around that time but it was not as severe.

The table below (data from Lidicker & McCollum 1997) contains counts of the number of individuals with a given genotype for six variable (polymorphic) two-allele loci.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | California | Alaska |
| Locus | Genotype | *n* | *n* |
|  | SS | 37 | 3 |
| EST | SF | 20 | 3 |
|  | FF | 7 | 2 |
|  | SS | 48 | 7 |
| ICD | SF | 4 | 2 |
|  | FF | 3 | 0 |
|  | SS | 20 | 3 |
| LA | SF | 11 | 2 |
|  | FF | 2 | 3 |
|  | SS | 16 | 1 |
| PAP | SF | 7 | 3 |
|  | FF | 10 | 2 |
|  | SS | 16 | 1 |
| ME | SF | 11 | 2 |
|  | FF | 5 | 1 |
|  | SS | 17 | 3 |
| NP | SF | 4 | 1 |
|  | FF | 5 | 0 |

We can use these data to calculate the allelic frequencies for a given locus, such as the EST locus in the California population (*n* = 64). Each individual with the genotype *SS* has two copies of the *S* allele; therefore the 37 individuals with this genotype have a count of 74 *S* alleles. Heterozygote individuals (SF) have one of each allele, so there are 20 *S* alleles and 20 *F* alleles among them. Like the *SS* homozygotes, individuals with the *FF* genotype have two copies of the *F* allele, so these seven individuals contribute 14 *F* alleles to our count. In other words, among the 64 individuals in this sample there are 94 *S* alleles and 34 *F* alleles. To calculate the allelic frequencies, we simply divide the number of *S* or *F* alleles by the total number of alleles: 94/128 = 0.734 = *p* = frequency of the *S* allele, and 34/128 = 0.266 = q = frequency of the *F* allele.

If this population were in Hardy-Weinberg equilibrium, we would expect the genotype frequencies for *SS*, *SF*, and *FF* to be *p*2, 2*pq*, and *q*2:

|  |  |
| --- | --- |
| $$p^{2}=\left(0.734\right)^{2}=0.539$$ | LaTeX Code: \[ p^2 = (0.734)^2 = 0.539\] |
| $$2pq=2\left(0.734\right)\left(0.266\right)=0.390$$ | LaTeX Code: \[ 2pq = 2(0.734)(0.266) = 0.390 \] |
| $$q^{2}=\left(0.266\right)^{2}=0.071$$ | LaTeX Code: \[ q^2 = (0.266)^2 = 0.071 \] |

For the 64 individuals in this sample, then, we would expect that approximately 34 individuals (*p*2 \* *n* = 0.539 \* 64 = 34.496) would have the *SS* genotype, 25 individuals (2*pq* \* *n* = 0.390 \* 64 = 24.960) would have the *SF* genotype, and 5 individuals (*q*2 \* *n* = 0.071 \* 64 = 4.544) would have the *FF* genotype. How do these expected values compare to the observed numbers for genotype frequencies at the EST locus?

|  |  |  |
| --- | --- | --- |
| genotype | observed | expected |
| SS | 37 | 34 |
| SF | 20 | 25 |
| FF | 7 | 5 |

Generally we would use a statistical test (see [CHI-SQUARE](http://www.tiem.utk.edu/~gross/bioed/bealsmodules/chi-square.html) module) to compare our expected and observed counts. In this case we can see that the numbers are fairly similar, and in fact the authors have used a chi-square test and concluded that the observed and expected counts are not significantly different from one another.

# Interpretation

We can check our math to ensure that we have calculated the correct genotype frequencies: *p*2 + 2*pq* + *q*2 should equal 1, and (0.734)2 + 2(0.734(0.266) + (0.266)2 does indeed equal 1. Similarly, *p* + *q* must equal 1 and 0.734 + 0.266 = 1. Our results suggest that for the California sea otter population, the allele and genotype frequencies at the EST locus are in Hardy-Weinberg equilibrium. In other words, we can expect these allele frequencies to remain constant over time (barring any specific evolutionary forces acting upon this locus), thus ensuring genetic variation in the population at the EST locus. This equilibrium in the genetic structure of the population at the EST locus does not necessarily imply, however, that the population is not evolving; it merely indicates that this particular locus is not changing. Even if the frequency of alleles at just a single locus is changing over the generations, the population is evolving.

# Conclusion

Natural populations with whole genotypes in Hardy-Weinberg equilibrium are rarely found; one or more of the assumptions are violated in most situations. If nothing else, most populations are under the influence of natural selection. Certainly, no population can be infinite, but many populations are not even large enough to be functionally infinite. Oftentimes populations are not completely isolated from one another, and migration of individuals into or out of one population can change its genetic makeup. Mutations can potentially alter the gene pool significantly, although the majority are thought to have little or no effect (neutral mutations). Finally, individuals often mate selectively rather than randomly; for example, humans show assortative mating by height (tall people tend to marry tall people and short people tend to marry short people).

# Additional Questions

1) For which loci are the genotypes apparently not in Hardy-Weinberg equilibrium (note that *n* is different for each locus investigated)? Is this true for both populations?

2) What might affect the validity of your conclusions about the Alaska population?

*Extra credit*: Confirm some or all of your conclusions for #1 by performing a chi-square test (see [CHI-SQUARE](http://www.tiem.utk.edu/~gross/bioed/bealsmodules/chi-square.html) module). The null hypothesis you are testing is that the observed and expected values are not significantly different from one another (because your expected values are calculated based on an assumption of Hardy-Weinberg equilibrium, this is the same as saying that the population is in H-W equilibrium for the genotype being tested). The critical value for the chi-square in this case is 3.841; if your calculated value of the chi-square is equal to or greater than that, the probability of the null hypothesis being correct (i.e., the probability of the population being in H-W equilibrium at that genotype) is 0.05, and the null hypothesis is rejected.

# Source

Campbell, N. A. 1996. *Biology, 4th ed*. Benjamin/Cummings Publishing Co., Menlo Park, CA.

Hartl, D. L. 1988. *A Primer of Population Genetics, 2nd ed*. Sinauer Associates, Inc., Sunderland, MA.

Lidicker, W. Z. and F. C. McCollum. 1997. *Allozymic variation in California sea otters*. Journal of Mammalogy 78:417-425

# About this Resource

This material was originally distributed as part of “Alternative Routes to Quantitative Literacy for the Life Sciences[[1]](#endnote-1)” - National Science Foundation Award DUE-9752339 to the University of Tennessee, Knoxville for August 1, 1998 - July 31, 2000. Principal Investigator: Louis J. Gross. Co-Principal Investigators: Beth C. Mullin and Susan E. Riechert.

This material is now being revised as part of the “Resources for Improving Quantitative Skills in Community College Biology[[2]](#endnote-2)” project. As part of that project is also aligned with the OpenStax Biology Textbook[[3]](#endnote-3).

It is published using the QUBES Open Education Resources publishing platform[[4]](#endnote-4).

1. http://www.tiem.utk.edu/~gross/bioed/ [↑](#endnote-ref-1)
2. https://qubeshub.org/community/groups/quantbioatcc/ [↑](#endnote-ref-2)
3. https://openstax.org/details/books/biology-2e [↑](#endnote-ref-3)
4. https://qubeshub.org/qubesresources/publications/1051/ [↑](#endnote-ref-4)