

Hardy-Weinberg Equilibrium

1. Pre-lab Reading

Chapter 19, “Evolution of Populations” *Biology2e*, OpenStax College

2. Purpose and Objectives

The purpose of this lab is to explore the theory of evolution by investigating the Hardy-Weinberg theorem.

Upon completion of this activity, you should be able to:

1. Explain the Hardy-Weinberg equilibrium and understand how allelic frequencies are calculated.
2. Describe conditions that must be met in order to maintain Hardy-Weinberg equilibrium.
3. Use the bean model to demonstrate conditions for evolution.

3. Background Information

One of the most influential theories to affect the field of biology has been the theory of evolution as proposed by Charles Darwin in his book, *On the Origin of Species by Means of Natural Selection* published in 1859. **Evolution** can be defined as changes that occur in a population of organisms over time. While Darwin was not the first person to devise a theory of evolution, the mechanism for evolution he proposed was unique. Darwin believed that all modern species descended from ancestral species and that evolution occurred by a process he termed **natural selection**. Natural selection is defined as the differential survival and reproduction of individuals in a population of organisms resulting from the interaction of organisms with their environment. This means that within a population, some organisms will be more successful than others at surviving and reproducing. The genes of these individuals will be passed on to the next generation.

Over time, the composition of the population's **gene pool** (sum of all genes within a population) may change to reflect those characteristics that were most beneficial to the survival and reproductive success of individuals within the population. When a change in the gene pool occurs, evolution may also take place.

Today we talk about evolution in terms of genes and changes in a population's gene pool. However, Darwin had no understanding of the mechanism involved in inheritance and could not explain adequately how natural selection occurred. It would take a better understanding of the principles of inheritance before a genetic explanation for natural selection would be possible. Ironically, the first paper to describe inheritance patterns was published by Gregor Mendel in 1865 but went unnoticed until the early 20th century. Once Mendel's paper was rediscovered, the door was open to describe the genetic basis of natural selection. **Population genetics**, the study of genetic changes in populations, helped blend the work of Darwin and Mendel, providing a genetic basis for variation and natural selection.

In order to understand how populations evolve, we will define a **population** as a group of the same species of organisms living in the same area. A **species** will be defined as a group of organisms within a population whose individuals can potentially interbreed and produce fertile offspring. The traits exhibited by any individual in the population are defined by the genes possessed by that individual. The **gene pool** of the population consists of the sum total of all the genes of all individuals within the population.

Since a gene pool includes all the genes of all individuals within the population it is made up of all the alleles of all individuals in the population. Diploid organisms possess two alleles for a given trait. If the alleles are identical, the individual is described as being **homozygous** for that trait, whereas **heterozygous** individuals have different alleles for that characteristic. If, in a population, all members are homozygous for the same allele, that allele is described as being **fixed** in the gene pool. Unless fixation occurs, all alleles will exist at various relative frequencies within the population. Keep in

mind that while individuals in a population are selected for or against, populations evolve.

The following explanation describes how to calculate allele frequencies in a population. In an imaginary wildflower population of 500 plants, there are two varieties of contrasting flower color where the allele for red color (R) is completely dominant over the allele for white color (r). In the total population of 500 plants, there are 1,000 alleles for flower color (remember, each plant has 2 alleles for flower color). Twenty of these plants are white and are therefore homozygous for the recessive allele (rr). Of the remaining 480 plants, suppose 320 are RR homozygous and 160 are heterozygous (Rr). The number of R alleles in the population would be calculated as follows:

$$(320 \times 2) + 160 = 800$$

The number of r alleles is calculated in similar fashion:

$$(20 \times 2) + 160 = 200$$

The frequency of R is therefore $800/1,000 = 0.8$ or 80% while the frequency of r is $200/1,000 = 0.2$ or 20%.

At the turn of the 20th century, British mathematician G.H. Hardy and German physician W. Weinberg independently developed a mathematical model of population genetics that described **non-evolving** populations of organisms, now known as the **Hardy-Weinberg theorem**. The model shows that the process of heredity itself does not affect the structure of a population. The theorem states that the frequency of alleles in a population remains the same over generations regardless of their starting frequencies. In other words, the distribution of alleles resulting from meiosis and random mating within a population will not alter the overall gene pool of that population. For the Hardy-Weinberg theorem to hold true, certain assumptions must be made concerning the population:

1. The population must be large.
2. Mating within the population occurs randomly.
3. Mutations do not cause any net changes in the gene pool.
4. No migration occurs.
5. All individuals are equal in reproductive success (i.e. no particular genotype is favored over any other).

The Hardy-Weinberg theorem can be expressed as a mathematical formula. In the simplest situation of describing the frequencies of two alleles where one is dominant over the other, p is used to represent frequency of the dominant allele and q the frequency of the recessive allele. In our example above, you can see that $p + q = 1$ ($0.8 + 0.2 = 1$). When sexual reproduction (pollination) occurs in our wildflower population, the probability of R gametes combining to form zygotes with the genotype RR is calculated as p^2 (0.64). The probability of forming zygotes with the genotype rr is calculated as q^2 or 0.04. The probable frequency of forming Rr zygotes is calculated as $2pq$ or $2 \times 0.8 \times 0.2 = 0.32$. If all these frequencies are added together, the resulting sum equals 1.

$$0.64 + 0.32 + 0.04 = 1$$

Our formula for calculating allele frequencies in succeeding generations is therefore:

$$p^2 + 2pq + q^2 = 1$$

The Hardy-Weinberg equation is useful because it allows us to estimate the percentage of the population that carries a specific allele, as in the cases of certain inherited human diseases. In addition, it suggests that if a population's gene pool is not in a state of equilibrium, it may be undergoing evolution.

4. Testing the Hardy-Weinberg Theorem Using a Bean Model

Working in pairs, you will test the Hardy-Weinberg theorem by simulating a population using red and white beans. The bag of beans represents the gene pool of a population. Each bean represents a single gamete while the bean's color represents a different allele of a single gene. Red (R) represents the dominant allele; white (r) the recessive allele.

You will use the Hardy-Weinberg theorem as your hypothesis.

Procedure

1. State your hypothesis and prediction in your lab book.
2. Begin by placing 70 red beans and 30 white beans into your bag.
3. Calculate the allele frequency for each allele:

$$\text{Frequency of red alleles (p)} = \frac{\# \text{ of red beans}}{\text{total number of alleles}} \\ (\# \text{ red beans} + \# \text{ white beans})$$

$$\text{Frequency of white alleles (q)} = \frac{\# \text{ white beans}}{\text{total number of alleles}}$$

Enter these values in Table 1 under "Parent Population".

Remember: $p + q$ should always equal 1.

4. Answer the following questions in your lab book:
 - a. How many diploid individuals are in this population?
 - b. What bean combination represents a homozygous dominant individual? A heterozygous individual? A recessive individual?
5. Calculate the expected genotypic numbers and genotypic frequencies for the next generation. The expected genotypic number is calculated by multiplying the expected

allele frequency by the total population size (50). Enter these values in Table 1 next to “Expected Population”.

Genotypic Frequency

Genotypic Number

$$RR_{\text{freq}} = p^2$$

$$\text{number of RR individuals} = (p^2) (50)$$

$$Rr_{\text{freq}} = 2pq$$

$$\text{number of Rr individuals} = (2pq) (50)$$

$$rr_{\text{freq}} = q^2$$

$$\text{number of rr individuals} = (q^2) (50)$$

Remember: $p^2 + 2pq + q^2$ should equal 1.

6. Create a table in your lab book to record results for genotypes of 50 individuals and label it Table 2.
7. Now, run the simulation. One lab partner should pick beans from the bag while the other records results. Without looking, remove two beans from your bag. These two beans represent one diploid individual in the next generation. Record the diploid genotype in the proper column of Table 2. **Return the beans to the bag.** Shake the bag to reinstate the gene pool.
8. This procedure is called **sampling with replacement**. By replacing the beans each time, the size of the gene pool remains constant and the probability of any allele being picked should remain equal to its frequency.
9. Repeat steps 6 and 7 until you have recorded the genotypes for 50 individuals (the next generation of the population). Add up the number of individuals for each genotype and record under “Genotypic Number” for the Observed Population in Table 1.

10. Calculate the observed genotypic frequencies and the observed allelic frequency of the new population. To calculate observed genotypic frequencies, divide each of the genotypic numbers from Table 1 by 50. Enter your values in Table 1 under Observed Population Genotypic Frequency.
11. Calculate the allelic frequency of the new population using the genotypic numbers as described in the following example:

If the genotypic number for $RR = 18$, $Rr = 24$ and $rr = 8$, the number of \underline{R} alleles in the population would equal $2(18) + 24 = 60$. The frequency of $\underline{R} = 60/100 = 0.60$. The number of \underline{r} alleles in the population would equal $24 + 2(8) = 40$. The frequency of $\underline{r} = 40/100 = 0.40$.

5. Simulating Evolutionary Change Genetic Drift

The term **genetic drift** describes the change in allelic frequencies observed in small populations as a **result of chance alone**. A large population may have been devastated by a natural disaster where only a small segment survives (also known as a **bottleneck effect**) or a small group may become separated from the larger group (the **founder effect**). In either case, the allele frequencies in the smaller group may no longer reflect those of the larger population and changes in allele frequencies may be observed over several generations. **Note that this is not an example of natural selection.** Survival during genetic drift is by chance not relative ability to reproduce.

Genetic fixation, the loss of all but one possible allele at a gene locus in a population, is a common result of genetic drift in small natural populations.

In this exercise, you will simulate the bottleneck effect using the Hardy-Weinberg theorem for your working hypothesis.

Procedure

1. State your hypothesis and prediction in your lab notebook

2. Establish a starting population of 50 individuals with a frequency of 0.5 for each allele (Generation 0). Calculate the expected genotypic numbers and genotypic frequencies for these allele frequencies and record in Table 3. Calculate the genotypic number based on a population of 5 individuals.
3. **Without replacement** (i.e., DO NOT put the beans you pick back in the bag this time; leave them on the table), randomly select five individuals (2 alleles per individual), or 10% of the population. Record the number of genotypes you selected in the space provided and number of R and r alleles for the new population in Table 3 (Generation 1).
4. Using your numbers from #3, determine the genotypic frequency observed for RR, Rr and rr. To do this, divide the number of individuals of each genotype by 5. For example, if 3 individuals had the genotype RR, genotypic frequency is calculated thus: $3/5 = 0.6$
5. Also calculate the new allelic frequencies for R (p) and r (q) for the surviving individuals. (Remember: the total number of individuals is 5 and the total number of alleles is 10.) These are your *observed* frequencies. Record these values in Table 3, Generation 1.
6. Reestablish the population to 50 individuals **using the new allelic frequencies**. For example, if your p value for Generation 1 is 0.6, 60 red beans should be in the bag for the next generation. This means your q value should be 0.4 ($p + q = 1$) and the total number of white beans in the bag is 40.
7. Repeat steps 3-6. Record your results in the appropriate generation in Table 3.
8. Reestablish the gene pool with new frequencies after each generation until one of the alleles becomes fixed in the population or until you have completed 5 generations.

9. Compare your results with your classmates.

6. Analyzing Results / Conclusion and Summary

1. To determine whether or not your results from Part 4, Table 1 are significantly different from your expected results for the Hardy-Weinberg simulation, you will use the chi-square (X^2) statistical test (see Appendix ii). Use the genotypic numbers for the Chi-square test.
2. Using your X^2 value, determine whether or not the data support or refute your hypothesis.
3. Using your data from the Genetic Drift simulation (Part 5, Table 3), graph the changes in p and q over time. Each set of values should be represented by a separate line on the same x-y plot.
4. Did the same allele go to fixation for all groups? If not, how many became fixed for R and how many for r?
5. Run the Chi-square test on the Genetic Drift simulation. Use the Expected Genotypic Numbers calculated for generation 0 in Table 3 and the Observed Genotypic Numbers from the last generation in Table 3. Do your results support or refute your hypothesis? Explain the significance of your results.
6. Were any of the results unexpected? Explain.
7. How reliable are your results? Were there any factors that could have affected the integrity of your data?

Source: Morgan, J.G and Cater, M.E.B. *Investigating Biology*, Third Edition, 1999