# Using the Hardy-Weinberg Equations:

# Quantifying Natural Selection and Allele Frequencies

**Objective:** Students will use the Hardy-Weinberg equations to determine if evolution is occurring in a real-world example. This will include calculating allele and genotypic frequencies, analyzing data sets and evaluating hypotheses through statistical analysis.

**Background:** In this activity, you will examine evolution by natural selection in a population of fruiting trees in Uganda and in Madagascar. Fruit has evolved to attract the attention of frugivores as a mechanism to disperse their seeds (which contain their offspring). When animals eat the fruit, the seeds are protected from digestion, thus when they defecate the offspring will grow far away from the parent tree. This increases the evolutionary fitness of fruiting plants by assisting dispersal. In both countries the main frugivores are primates, however there are important differences in their vision and the colors they see. The primates of Uganda (monkeys and apes) have trichromatic vision and can differentiate red and green, while the primates of Madagascar (lemurs) have dichromatic vision and are red/green color blind. The trees have evolved to best attract the attention of the primate frugivores that live in their habitat. As a result, in Uganda fruit tends to be red, while in Madagascar fruit tends to be yellow. The evolutionary scenario in this lab is based on research done in Uganda and Madagascar, but the genetics have been made up to simplify the example. (Please refer back to background reading for detailed discussion).

**Note:** In this activity, the dominant allele for red fruit is represented by *R* and the recessive allele for yellow fruit is represented by *r.* Trees that inherit two *R* alleles or one *R* and one *r* allele have red fruit, while trees that inherit two *r* alleles have yellow fruit. In this worksheet you will review some important principles and terms applicable to genetics. You will also utilize the Chi-squared statistical test to determine if there are deviations from Hardy Weinberg equilibrium. Your lab instructor will either assign one or both environments to your group.

**Materials per Group**

Uganda (Kibale National Park)

* 1 background (green)
* 1 container of 50 “alleles” – 1 x 2 LEGO (20 red and 30 yellow)
* 1 container of 25 “trees” – 2 x 4 LEGO (green)

Madagascar (Ranomafana National Park)

* 1 background (grey)
* 1 container of 50 “alleles” – 1 x 2 LEGO (20 grey and 30 yellow)
* 1 container of 25 “trees” – 2 x 4 LEGO (grey)

# Review of The Hardy-Weinberg (HW) Principle

Evolution is the change in allele frequency of a population over time. In the absence of evolution, allele frequencies will stay the same in a population from generation to generation. This constant allele frequency is the fundamental idea described by the Hardy-Weinberg (HW) principle. The HW model allows us to predict the genotype frequencies in the next generation of a population from the existing population’s allele frequencies as long as evolution is not occurring. The HW model has specific assumptions that describe a non-evolving population. These assumptions are: a large population (no genetic drift), no immigration or emigration (gene flow), no mutation, no natural selection, and random mating (no sexual selection). In this way, HW provides a null hypothesis for evolution. If a population’s genotypic frequencies change over generations, then the population is not in HW equilibrium and evolution is occurring.

What if a population’s genotypic frequencies remain the same from one generation to the next? Can we say that the population is in HW equilibrium? No, unfortunately. The population could be in HW equilibrium but not necessarily. HW has specific assumptions of no evolutionary processes acting on the population. Some evolutionary processes maintain stable frequencies in populations, reaching a type of equilibrium (such as balancing selection). It is very challenging to positively state that a population is in HW equilibrium. Even with that limitation, the Hardy-Weinberg model remains a very powerful tool to definitively recognize a population undergoing evolution.

# The following variables and equations are used to model the HW principle:

*p* + *q* = 1, where

* *p* is the frequency of dominant allele
* *q* is the frequency of recessive allele

*p2* + *2pq* + *q2* = 1, where

* *p2* is the expected genotype frequency of homozygous dominant individuals (*RR*)
* *2pq* is the expected genotype frequency of heterozygous individuals (*Rr*)
* *q2* is the expected genotype frequency of homozygous recessive individuals (*rr*)

**Hypotheses and Predictions**

We are testing for whether or not the genotypes are changing over time and by extension, if the population is evolving. Given the background reading, the context of the experiment and the Hardy-Weinberg principle, write a null hypothesis, alternative hypothesis, and prediction for both environments.

Uganda

Null Hypothesis:

Predictions for allele and genotype frequencies based on the Null Hypothesis:

Alternative Hypothesis:

Predictions for allele and genotype frequencies based on the Alternative Hypothesis:

Madagascar

Null Hypothesis:

Predictions for allele and genotype frequencies based on the Null Hypothesis:

Alternative Hypothesis:

Predictions for allele and genotype frequencies based on the Alternative Hypothesis:

# Procedure

Your instructor will assign you one or both environments: Uganda and/or Madagascar.

If you were assigned the Madagascar environment, remember the frugivores are red/green color blind, and the red LEGOs will appear grey.

1. The red LEGOs represent the allele for red fruit (*R*), and the yellow LEGOs represent the allele for yellow fruit (*r*). The background represents the environment where the trees live.
2. Use the LEGOs (alleles) in your container (Uganda or Madagascar) to count and record your starting allele frequencies. (Note: Don’t trust the numbers written on the container. Count them yourselves.) Record the number of red LEGOs (*R*) and the number of yellow LEGOs (*r*) in the “Generation 0” row in the columns labeled "Number of *R* Alleles;" and "Number of *r* Alleles", respectively. Calculate the frequency of each allele by dividing the number of each allele by the total number of alleles. Then, place all LEGOs (alleles) back in the container and thoroughly mix.
3. Without looking at the Legos, select two at a time, snap them on a “tree” LEGO brick, and record the results on the data table next to "Generation 1." For example, if you draw one red/grey and one yellow Lego, place a mark in the chart under "Number of *Rr* individuals." This process represents random mating and fusing gametes to produce zygotes. Continue drawing pairs of LEGOs and recording the results in your chart until all LEGOs have been selected and sorted. Place the diploid "trees" onto the background.
4. Uganda: The *rr* trees have yellow fruit and do not stand out well against the green foliage compared to red fruit. Therefore, they do not get eaten and dispersed. The seeds inside do not survive to adulthood and are removed from the population. Take all the rr trees and place them aside. (Remember where you put them! All LEGOs must be returned to the instructor). All trees with red fruit will survive to reproduce (*RR* and *Rr*).

Madagascar: The fruit of trees with genotype *RR* and *Rr* appear grey and are difficult to distinguish from the foliage, severely impacting the ability of the offspring (seeds) to reach adulthood. However, trees in Madagascar depend more on the smell of ripe fruit to attract frugivores. For the purpose of this activity we will say that the heterozygotes (*Rr*) have stronger smelling fruit than the homozygous dominant (*RR*) trees. Therefore, all trees with yellow fruit survive (*rr*) as will ½ of the heterozygotes (*Rr*). If there is an odd number of heterozygotes, divide the number by 2 and round up to the nearest whole number to obtain the number of survivors. The homozygous dominant trees do not survive (*RR*). Take all dead trees and place them aside.

1. Now count the *R* and *r* alleles (Legos) that remain on the surviving trees. These alleles will contribute to the next generation. Enter the number of each allele and their frequencies in the right side of Table 1 of your environment.
2. Remove the alleles from the "tree” LEGO blocks. Place the alleles of the surviving trees (which have grown, survived, and reached reproductive age) back into the container.
3. Repeat steps 3 through 6 to obtain the remaining generations listed in the tables below.

**Table 1: Uganda**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Parents** |   | **Alleles contributing to next generation** |
| Gen. | Number of *RR* individuals | Number of *Rr* individuals | Number of *rr* individuals |   | Number of *R* alleles | Number of *r* alleles | Frequency of *R* | Frequency of *r* |
| 0 |  XXXXXX |  XXXXXX |  XXXXXX |   |   |   |   |   |
| 1 |   |   |   |   |   |   |   |   |
| 2 |   |   |   |   |   |   |   |   |
| 3 |   |   |   |   |   |   |   |   |
| 4 |   |   |   |   |   |   |   |   |

**Table 1: Madagascar**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Parents** |   | **Alleles contributing to next generation** |
| Gen. | Number of *RR* individuals | Number of *Rr* individuals | Number of *rr* individuals |   | Number of *R* alleles | Number of *r* alleles | Frequency of *R* | Frequency of *r* |
| 0 |  XXXXXX |  XXXXXX |  XXXXXX |   |   |   |   |   |
| 1 |   |   |   |   |   |   |   |   |
| 2 |   |   |   |   |   |   |   |   |
| 3 |   |   |   |   |   |   |   |   |

**Discussion Questions**

1. For the environment/s assigned, prepare a line graph with allele frequency as a function of time (generation number) and graph your allele frequencies. (Allele frequency is plotted along the y-axis, and time [generation] is plotted on the x-axis.) Frequency should be represented in decimals. Plot all frequencies on one graph. Your instructor will inform you how to share your graph with the class.

Compare the frequencies of the alleles in the last generation to the starting generation. Would you say this particular tree population has evolved over multiple generations? Explain your reasoning.

1. Compare the change in the frequency of *R* across different environments. Describe what you see and explain the different results.
2. Calculate the *expected* Hardy-Weinberg **genotypic** frequencies in the starting generation for all three possible genotypes.
3. Allele frequencies given in your starting population (Generation 0):

*p* = frequency of **allele** *R*: *f*(*R*) =

*q* = frequency of **allele** *r*: *f*(*r*) =

1. **Expected** Hardy-Weinberg genotypic frequencies in the starting population (Generation 0):

*p2* = frequency of **genotype** *RR*: *f*(*RR*) =

2pq = frequency of **genotype** *Rr*: *f*(*Rr*) =

q2 = frequency of **genotype** *rr*: *f*(*rr*) =

1. If the population is under Hardy-Weinberg equilibrium, what will the genotypic frequencies be in the last generation?

Expected *f*(*RR*) =

Expected *f*(*Rr*) =

Expected *f*(*rr*) =

**Chi-squared (*χ2*) Statistic**

1. You will now use the Chi-squared (χ2) statistic to test the hypothesis that your observed genotypic frequencies are different than the expected HW genotypic frequencies. Did evolution occur? Follow the steps below, which will guide you in answering this question. *You will do this for both environments*.
2. In the table below use the *expected* frequencies you calculated in **question 3b** and the *observed* total number of trees you had in the last generation to determine how many trees of each genotype you *expected* see to see in the last generation.

Expected # of *RR* trees:

*Expected f*(*RR*)x (total # of adult trees in the last generation)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Expected *f*(*RR*) | x | Total # of adult trees in last generation | = | Expected # of *RR* trees |
| Uganda |  | x |  | = |  |
| Madagascar |  | x |  | = |  |

Expected # of *Rr* trees:

*Expected f*(*Rr*)x (total # of adult trees in the last generation)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Expected *f*(*Rr*) | x | Total # of adult trees in last generation | = | Expected # of *Rr* trees |
| Uganda |  | x |  | = |  |
| Madagascar |  | x |  | = |  |

Expected # of *rr* trees:

*Expected f*(*rr*)x (total # of adult trees in the last generation)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Expected *f*(*rr*) | x | Total # of adult trees in last generation | = | Expected # of *rr* trees |
| Uganda |  | x |  | = |  |
| Madagascar |  | x |  | = |  |

Next, enter these data into the “Expected” column in the table provided below for the Chi-Squared analysis.

For Chi-squared analysis, you will need to enter count data (number of trees) in the table below. In the “Observed” column enter the actual number of trees you saw for each genotype in the final generation. Note that the sum of your “Expected” column should be the same as the sum of your “Observed” column. (Note: If the sums are not the same, something is probably wrong with how you filled in your “Expected” column.) Use these calculated count data to solve the equation (O-E)2 /E for each row, and sum those values to produce your Chi-squared value.

$$χ^{2}=\sum\_{all cells}^{}\frac{\left(Observed-Expected\right)^{2}}{Expected}$$

Uganda:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed (O) | Expected (E) | (O-E)2/E |
| *RR* |  |  |  |
| *Rr* |  |  |  |
| *rr* |  |  |  |
|  | *Χ2*= |

Madagascar:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Observed (O) | Expected (E) | (O-E)2/E |
| *RR* |  |  |  |
| *Rr* |  |  |  |
| *rr* |  |  |  |
|  | *Χ2*= |

1. Using the *χ2* Critical Values table below with 1 degree of freedom (see note below), what is the critical value for *χ2* to reach a significance level of 0.05?

 \_\_\_\_\_\_\_\_\_\_\_

In order to reject the null hypothesis that our outcome was due to random chance (no evolution) we would like a p-value < 0.05 (see note below). To reach this level of significance our calculated *χ2* needs to be larger than the critical *χ2*.

Reject the null hypothesis if $χ\_{calc}^{2}>χ\_{crit}^{2}$

|  |
| --- |
| *χ2* Critical Values |
|  | Significance levels  |
| **Degrees of Freedom** | **.99** | **.80** | **.50** | **.20** | **.10** | **.05** | **.01** | **.001** |
|  **1**  | .00016 | .064 | .46 | 1.6 | 2.7 | 3.8 | 6.6 | 10.8 |
| **2** | .20 | .45 | 1.4 | 3.2 | 4.6 | 6.0 | 9.2 | 13.8 |
| **3** | .12 | 1.0 | 2.4 | 4.6 | 6.3 | 7.8 | 11.3 | 16.3 |
| **4** | .30 | 1.6 | 3.4 | 6.0 | 7.8 | 9.5 | 13.3 | 18.5 |

1. Did natural selection significantly alter genotype frequencies in this scenario? Explain.
2. Do you reject or fail to reject your null hypothesis? (In other words, is there a significant difference between observed and expected values?) Explain and justify your answer.
3. Did evolution occur? Evaluate your Alternative Hypothesis. Explain and justify your answer.

**Note: Degrees of freedom**: We are using a Chi-squared (*χ2*) test for a contingency table with two alleles. Because we have two alleles (two categories) our degree of freedom is 2-1 =1. We subtract 1 because we had to know the frequency of one allele (*p*) to calculate the other (1-*p*=*q*). Therefore, *p* is the only category that can vary, since *q* is dependent on *p*. This gives us one degree of freedom. Another way to approach it is - in a contingency table (expected offspring from gene pool with two alleles for gene) with *r* rows and *c* columns there are (*r*-1) x (*c*-1) degrees of freedom. Therefore, in a 2 x 2 contingency table (2-1) x (2-1) = 1 *df*.

**Note: P-value - Biological explanation for this scenario:** P-Value is the probability that our observed values (our results) can be explained by the null hypothesis (randomness – no evolution, no significant change in frequencies). Our null hypothesis is Hardy-Weinberg Equilibrium – no evolution. That would mean we expect our observed frequencies to be as predicted by the HW model – the Expected values. If the p-value is really small, less than 0.05, that means that the probability our observed values can be explained by randomness is really, really small - so small that we can reject randomness as our explanation (null hypothesis). Therefore, we can accept our alternative hypothesis, that evolution did occur.