# Lab 4. The Hardy Weinberg Model

**Notes:**

- Pre-Lab Exercise on Page 8
- Bring a calculator to lab.
- This lab is adapted from Bixler and Schee, 2005, "Application of the Hardy-Weinberg model to a mixed population of *Bar* and wild-type *Drosophila*. Pages 175-191, *in* Tested Studies for Laboratory Teaching, Volume 26 (MA O'Donnell, Editor). Proceedings of the 26th Workshop/Conference of the Association for Biology Laboratory Education (ABLE), 452 pages. Your instructors have modified the original exercise for use in this class.

**Objectives:**

1. Determine whether a population is in Hardy-Weinberg proportions.
2. Evaluate which of the five assumptions of the Hardy-Weinberg model may have been violated when a population is not in Hardy-Weinberg equilibrium.
3. Perform a chi-square test.

**KEY WORDS: allele frequency; chi-square test; genetic drift; genotype frequency; Hardy-Weinberg Model; mutation; migration; selection**

# Contents

# 1 Overview of the Hardy-Weinberg Model

The genetic diversity of a population can be characterized by the frequency of the alleles they carry for a given gene and the distribution of those alleles into genotypes. The frequency of alleles or genotypes can change from generation to generation. This change in allele and genotype frequency is one of the primary ways that we detect evolutionary *genetic* changes that occur in populations. Today, we'll work with one method that scientists use to detect such evolutionary changes.

The Hardy-Weinberg model is a valuable mathematical tool for studying how changes occur to a population's gene pool, and is covered in *Chapter 19: The Evolution of Populations* in the Biology 2e text. The model makes five assumptions about the population it describes:

1. the population is large in size and **genetic drift** does not affect allele frequencies for the gene of interest
2. **mutations** do not affect allele frequencies for the gene of interest
3. migration (**gene flow**) does not affect allele frequencies for the gene of interest

4. **natural selection** is not acting on the gene of interest
5. **mating** within the population is random with regard to the gene of interest

If these five assumptions are met, we infer that no forces are acting on the population to cause changes in the allele or genotype frequencies of this gene in this population. This conclusion allows us to make predictions about the allele and genotype frequencies we expect to find in the population in the present and future generations.

If we consider a single gene with two alleles ($A$ and $a$), we can describe the allele frequencies in the population as $p$ = frequency of the $A$ allele and $q$ = frequency of the $a$ allele. Given that the above assumptions are met (especially number 5!), we can predict the frequency with which we expect to find each diploid genotype by combining pairs of alleles at random (i.e., according to their frequency in the gene pool):

Expected frequency of the $AA$ genotype = (frequency of finding $A$ as the first allele) x (frequency of finding $A$ as the second allele) which simplifies to:

$$\text{expected frequency of genotype } AA = p \times p = p^2$$

Likewise, the expected frequency for the $Aa$ genotype $= 2pq$ and the expected frequency for the $aa$ genotype $= q^2$. Because this example has only two alleles, only these three genotypes are possible. Because no other genotypes are possible, we know that the three genotype frequencies must sum to 1.

$$p^2 + 2pq + q^2 = 1 \qquad \text{this equation may also be expressed as} \quad (p+q)^2 = 1$$

We predict these expected genotype frequencies, if the five assumptions described earlier hold. In other words, if a population is in Hardy-Weinberg Proportions, then the expected genotype frequencies will be equal to the observed genotype frequencies (which we figure out by actually genotyping the individuals in the population). In addition, if the five assumptions are met, then the allele and genotype frequencies will not change across generations, meaning the population is in Hardy-Weinberg Equilibrium.

# 2 The study organism and mutation

In this lab, we will examine the Hardy-Weinberg model and see how it can be used to study the factors that affect a population's evolution. The populations we will study are experimental populations of *Drosophila melanogaster*. We'll examine frequencies for two alleles of the *Bar* gene, which you were introduced to last week. Recall that the *Bar* gene is **sex-linked** (on the $X$ chromosome) and shows **incomplete dominance**, allowing us to determine the genotype of a fly by examining its phenotype. Remember that the wild-type round eye is characteristic of female flies that are wild-type homozygous ($X^+X^+$) and males that are wild-type hemizygous ($X^+Y$; Figure 1a). In contrast, the slit-shaped *Bar* eye is found in females that are *Bar*-homozygous ($X^BX^B$)and males that are *Bar*-hemizygous ($X^BY$; Figure 1c). Heterozygous females ($X^+X^B$) have a kidney- or heart-shaped eye (Figure 1b).

A couple of weeks ago, your instructors created several experimental populations of flies with the composition described in Table 1. After allowing the flies to mate for about a week, we removed the parents and allowed the offspring to continue their development. This week in lab, we're going to see the results of setting up those identical populations when we phenotype all of the offspring from each of the replicate populations—what do we expect the offspring allele frequencies to be? Their genotype frequencies?
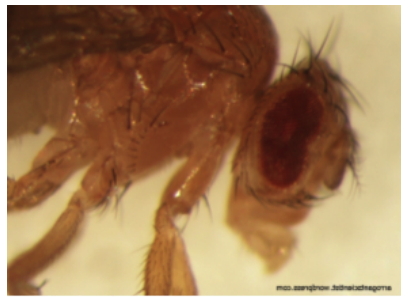
**(a)** Wild-type round eye      **(b)** Kidney- or heart-shaped eye      **(c)** *Bar* slit-shaped eye

**Figure 1:** Comparison of three eye phenotypes. [Images: Bixler and Schnee 2005]

**Table 1:** Composition of Parent generation crosses set up by instructors.

| Genotype | $X^+X^+$ | $X^+Y$ | $X^BX^B$ | $X^BY$ |
|---|---|---|---|---|
| Number of flies | 5 | 5 | 5 | 5 |

# 3     Allele and Genotype Frequencies in the Offspring Generation

Count the number of flies of each genotype in your experimental population. *To do so, you will need to distinguish males from females. You will also need to distinguish heterozygous females with kidney-shaped eyes ($X^BX^+$) from wild type females with round eyes ($X^+X^+$). (Remember that last week in lab you learned how to distinguish between the sexes.)* Record the data in Table 2 below. After you have counted the flies, return them to their original vials.

Once you have the counts for each genotype, calculate the frequency of each sex in the population (we cannot assume it will be 50:50 male:female) and the frequency of each genotype. Note that you will calculate genotype frequency both within each sex (i.e., what percentage of females has each female genotype?) and overall (i.e., what percentage of the total population has each genotype?). Record your results in the table. (Reminder: make sure your genotype frequencies sum to 1!)

Note: Record frequencies in decimal form, with 3 decimal places.

**Table 2:** Observed genotype numbers and frequencies for the first offspring generation ($O_1$).

| Phenotype | Genotype | Number of flies | Observed genotype frequency | |
|---|---|---|---|---|
| | | | *within-sex* | *overall* |
| *Bar* female | $X^BX^B$ | | | |
| Kidney eyed female | $X^BX^+$ | | | |
| Wild-type female | $X^+X^+$ | | | |
| *Bar* male | $X^BY$ | | | |
| Wild-type male | $X^+Y$ | | | |

Population ID # = _____

Total number of female flies = _____    Frequency of female flies = _____

Total number of male flies = _____    Frequency of male flies = _____

Total number of flies = _____

*Work space:*

---

**Post your observed genotype numbers and within-sex frequencies on the board when you are finished and be sure to copy down the values for the other populations in the table in the Post-Lab Assignment.**

---

From the information above in Table 2, you can calculate the allele frequencies in your population. Like in the pre-lab assignment, we have to account for the fact that males have only one allele for this gene. Use the equations below to calculate the allele frequencies and enter them in Table 3. Express each proportion as a decimal, with 3 decimal places. Please neatly show your work.

*Note: The calculations are done this way because females have two alleles, whereas males only have one (because the gene is on the X chromosome). So for q, the numerator represents all the Bar alleles that occur in homozygous females (2 per female) + the Bar alleles that occur in Bar-eyed males (1 per male), plus the Bar alleles that occur in heterozygous females (1 per female).*

As in the pre-lab assignment, let $q$ be the frequency of the *Bar* allele and $p$ be the frequency of the wild-type allele. Then,

$$q = \frac{(\text{\# } Bar \text{ females } \times 2 \text{ alleles}) + (\text{\# } Bar \text{ males} \times 1 \text{ allele}) + (\text{\# heterozygous females} \times 1 \text{ allele})}{(\text{\# females} \times 2) + (\text{\# males})}$$

$$p = \frac{(\text{\# wild-type females } \times 2) + (\text{\# wild-type males}) + (\text{\# heterozygous females})}{(\text{\# females } \times 2) + (\text{\# males})}$$

**Table 3:** Observed allele frequencies in the offspring generation.

| Allele | Frequency in $O_1$ |
|:------:|:------------------:|
| B      |                    |
| +      |                    |

Now, let's compare the $P$ generation and the $O_1$ generation.

If these populations are in Hardy-Weinberg equilibrium, what would we predict to be true about the allele frequencies in the $O_1$ generation compared to the allele frequencies in the $P$ generation?

Remember that in the $P$ generation, allele frequencies were $p = q = 0.5$ for all populations. How do the allele frequencies in your $O_1$ population compare to those of the $P$ generation? Does it look like the allele frequencies have changed over this one generation? Do you think your population is close to what you'd predict if the population is in Hardy-Weinberg equilibrium?

# 4   Testing for Hardy-Weinberg Proportions in the Offspring Generation

Lack of change in allele frequencies is only one part of the Hardy-Weinberg model – the other expectation of the HW model is that genotype frequencies within a generation will be a direct function of the allele frequencies. To check this prediction, we need to use the observed allele frequencies (in the offspring generation) to calculate the expected numbers of *Bar* and wild-type males, *Bar* females, wild-type females and heterozygous females in the offspring generation. Remember that *Bar* is sex-linked, so the equations for calculating genotype frequencies from allele frequencies will be different than the usual: the frequencies of males will be calculated differently from the frequencies of females AND you need to consider the frequency of each sex in the population. Discuss with your group how you would do this.

Complete the table below with the data that you collected in lab.

**Table 4:** Expected genotype frequencies and numbers for $O_1$ flies.

| Genotype | Expected genotype frequency | | Expected number of flies |
| --- | --- | --- | --- |
| | *within-sex* | *overall* | |
| $X^B X^B$ | | | |
| $X^B X^+$ | | | |
| $X^+ X^+$ | | | |
| $X^B Y$ | | | |
| $X^+ Y$ | | | |

*Work Space:*

# 5    Statistically evaluating our results with a chi-square test

Now we can do a statistical test to determine whether our observed numbers for the offspring generation differ significantly from the numbers we expected if the population was in Hardy-Weinberg proportions. If the observed numbers of genotypes ARE statistically different from those expected then evolution has occurred! We will use a chi-square ($\chi^2$) goodness-of-fit test. Briefly, the test calculates how much difference there is between the observed and expected numbers (NOT the frequencies), accounting for variation in sample size, and we use the $\chi^2$ test statistic to determine the probability of getting a difference this size just by chance (i.e., if in truth the population is actually in HW proportions). The chi-square test statistic is calculated as follows:

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where,

$$\chi^2 = \text{chi-square, the name of the test statistic}$$
$$\sum = \text{ sum of, in our case sum of mean differences for each genotype}$$
$$O = \text{ the observed number of individuals of a given genotype}$$
$$E = \text{ the expected number of individuals of that genotype}$$

The value that you calculate for the chi-square statistic is interpreted with the help of the chi-square distribution table, which tells us how often we would expect to randomly observe a given chi-square statistic value. To use the table to determine whether our deviation is statistically significant, we first need a little more information about our data: we need to know the degrees of freedom (df) and choose our alpha-level (i.e., significance cutoff).

- **Degrees of freedom** are directly related to the number of variables that you measure in the system you are working on. Each independent variable (meaning each variable whose value does NOT depend on the other variables in the system we are measuring) is equivalent to one degree of freedom. In most systems, however, the values of one or more variables are dependent on the values of others. For example, when measuring genotype frequencies at a given locus (as we are doing today), all genotype frequencies must add up to 1.0. Thus, if we measure the frequencies of *Bar*-eyed flies and kidney-eyed flies, we can determine the frequency of wild-type flies without actually observing it (since the sum of all three frequencies must be 1.0); given 2 values, we automatically know the third. Because of this, there is only one degree of freedom in this example. In general, the degrees of freedom for chi-square analysis are equal to the number of measured categories minus 1 (because the observed and expected totals must be equal) minus the number of parameters estimated from the data. **In our case, the number of categories = 5 (the number of genotypes) and the number of parameters = 2 (the frequency of one allele and the frequency of one sex). Thus, we have 5 - 1 - 2 = 2 degress of freedom.**

- The **alpha-level** or *P*-value sets the cutoff that we use to decide whether observing our deviation is improbable enough that we choose to reject the null hypothesis and therefore conclude that the population is not in HW proportions. The cutoff that we choose is arbitrary, at some level, but it is standard in biology to use $P < 0.05$. Essentially, this means that even if no evolution has occurred, there is less than a 5% chance of seeing our actual data.

Now that we know our df = 2 and our alpha-level ($P < 0.05$), we can use the chi-square distribution table to see how likely it would be that we observed our data, if the null hypothesis is true (i.e., the population

is in HW proportions). To do this, we find the row that corresponds to our df and then identify where our chi-square statistic falls relative to the columns for the *P*-value. If our value is larger than the one in the $P = 0.05$ column, then we conclude that if no evolution has occurred (the pop is in HW proportions) there is less than a 5% chance of observing this data. As a result, we would suspect that evolution HAS in fact occurred!

Let's consider a hypothetical example to understand how we interpret the values in the table. If our calculated chi-square statistic was 9.2, and we had df = 2, we would check the table to see that the value of 9.2 falls between the values for $P = 0.05$ and $P = 0.01$. So, in a population not experiencing any evolution, we would expect to see this hypothetical data, with a chi-square statistic of 9.2, approximately 1 out of every 100 times we sampled the population (i.e., 1% of the time). In other words, if we had 100 *Drosophila* populations that started with identical genotype frequencies and population sizes, we'd expect to observe a chi-square of 9.2 or greater in one of the 100 replicate populations, just by chance.

Chi-square distribution table.

| df | $P = 0.05$ | $P = 0.01$ | $P = 0.001$ |
|---|---|---|---|
| 1 | 3.84 | 6.64 | 10.83 |
| 2 | 5.99 | 9.21 | 13.82 |
| 3 | 7.82 | 11.35 | 16.27 |
| 4 | 9.49 | 13.28 | 18.47 |

For your population, perform the chi-square calculations by hand with a calculator. You should use the overall observed and expected genotype numbers in the chi-square equation (i.e., do NOT use the within-sex values).

**Show your work below** (use the last page of this handout if you need extra space). ***Then, write out a one-sentence summary of the statistical results, indicating whether they are statistically significant.***

**Summary of your statistical results:**

Show your chi-square results to your instructor or TA before leaving lab for the day.

# 6   Pre-Lab Exercise for Lab 4

Be sure to read sections *1 Overview of the Hardy-Weinberg Model* and *2 The study organism and mutation* before completing the exercise below.

### 6A   Calculating genotype and allele frequencies for the *P* generation

In lab, we will phenotype (and therefore genotype) the offspring of the parent population just described and then we'll test whether the populations are in Hardy-Weinberg proportions. As preparation, you will calculate the genotype and allele frequencies for the Parent generation (*P*)—these values will be the same for all of the replicate populations. Below are the parental generation numbers of adults of each genotype that were used to start each of the replicate populations.

1. Complete Table 5 below by filling in the genotype that corresponds to each phenotype and calculating the frequency of each genotype in the *P* generation, based on the data in Table 1. Express each proportion as a decimal, with 3 decimal places. Please neatly show your work.

**Table 5:** Observed phenotype/genotype numbers and frequencies for the *P* generation.

| *Phenotype* | *Genotype* | *Number of flies* | *Genotype frequency* | |
| --- | --- | --- | --- | --- |
| | | | *within sex* | *overall* |
| Wild-type female | | | | |
| Heterozygous female (kidney eyes) | | | | |
| *Bar* female | | | | |
| Wild-type male | | | | |
| *Bar* male | | | | |

2. Now, use the data from Table 5 to calculate the allele frequencies of the *P* generation and fill in the table below. Remember that because this gene is sex-linked, males have only 1 allele while females have 2. Because the *Bar* gene is sex-linked, we have to pay attention to how we count up our alleles here; we have to account for the fact that males can only have 1 allele for this gene while females have the usual 2 alleles. Use the equations below to calculate the allele frequencies and enter them in Table 6. Express each proportion as a decimal, with 3 decimal places. Please neatly show your work.

   *Note: For q, the numerator represents all the Bar alleles that occur in homozygous females (2 per female) + the Bar alleles that occur in Bar-eyed males (1 per male), plus the Bar alleles that occur in heterozygous females (1 per female).*

   Let $q$ be the frequency of the *Bar* allele and $p$ be the frequency of the wild-type allele. Then,

   $$q = \frac{(\text{\# } Bar \text{ females} \times 2 \text{ alleles}) + (\text{\# } Bar \text{ males} \times 1 \text{ allele}) + (\text{\# heterozygous females} \times 1 \text{ allele})}{(\text{\# females} \times 2) + (\text{\# males})}$$

   $$p = \frac{(\text{\# wild-type females} \times 2) + (\text{\# wild-type males}) + (\text{\# heterozygous females})}{(\text{\# females} \times 2) + (\text{\# males})}$$

**Table 6:** Allele frequencies of the *P* generation. Use 3 decimal places!

| Allele | Frequency |
|--------|-----------|
| $X^B$ | $q =$ |
| $X^+$ | $p =$ |

Next, we can ask: given the frequency of each allele in the *P* generation, what genotype frequencies would we expect to have seen, if the parent population was meeting the Hardy-Weinberg assumptions? To do this, we will calculate the *expected* genotype frequencies.

Remember, for a non-sex-linked gene with 2 alleles ($A^+$ and $A^-$), we'd expect:

| Genotype | Expected frequency |
|----------|-------------------|
| $A^+A^+$ | $p^2$ |
| $A^+A^-$ | $2pq$ |
| $A^-A^-$ | $q^2$ |

Of course, in our case the *Bar* gene is sex-linked, so we'll have to calculate our expected genotype frequencies a little bit differently than this (males have only 1 allele). After we calculate the expected frequencies using the HW relationships, then we also need to multiply those values by the proportion of the population that is male or female (depending on the genotype).

3. Fill in the following table with your predictions of how we should calculate expected genotype frequencies for our sex-linked locus and then calculate the expected frequencies for the *P* generation using the allele frequencies you calculated above. *Note that the values in Column b should sum to 2.0 while the values in Column d should sum to 1.0.*

| (a) Genotype | (b) Equation for HW expected frequency (within sex) | (c) Observed frequency of males or females | (d) Expected genotype frequency (overall) |
|--------------|-----------------------------------------------------|--------------------------------------------|-------------------------------------------|
| $X^+X^+$ | | | |
| $X^+X^B$ | | | |
| $X^BX^B$ | | | |
| $X^+Y$ | | | |
| $X^BY$ | | | |

4. Describe how your sex-linked expected values are different from the values expected for an autosomal (non-sex-linked) gene

5. Comparing the observed and expected genotype frequencies for the *P* generation, do you think the *P* generation was in Hardy-Weinberg proportions? Explain. (No need to do a statistical test.)

**End of Pre-lab assignment!**

# 7 Post-Lab Assignment: Turn in at start of Lab 5

**Determining which assumptions of HW equilibrium are violated:** The table below shows the results of three experiments (A, B, C) that were done exactly the same way as our populations (numbers in parentheses indicate the frequency of genotype by sex).

| Population | Observed Numbers by Genotype (frequency within each sex) | | | | | | |
|---|---|---|---|---|---|---|---|
| | $X^+X^+$ | $X^BX^+$ | $X^BX^B$ | Total # Females | $X^+Y$ | $X^BY$ | Total # Males |
| Experiment A | 86 (0.54) | 74 (0.46) | 0 (0) | 160 | 150 (0.88) | 21 (0.12) | 171 |
| Experiment B | 111 (0.56) | 68 (0.35) | 18 (0.09) | 197 | 83 (0.68) | 39 (0.32) | 122 |
| Experiment C | 55 (0.47) | 61 (0.53) | 0 (0) | 116 | 62 (0.74) | 22 (0.26) | 84 |
| Pop 1 | | | | | | | |
| Pop 2 | | | | | | | |
| Pop 3 | | | | | | | |
| Pop 4 | | | | | | | |
| Pop 5 | | | | | | | |
| Pop 6 | | | | | | | |
| Pop 7 | | | | | | | |
| Pop 8 | | | | | | | |
| Pop 9 | | | | | | | |
| Pop 10 | | | | | | | |
| Pop 11 | | | | | | | |
| Pop 12 | | | | | | | |

1. Do you notice any trends in the data? Are all the replicate populations similar in genotype frequencies? What about sex ratio? Do you think chance could explain the variation among different populations? How could you decide whether differences are due to chance or something else?

2. The HW assumptions would be broken if mating was not random. Think about the flies and their phenotypes – do you think that the behavior or senses of flies with *Bar* eyes would be the same as those of flies with wild-type eyes? The table below shows the results of experiments that tested the mating abilities of *Bar* and wild-type flies.

**Table 7:** Courtship behavior of *Bar* and wild-type flies.

| Phenotype of fly tested | | Percent time male spent courting | Copulation success | | |
|---|---|---|---|---|---|
| Male | Female | | Pairs copulated | Pairs tested | % Copulated |
| *Bar* | *Bar* | 24 | 5 | 54 | 9 |
| *Bar* | wild-type | 22 | 5 | 48 | 10 |
| wild-type | *Bar* | 59 | 17 | 52 | 33 |
| wild-type | wild-type | 68 | 23 | 50 | 46 |

Based on these data, do you think that the mating ability of flies is the same for all genotypes? Briefly explain.

3. What prediction(s) would you make with respect to changes in the genetic make-up of the experimental population, given these data? Briefly explain.

4. In another experiment, heterozygous females $(X^+X^B)$ were crossed to *Bar* males $(X^BY)$ and the following results were obtained. Based upon the above data, is there any evidence that the *Bar* allele affects survival of the flies? Do you think selection on survival affects the genotype frequencies in the populations you have observed? Explain.

**Table 8:** Genotype frequencies and numbers from offspring of an $F_2$ cross $(X^+X^B \times X^BY)$.

| Genotype | Frequency (within each sex) | Number |
|---|---|---|
| $X^BX^B$ | 0.47 | 43 |
| $X^BX^+$ | 0.53 | 49 |
| $X^BY$ | 0.50 | 38 |
| $X^+Y$ | 0.50 | 38 |

5. In the previous 3 questions, you considered the effects of drift (chance), non-random mating, and selection (differences in survival) on your populations. What are the other two Hardy-Weinberg assumptions? Can they be ruled out in your populations? Why or why not?

6. Which of the Hardy-Weinberg assumptions do you think was/were violated in the experimental populations we studied? Which factor(s) do you think were important in producing the allele and genotype frequencies found in your population, as well as in the other populations in the table of our observations? Explain your reasoning.