

30

MULTILOCUS HARDY-WEINBERG AND LINKAGE DISEQUILIBRIUM

Objectives

- Develop a spreadsheet model of allele and genotype frequencies at two loci.
- Examine properties of independent assortment of alleles.
- Use the chi-square test to determine if an offspring population is in Hardy-Weinberg equilibrium.
- Calculate D , the linkage disequilibrium coefficient.
- Graphically determine whether the population is in linkage equilibrium.

Suggested Preliminary Exercise: Hardy-Weinberg Equilibrium

INTRODUCTION

Now that you have been introduced to the Hardy-Weinberg equilibrium principle, it's time to explore the model in greater detail. Recall that this "null model" of evolution specifies algebraically what will happen across generations to the frequencies of alleles and genotypes. The bottom line is that in the absence of natural selection, genetic drift, mutation, and gene flow (and given a population of infinite size where mating is random), allele and genotype frequencies will not change over generations. That is, populations will not evolve. If the allele frequencies for a given locus in a population are given by p and q , the genotype frequencies will be p^2 , $2pq$, and q^2 if the population is in Hardy-Weinberg equilibrium.

In a previous exercise, you developed a single-locus model of the Hardy-Weinberg principle for locus A where p_1 was the frequency of the A_1 allele and q_1 was the frequency of the A_2 allele. In reality, organisms may have hundreds of loci on each of their chromosomes, and thus we need to start thinking about evolution at *multiple* loci.

In this exercise, you will learn that when multiple loci are involved, there are two kinds of equilibrium states: one is *Hardy-Weinberg* equilibrium, in which allele frequencies remain constant from generation to generation, and the second is *linkage* equilibrium. You will extend your single-locus model to examine two loci, loci A and B , simultaneously and to discover whether they are in fact in linkage equilibrium.

Hardy-Weinberg Equilibrium for Two Loci

Let's assume that the two alleles at locus *B* have the frequencies p_2 for the B_1 allele and q_2 for the B_2 allele. Furthermore, let's assume that the locus *B* is located on a different chromosome than locus *A*. Since the *A* and *B* loci each have only two alleles present in the population, the frequencies for *each* locus (p and q) must add to 1:

$$p_1 + q_1 = 1$$

Equation 1

and

$$p_2 + q_2 = 1$$

Equation 2

When two loci are considered, the genotype of an organism is characterized by its genotype at both loci, and 9 different genotypes are possible:

$A_1A_1B_1B_1$	$A_1A_1B_1B_2$	$A_1A_1B_2B_2$
$A_1A_2B_1B_1$	$A_1A_2B_1B_2$	$A_1A_2B_2B_2$
$A_2A_2B_1B_1$	$A_2A_2B_1B_2$	$A_2A_2B_2B_2$

Now suppose our hypothetical population mates randomly to produce a new generation of offspring. Individuals produce gametes (sex cells) through the process of meiosis. The end result is an egg or sperm cell that contains a single allele for the *A* locus and a single allele for the *B* locus. When an egg and sperm unite via sexual reproduction, the offspring zygote will regain its full complement of alleles. Depending on their genotype, individuals can produce between 1 and 4 different kinds of gametes (called **gamete classes**). The $A_1A_1B_1B_1$ individual can produce only 1 kind of gamete: A_1B_1 . The $A_1A_2B_1B_2$ individual can produce 4 kinds of gametes: A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . In the space provided in Figure 1, write in the kinds of gametes that each genotype can produce.

Genotype	Gametes
$A_1A_1B_1B_1$	
$A_1A_1B_1B_2$	
$A_1A_1B_2B_2$	
$A_1A_2B_1B_1$	
$A_1A_2B_1B_2$	
$A_1A_2B_2B_2$	
$A_2A_2B_1B_1$	
$A_2A_2B_1B_2$	
$A_2A_2B_2B_2$	

Figure 1

The frequencies of each gamete class (A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2) in a population depend on the genotype frequencies in the adult population. Thus, the gamete frequencies in the total population must be related in some way to the allele frequencies in the population. Indeed, the frequency of a gamete class is the product of the frequencies of the alleles that make up the gamete (Hartl 2000):

Frequency of the A_1B_1 gamete = $p_1 \times p_2$

Equation 3

Frequency of the A_1B_2 gamete = $p_1 \times q_2$

Equation 4

Frequency of the A_2B_1 gamete = $q_1 \times p_2$

Equation 5

Frequency of the A_2B_2 gamete = $q_1 \times q_2$

Equation 6

If we assume that p and q are known for each locus, Equations 3–6 allow us to predict the genetic makeup of the offspring population. Let's walk through an example. If our parental population has initial frequencies of $p_1 = 0.5$ and $q_1 = 0.5$ for the first locus, and $p_2 = 0.25$ and $q_2 = 0.75$ for the second locus, the frequencies of the gamete classes are:

$$\text{Frequency of the } A_1B_1 \text{ gamete} = p_1 \times p_2 = 0.5 \times 0.25 = 0.125$$

$$\text{Frequency of the } A_1B_2 \text{ gamete} = p_1 \times q_2 = 0.5 \times 0.75 = 0.375$$

$$\text{Frequency of the } A_2B_1 \text{ gamete} = q_1 \times p_2 = 0.5 \times 0.25 = 0.125$$

$$\text{Frequency of the } A_2B_2 \text{ gamete} = q_1 \times q_2 = 0.5 \times 0.75 = 0.375$$

Note that the sum of the gamete frequencies is 1, as it should be. Now that we know what the gamete frequencies are, we can predict the genotype frequencies of the offspring population by multiplying the probability that two gamete types will join to form a zygote. For example, an $A_1A_1B_1B_1$ genotype in the offspring population is the result of combining an A_1B_1 egg with a A_1B_1 sperm. The frequency of this genotype should be $0.125 \times 0.125 = 0.015625$ in the offspring population.

Because the gamete frequencies are related to the allele frequencies in the parental population, a second way of predicting the genotype frequencies of the offspring populations is to multiply their *independent* allele probabilities together. For example, if we want to estimate the proportion of $A_1A_1B_1B_1$ individuals in the next generation, we would multiply the probability that the offspring would inherit two A_1 alleles,

$$\text{Probability} = p_1 \times p_1 = p_1^2$$

by the probability of inheriting two B_1 alleles, or

$$\text{Probability} = p_2 \times p_2 = p_2^2$$

In our example, the proportion of $A_1A_1B_1B_1$ individuals is expected to be $(0.5 \times 0.5) \times (0.25 \times 0.25) = 0.015625$, or about 1.5% of the population. This is the same answer obtained by the gamete probability method. As a second example, if we want to estimate the proportion of $A_1A_2B_2B_2$ individuals in the population or in the next generation, we would multiply the probability of being heterozygous at the A locus ($2 \times p_1 \times q_1$, or $2 \times 0.5 \times 0.5$) by the probability of being homozygous B_2B_2 at the B locus ($q_2 \times q_2$, or 0.75×0.75). This generates a probability of $(2 \times 0.5 \times 0.5) \times (0.75 \times 0.75)$, which is 0.28125, or about 28% of the population. It's really that simple ... or is it?

Linkage Disequilibrium

A key assumption in calculating Hardy-Weinberg frequencies for two or more loci is that the loci are independent of each other. Essentially, this means that if you know what genotype the organism has at the first locus, you can't necessarily predict what its genotype will be at the second locus beyond what Hardy-Weinberg predicts. Knowing that an individual is A_1A_1 at the first locus doesn't tell us what the genotype at the second locus will be. Given that $p_2 = 0.25$ and $q_2 = 0.75$, Hardy-Weinberg tells us it has a 0.0625 chance of being B_1B_1 at the second locus, a 0.375 chance of being B_1B_2 at the second locus, and 0.5625 chance of being B_2B_2 at the second locus. Note that these frequency probabilities for this locus would be the same regardless of the genotype at the first locus. When alleles at different loci associate independently (at random), they are said to be in **linkage equilibrium**.

Sometimes, however, the two loci are *not* independent. For example, the A_1 allele may always associate with the B_1 allele and the A_2 allele with the B_2 allele. When this happens, the population is said to be in **linkage disequilibrium**. Linkage disequilibrium means, for example, that the different B genotypes are not distributed randomly among the different A genotypes and that, generally speaking, if you know the genotype at the A locus, you have a good idea of what the genotype at the B locus will be. Figure 2, for instance, shows that the B_1B_1 genotype occurs more commonly with the A_1A_1 genotype, and the B_2B_2 genotype occurs more frequently with the A_2A_2 genotypes.

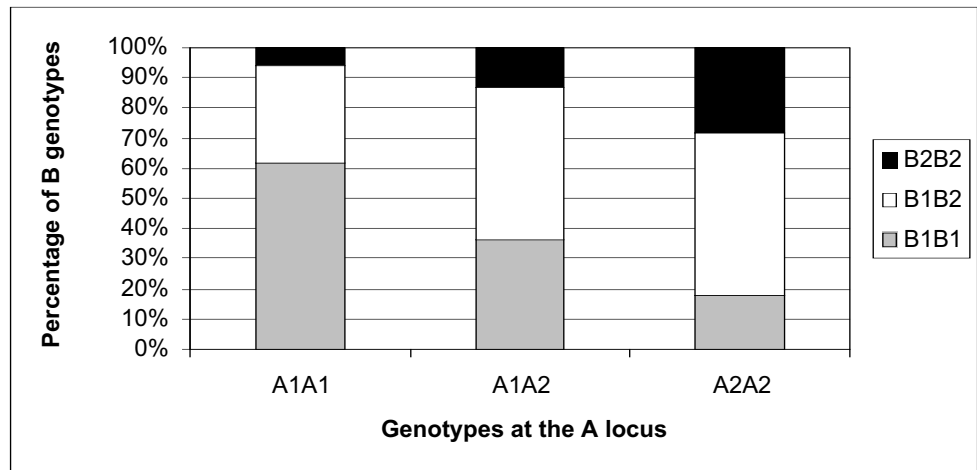


Figure 2 The bar graph shows linkage disequilibrium between the *A* and *B* alleles. If the population were in linkage equilibrium, the three different *B* genotypes would be distributed more or less equally among the three genotypes of the *A* allele.

Linkage disequilibrium can occur when the two loci are physically linked, meaning that they must be located close to each other on the same chromosome. During meiosis, the two alleles on the same chromosome may tend to segregate into the same gamete because of this physical linkage (Be aware, however, that not all alleles on the same chromosome are physically linked.)

Alleles can also be associated with each other because they are *coadapted*. **Coadaptation** is a beneficial interaction between alleles at different loci. For instance, if the A_1 and B_1 allele “work well” together to benefit an organism in its environment, they are said to be coadapted. Ayala (1982) gives this analogy to illustrate coadaptation of alleles at different loci:

A successful performance by a symphony orchestra requires not only that each player know how to play his instrument (a gene must be able to function), but also that he master his part in the piece being performed (a gene type must interact well with the other genes). A violinist playing his part for Beethoven’s Sixth Symphony while the rest of the orchestra was playing Ravel’s *Bolero* would be cacophonous.

Linkage disequilibrium can be quantified as the difference between the probability that A_1B_1 gametes unite with A_2B_2 gametes (these are called the **coupling gametes**) and the probability that A_1B_2 and A_2B_1 gametes unite (the **repulsion gametes**). The linkage disequilibrium coefficient is

$$D = G_{A_1B_1}G_{A_2B_2} - G_{A_1B_2}G_{A_2B_1} \quad \text{Equation 7}$$

where $G_{A_1B_1}$ is the frequency of the A_1B_1 gamete, $G_{A_2B_2}$ is the frequency of the A_2B_2 gamete, etc. The value of D ranges from 0 to 0.25. When the two alleles associate randomly, D will be 0. If the alleles are not randomly associated, D will increase. Assuming that the *A* and *B* loci are situated on different chromosomes, and assuming that the population mates at random without natural selection, gene flow, or mutation, the “level” of linkage disequilibrium breaks down with every passing generation. Unlike the single-locus Hardy-Weinberg model, which demonstrated that populations that are out of equilibrium go back into equilibrium after a single generation, several generations may be required for a population that is in linkage disequilibrium to acquire low levels of D .

PROCEDURES

The spreadsheet model you are about to develop is intended to give you some insights into how allele and genotype frequencies change over time when multiple loci are considered, and to help you determine whether or not a population is in linkage equilibrium. In this exercise, you will set up a population of 1000 individuals with a specified genotype frequency, let them mate at random, and then examine the genotype and allele frequencies of the offspring population. You will also calculate D , the linkage disequilibrium coefficient, and graphically determine whether populations are in linkage equilibrium. The approach in assigning genotypes to individuals in the population will be different than in the single-locus Hardy-Weinberg exercise, so that you can easily see how linkage disequilibrium works.

As always, save your work frequently to disk.

INSTRUCTIONS

A. Set up the model population.

1. Open a new spreadsheet and set up headings as shown in Figure 3.

2. In cells B5–B13, enter genotype frequency values shown.

3. In cells C4–C13, enter formulae to keep a running tally of the total genotype frequencies.

ANNOTATION

	A	B	C	D
1	Multilocus Hardy-Weinberg			
2				
3	Genotype	Frequency	Tally count	
4			0	
5	A1A1B1B1	0		
6	A1A1B1B2	0		
7	A1A1B2B2	0		
8	A1A2B1B1	0		
9	A1A2B1B2	1		
10	A1A2B2B2	0		
11	A2A2B1B1	0		
12	A2A2B1B2	0		
13	A2A2B2B2	0		<= This number MUST = 1
14				
15				
16	Individual	Random #	Genotype	Gamete

Figure 3

Cells B5–B13 give the genotype frequencies for the population. Enter the number 1 in cell B9, and 0s in the remaining cells. This indicates that our population will consist solely of $A_1A_2B_1B_2$ genotypes. Later in the exercise you will modify the values in these cells. Remember that the sum of the genotype frequencies in the population must equal 1.

Enter the number 0 in cell C4.

Enter the formula **=B5** in cell C5.

Enter the formula **=SUM(\$B\$5:B6)** in cell C6 and copy this formula down to cell C13. Cell C5 gives the running tally of genotype frequencies when only the first genotype, $A_1A_1B_1B_1$, has been considered. When you use the **SUM** function in cell C6 and copy the formula down to cell C13, it keeps a running tally of the genotype frequencies in your total population. Note that **\$B\$5** is an absolute reference, whereas the other cells

are relative references. This “anchors” cell B5 in the **SUM** so that the tally is a running tally. If cell C13 does not equal 1, it means that cells B5–B13 don’t add to 1. If so, make the necessary adjustments.

	A	B	C	D
3	Genotype	Frequency	Tally count	
4			0	
5	A1A1B1B1	0	0	
6	A1A1B1B2	0	0	
7	A1A1B2B2	0	0	
8	A1A2B1B1	0	0	
9	A1A2B1B2	1	1	
10	A1A2B2B2	0	1	
11	A2A2B1B1	0	1	
12	A2A2B1B2	0	1	
13	A2A2B2B2	0	1	<= This number MUST = 1

Figure 4

Your spreadsheet should now look like Figure 4. This tally will allow you to assign genotypes to individuals in a later step, and will help you determine quickly if your genotype frequencies add to 1.

4. Save your work prior to assigning genotypes to individuals in the next step.

5. In cells A17–A1016, set up a linear series from 1 to 1000.

6. In cells B17–B1016, generate a random number between 0 and 1.

7. In cells C17–C1016, enter a formula to assign a genotype to each individual

Enter 1 in cell A17.

Enter the formula **=A17+1** in cell A18. Copy this formula down to cell A1016.

Your population now consists of 1000 individuals.

Enter the formula **=RAND()** in cell B17 and copy this formula down to cell B1016.

When you press F9, the calculate key, the spreadsheet will generate new random numbers that will be used to assign a genotype to individuals in the population.

Enter the formula **=LOOKUP(B17,\$C\$4:\$C\$13,\$A\$5:\$A\$13)** in cell C17. Copy this formula down to cell C1016.

Here we use the **LOOKUP** function to assign genotypes based on the random number generated for each individuals, the frequencies you entered in cells B5–B13, and the tally of genotype frequencies in cells C4–C13. The function looks up a value (**B17**) in a vector that you specify (cells **\$C\$4:\$C\$13**) and returns a genotype for the individual given in the vector **\$A\$5:\$A\$13**. (Remember that a *vector* is a single row or column of values.) The **LOOKUP** function is handy because if it can’t find the exact lookup value (the random number given in cell B17), it matches the largest value in lookup vector (cells **\$C\$4:\$C\$13**) that is less than or equal to **lookup_value**. The result is that genotypes are assigned to individuals in approximately the proportions that you specified.

Examine your first 10 genotypes. They should all be $A_1A_2B_1B_2$ if the **LOOKUP** function worked properly. To see how the function works, change cells B5 and B13 to 0.5, and set cell B9 to 0. (Remember that the final tally of genotype frequencies must equal 1 in cell C13.) Now examine the genotypes of your first 10 individuals. The genotypes should be either $A_1A_1B_1B_1$ or $A_2A_2B_2B_2$. When you feel you have a handle on how this function works, return cells B5 and B13 to 0, and return cell B9 to 1.

8. Save your work.

B. Calculate allele frequencies, and determine gamete probabilities.

1. Set up new column headings as shown in Figure 5.

2. Enter formulae in cells F5–F6 and H5–H6 to calculate the allele frequencies for the two loci.

3. In cells E10–H10, enter formulae to calculate the expected gamete proportions.

4. In cell I10, enter a formula to sum the gamete probabilities.

	E	F	G	H
3	Allele frequencies			
4	Locus 1		Locus 2	
5	$p_1 = A1 =$		$p_2 = B1 =$	
6	$q_1 = A2 =$		$q_2 = B2 =$	
7				
8	Gamete probabilities			
9	A1B1	A1B2	A2B1	A2B2
10				
11				

Figure 5

In cell F5 enter the formula `=(COUNTIF(C17:C1016,"A1A1")*2+COUNTIF(C17:C1016,"A1A2"))/(2*A1016)`.

In cell F6 enter the formula `=1-F5`.

In cell H5 enter the formula

`=(COUNTIF(C17:C1016,"*B1B1")*2+COUNTIF(C17:C1016,"*B1B2"))/(2*A1016)`.

In cell H6 enter the formula `=1-H5`.

Recall from your first Hardy-Weinberg exercise that the frequencies of the A_1 and A_2 alleles are

$$\text{Frequency } (A_1) = 2N_{A_1A_1} + N_{A_1A_2})/2N$$

$$\text{Frequency } (A_2) = 2N_{A_2A_2} + N_{A_1A_2})/2N$$

There are 1000 individuals in the population, so the denominator will be 2000, which means that there are 2000 total "gene copies" present in the population. To obtain the frequency of the A_1 allele, we need to know how many of those gene copies are A_1 . Since this locus has only two alleles, the remainder of the gene copies will carry allele A_2 , so its frequency can be obtained by subtraction.

The * in the **COUNTIF** formulae is a "wildcard" that represents one or more unspecified characters. The F5 formula, for example, tells the spreadsheet to search for and count the number of A_1A_1 individuals regardless of what the remaining text in the cell is. Similarly, the H5 formula tells the spreadsheet to search for and count the number of B_1B_1 individuals regardless of what their genotype was at the A locus.

Enter the following formulae:

Cell E10 `=F5*H5`.

Cell F10 `=F5*H6`.

Cell G10 `=F6*H5`.

Cell H10 `=F6*H6`.

These formulae correspond to Equations 3–6. Gametes contain a single allele for the A locus and a single allele for the B locus. There are four possible gamete combinations: A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . The expected proportions of each combination are calculated by multiplying the appropriate allele frequencies together. For example, the expected proportion of A_1B_1 gametes in the population is the product of the A_1 allele frequency times the B_1 allele frequency.

Enter the formula `=SUM(E10:H10)` in cell I10.

The sum of the gamete probabilities will always be 1.

5. In cell D15, enter a formula using the **MID** function to generate a gamete type for each individual.

6. In cell D17–D1016, enter a combination of the **RAND()** and **MID** functions to generate a random gamete for each individual.

7. In cells E11–H11, use the **COUNTIF** formula to calculate the observed gamete frequencies.

8. Save your work.

C. Simulate sexual reproduction.

1. Set up new column headings as shown in Figure 6.

2. Use the **RAND()** and **VLOOKUP** functions to select random parents and lookup their gametes as you did in the Hardy-Weinberg equilibrium exercise.

In cell D15 enter the formula **=MID(C17,1,2)**. The **MID** function has the syntax **MID(text,start_num,num_chars)**. The formula in cell D15 tells the spreadsheet to examine the text in cell C17 and, starting with the first character, return 2 characters. If the formula were **=MID(C17,3,2)**, the spreadsheet would examine the text in cell C17 and would return 2 characters starting with the *third* character. In the next step, the **MID** function will allow us to generate a single gamete (selected randomly from the possible gametes that can be produced by an individual) for each individual in the population. If an individual is selected for mating, this gamete will be incorporated into the offspring’s gene pool. The gamete will contain either the first allele (A_1) or the second allele (A_2) for the A locus, and either the first allele (B_1) or second allele (B_2) for the B locus.

In cell D17 enter the formula **=IF(RAND()<0.5,MID(C17,1,2),MID(C17,3,2))&IF(RAND()<0.5,MID(C17,5,2),MID(C17,7,2))**. Copy this formula down to cell D1016.

The first part of this formula (to the left of the **&**) generates the *A* allele in the gamete, and the second part (to the right of the **&**) generates the *B* allele. The first part draws a random number between 0 and 1; if this random number is less than 0.5, the spreadsheet returns the first and second values from cell C17; otherwise, it returns the third and fourth values from C17. The second part of the formula draws a random number, and returns the fifth and sixth values from C17 or returns the seventh and eighth values. Joining the two parts with the **&** symbol results in a gamete for the individual.

Enter the following formulae:
Cell E11 **=COUNTIF(\$D\$17:\$D\$1016,E9)/1000**.
Cell F11 **=COUNTIF(\$D\$17:\$D\$1016,F9)/1000**.
Cell G11 **=COUNTIF(\$D\$17:\$D\$1016,G9)/1000**.
Cell H11 **=COUNTIF(\$D\$17:\$D\$1016,H9)/1000**.

Note that when you press F9, the calculate key, new random numbers are generated. This action generates new genotypes, and also generates a new gamete for each individual in the population.

	E	F	G	H	I	J	K	L	M
15	Random	Random	Random	Random	Offspring genotype				
16	mom	mom's egg	dad	dad's sperm	Locus A		Locus B		Genotype

Figure 6

In cells E17–E1016 and cells G17–G1016 you can enter either one of the follow formulae:
=ROUNDUP(RAND()*1000,0)
=RANDBETWEEN(1,1000)
In cells F17–F1016 enter the formula **=VLOOKUP(E17,\$A\$17:\$D\$1016,4)**.
In cells H17–H1016 enter the formula **=VLOOKUP(G17,\$A\$17:\$D\$1016,4)**.
Refer to Exercise 29, “Hardy-Weinberg Equilibrium,” if needed. Your spreadsheet should look similar to Figure 7, although your numbers will be different.

	E	F	G	H
15	Random	Random	Random	Random
16	mom	mom's egg	dad	dad's sperm
17	654	A2B1	528	A1B2
18	367	A1B1	568	A2B2
19	175	A2B2	70	A2B1
20	643	A2B2	692	A2B1
21	956	A1B2	488	A1B2

Figure 7

In Figure 7, the first random Mom was individual 654, and the first random Dad was individual 528. Since the population has a genotype frequency of $A_1A_2B_1B_2 = 1$, all individuals in the population have the genotype $A_1A_2B_1B_2$. This type of individual can produce four kinds of gametes: A_1B_1 , A_1B_2 , A_2B_1 , and A_2B_2 . Although four different kinds of gametes can be produced, a single randomly chosen gamete from an individual will fuse with a gamete from another individual, producing a zygote. Mom 654 has a gamete A_2B_1 , while Dad 528 has a sperm gamete A_1B_2 . The zygote offspring from this union will have the genotype $A_1A_2B_1B_2$. The next few steps will generate the genotypes of the offspring.

3. In cells I17–I1016 enter a formula to determine the offspring's genotype at the *A* locus.

Enter the formula `=LEFT(F17,2)&LEFT(H17,2)` in cell I17. Copy this formula down to cell I1016.

4. In cells K17–K1016 enter a formula in cell K17 to determine the offspring's genotype at the *B* locus.

Offspring 1 in cell I17 will inherit one *A* allele from its mother and one *A* allele from its father. The formula in cell I17 takes the left two characters from cell F17 and combines them with the left two characters from cell H17.

Enter the formula `=RIGHT(F17,2)&RIGHT(H17,2)` in cell K17. Copy this formula down to cell K1016.

5. In cells J17–J1012 and L17–L1012, enter a formula to adjust the genotypes so that all heterozygotes are described as either A_1A_2 or B_1B_2 (not A_2A_1 or B_2B_1).

Enter the formula `=IF(I17="A2A1","A1A2",I17)` in cell J17 and copy it down to cell J1012.

Enter the formula `=IF(K17="B2B1","B1B2",K17)` in cell L17 and copy it down to cell L1012.

This step is necessary because an A_1A_2 heterozygote is the same thing as an A_2A_1 heterozygote, but the spreadsheet "interprets" them as being different.

6. In cells M17–M1016 enter a formula to determine the genotype of each offspring.

Enter the formula `=J17&L17` in cell M17. Copy your formula down to cell M1016. The genotype of the offspring is the combination of genotypes at the *A* and *B* loci.

7. Save your work.

D. Determine if the population is in Hardy-Weinberg equilibrium and linkage equilibrium.

1. Set up new column headings as shown in Figure 8.

	J	K	L	M
1		Observed genotype frequencies		
2		A1A1	A1A2	A2A2
3	B1B1			
4	B1B2			
5	B2B2			
6		Expected genotype frequencies		
7		A1A1	A1A2	A2A2
8	B1B1			
9	B1B2			
10	B2B2			
11	Chi-square test P value =			
12	Hardy-Weinberg equilibrium?			
13	Linkage disequilibrium coefficient = D =			

Figure 8

2. Enter formulae in cells K3–M5 to calculate the observed genotype frequencies in the offspring population.

Double-check your results. Your frequencies should add to 1.

3. Enter formulae in cells K8–M10 to calculate the expected genotype frequencies in the offspring population.

Double-check your results. Your frequencies should add to 1.

Enter the following formulae:

- Cell K3 =COUNTIF(\$M\$17:\$M\$1016,"A1A1B1B1")/1000
- Cell K4 =COUNTIF(\$M\$17:\$M\$1016,"A1A1B1B2")/1000
- Cell K5 =COUNTIF(\$M\$17:\$M\$1016,"A1A1B2B2")/1000
- Cell L3 =COUNTIF(\$M\$17:\$M\$1016,"A1A2B1B1")/1000
- Cell L4 =COUNTIF(\$M\$17:\$M\$1016,"A1A2B1B2")/1000
- Cell L5 =COUNTIF(\$M\$17:\$M\$1016,"A1A2B2B2")/1000
- Cell M3 =COUNTIF(\$M\$17:\$M\$1016,"A2A2B1B1")/1000
- Cell M4 =COUNTIF(\$M\$17:\$M\$1016,"A2A2B1B2")/1000
- Cell M5 =COUNTIF(\$M\$17:\$M\$1016,"A2A2B2B2")/1000

Remember that you can calculate the expected genotype frequencies of the offspring in either one of two ways:

- Multiply the expected gamete frequencies *or*
- Multiply the allele frequencies in the adult population

Both methods should both yield the same results.

If you calculate the expected frequencies based on expected gamete frequencies in the adult population, remember to calculate the variety of ways in which gametes from Mom and Dad can combine. For example, if the frequency of an offspring genotype of $A_1A_1B_1B_2$ can be generated in two ways: Mom's egg can be A_1B_1 and Dad's sperm can be A_1B_2 , or Mom's egg can be A_1B_2 and Dad's sperm can be A_1B_1 . Both possibilities need to be accounted for to generate correct offspring genotype frequencies. Enter the following formulae:

- Cell K8 =E10*E10
- Cell K9 =E10*F10+F10*E10
- Cell K10 =F10*F10
- Cell L8 =E10*G10+G10*E10
- Cell L9 =E10*H10+H10*E10+F10*G10+G10*F10
- Cell L10 =F10*H10+H10*F10
- Cell M8 =G10*G10
- Cell M9 =G10*H10+H10*G10
- Cell M10 =H10*H10

4. Save your work.

5. In cell M11, conduct a chi-square test on observed and expected frequencies.

6. In cell M12, enter a formula to answer “yes” or “no” to the question “Is the population in Hardy-Weinberg equilibrium?”

7. In cell M13, enter a formula to calculate D , the linkage disequilibrium coefficient.

E. Create graphs.

1. Create a column graph of the genotypes observed in the offspring population. Label your axes fully.

If you calculate the expected frequencies based on allele frequencies in the parental population, enter the following formulae:

- Cell K8 = $F5 * F5 * H5 * H5$
- Cell K9 = $F5 * F5 * 2 * H5 * H6$
- Cell K10 = $F5 * F5 * H6 * H6$
- Cell L8 = $2 * F5 * F6 * H5 * H5$
- Cell L9 = $2 * F5 * F6 * 2 * H5 * H6$
- Cell L10 = $2 * F5 * F6 * H6 * H6$
- Cell M8 = $F6 * F6 * H5 * H5$
- Cell M9 = $F6 * F6 * 2 * H5 * H6$
- Cell M10 = $F6 * F6 * H6 * H6$

Enter the formula =CHITEST(K3:M5,K8:M10) in cell M11.

Refer to Exercise 29, on “Hardy-Weinberg Equilibrium,” for the information on this test and its interpretation.

Enter the formula =IF(M11>0.05,“yes”,“no”) in cell M12.

In cell M13 enter the formula =E11*H11-F11*G11.

Equation 7 gave the formula for the disequilibrium coefficient D as

$$D = G_{A_1B_1}G_{A_2B_2} - G_{A_1B_2}G_{A_2B_1}$$

where G represents the frequency of the different kinds of gametes observed in the population. Remember that D ranges between 0 and 0.25. When the population is in linkage equilibrium, $D = 0$. Your result for this exercise should be very close to 0, indicating that your population *is* in linkage equilibrium.

Select cells J2–M5. Use the bar graph option and label your axes fully. Your graph should resemble Figure 9, although your frequencies will likely be a bit different than the ones shown.

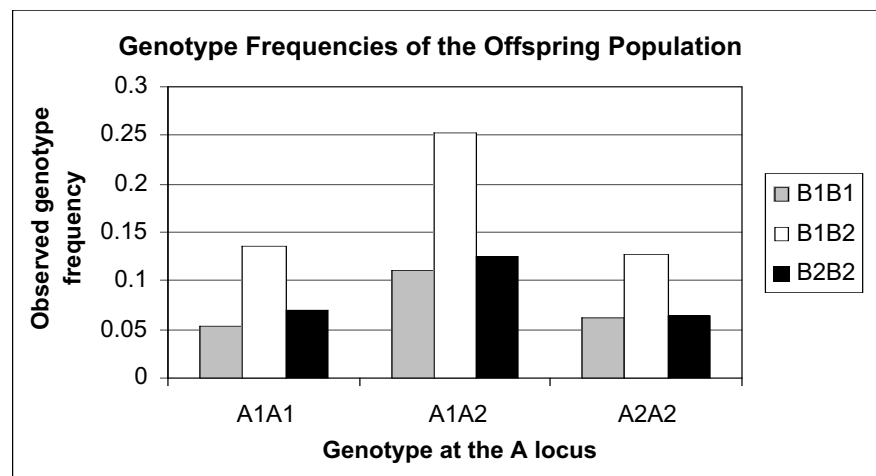


Figure 9

2. Graphically determine whether the various *B* genotypes are distributed more or less equally among the various *A* genotypes.

3. Save your work.

Select cells J2–M5 again. Create a new bar graph and choose the 100% stacked column option. Your graph should resemble Figure 10. This graph breaks down the percentage of each *B* genotype within each *A* genotype. Since the percentages are relatively equal, this population is in linkage equilibrium.

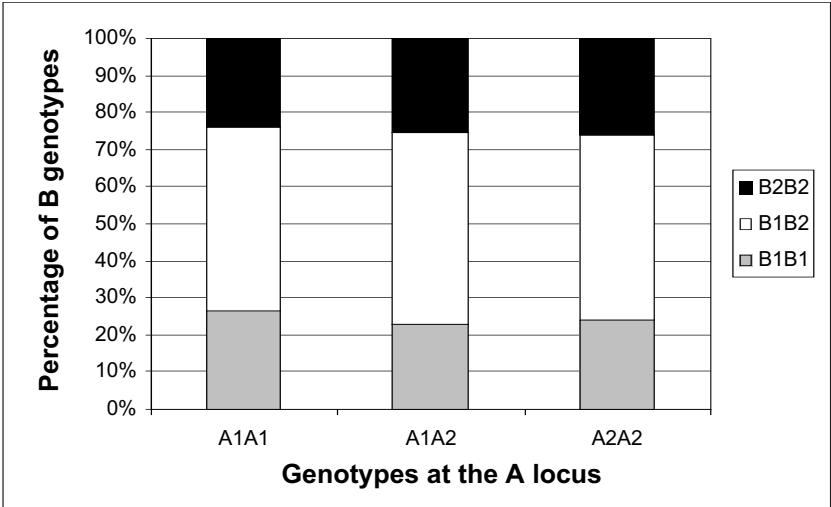


Figure 10

QUESTIONS

- 1. Interpret the graph you generated in the very last step. In particular, comment on whether the frequencies of *B*₁*B*₁, *B*₁*B*₂, and *B*₂*B*₂ are proportionately the same for *A*₁*A*₁, *A*₁*A*₂, and *A*₂*A*₂ individuals. Is your population in linkage equilibrium? Why or why not?
- 2. Alter allele frequencies as shown below. Update your graphs and calculate *D*. Comment on whether the frequencies of *B*₁*B*₁, *B*₁*B*₂, and *B*₂*B*₂ are the same for all of the *A*₁*A*₁, *A*₁*A*₂, and *A*₂*A*₂ individuals.

	A	B
3	Genotype	Frequency
4		
5	A1A1B1B1	0.1
6	A1A1B1B2	0
7	A1A1B2B2	0.1
8	A1A2B1B1	0.7
9	A1A2B1B2	0
10	A1A2B2B2	0
11	A2A2B1B1	0
12	A2A2B1B2	0
13	A2A2B2B2	0.1

- 3. Assume that alleles *A*₁ and *B*₁ interact well with each other and thus are coadapted, and that the *A*₂ and *B*₂ alleles are also coadapted. Assume also that other combinations of alleles (*A*₁*A*₂*B*₁*B*₁, etc.) yield a poorly adapted phenotype. In

this case, $A_1A_1B_1B_1$ and $A_2A_2B_2B_2$ individuals will dominate the population. Alter allele frequencies so that $A_1A_1B_1B_1 = 0.5$, and $A_2A_2B_2B_2 = 0.5$. Modify values in cells B5 and B13, and set the remaining genotype frequencies to 0. Is the parental population in Hardy-Weinberg equilibrium? Is the offspring population in Hardy-Weinberg equilibrium? What is D ? Graph your results and interpret D .

4. If your offspring population from question 3 were to reproduce, how would D change over time? How does the frequency of the A_1 allele (p_1) and the frequency of the B_1 (p_2) allele change over time? Simulate the reproduction of individuals over three generations. Set up column headings as shown in the figure below. Start with the genotype frequencies shown for generation 1. Enter 0.5 in cells B5 and B13. Set the remaining genotypes to 0. Calculate D , p_1 , and p_2 . Record this information in cells V15–V17. Examine the genotypes of the offspring. Enter those genotype frequencies in cells W5–W13 (as shown in the figure below; your numbers will be slightly different). Enter them again in cells B5–B13. Calculate D , p_1 , and p_2 for the second generation. Record your results in cells W15–W17. Repeat the process for generation 3. For each generation, examine the 100% column graph (as in Figure 9). Graphically show how D , p_1 , and p_2 change over generations.

	U	V	W	X
4		Generation 1	Generation 2	Generation 3
5	A1A1B1B1	0.5	0.266	
6	A1A1B1B2	0	0	
7	A1A1B2B2	0	0	
8	A1A2B1B1	0	0	
9	A1A2B1B2	0	0.494	
10	A1A2B2B2	0	0	
11	A2A2B1B1	0	0	
12	A2A2B1B2	0	0	
13	A2A2B2B2	0.5	0.24	
14				
15	$D =$	0.25		
16	$A1 =$	0.5		
17	$B1 =$	0		

LITERATURE CITED

- Ayala, F. J. 1982. *Population and Evolutionary Genetics: A Primer*. Benjamin/Cummings, Menlo Park, CA.
- Hartl, D. L. 2000. *A Primer of Population Genetics*, 3rd Edition. Sinauer Associates, Sunderland, MA.