

LAB 19 – Population Genetics and Evolution II

Objectives:

- To use a data set that reflects a change in the genetic makeup of a population over time and to apply mathematical methods and conceptual understandings to investigate the cause(s) and effect(s) of this change
- To apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future
- To evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time
- To use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and the effect of selection in the evolution of specific populations
- To justify data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and the effects of selection in the evolution of specific populations

Introduction:

Evolution occurs in populations of organisms and involves variation in the population, heredity, and differential survival. One way to study evolution is to study how the frequency of alleles in a population changes from generation to generation. In other words, you can ask *What are the inheritance patterns of alleles, not just from two parental organisms, but also in a population?* You can then explore how allele frequencies change in populations and how these changes might predict what will happen to a population in the future.

Mathematical models and computer simulations are tools used to explore the complexity of biological systems that might otherwise be difficult or impossible to study much more data than our previous Hardy-Weinberg simulation. Several models can be applied to questions about evolution. In this investigation, you will build a spreadsheet that models how a hypothetical gene pool changes from one generation to the next. This model will let you explore parameters that affect allele frequencies, such as selection, mutation, and migration.

This investigation also provides an opportunity for you to review concepts you might have studied previously, including natural selection as the major mechanism of evolution; the relationship among genotype, phenotype, and natural selection; and fundamentals of classic Mendelian genetics.

The second part of the investigation asks you to generate your own questions regarding the evolution of allele frequencies in a population. Then you are asked to explore possible answers to those questions by applying more sophisticated computer models. These models are available for free.

There are some important things to remember when computer modeling in the classroom. To avoid frustration, periodically save your work. When developing and working out models, save each new version of the model with a different file name. That way, if a particular strategy doesn't work, you will not necessarily have to start over completely but can bring up a file that had the beginnings of a working model.

If you have difficulty refining your spreadsheet, consider using the spreadsheet to generate the random samples and using pencil and paper to archive and graph the results.

As you work through building this spreadsheet you may encounter spreadsheet tools and functions that are not familiar to you. Today, there are many Web-based tutorials, some text based and some video, to help you learn these skills. For instance, typing “How to use the SUM tool in Excel video” will bring up several videos that will walk you through using the SUM tool. You will need to figure out things with your own searches and successful navigation of the help files of all programs we use.

Procedure:

Quantitatively Describing the Biological System

Spreadsheets are valuable tools that allow us to ask *What if?* questions. They can repeatedly make a calculation based on the results of another calculation. They can also model the randomness of everyday events. Our goal is to model how allele frequencies change through one life cycle of this imaginary population in the spreadsheet. Use the diagram in **Figure 1** as a guide to help you design the sequence and nature of your spreadsheet calculation. The first step is to randomly draw gametes from the gene pool to form a number of zygotes that will make up the next generation.

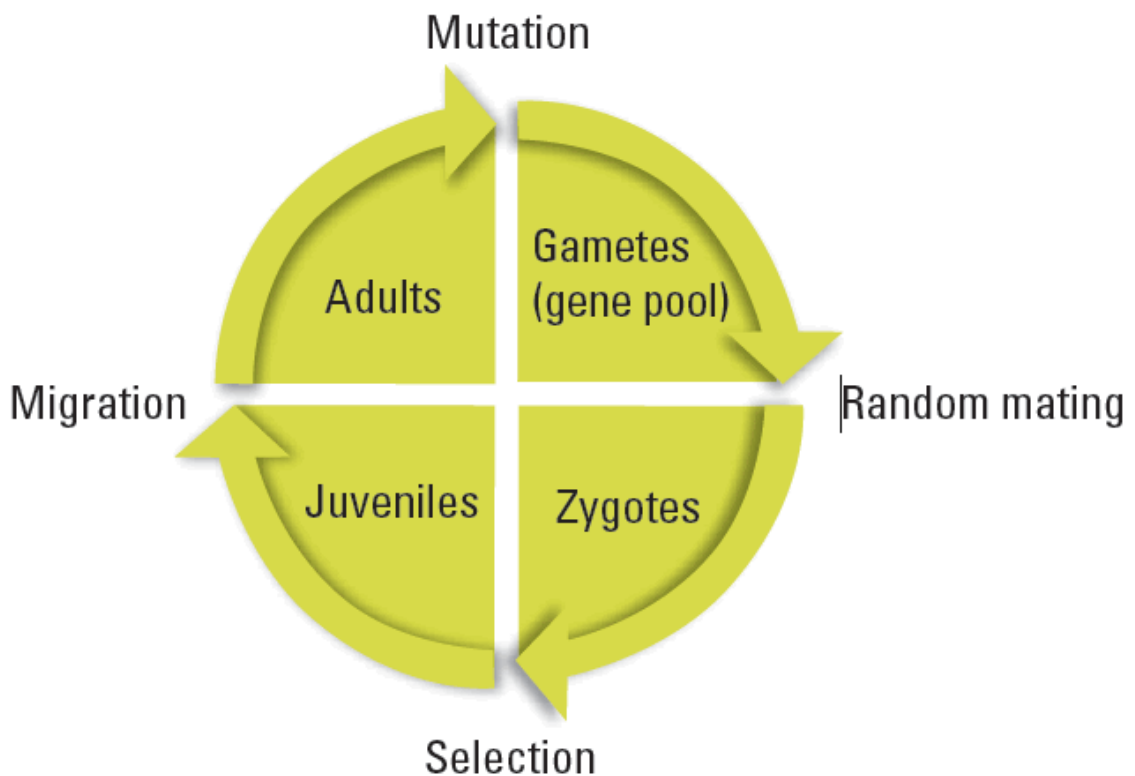


Figure 1: Life Stages of a Population of Organism

To begin this model, let's define a couple of variables.

Let:

p = the frequency of the *A* allele
and let q = the frequency of the *B* allele

1. Bring up the spreadsheet on your computer. The examples here are based on Microsoft Excel, but almost any modern spreadsheet can work, including Google's online Google Drive) and Zoho's online spreadsheet (<http://www.zoho.com>).

HINT: If you are familiar with spreadsheets, the RAND function, and using IF statements to create formulas in spreadsheets, you may want to skip ahead and try to build a model on your own. If these are not familiar to you, proceed with the following tutorial.

2. In cell D2, enter a value for the frequency of the *A* allele. This value should be between 0 and 1. Go ahead and type in labels in your other cells and, if you wish, shade the cells as well. This blue area will represent the gene pool for your model. (Highlight the area you wish to format with color, and right-click with your mouse in Excel to format.) This is a spreadsheet, so you can enter the value for the frequency of the *B* allele; **however**, when making a model it is best to have the spreadsheet do as many of the calculations as possible. All of the alleles in the gene pool are either *A* or *B*; therefore $p + q = 1$ and $1 - p = q$. In cell D3, enter the formula to calculate the value of q . In spreadsheet lingo it is:

$$=1-D2$$

Your spreadsheet should now look something like **Figure 2**.

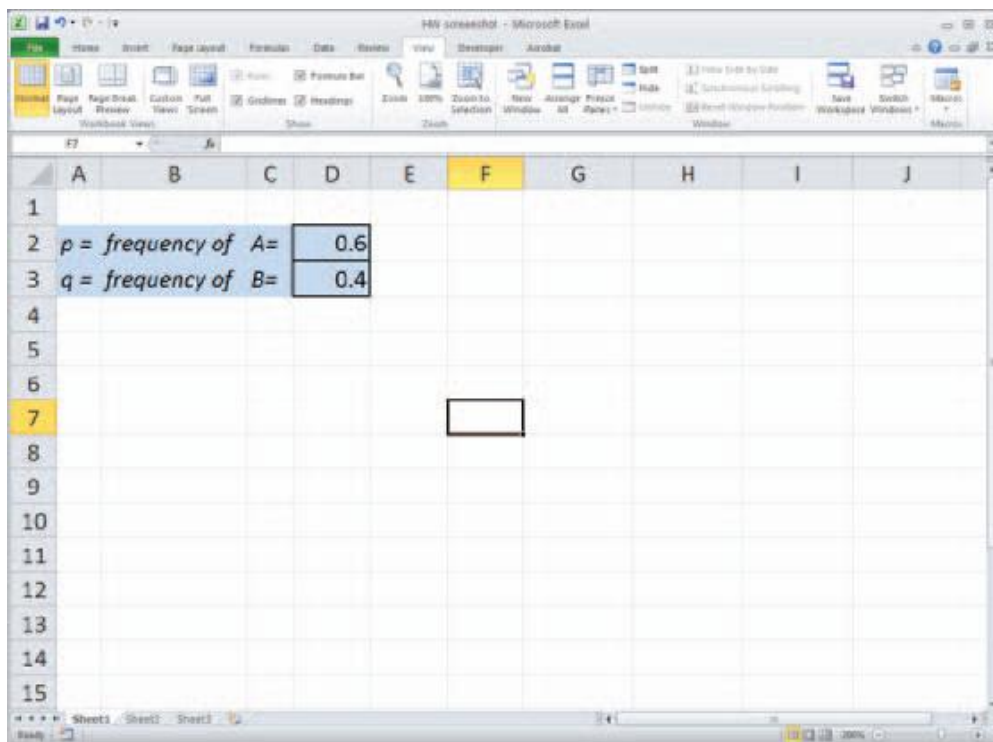


Figure 2

3. Let's explore how one important spreadsheet function works before we incorporate it into our model. In cell G2, enter the following function (we will remove it later).

$$=RAND()$$

Note that the parentheses have nothing between them. After hitting *return*, what do you find in the cell? If you are on a PC, try hitting the F9 key several times to force recalculation. Notice what happens to the value in the cell.

The RAND function returns random numbers between 0 and 1 in decimal format. This is a powerful feature of spreadsheets. It allows us to enter a sense of randomness to our calculations if it is appropriate – and here it is when we are “randomly” choosing gametes from a gene pool. Go ahead and delete the RAND function in the cell.

4. Let's select two gametes from the gene pool. In cell E5, let's generate a random number, compare it to the value of p , and then place either an A gamete or a B gamete in the cell. We'll need two functions to do this, the RAND function and the IF function.

Check the help menu if necessary. Note that the function entered in cell E5 is

$$=IF(RAND()<=D2, "A", "B")$$

Be sure to include the $\$$ in front of the D and the 2 in the cell address D2. It will save time later when you build onto this spreadsheet by always referring to the value in cell D2, even when you cut and paste this formula around.

The formula in this cell basically says that if a random number between 0 and 1 is less than or equal to the value of p , then put an A gamete in this cell, or if it is not less than or equal to the value of p , put a B gamete in this cell. IF functions and RAND functions are very powerful tools when you try to build models for biology.

5. Now create the same formula in cell F5, making sure that it is formatted exactly like E5. When you have this completed, press the recalculate key (F9) to force a recalculation of your spreadsheet. If you have entered the functions correctly in the two cells, you should see changing values in the two cells. Your spreadsheet should look like **Figure 3**.

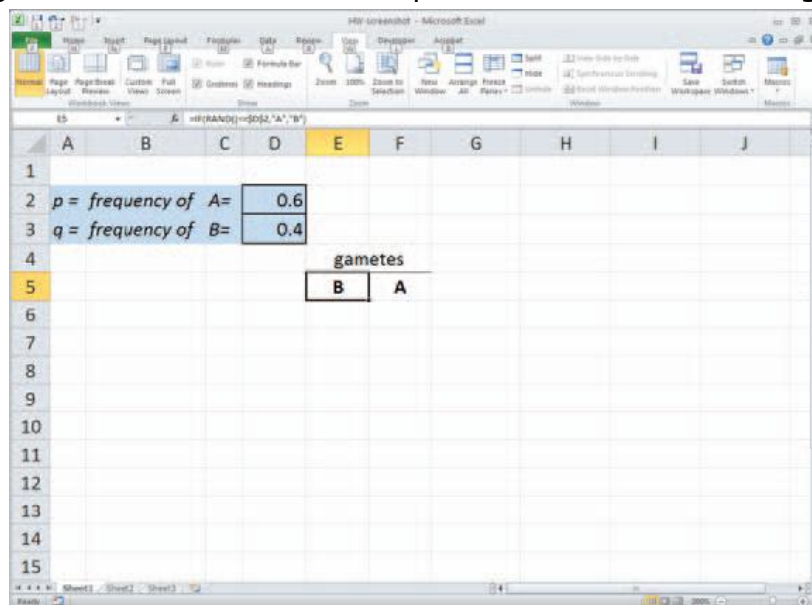


Figure 3

6. Recalculating by hitting F9 20 times. After each recalculation, observe the letters in cells E5 and F5. Are your results consistent with what you expect? Try changing your p value to 0.8 or 0.9. Does the spreadsheet still work as expected? Try lower p values. If you don't get approximately the expected numbers, check and recheck your formulas now, while it is early in the process.

You could stop here and just have the computer recalculate over and over — similar to tossing a coin. However, with just a few more steps, you can have a model that will create a small number or large number of gametes for the next generation, count the different genotypes of the zygotes, and then be able to graph the results.

7. Copy these two formulas in E5 and F5 down for 20 rows (down to E24 and F24) to represent gametes that will form 16 offspring for the next generation, as in Figure 4. (To copy the formulas, click on the bottom right-hand corner of the cell and, with your finger pressed down on the mouse, drag the cell downward. This is just one of many ways to copy cells – feel free to use whichever you are most comfortable with.)

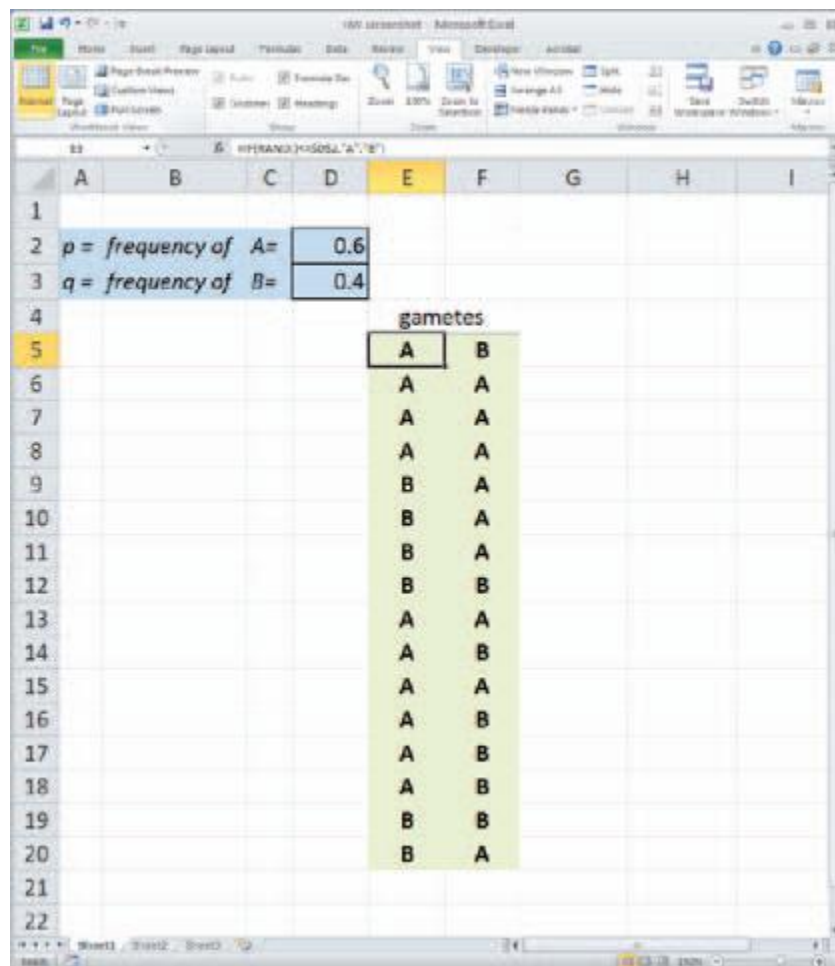


Figure 4

8. Next, we will combine the two gametes in cells E5 and F5 and put the zygote's genotype in cell G5. The zygote is a combination of the two randomly selected gametes. In spreadsheet vernacular, you want to concatenate the values in the two cells. In cell G5 enter the function

$$=CONCATENATE(E5,F5)$$

and then copy this formula down for the 20 rows that you have gametes, as in **Figure 5** below. Concatenate simply means "to put two values together" in Excel.

| | A | B | C | D | E | F | G | H | I | J |
|----|---|---------------------|---|---|---|---------|--------|---|---|---|
| 1 | | | | | | | | | | |
| 2 | | p = frequency of A= | | | | 0.6 | | | | |
| 3 | | q = frequency of B= | | | | 0.4 | | | | |
| 4 | | | | | | gametes | zygote | | | |
| 5 | | | | | B | A | BA | | | |
| 6 | | | | | B | A | BA | | | |
| 7 | | | | | B | B | BB | | | |
| 8 | | | | | A | B | AB | | | |
| 9 | | | | | A | B | AB | | | |
| 10 | | | | | A | A | AA | | | |
| 11 | | | | | A | A | AA | | | |
| 12 | | | | | A | A | AA | | | |
| 13 | | | | | A | B | AB | | | |
| 14 | | | | | A | A | AA | | | |
| 15 | | | | | A | B | AB | | | |
| 16 | | | | | A | A | AA | | | |
| 17 | | | | | A | B | AB | | | |
| 18 | | | | | A | B | AB | | | |
| 19 | | | | | A | A | AA | | | |
| 20 | | | | | A | B | AB | | | |
| 21 | | | | | | | | | | |
| 22 | | | | | | | | | | |
| 23 | | | | | | | | | | |

Figure 5

9. The next columns on the sheet, H, I, and J, are used for bookkeeping — that is, keeping track of the numbers of each zygote's genotype. They are rather complex functions that use IF functions to help us count the different genotypes of the zygotes.

The function you will enter in cell H5 is

$$=IF(G5="AA",1,0)$$

which basically means that if the value in cell G5 is AA, then put a 1 in this cell; if not, then put a 0.

10. Enter the following very similar function in cell J5:

$$=IF(G5="BB",1,0)$$

Your spreadsheet now should resemble **Figure 6** below.

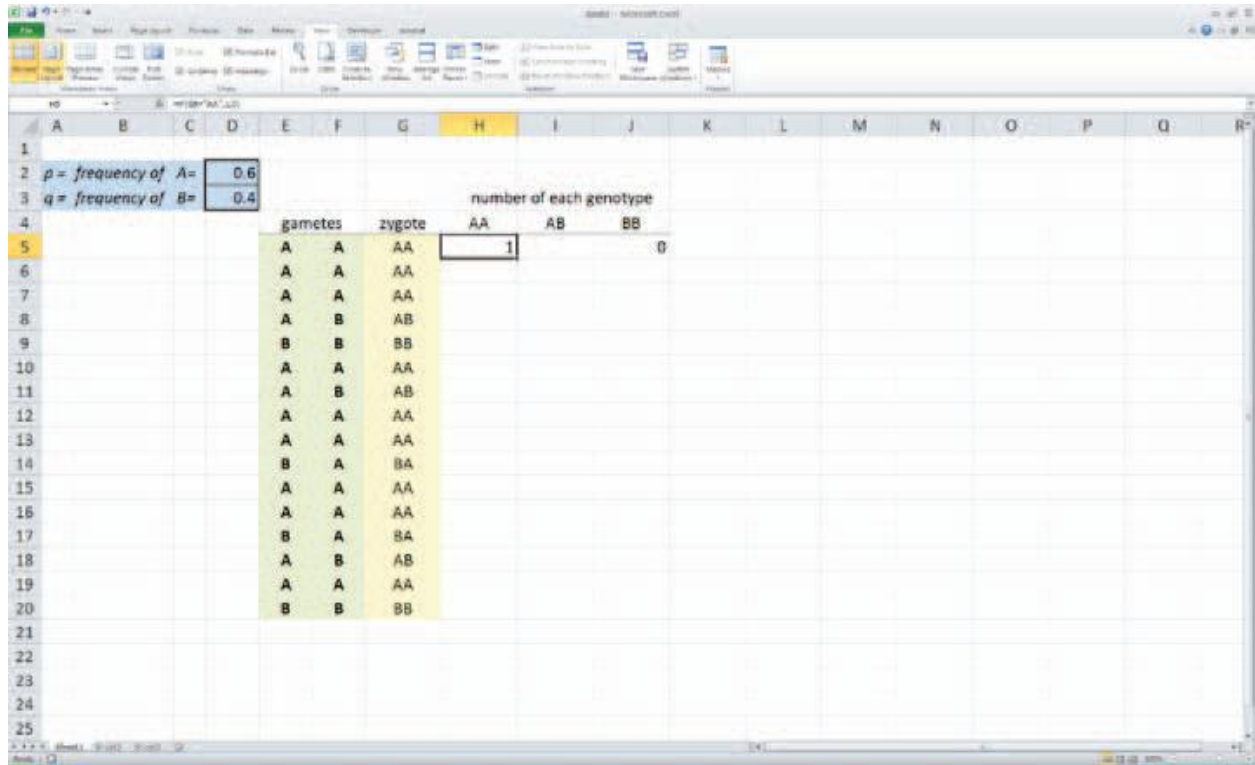


Figure 6

11. Now let's tackle the nested IF function. This is needed to test for either AB or BA . In cell I5, enter the nested function:

$$=IF(G5="AB",1,(IF(G5="BA",1,0)))$$

This example requires an extra set of parentheses, which is necessary to nest functions.

This function basically says that if the value in cell G5 is exactly equal to AB , then put a 1; if not, then if the value in cell G5 is exactly BA , put a 1; if it is neither, then put a 0 in this cell. Copy these three formulas down for all the rows in which you have produced gametes.

12. Copy the cells H5 through J5 for the 20 zygotes that already have.

13. Enter the labels for the columns you've been working on — first merge cells E4 and F4, and then label that merged cell *gametes*, label *zygote* in cell G5, AA in cell H4, AB/BA in cell I4, and BB in cell J4, as shown in **Figure 7** on the next page.

| | gametes | | zygote | number of each genotype | | |
|---|---------|---|--------|-------------------------|----|----|
| | A | B | | AA | AB | BB |
| A | A | A | AA | 1 | 0 | 0 |
| B | B | A | BA | 0 | 1 | 0 |
| A | A | B | AB | 0 | 1 | 0 |
| A | A | A | AA | 1 | 0 | 0 |
| B | B | A | BA | 0 | 1 | 0 |
| B | B | B | BB | 0 | 0 | 1 |
| A | A | A | AA | 1 | 0 | 0 |
| B | B | A | BA | 0 | 1 | 0 |
| B | B | B | BB | 0 | 0 | 1 |
| A | A | A | AA | 1 | 0 | 0 |
| A | A | A | AA | 1 | 0 | 0 |
| A | A | A | AA | 1 | 0 | 0 |
| B | B | B | BB | 0 | 0 | 1 |
| A | A | A | AA | 1 | 0 | 0 |
| A | A | B | AB | 0 | 1 | 0 |
| B | B | A | BA | 0 | 1 | 0 |

Figure 7

14. As before, try recalculating (F9) a number of times to make sure everything is working as it should. What is expected? If you aren't sure yet, keep this question in mind as you complete the sheet. Use a p value of 0.5, and you'd see numbers similar to the ratios you would get from flipping two coins at once. Don't go on until you are sure the spreadsheet is making correct calculations.
15. Go to cell G25 and format the text to align right. Label this cell "Number of Each Genotype".
16. Use the SUM function to calculate the numbers of each genotype in columns H, I, and J. Remember, there are multiple ways to do this. You can type in =sum and then highlight the boxes, go to the Insert>Equations tab, or you can type in the following:

$$=SUM(H5:H24)$$
17. Look at cells H25 through J25. These numbers are the genotypes of the next generation. We will now show this graphically as well.
18. Add a bar graph to the right of your data. Follow the basic flow here – remember, you can always adjust the formatting, axis range, etc. Go to the Insert tab, then for charts, choose Column > the first 2D.
19. Position this new 'window' to the right of your data. You can resize it accordingly.
20. Right click in this 'window' and choose "Select Data..."

21. A new window will pop up. "Chart data range:" will be highlighted. Choose the cells H25 through J25.
22. Right click on the on the newly formed chart. Again, choose "Select Data..." Now, to the right where it says Horizontal (Category) Axis Label, click the "Edit" button. Select the cells H4 through J4. This will label each bar with its respective genotype.
23. Then select and delete the Chart Legend that says "series 1".
24. Now we will give the chart a title and label our axes.
 - a) Go to the "Chart Tools" set of tabs. And choose "Layout".
 - b) Click on the "Chart Title" option and select one. Label the chart "Distribution of Next Generation Genotypes"
 - c) Click on the Axis Titles and for both the horizontal (X) and vertical (Y) axis; add an appropriate title.
 - d) Format each of the newly added titles (as well as the axis labels) to make them easier to read. 20pt font should be easier to read.
 - e) You can change the appearance of this chart in many ways to make it easier to view by manipulating the size of the window, axis range, axis spacing, color of the bars, etc. Your screen should look somewhat like **Figure 8**.

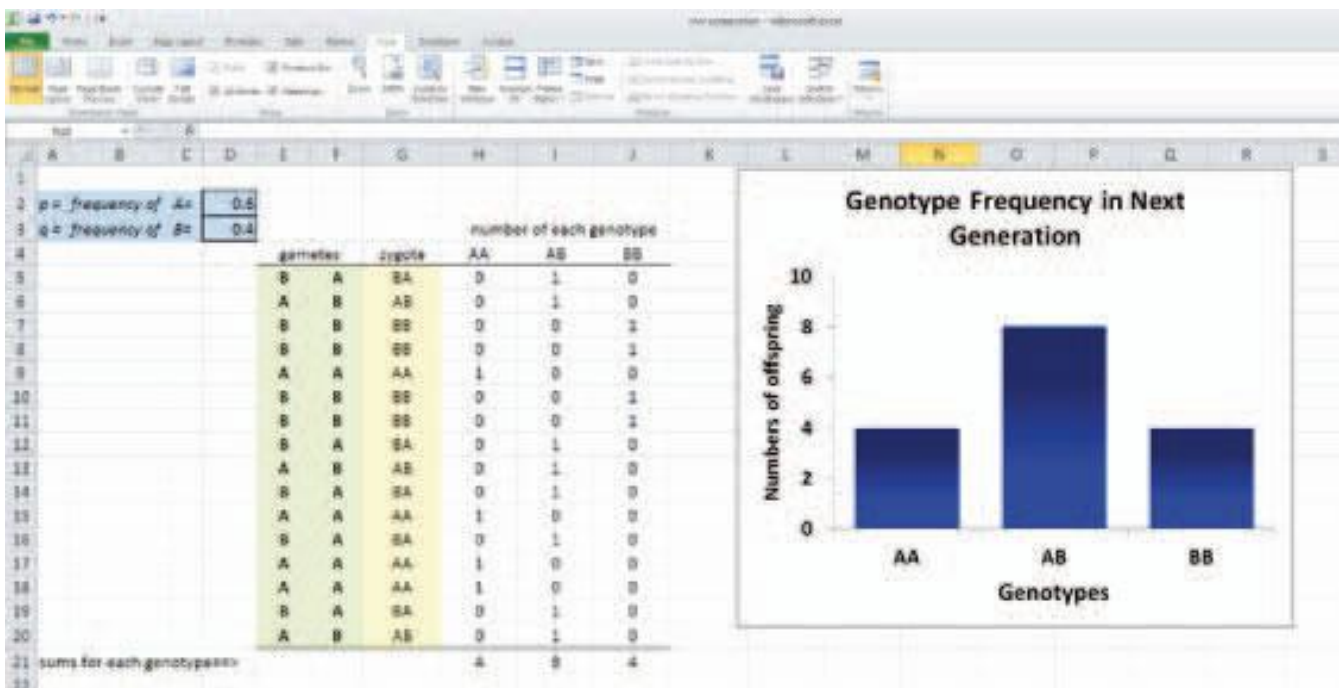


Figure 8

25. The next steps will give us an easy way to see what we would expect the allelic/genotypic frequencies of the next generation to be. Furthermore, we will then set up the spreadsheet to compare our expected frequencies to what actually occurred.

26. Merge cells B6, C6, and D6. Label this newly merged cell "If H-W Equilibrium".

27. Label B7: " p^2 ", C7: " $2pq$ ", and D7: " q^2 ".

28. In the cells B8 through D8, we will have the spreadsheet calculate what the expected genotypic frequencies would be in the next generation based on our starting allelic frequencies **IF** the Hardy-Weinberg equilibrium conditions are met. REMEMBER: If the Hardy-Weinberg conditions are met, the *allelic frequencies* (**p** and **q**) should be the same! Enter the following formulas in their respective cells (if you cannot translate these into 'English', please let me know):

$$B8: =D2*D2$$

$$C8: =2*D2*D3$$

$$D8: =D3*D3$$

29. Highlight cells C12 down to C21. Format the text to 'align right' in the cell. Enter the following information into each cell:

C12: Total Alleles =

C13: Number of p alleles =

C14: Number of q alleles =

C16: NEW Frequency of p =

C17: NEW Frequency of q =

C19: NEW Frequency of p^2 =

C20: NEW Frequency of $2pq$ =

C21: NEW Frequency of q^2 =

30. Finally, we will enter the formulas to calculate the above listed values from our data (again, if you cannot translate these into 'English', please let me know).

$$D12: =SUM(H25:J25)*2$$

$$D13: =(H25*2)+(I25)$$

$$D14: =(J25*2)+(I25)$$

$$D16: =D13/D12$$

$$D17: =D14/D12$$

$$D19: =D16*D16$$

$$D20: =2*D16*D17$$

$$D21: =D17*D17$$

Quantitatively Describing the Biological System

This mathematical model can predict allele frequencies from generation to generation. In fact, it is a *null* model. That is, in the absence of random events or other real-life factors that affect populations, the allele frequencies do not change from generation to generation. This is known as the Hardy-Weinberg equilibrium (H-W equilibrium). The H-W equilibrium is a valuable tool for population biologists because it serves as a baseline to measure changes in allele frequencies in a population. If a population is not in H-W equilibrium, then something else is happening that is making the allele frequencies change.

At this point, we have a new population of 20 individuals from an original population with given starting allelic frequencies (the values in D2 and E2). Run this simulation (recalculate – F9) 5 times with the existed spreadsheet population of 20 *for each* starting allelic frequency. Record your data in **Table 1**.

Table 1: Comparing Frequencies of p (allele A) with a Population of 20.

| Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p |
|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |

Highlight cells E9:J24. Now delete these values. This will give you a population of just 4 individuals. Rerun (F9) this simulation an additional 5 times for each starting allelic frequency and record this data in **Table 2**.

Table 2: Comparing Frequencies of p (allele A) with a Population of 4.

| Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p |
|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |

We will now attempt this simulation with a population of 10000. Highlight cells E25 to J25. Cut and paste these "SUM" cells to N38 to S38 (or somewhere just below your chart). Now highlight cells E5 to J5. Click and hold the small black box at the bottom right of these highlighted cells. Drag all the way down to row 10004. This will give us a population of 10000. You will now have to adjust your "SUM" cells that you moved to read:

$$=SUM(H5:H10004)$$

$$=SUM(I5:I10004)$$

$$=SUM(J5:J10004)$$

Your chart should automatically update the data. Rerun (F9) this simulation an additional 5 times for each starting allelic frequency and record this data in **Table 3**.

Table 3: Comparing Frequencies of p (allele A) with a Population of 10000.

| Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p | Starting Frequency of p | New Frequency of p |
|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|---------------------------|----------------------|
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |
| 0.1 | | 0.3 | | 0.5 | | 0.7 | | 0.9 | |

Carry out the statistical analysis for the three varied populations below. Determine the range, mean, and standard deviation for each population used. Then answer the question on the next page.

Table 4: Statistical Analysis of p Frequencies (allele A) with Varied Population.

| Population of 4 | | | | Population of 20 | | | | Population of 10000 | | | |
|---------------------------|--------------------------------------|-------------------------------------|---|---------------------------|--------------------------------------|-------------------------------------|---|---------------------------|--------------------------------------|-------------------------------------|---|
| Starting Frequency of p | <u>RANGE</u> New Frequency of p | <u>MEAN</u> New Frequency of p | <u>STANDARD DEVIATION</u> New Frequency of p | Starting Frequency of p | <u>RANGE</u> New Frequency of p | <u>MEAN</u> New Frequency of p | <u>STANDARD DEVIATION</u> New Frequency of p | Starting Frequency of p | <u>RANGE</u> New Frequency of p | <u>MEAN</u> New Frequency of p | <u>STANDARD DEVIATION</u> New Frequency of p |
| 0.1 | | | | 0.1 | | | | 0.1 | | | |
| 0.3 | | | | 0.3 | | | | 0.3 | | | |
| 0.5 | | | | 0.5 | | | | 0.5 | | | |
| 0.7 | | | | 0.7 | | | | 0.7 | | | |
| 0.9 | | | | 0.9 | | | | 0.9 | | | |

Quantitatively Describing Generational Studies

The next part of this simulation will deal with multiple generation simulations. We will start with the population of 10000 first as your spreadsheet should already be configured for this large sample size. Reset cell D2 (frequency of p) to 0.5. After you hit return, the simulation will have run and the value in D16 (new frequency of p) will have changed. Record this in the "Generation 2 row of $n=10000$ " place in **Table 5**.

Reseed the next generation with the new value by changing the value of D2 with what is in D16. (NOTE: cell D3, the frequency of q , should also change so that $D2+D3=1$. If this is not the case, check your cell functions!) Repeat this step for a total of 10 generations.

Highlight cells E105:J10004. Delete the contents. This will give you a starting population of 100. Reset cell D2 to 0.5 and repeat the same procedure as listed above for 10 generations.

Finally, highlight cells E15:J104. Delete the contents. This will give you a starting population of 10. Reset cell D2 to 0.5 and repeat the above procedure for 10 generations. Answer the question on the following page.

Table 5: Comparing Generational Frequencies of p (allele A) with Varied Populations.

| Generation Number | n = 10000 Starting Frequency of p | n = 100 Starting Frequency of p | n = 10 Starting Frequency of p |
|-------------------|---|---|--|
| 1 | 0.5 | 0.5 | 0.5 |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| 7 | | | |
| 8 | | | |
| 9 | | | |
| 10 | | | |

Other Population Genetics Simulations

There are many decent simulations that can be found online. One of them (that I like) can be found here: http://www.radford.edu/~rsheehy/Gen_flash/popgen/ Open your browser (I personally like Chrome) and type the above in the address window. (For some reason, you might have to "de-maximize" the screen size, so you can see all the buttons – I don't know why this is the case – most likely it was a set window size configured with Flash.)

At the lower right of this website is a green circle with a red question mark. Click on this and read through the instructions. I will not be redundant and recopy them here. But I will clarify a few terms/items/variances that he uses.

- A_1 will be the same as p . A_2 will be the same as q .
- When more than one population is simulated the upper graph depicts the frequency of the A_1 allele (p) and the lower graph mirrors the upper graph with the frequency of the A_2 allele (q).
- When a single population is simulated the upper graph depicts the frequencies of both the A_1 and A_2 alleles while the lower graph depicts the genotype frequencies (A_1A_1 , A_1A_2 , and A_2A_2). This will be useful to see if there are any 'favorable' genotypes of the three.
- Stochastic = finite. The top right-box can be clicked. If the checkmark shows, then it is an infinite population. When you use this option, you can start A_1 (p) at different levels for up to 5 different populations. When this box is unchecked, you can modify the population size (once you go past 250, it gets really slow) as well as starting A_1 (p) frequency.
- For Selection/Fitness, I wouldn't be overly concerned with the formulas provided (unless you want to tinker with our Excel model ☺). But you can modify the survivability for the different genotypes in different populations – remember, not every population lives in the same ecological environment!
- Migration, Mutation, and Bottlenecks are fun to play with. Experiment with modifying these variables to see their result on the populations.

For this part of the lab, I would like you to choose two of the following conditions that you can manipulate: Fitness, Migration, Mutation, and Bottleneck Effect. For each, come up with an hypothesis of what you think will happen based on the change you will make **and explain why you think this will happen!** Be sure to do this part first, before the simulation actually runs. Then, run the simulation and explain the data. You can take a screenshot of the simulation (using the Print Screen button – then pasting into a photo editor like Photoshop or Fireworks to modify – or just paste into a Word document) to help explain your outcome.