

17. MATHEMATICAL MODELING OF EVOLUTION

Background

Evolution in a biological context refers specifically to changes in the genetic makeup of populations over time. Evolution involves variation in the population, heredity, and differential survival. One way to study if a population is evolving is to monitor the frequencies of alleles in a population over time, from generation to generation.

This leads to the question, “What are the inheritance patterns of alleles, not just from two parental organisms but also in a population?” and to the exploration of how allele frequencies change in populations and how these changes can be predicted in a population for future generations.

Because biological systems are complex and very difficult to study, mathematical models and computer simulations are perfect tools for exploring changes in the genetic makeup of a population from generation to generation. In this investigation, a spreadsheet programmed to model changes of a hypothetical gene pool from one generation to the next will be used to explore parameters that affect allele frequencies, such as selection, mutation, and migration.

Driving Question

How can mathematical models be used to investigate factors that affect changes in allele frequencies within a population? How do these factors contribute to evolutionary change in populations?

Materials and Equipment

Use the following materials to complete the initial investigation. For conducting an experiment of your own design, check with your teacher to see what materials and equipment are available.

- Computer
- Spreadsheet program (such as Microsoft Excel®, Numbers¹, or Google Sheets^{TM2})
- Mathematical model spreadsheet file:
ABI Mathematical Modeling Spreadsheet.xlsx

¹ Numbers is a trademark of Apple Inc., registered in the U.S. and other countries.

² © 2012 Google Inc. All rights reserved. Google Sheets is a trademark of Google Inc.

Initial Investigation

Complete the following investigation before designing and conducting your own experiment. Record all observations, data, explanations, and answers in your lab notebook.

The spreadsheet for modeling evolution has four sheets: Small Population, Large Population, Selection, and Multiple Generations. This investigation starts on the “Small Population” sheet, which shows the initial gene frequencies in the population of 500 individuals. The alleles are represented by “A” and “B.”

The Small Population model makes several assumptions, including:

- No mutations
 - No selection
 - No migration
 - Sexual reproduction and alleles of each gamete are randomly selected
 - Two-allele, simple-dominance pattern of inheritance
 - All gametes form a viable zygote that lives to become an adult
 - Population size remains constant
1. Open the mathematical model spreadsheet file. Save the file locally with a different filename. Select the Small Population sheet (click on the tab at the bottom of the screen).
 2. Enter some different frequencies for the A allele in cell E2.

	A	B	C	D	E	F
1						
2		Frequency of A allele			0.20	(p)
3		Frequency of B allele			0.80	(q)

- a. When the frequency of the A allele changes, what happens in cell E3? How does the spreadsheet determine the value for E3?
NOTE: To see the formula for any part of the spreadsheet, simply select the cell of interest and look in the formula bar (f_x).
 - b. If a number greater than 1.00 is entered in cell E2, the message box labeled “Allele Frequency Invalid” appears. Use the Hardy–Weinberg equation to explain why the model requires a value less than or equal to 1.00 entered in cell E2.
3. In the model population, a gamete can contain either an A allele or a B allele. Which of the two alleles is in any given gamete is random, so columns B and C have been set up using the spreadsheet’s RANDOM function; this function generates a random number between 0 and 1.

Observe how the RANDOM function works by entering the formula “=RAND()” into cell G2. Press the “F9” key on a Windows® computer or the “command” and “=” keys simultaneously on a Mac® computer to randomly generate a new number in this cell. Repeat this command three or more times.

4. Notice that the command not only generated a new value in G2, but data in other parts of the spreadsheet also changed. This is because some cells contain formulas that involve the RAND() function. For example, this function is part of the formula the spreadsheet uses to randomly determine if an A or a B allele is in a gamete. In this case, the function is not dependent on the F9 or “command =” commands, though. Change the frequency of the A allele in cell E2 and press the *enter* (or *return*) key; the gamete columns change, as does the summary chart, and a new value appears in G2.

NOTE: Delete the RANDOM function formula from cell G2 before continuing.

5. A formula using the RAND() function has been set up in cells B6 to B55 and C6 to C55 as follows:

$$=IF(RAND()<=E\$2, "A", "B")$$

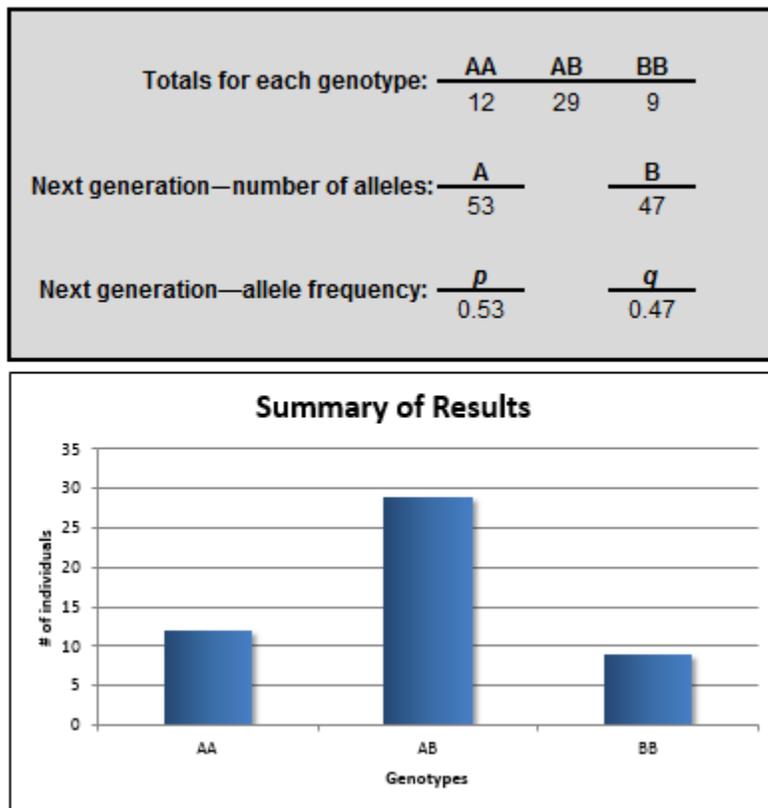
Click on cell B6 and look for the equation in the f_x window. This equation directs the spreadsheet program to first generate a random number between 0 and 1. Each time a random number is generated, the IF statement checks to see if it is less than or equal to the frequency of the A allele. If so, an “A” gamete is produced, whereas a random number that is greater than the frequency of the allele causes a “B” gamete to be produced.

In the 50 rows of the table, gametes are randomly selected to be “A” or “B” and 50 zygotes are generated from the combination of these gametes.

5	Gametes		Zygote	AA	AB	BB
6	B	B	BB	0	0	1
7	B	B	BB	0	0	1
8	A	B	AB	0	1	0

6. If the frequency of the A allele is 0.50 and the RAND() function generates a value of 0.49, the gamete will be assigned an A allele. Why does the model use the current allele frequency to determine which allele is assigned to a gamete?
7. Is the spreadsheet model an accurate representation of reproduction in real populations? In other words, does the random selection of alleles and random combining of gametes occur in real populations? Explain your answer.

Another part of the spreadsheet is the results summary—a summary of the population's gene pool. Formulas in cells N5, O5, and P5 determine the number of zygotes of each genotype and this information is used to determine the number and frequency of A and B alleles in the new generation. The bar graph provides a visual summary of the genetic makeup of the population.



8. Click on cell N8.
- What formula is the spreadsheet program using to calculate the value for this cell?
 - Explain what this formula is doing and explain the reason for using this formula to determine the number of A alleles.
9. Set the starting A allele frequency in the model to 0.50. Create a data table to record the frequencies of the A and B alleles (p and q) for 5 generations, as instructed below.
- Record in the table the initial frequency of each allele and the next generation allele frequencies from the summary chart.
NOTE: The next generation was automatically calculated when you pressed enter/return after setting the initial p frequency to 0.50.
 - Enter the value from N11 (the frequency of p in the next generation) into E2. Record the A allele frequency of the 2nd generation into the data table. Recalculate the model until you have created five generations. Do not forget to enter the new allele frequency in E2 *each time*.
10. How did the allele frequencies change in the population over five generations? Are changes in allele frequency evidence that the population is evolving? Explain your answer.
11. Repeat the process to produce four more data sets showing the change in allele frequency over five generations. For each data set, start with initial frequencies of 0.50 as you did before. (Create additional data tables for the new data sets.)

12. Are the results the same in all five data sets? If yes, why do you think this is the case? If no, what might be a reason for the differences? Do you detect any trends?
13. Repeat the above test showing the change of allele frequency, but set the initial p frequency to 0.90.
14. After reviewing your results from each scenario, what can you conclude about the effect of the initial allele frequency on the change in frequency over five generations?
15. Do the changes in frequencies generation-to-generation progress in one direction? Why might a high A allele frequency in a small population be detrimental to the survival of that population?
16. Move to the next tab in the spreadsheet, Large Population. The Large Population sheet is identical to the Small Population sheet except the population size has been increased to 5,000 individuals.

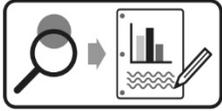
A	A	AA	1	0	0
A	A	AA	1	0	0

Set the starting frequency of the A allele to 0.90 and create three data sets of five generations each, as you did before. Create data tables to record the results.

17. Did the allele frequencies change similarly in the small and large populations over five generations? How do you explain the difference (if any) between the populations?
18. Move to the Selection sheet in the spreadsheet. On this sheet, the survival rate of each genotype can be manipulated and the gene frequencies are shown for a population of 500 individuals. Set the starting frequency to 0.50 for the A allele and set the survival rate of the BB genotype to 0%. The survival rate of AA and AB should be set at 100%.
- Run the model for five generations, again setting the allele frequency in E2 after each run. Record the new allele frequency of each generation in a table.
19. In some populations there is a *heterozygote advantage*. Change the spreadsheet in a way you think models heterozygote advantage. Describe what you change and record the results in a table for five generations. Record both the allele frequencies for each generation and the number of individuals with each genotype.
20. How did the selection against the BB genotype affect the allele frequencies over five generations? How do these results compare to those of the model reflecting heterozygote advantage?
21. Which of the four situations, if any, represented a population in Hardy–Weinberg equilibrium? Justify your answer with evidence. For any population that was not in equilibrium, identify the factor or factors that disrupted the equilibrium.
22. Move to the final sheet in the spreadsheet, Multiple Generations, which models a population of 500 organisms for 20 generations. In this sheet, the allele frequencies are shown in the table and on the graph. Set the initial frequency to 0.80 for the A allele and observe the population changes for 20 generations. Instead of creating a data table, graph your results.
23. Compare your results to those of several other classmates or groups. How do they differ? Do differences in the results indicate a flaw in the model? Explain your reasoning.
24. Computer modeling is a powerful tool for biologists but, as with all models, it has limitations. For the spreadsheet models you used to investigate population genetics and evolution, identify two useful aspects of computer modeling and two limitations.

Design and Conduct an Experiment

The Hardy–Weinberg equilibrium model states that both allele and genotype frequencies in a population remain constant from generation to generation unless specific disturbing influences are introduced into the population. Use the mathematical model provided for this investigation to simulate a population over multiple generations where a disturbance is introduced into the population.



Design and carry out your experiment using either the Design and Conduct an Experiment Worksheet or the Experiment Design Plan. Then complete the Data Analysis and Synthesis Questions.

Design and Conduct an Experiment: Data Analysis

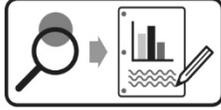
1. Based on your model and data, describe how the independent variable in the models affected allele frequencies in the population.
2. Is there any evidence in your data or from your observations that experimental error or other uncontrolled variables affected your results? If yes, is the data reliable enough to determine if your hypothesis was supported?

Synthesis Questions

1. A great deal of genetic variation is found within healthy populations. This goes against the logic that natural selection works toward genetic uniformity in a population—the most fit genotypes produce the most offspring, thereby increasing the frequency of some alleles. Describe factors that could cause this genetic variation.
2. The assumptions of the Hardy–Weinberg equilibrium are stringent and are seldom observed or never completely met in real populations, so why do genotype frequencies of many populations not deviate significantly from Hardy–Weinberg expectations?
3. Changes in allele frequencies lead to an evolving population. In a population of 200 peppered moths where white coloration is dominant over dark coloration, there are 32 dark peppered moths.
 - a. What is the genotype breakdown of the population for the peppered moths?
 - b. Environments by nature are consistently changing. Identify a particular change and describe how it would affect the allele frequency of the peppered moth. Explain how the Hardy–Weinberg principle would be affected.

Design and Conduct an Experiment Worksheet

The Hardy–Weinberg equilibrium model states that both allele and genotype frequencies in a population remain constant from generation to generation unless specific disturbing influences are introduced into the population. Use the mathematical model provided for this investigation to simulate a population over multiple generations where a disturbance is introduced into the population.



Develop and conduct your experiment using the following guide.

1. For a population in Hardy–Weinberg equilibrium, what five factors or conditions must be taking place in order to maintain the equilibrium of the population?

2. Create a driving question: choose one of the factors you've identified that can be controlled in the lab and develop a testable question for your experiment.

3. What is the justification for your question? That is, why is it biologically significant, relevant, or interesting?

4. What will be the independent variable of the experiment? Describe how this variable will be manipulated in your experiment.

5. What is the dependent variable of the experiment? Describe how the data will be collected and processed in the experiment.

6. Write a testable hypothesis (If...then...).

7. What conditions will need to be held constant in the experiment? Quantify these values where possible.

8. How many trials will be run for each experimental group? Justify your choice.

9. What will you compare or calculate? What analysis will you perform to evaluate your results and hypothesis?

10. Describe at least 3 potential sources of error that could affect the accuracy or reliability of data.

11. Use the space below to create an outline of the experiment. In your lab notebook, write the steps for the procedure of the lab. (Another student or group should be able to obtain similar results using this procedure.)

12. Have your teacher approve your answers to these questions and your plan before beginning the experiment.

