Big_{Idea}

INVESTIGATION 2 MATHEMATICAL MODELING: HARDY-WEINBERG*

How can mathematical models be used to investigate the relationship between allele frequencies in populations of organisms and evolutionary change?

BACKGROUND

"Mathematics is biology's next microscope, only better ..." (Cohen 2004)

It is not hard to understand the value of microscope technology to biology and how this technology opened up entire new worlds of biological understanding. However, for some, it is not as easy to see the value of mathematics to the study of biology. Like the microscope, math and computers provide tools to explore the complexity of biology and biological systems, offering deeper insights to and understanding of what makes living systems work. Even the incredible complexity of evolution in populations is illuminated by relatively simple mathematical equations, several of which are based on the Hardy-Weinberg (H-W) equilibrium. Students (and their teachers) have traditionally found the topic of population genetics in an introductory biology class to be challenging, due in part to the fact that for the last couple of generations biology has been thought of as the science with only a minimal mathematics foundation — particularly in comparison to chemistry or physics. Modern biology, however, is vastly different.

One of the specific difficulties of the H-W null hypothesis is that it is the null hypothesis — it is what would happen to allele frequencies in the absence of any evolutionary parameter. This is counterintuitive for most students. H-W is the standard by which evolution can be measured. To that end, most simulations that try to create a population manipulated by students to model H-W are flawed from the beginning. Student classroom populations, by definition, are so small that genetic drift will swamp any other factors that the simulation is trying to model. By starting with a population that is modeled on a computer spreadsheet with explicit randomness, students are able to build their knowledge toward an inference on allele inheritance patterns in a theoretically infinite population. They do this by creating larger and larger populations to minimize fluctuations from the expected probabilities.

Having students build their own models (even if extensively guided by the teacher) is a requisite component to this investigation. It is during the model building, testing, and corrective phase of construction that the reflective and analytical skills of model building are learned. These skills, by their very nature, have a broad application to learning.

^{*} Transitioned from the *AP Biology Lab Manual* (2001)

In this investigation, the students will build a spreadsheet that models how a hypothetical gene pool changes from one generation to the next. This model will allow for the exploration of parameters that affect allele frequencies, such as selection, mutation, and migration.

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

This investigation also provides an opportunity for students to review concepts they 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.

Without an opportunity to build these skills, students, when faced with more sophisticated, prebuilt models, tend to click randomly, with no particular plan in mind, rather than use a systematic exploration of the model's parameters and output. Likewise, without taking the time to develop modeling skills, students cannot fully appreciate the limitations and the strengths of modeling natural phenomena.

PREPARATION

Materials and Equipment

- Laboratory notebooks
- Miniposters and miniposter supplies
- Computers with spreadsheet software like Microsoft[®] Excel, Macintosh[®] Numbers, or OpenOffice Calc (An alternative to the installed spreadsheet software is an online spreadsheet, such as Google Docs Spreadsheet.)

http://docs.google.com or Zoho Sheet *http://www.zoho.com*. In addition, you'll need to download a more sophisticated simulation, such as the ESTEEM module: Deme 1.0 or 2.0 at *http://bioquest.org/esteem/ esteem_result.php* [You'll need to register for the site, but it is free.] or Jon Herron's Allelle1 at *http://faculty.washington.edu/ herronjc/SoftwareFolder/AlleleA1. html.* An alternative is an online Web browser-based model at *http://www.radford.edu/~rsheehy/ Gen_flash/popgen/.*)

Timing and Length of Lab

Generally, the model building part of this investigation can be accomplished in two class periods (i.e., depending on school setting, class periods could be 45, 50, or 60 minutes in duration), with at least one additional class period required for model exploration. After completing the model exploration, the class should consider exploring applications of the null Hardy-Weinberg model in the real world by following up with the supertasters lab suggested in the section Where Can Students Go from Here?

Safety and Housekeeping

Typical laboratory safety concerns do not apply to computer modeling. However, there are some important things to consider when managing such work in the classroom. To avoid frustration, develop a system whereby the students periodically save their work. When developing and working out models, it is often a good idea to save each new version of the model with a different file name. That way, if a particular strategy doesn't work, a student will not necessarily have to start over completely but can bring up a file that had the beginnings of a working model.

Encourage idea sharing but not necessarily file swapping. Students should generate their own work to get the maximal benefit from this exercise.

Finally, your students may be able to produce a rudimentary spreadsheet but have difficulty refining the sheet. Consider having these students rely on a hybrid approach — use the spreadsheet to generate the random samples, but use reliable pencil and paper to archive and graph the results.

The other issue to consider is the ease of digital reproduction and sharing. Without a doubt, the students will be able find a Hardy-Weinberg spreadsheet model on the Internet. Likewise, it will be easy for them to share with one another. Keep this in mind as you make the assignment, and consider having most of the work done in class if you have access to computers.

ALIGNMENT TO THE AP BIOLOGY CURRICULUM FRAMEWORK

The investigation will fit best with content from the process of evolution (big idea 1), but you'll find that this lab will also fit nicely in genetics and information transfer (big idea 3), particularly after covering individual Mendelian inheritance patterns. As always, it is important to make connections between big ideas and enduring understandings, regardless of where in the curriculum the lab is taught. The concepts align with the enduring understandings and learning objectives from the AP Biology Curriculum Framework, as indicated below.

Enduring Understandings

- 1A1: Natural selection is a major mechanism of evolution.
- 1A2: Natural selection acts on phenotypic variations in populations.
- 1A3: Evolutionary change is also driven by random processes.
- 1C3: Populations of organisms continue to evolve.

Learning Objectives

- The student is able to convert a data set from a table of numbers that reflect 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 (1A1 & SP 1.5, SP 2.2).
- The student is able to apply mathematical methods to data from a real or simulated population to predict what will happen to the population in the future (1A1 & SP 2.2).
- The student is able to evaluate data-based evidence that describes evolutionary changes in the genetic makeup of a population over time (1A1 & SP 5.3).
- The student is able to use data from mathematical models based on the Hardy-Weinberg equilibrium to analyze genetic drift and effects of selection in the evolution of specific populations (1A3 & SP 1.4, SP 2.1).
- The student is able 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 (1A3 & SP 2.1).
- The student is able to describe a model that represents evolution within a population (1C3 & SP 1.2).
- The student is able to evaluate given data sets that illustrate evolution as an ongoing process (1C3 & SP 5.3).

ARE STUDENTS READY TO COMPLETE A SUCCESSFUL INQUIRY-BASED, STUDENT-DIRECTED INVESTIGATION?

Students will be challenged to think about an idealized life cycle of a population of hypothetical organisms that serves as the basis for the modeling exercise. You might wish to review life cycles, even though this topic is generally taught in elementary grades. It is not unusual for students to have an incomplete understanding of an organism's life cycle.

Try working through an idealized life cycle and population as a class before the students view the investigation. Specifically, use the modeling guidelines and assumptions described in the student investigation as a road map for your instruction as you and your students try to conceptualize this idealized population. Use questions such as the following:

- Can you describe the life cycle of an organism?
- How does this life cycle work in a population?
- How would you track one trait through a population's life cycle?
- Can you describe an idealized life cycle that would work the best for keeping track of a genetic trait in a population?

Help the students arrive at the simplifying assumptions and modeling solutions.

Skills Development

From a biology standpoint, students should carry out this investigation after they have studied Mendelian genetics, perhaps just as they start studying evolution. They should have a good handle on the differences between genes and alleles. Traditional Hardy-Weinberg activities, such as those in Lab 8 of the *AP Biology Lab Manual* (2001) and exercises like the "M&M" lab (*http://www.woodrow.org/teachers/bi/1994/hwintro.html*), provide a useful transition for students as they enter into computer-based models. Students should also be familiar with Excel and spreadsheet operations.

Potential Challenges

Students' and teachers' unfamiliarity with spreadsheet operations is a major limiting factor for this investigation. The degree of familiarity is difficult to measure. Generally, most students have been exposed to spreadsheets by the time they are in high school, but they typically have memorized the steps of their operations. Some of this will come back to them as they work on the spreadsheets, but unless they use spreadsheets often, most students will need quite a bit of help using the logic of IF statements.

There are many ways to create a spreadsheet model. For example, a simpler alternative to using the nested IF statements in the example model in the Student Manual is to use the function COUNTIF to count the AB string, plus COUNTIF again to count the BA string. While this may be more intuitive for some students, it does introduce a new, unique spreadsheet function. Encourage students to develop and find their own solutions to spreadsheet design.

One of the distinct advantages of using online spreadsheet applications is the ability to adjust the language of the spreadsheet to many different languages. The calculations are the same, but teachers and students can choose the language in which the instructions and navigation are displayed by changing the account settings.

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.

THE INVESTIGATIONS

Getting Started: Prelab Assessment

Investigating biology requires a variety of skills. The skills that are reinforced and introduced vary across the laboratories in this manual. The skills emphasized in a laboratory dictate whether a prelab assessment is appropriate.

This particular investigation provides a lab environment, guidance, and a problem designed to help students understand and develop the skill of modeling biological phenomena with computers. There are dozens of computer models already built and freely available. The idea of this laboratory is for students to experience the benefits of building their own model from scratch. To gain the maximum benefit from this exercise, students should not do too much background preparation. By building and exploring their own models, students should develop a more thorough understanding of how genes behave in populations.

Many classroom investigations of the Hardy-Weinberg equilibrium involve simulated populations, such as M&M candies or beans. Students can model gene inheritance in a simulated population and apply the H-W equation. While all models are simplifications and approximations, these particular models are generally based on such small populations that students often develop misconceptions regarding H-W equilibrium. However, for students who are simply stuck on how to get started, these pencil-and-paper simulations can provide good preparation for the computer modeling featured here.

Designing and Conducting Independent Investigations

The most important preparation teachers can do for this laboratory investigation is to work through the model building procedure themselves in their home or office. The author of this investigation has made this model dozens of times from scratch and usually learns something new, thinks of a different way to accomplish a task, or simply improves the model with every new build. Each time you work out this model with students you develop a richer repertoire of methods to apply to modeling. It is that rich environment that produces an authentic learning experience for students — a learning experience that transfers generally to a deeper understanding of the Hardy-Weinberg equilibrium and its application in population genetics. The following is a worked example of one version of the spreadsheet:

http://www.nabt.org/blog/2009/10/04/teaching-hardy-weinberg-andpopulation-genetics-using-spreadsheet-models-part-1/ There are two ways to carry out the Designing and Conducting Independent Investigations section.

- You can have the students build their models to explore other aspects of H-W equilibrium. This is suggested in the Student Manual. Have the students build models that explore selection and multigenerations, mutation, or migration. However, because students will have a limited skill set for working in the spreadsheet environment even at this point, this option might not be time effective.
- **2.** The alternative is to have students move from their spreadsheet models to more sophisticated models available on the Internet to answer their own questions; see the suggested online options that follow. In this option, while the students are not building the model, they are applying their knowledge of models to explore questions about population genetics.

Options for More Advanced Modeling Experiences

Have students generate their own questions regarding the evolution of allele frequencies in a population. Encourage them to experiment a bit with one of these more sophisticated and powerful models, trying out various combinations of changes to parameters or various extremes of one parameter at time. This is exploration, and it should allow the students to generate questions that have direct implications to the real world. Have them record their questions in their lab notebook. Encourage them to systematically investigate the consequences of changing variables (parameters) on the system they are studying — in this case, population genetics. From these questions have the students develop hypotheses that can be tested with the use of more sophisticated models — those that allow more easily manipulated parameters of population size, number of generations, selection (fitness), mutation, migration, and genetic drift.

Any number of computer-based Hardy-Weinberg simulations can be found with a Google search. Here are some suggestions:

- Deme 1.0 and 2.0: Another Excel model with more sophistication than the model you built in class, Deme 1.0 and 2.0 with documentation are available at *http://bioquest.org/esteem/esteem_details.php?product_id=193*, where you need to establish an account (free) before you can download it. It works in Excel just as the spreadsheet model you created earlier. See Figure 1.
- AlleleA1: Jon Herron from the University of Washington has created a simulation called AlleleA1 along with documentation. It is available for free at *http://faculty.washington.edu/herronjc/SoftwareFolder/AlleleA1.html*. See Figure 2.



Figure 1. Deme 1.0

Figure 2. AlleleA1

• Population genetics simulation program: Bob Sheely from Radford University has created a simulation and documentation in the form of a Web application. It is available for free at *http://www.radford.edu/~rsheehy/Gen_flash/popgen/*. See Figure 3.



Figure 3. Population Genetics Simulation Program

Summative Assessment

The ideal postlab assessment would be the application of model and theory to a realworld situation, such as that referenced in the section Where Can Students Go from Here? Be sure that all students/student teams treat their model spaces as experimental spaces. Have them propose, test, and analyze specific hypotheses and report these in a miniposter presentation that is designed to generate a rigorous peer review before the teacher evaluation.

Where Can Students Go from Here?

The ultimate laboratory learning experience would be to develop a model and then try out that model in the real world. An excellent extension to the modeling and theory work in this laboratory would be to have the students work through the following laboratory reported in the ABLE Proceedings: "Supertasters — Updating the Taste Test for the A & P Laboratory." There are few human traits that express the intermediate dominance necessary for testing for the null hypothesis. The supertaster trait described in this laboratory does express an intermediate phenotype; therefore, it creates an exemplary investigative population genetics laboratory.

SUPPLEMENTAL RESOURCES

- Cohen, 2004, "Mathematics is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better." PLoS Biol 2(12): e439. doi:10.1371/journal. pbio.0020439
- McMahon, K. A. 2008. "Supertasters Updating the Taste Test for the A & P Laboratory." In *Tested Studies for Laboratory Teaching*, Vol. 29, ed. K. L. Clase, 398– 405. Proceedings of the 29th Workshop/Conference of the Association for Biology Laboratory Education (ABLE).
- Otto, S. P. and T. Day (2007). A Biologist's Guide to Mathematical Modeling in Ecology and Evolution. Princeton University Press.

http://www.zoology.ubc.ca/biomath/