An Invitation to Modeling: Building a Community with Shared Explicit Practices

Kam D. Dahlquist*, Department of Biology, Loyola Marymount University, Los Angeles, CA 90045
Melissa L. Aikens, Department of Biological Sciences, University of New Hampshire, Durham, NH 03824
Joseph T. Dauer, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, 68503
Samuel S. Donovan, Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260
Carrie Diaz Eaton, Environmental Literacy Program, Unity College, Unity, ME 04988
Hannah Callender Highlander, Department of Mathematics, University of Portland, Portland, OR 97203
Kristin P. Jenkins, BioQUEST, Boyds, MD 20841
John R. Jungck, Department of Biological Sciences, University of Delaware, Newark, DE 19716
M. Drew LaMar, Department of Biology, College of William and Mary, Williamsburg, VA, 23187
Glenn Ledder, Department of Mathematics, University of Nebraska-Lincoln, Lincoln, NE 68588
Robert L. Mayes, College of Education, Georgia Southern University, Statesboro, GA 30460
Richard C. Schugart, Department of Mathematics, Western Kentucky University, Bowling Green, KY 42101

*Corresponding author:
Kam D. Dahlquist
Department of Biology
Loyola Marymount University
1 LMU Drive, MS 8888
Los Angeles, CA 90045 USA
Tel: 310-338-7697
Fax: 310-338-5317
kdahlquist@lmu.edu

Article type: Essay

Character count (with spaces): 48,271 (without title page or references); 65,051 (entire document)

Running title: An Invitation to Modeling

Keywords: models and modeling, model-based reasoning, process of science, quantitative skills, undergraduate

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

Abstract

Models and the process of modeling are fundamental to the discipline of biology, and therefore should be incorporated into undergraduate biology courses. In this essay, we draw upon the literature and our own teaching experiences to provide practical suggestions for how to introduce models and modeling to introductory biology students. We begin by demonstrating the ubiquity of models in biology, including representations of the process of science itself. We advocate for a model of the process of science that highlights parallel tracks of mathematical and experimental modeling investigations. With this recognition, we suggest ways in which instructors can call students' attention to biological models more explicitly by using modeling language, facilitating metacognition about the use of models, and employing model-based reasoning. We then provide guidance on how to begin to engage students in the process of modeling, encouraging instructors to scaffold a progression to mathematical modeling. We use the Hardy-Weinberg Equilibrium model to provide specific pedagogical examples that illustrate our suggestions. We propose that by making even a small shift in the way models and modeling are discussed in the classroom, students will gain understanding of key biological concepts, practice realistic scientific inquiry, and build quantitative and communication skills.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

Overview

The central role of models and modeling in scientific practice should be reflected in the ways we teach science. Teaching with models and modeling provides an opportunity to engage students in authentic scientific practices. Indeed, national reports and proposed standards for improving biology education have advocated for the development of competencies and skills related to using models and modeling (AAMC-HHMI, 2009; AAAS, 2011; NGSS, 2013; College Board, 2015). Biologists use models to study complex systems, make predictions, test ideas which are experimentally difficult or impossible to test, develop conceptual frameworks, and generate causal relationships (Odenbaugh, 2005; Svoboda and Passmore, 2011). Biologists' use of models to communicate ideas and explore theories is an integral component of the scientific process (Tomasi, 1988; Gilbert, 1991; Lander, 2010; Jungck, 2011), and as computational power and access to data increases, biologists are able to employ quantitative models more frequently and to greater effect. For students, working with models can enhance understanding of key biological concepts, provide practice in realistic modes of scientific inquiry, and build quantitative and communication skills (Lehrer and Schauble, 2005; Garfunkel and Montgomery, 2016). However, it can seem a daunting challenge to incorporate the teaching of these skills into already overcrowded course curricula.

Who we are and our goals for this essay

We, the authors, are an interdisciplinary group of biologists, mathematicians, mathematical biologists, and education researchers who came together for a working group at the National Institute for Mathematical and Biological Synthesis (NIMBioS), organized by the leadership team from the Quantitative Undergraduate Biology Education and Synthesis project (QUBES; https://qubeshub.org; Donovan et al., 2015), to address the challenges of teaching modeling. This working group has provided us the opportunity over two years to explore the education research on modeling, share how modeling is applied and taught in our various disciplines, and examine our individual teaching experiences for best practices. In the process, we have become more thoughtful in our practice and have developed more

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

granular, nuanced, and inclusive definitions of models and modeling in a broad sense, and mathematical modeling in particular (Eaton et al., 2017). This has led to insight about how we might improve our approach to teaching with models in biology courses, which we share in this essay. We have three main goals in mind:

1. We ask readers who may not have previously thought of themselves as modelers to consider that models and modeling are ubiquitous in biology, and, as a result, *we are all modelers*.
2. We ask those new to modeling, and modeling veterans alike, to use our suggested pedagogical strategies, beginning with explicitly using the language of models and the process of modeling in the classroom in order to raise student awareness of the role of models and modeling in biology.
3. We ask instructors to begin to incorporate models and modeling activities in the classroom to lay the foundation for incorporating more mathematical modeling both within courses, and, eventually, throughout the biology curriculum.

What is a model? What is modeling?

For clarity, we will begin by defining our terms (Box 1). Our working definition is that a *model* is a simplified, abstract or concrete representation of relationships and/or processes in the real world, constructed for some purpose (Eaton et al., 2017). Model representations can be experiential (physical manipulatives, animations/simulations, experiments), visual (schematics, diagrams, flowcharts), verbal (hypotheses, predictions, descriptions, assumptions), numerical (data tables, graphs), or symbolic (equations, formulas). The symbolic representation, either alone or in combination with one of the other representations, is what is usually described as a mathematical model (Gilbert, 2004; Eaton et al., 2017). Whatever the representation, these all share features as models. The representation is chosen based on the problem at hand. For example, genes can be represented in different ways: a molecular biologist might use a visual gene diagram that accounts for the mechanisms of gene expression, whereas a population geneticist might incorporate genes symbolically into a mathematical model to understand the evolutionary

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

processes occurring (Figure 1). Both of these gene models are abstractions, including some details and leaving others out, appropriate to the desired application, question being asked, or concept being taught. In other words, they are appropriate to their "model utility," the purpose of constructing and using a model, an important notion since not all models are appropriate for all purposes (Odenbaugh, 2005; Svoboda and Passmore, 2011). Table 1 describes five uses of models in biology, giving examples for each.

Models are ubiquitous in biology, taking each of the aforementioned forms (Figure 1): a physical model of the DNA double helix, a visual model of the organelles within a eukaryotic cell, or a mathematical model of population growth. Hypotheses can be thought of as models because they are verbal representations of relationships. Experiments themselves are also models as they are based on interventions and/or observations of the real world by a scientist. They reflect a particular model of the world and are not the real world itself. Or, consider the figures in an introductory biology text; most, if not all, of them are visual model representations of biological concepts. As such, figures need to be viewed critically in terms of intent: Which features of the model were chosen to be foregrounded? Which features were chosen to be backgrounded or left out entirely? Does the figure represent a summary of features, which collectively have no actual counterpart in the real world? Thus, only a small shift is needed to see that models are everywhere in biology.

Having defined models as representations, *modeling* is then an iterative process in which a model is proposed, explored, and refined (Figure 2). Modeling allows us to explore and explain complex, real-world problems. Scientific practices such as experimental or observational studies are one form of modeling (right side of Figure 2), as is mathematical modeling (left side of Figure 2). Biological systems present extremely complicated and difficult problems, and with the increase in computational power, mathematical modeling is proving to be a valuable tool for exploring these problems.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

Process of modeling as process of science

In Figure 2, we use model-based reasoning (Windschitl et al., 2008) as a framework to show that the process of modeling is an instance of the process of science. We argue that various types of modeling present different, and complementary ways, to reason about scientific questions. Understanding of a biological concept may be enhanced by exploring and reasoning about the concept with models. In this framework, to address a biological problem with a model we must identify the problem in its biological context, organize the relevant information (what we know and what we would like to know), ask questions, and formulate hypotheses (Figure 2, top). We then seek evidence through experimental or mathematical modeling (Figure 2, middle, discussed below). Finally we analyze the results, constructing an argument for a biological interpretation of the results within a larger disciplinary conceptual framework that is then subjected to review by the scientific community (Figure 2, bottom). Feedback from other scientists leads to a reframing of the biological problem itself, kicking off the next iteration of the entire process. Within this framework, experimental modeling (Figure 2, right) and mathematical modeling (Figure 2, left) are two parallel tracks that can be pursued when seeking evidence. While we highlight experiments in Figure 2, we note that observational field studies, evolutionary reconstructions, and meta-analyses are other modes of investigation carried out by biologists that can be found under the umbrella of model-based reasoning.

When designing either an experimental model or a mathematical model, a key step is deciding which aspects of the biological system are relevant to the question and which can be safely ignored, or in modeling terms, making simplifying assumptions (Figure 2, middle). We might, of course, be wrong, but the intentional process of identifying the assumptions and explaining why they are acceptable for this particular situation and question is itself one way to learn about the system (Levins, 1966). Note that in the experimental modeling process (Figure 2, right track), simplifying assumptions include the choice of a tractable experimental system such as a model organism that is well-suited to study the phenomenon of interest and easy to grow in the lab *(e.g.*, Tetrahymena for the study of telomeres; Kain, 2009). For

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

experimental modeling, we next select the appropriate materials and methods, bearing in mind the advantages and disadvantages of each, and design the appropriate positive and negative controls. On the other hand, to construct the mathematical model (Figure 2, left track), we progress by mathematizing the biological problem, starting with choices about the type of mathematical technique to use (*e.g.*, deterministic vs. probabilistic) and then creating a mathematical representation of the system containing some (but not all!) elements, interactions, and mechanisms (formalized as variables, parameters, and formula or equations). The next step in the process is to run the experiment or explore the mathematical model, generating data to analyze.

Model validation (Figure 2, dashed arrow) occurs when experimental and mathematical modeling results are compared to assess each model's utility and accuracy. When validating model results, we ask "Does this choice of model give us explanatory power to make sense of the biological phenomena we are observing?", "Does the model allow us to make predictions of what might happen in areas we have not yet explored?", "Does it promote further inquiry?". If we fail to validate the model, we can go back to our experimental or mathematical model and refine it by modifying the experiment or observational study in the former case or by changing the parameters, variables, formulas, or equations in the latter case. It is worth noting that this process of formulating hypotheses, building models, constructing an argument from data, and refining models is a process of theorizing from observation and requires genuine creativity (Windschitl et al. 2008).

In summary, model-based reasoning is not just a pedagogical approach, it is an important part of the process of science. Windschitl and colleagues (2008) stress the idea of viewing science as the process of creating and refining explanatory models of the world. The model-based reasoning view of science is very general and encompasses the practice of seeking multiple lines of evidence in order to choose one hypothesis from a set of competing hypotheses. It gives us a framework to draw parallels between multiple lines of reasoning (mathematical and experimental). Thus, as can be seen in Figure 2, mathematical models and experimental models are both different ways to construct evidence to test a hypothesis, with parallels even within the individual steps of each. *As scientists, we are all modelers.*

Aspects of modeling that become hidden in expert practice

When we as scientists and mathematicians use models, mathematical or otherwise, we recognize key features of the modeling process that define the utility and value of models to solving biological problems. For example, we understand that the model is a simplified version of the biological system under study, and as such, it was built with a set of assumptions and carries with it certain limitations. Those assumptions may be appropriate under some circumstances and inappropriate under others (Levins, 1966). Limitations mean that the model may describe one aspect of the biological system well but neglect or perform poorly on another aspect. Understanding the assumptions and limitations of a model allows us to infer under what circumstances a model is useful. For example, when biologists recognize a box representation as a gene (Figure 1B), we do not think this is actually what a gene looks like and can assimilate this model with a different representation of a gene, such as a chromosomal diagram, or the abstraction of genes and alleles in the Hardy-Weinberg Equilibrium mathematical model (Figure 1C-E). Or, when a textbook diagram of a gene includes a start codon but not a stop codon, as experts in discipline-specific knowledge, we can fill in that missing piece. *When we become familiar with a model, its assumptions and limitations may recede from our immediate consciousness, and become part of the implicit "expert knowledge" and skills we bring to bear in our discipline*. In addition, as experts, our model-based reasoning skills make us adept at layering disciplinary information on models and using models to explore problems. For example, we can visualize population growth as a logistic curve, consider the impact a change in variables would have on the shape of the curve, and interpret what this means to a population. These layers of understanding and intuitive use of models are based on our expert knowledge of the discipline, these models, and the modeling process. However, even though science students are constantly making mental models as they make sense of the world (Gilbert, 2004; Louca and Zacharia, 2012), they are less likely to have the depth of scientific modeling experience or understanding

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

of the norms of disciplinary reasoning with models. Modeling instruction should encourage students to reflect on: (1) the attributes of models, (2) the strategies for their use, and (3) the connections between models, data, and inferences about biological systems. Furthermore, by describing experiments as models, students should have a better understanding of what the experiment is doing, *i.e.*, experiments are giving them as much information as their tools were able to measure, but the results still may not be the complete picture of reality. That is why scientists use different techniques, such as complementing *in vitro* experiments with those conducted *in vivo*.

What do students need to do to become mathematical modelers?

Mathematical modeling requires a complex suite of skills including higher-order thinking, quantitative skills, communication and collaboration skills, and foundational knowledge of the biological problem. Furthermore, as previously mentioned, modeling is an iterative process – there is no fixed end since models are constantly revised. To develop proficiency in modeling, students must have multiple opportunities to practice. However, few curricular resources are designed to provide this opportunity, especially at introductory levels (NRC, 2003; AAMC-HHMI, 2009). Yet, the time invested in teaching and learning mathematical modeling develops skills that increase student success in any discipline (P21 Partnership, 2009; AAAS, 2011; Garfunkel and Montgomery, 2016; Schuchardt and Schunn, 2016). Here we share ideas and best practices we have gleaned from the literature and our experience to make incorporating modeling activities more feasible.

Recommendations for incorporating models and modeling in the classroom

Our working group includes both veterans and relative newcomers to modeling whose experience in teaching modeling ranges from in-depth testing and refining of techniques to just recently trying these ideas for the first time. The discussion of our collective experiences has given us a more practical understanding of the problems that students and faculty face when including modeling as part of the curriculum. For the rest of the examples presented in this essay, we focus primarily on lower level courses

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

where exploring models and modeling is not necessarily an emphasis. Skills learned there can translate to upper division courses specifically focused on mathematical modeling, which present a different, but overlapping, set of challenges. Expanding students' experiences with models and modeling early in their course of study will increase their understanding of biological concepts and prepare them for more advanced mathematical modeling in the future (AAAS, 2011; Schuchardt and Schunn, 2016). Because mathematical modeling can be a complex and lengthy process which requires time not often available in our biology curriculum, we do not expect that in every course, students will engage in the entire modeling process from conception to dissemination (Figure 2, top to bottom). Instead, we suggest that having students engage in more granular "modeling activities" (individual arrows of Figure 2) forms part of a longer learning trajectory that *can* be supported in every class; an approach we advocated in Eaton et al. (2017). Although we acknowledge that breaking down modeling into individual activities does not provide students with an experience of the full modeling process, it does provide opportunities to gain needed skills (for more detail, see Eaton et al., 2017). Our suggestions for increasing student facility with models and the modeling process in lower division biology courses includes the following:

1. Being explicit about model use, model utility, and the modeling process by adopting consistent and detailed modeling language and concepts in the classroom (such as assumptions and limitations; Brewe, 2008);
2. Facilitating metacognition by giving students opportunities to reflect about models and modeling to help them develop an awareness and evaluation of their thinking (Schwarz and White, 2005; Papaevripidou et al., 2007);
3. Using an anchor, a student-accessible concrete base to launch the modeling process (Schwarz et al., 2009);
4. Encouraging students to ask their own questions (Jungck, 1985; Peterson and Jungck, 1988; Rothstein and Santana, 2011);

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

5. Scaffolding the progression to mathematical modeling by practicing moving between model representations (experiential, verbal, visual, numerical, and symbolic), and by refining models (Mayes et al., 2013; Eaton et al., 2017); and
6. Providing opportunities for students to develop their modeling abilities through repeated practice (Weisstein, 2011).

We placed these recommendations in this particular order so that they progress from relatively easy to implement for newcomers to modeling to more advanced activities for those who wish to deepen their practice. Of course, as with the process of modeling and process of science, the real-life implementation of these pedagogical strategies may come in a different order and can be messy; consider these a menu of options. In the sections that follow, we flesh out these recommendations in the context of the commonly taught Hardy-Weinberg Equilibrium (HWE) population genetics model (Soderberg and Price, 2003).

Be explicit by using modeling language and concepts

A simple step we can take to increase students' awareness of the utility of models is to be explicit in our language about models and the process of modeling throughout the course (Brewe, 2008; Windschitl et al. 2008) as follows:

- First, draw students' attention to the models used in class (*e.g.*, Figure 1 and textbook figures).
- Show the parallels between the process of modeling and process of science (Figure 2).
- In a think-write-pair-share exercise, have students write and discuss their own definitions of models and modeling. Write some of these on the board and together develop the definitions that we have proposed in this essay (Box 1).
- Verbally describe the assumptions used to build a particular model and how that relates to the limitations of the model.
- Compare different models of the same phenomenon, reviewing how models have been revised through time and exploring connections between biological and mathematical or statistical models.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

- Include the process of modeling in course learning objectives, as an indicator of the importance of using modeling practices throughout the course (*e.g.*, students will analyze and interpret diverse models used in science; students will construct models to represent biological systems; or, students will test predictions using mathematical models).

These suggestions do not require substantial alteration of the existing curriculum, just a small shift that will draw attention to the use of models in biology. By changing how we think about presenting this information as part of a lecture or activity, we can achieve the goal of raising student awareness and understanding of models.

If we consider the population genetics model, Hardy-Weinberg Equilibrium (HWE), we can see how it can be used to foreground the modeling process with students, developing their modeling skills, while at the same time helping them learn important evolutionary concepts (Box 2). The utility of HWE as a model is in providing a simple conceptual framework for inferring when evolutionary forces are acting on a population, one that can be expanded upon to explore which forces may be at play (Table 1). This type of theoretical exploration provided by the model is an opportunity for students both to engage in quantitative interpretation of a mathematical model, and to work with the various models of the biological concepts of genes and alleles, which can be elusive for students to grasp (Speth et al., 2014). First, be explicit that HWE *is* a model (something potentially hidden by the distractor of "equilibrium" in the name), and a mathematical model at that. Then be explicit with how the assumptions built into the HWE model (one gene, two alleles, a large, randomly mating population with no overlapping generations, and no evolutionary forces acting upon the population) both make it useful as a null model and also limit its utility to situations unlikely to be found in real populations. Have the students describe the model (Box 2) by identifying the variables in the model, their attributes (such as whether they are continuous or discrete), and measures. Have students explain why these variables are included in the model, and how the attributes of the variables lead to relationships that result in the model, *i.e.*, link the verbal description of the assumptions of the model to the symbolic equation. Deliberately emphasizing these aspects of the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

modeling process can help students move beyond focusing simply on the "plug-and-chug" nature of inputting numbers into $p^2 + 2pq + q^2 = 1$ (what Stewart et al., 2010 called "model-less problem solving") and towards using model-based reasoning to see how the model fits into a larger explanatory framework of the evolutionary forces at work ("model-using problem solving" in the language of Stewart et al., 2010). This type of model description is a gateway to later having students develop their own models.

Facilitate metacognition

The previous sections described how model-based reasoning informs our approach to biological problems and how we can introduce the practices of model use and modeling if we are more explicit in the language we use in our teaching. Next, by structuring students' opportunities to reflect about models and modeling, we encourage them to make new connections between ideas and recognize generalizations about their experiences. Of course, the goal of having students reflect about using models is to help them develop an awareness and evaluation of their thinking (a.k.a., metacognition) about models when they encounter new scientific problems. Some have defined this level of awareness about modeling as "metamodeling", which is the ability to be metacognitive about the process of modeling (Box 1, 2; Schwarz and White, 2005; Papaevripidou et al., 2007). Metamodeling can improve students' understanding of practices like predicting, observing, and explaining phenomena (Barab et al., 2000; Schwarz and White 2005; Sins et al., 2005) and the ability to make mechanistic explanations (Fretz et al., 2002; Louca and Zacharia, 2012). Importantly, metamodeling enhances students' abilities to regulate their own learning (Papaevripidou and Zacharia, 2015). As students gain awareness of where they are relative to a learning progression of modeling (Schwarz et al., 2009), they can be more aware of how they are using models to address biological problems. For example, a couple of the authors have had success with journal assignments that ask students to reflect on what they have learned through the modeling activity. Journal prompts can include questions like:

- How did clarifying your assumptions help you develop a better model?
- How did you determine if the results were biologically valid?

- What was most surprising about your findings?
- How has the modeling activity given you different insight into the biological problem?

Use anchors

A good pedagogical approach for engaging students more deeply with modeling is to provide an "anchor" – a personally accessible puzzling event or observation, rooted in a complex phenomenon, that acts as a concrete base for exploring scientific concepts (Schwarz et al., 2009). Anchoring provides an opportunity for students to engage in active learning, and importantly, provides a larger, compelling problem for students to solve by applying their knowledge. Introducing a relevant, biological problem and providing the opportunity to explore the phenomenon through the modeling process sets students up to learn both disciplinary content and modeling skills (Garfunkel and Montgomery, 2016). The relevance of the problem provides students with a "need to know" that drives their interest in the problem (*e.g.*, Dohn et al., 2009), motivating students as they struggle to learn new skills and new information (Hidi and Harackiwiecz, 2000).

In the case of the HWE model, the anchor could be a familiar case in which organismal phenotypes change through time, such as artificial selection for agricultural purposes or experiments using Wisconsin Fast Plants that they may have conducted in biology labs (Williamson, 2015). This allows students to draw on what they already know. Another type of anchor is to have students conduct their own research by collecting observations on the phenomenon (Box 2). There is a myriad of simulations, both computer-based and physical, that demonstrate changes in allele frequency in a population (Jungck et al., 2010; Brewer and Gardner, 2013; Williamson, 2015). Finally, students might read the original paper in *Science* by the mathematician G. H. Hardy (Hardy, 1908). This paper was motivated by providing an explanation for the anchoring phenomenon that dominant alleles causing bradydactyly in humans do not increase in frequency in a population, a phenomenon still misunderstood today.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

Encourage students to ask their own questions

Anchoring introduces students to a relevant, real-world, and messy biological problem and provides some connection to personal experience with the phenomena of evolutionary forces and population genetics. Beyond simply providing an example for data collection, this is an opportunity for students to engage in both the scientific process and the modeling process by asking questions about the phenomenon (Figure 2, top, Box 2). Students rarely have experience in asking questions, let alone refining questions to be "good" scientific questions (Jungck, 1985; Peterson and Jungck, 1988; Windschitl et al., 2008). This fundamental process-of-science skill requires practice and is worth the time to develop. Encourage students to not just observe *what* is happening, but also ask *why* it is happening. Initial student questions are likely to need revision. Staging the process so that students have an opportunity to brainstorm unrefined questions, followed by working as a group to hone questions to be more relevant will help students develop this skill (Rothstein and Santana, 2011). As students settle on questions they would like to pursue, have them sketch a visual model representing the phenomenon (Dauer et al., 2013). As with their initial questions, this model is likely to also require refinement. Engaging in the modeling process to develop a visual model helps students (and their instructors) identify where they are missing information (Pearsall et al., 1997; Long et al., 2014; Speth et al., 2014). The process of asking questions, evaluating and refining questions, and making a first pass at a model representing the phenomenon are valuable learning activities, and provides instructors with insight into how students are thinking.

In the case of the HWE model, ask students to come up with questions about the observed anchoring phenomenon of change in a population, and draw a visual model. It is important for students to have time to work together to clarify and focus their questions and their models. It is not important for the questions to be good scientific questions, or for the model to be a reliable representation. The educational value lies in the process, not the product (Garfunkel and Montgomery, 2016). Building in time for students to ask and refine questions, and develop a conceptual model does require course time. Students will need some support in the first iteration of the modeling process, and it is essential they have time to struggle with the problem. After students have produced their first model, focus discussion on what quantitative aspects they included in the model (Weisstein, 2011). If the initial models are only qualitative and visual, then ask the students to brainstorm what quantitative aspects could be added.

Move between multiple representations

In the previous section, we suggested helping students move from a verbal model to a visual model to a mathematical model representation by asking them what quantitative aspects could be added to their model. In general, an important modeling skill that can be scaffolded on the way to mathematical modeling is the ability to represent a model in multiple modalities (experiential, visual, verbal, numerical, and symbolic, cf. Eaton et al. 2017). Students should practice moving between representations and be able to explain how the representations are related. We encourage students to develop both qualitative and quantitative models, moving back and forth between them as different forms of evidence that strengthen understanding of the phenomena. Having students use multiple representations of a model can support their learning (Ainsworth, 1999). Multiple representations of a model may complement each other, each providing different information that allows a student to have a more comprehensive understanding of the biological problem (Ainsworth, 1999). In addition to translation among multiple representations, it is vital to stress meaningful qualitative and quantitative interpretation of these models when determining trends and making predictions, making explicit the link between the models in the biological context (Mayes et al., 2013).

Shifting between a qualitative verbal or visual model and a quantitative mathematical model (*mathematization*, Figure 2, left track) can be particularly challenging for students. Scaffolding the process by starting with small, accessible steps, may help build student confidence. The first steps in mathematizing a problem involve identifying variables, a unit measure for each variable, and attributes of each variable that help determine covariation between variables (Thompson, 2011). Students can use a qualitative verbal or visual model to identify variables related to the phenomenon and to identify the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

relationships between those variables. With an understanding of the variables and their relationships, students will be better prepared to move forward with developing a mathematical model.

To derive the HWE mathematical model, students must identify allele frequency as a measure of evolutionary change in the population. The ease with which we, as instructors, throw around terms like gene, allele, *p*, and *q*, may reproduce textbook conflations of genes and alleles that generate ambiguity and hide uncertainty on the part of students as to what these terms mean. Connect students to what they know about the life cycle of an organism (College Board, 2012) and then have them work from their questions and qualitative verbal or visual models of HWE to identify the variables (alleles) and their metrics (frequency). Mathematizing the problem themselves gives students the important role of 'owners' of the modeling process because it makes them responsible for learning about phenomena they discovered (Papaevripidou et al., 2015). In other words, they become self-regulated modelers.

Taking this further, moving back and forth between a symbolic mathematical model and a simulation (an experiential model) may be optimal for demonstrating how each type of evolutionary force results in changes in the genetic variation of a population and for preparing students to refine their initial models (next section). PopGen (http://www.radford.edu/~rsheehy/Gen_flash/popgen/) or the Biological ESTEEM Project modules Deme 2.0 (http://bioquest.org/esteem/esteem_details.php?product_id=193), DeFinetti 1.0 (http://bioquest.org/esteem/esteem_details.php?product_id=204), or Evolution Through Natural Selection (http://bioquest.org/esteem/esteem_details.php?product_id=7080) can be used by students to manipulate the strength of evolutionary forces and observe changes in allele frequencies over generations, generating a numerical model. Note that in this pedagogical strategy, we did not start by just giving the students previously collected data. A data table already indicates the variables that are important and leapfrogs the student to interpreting trends and even making predictions, both relevant and irrelevant. In this more advanced example, students should practice constructing the HWE mathematical model for themselves.

Refine models

Two frequently overlooked modeling activities that can be performed with students are the validation and refinement of models (Figure 2). An important aspect of model-based reasoning is model validation, that the model should be tested empirically against the observations of the phenomena (Windschitl et al., 2008; Schwarz et al., 2009) by evaluating the data collected to determine if the model fits the data. Student models should also be tested conceptually against other models, by comparing them with alternate models from their peers and with established models of the phenomenon. This provides the opportunity for students to communicate and collaborate to identify weaknesses in their models and sets the stage for the need for model refinement.

In the case of the HWE model, students can validate the model by making predictions and explaining outcomes by testing the model on data collected from existing populations. Starting with a pool of F_0 genotypes that are in Hardy-Weinberg equilibrium, give students several possible F_1 pools and ask if evolutionary forces impacted the allele frequencies observed in each pool. For example, a much smaller number of **aa** genotypes and much larger number of **AA** genotypes is consistent with a population in which the homozygous recessive individuals fail to survive to adulthood. For some courses it will be appropriate to introduce students to the Chi-square test to quantify the probability that a given F_1 population is in HWE. In doing so, we are layering the Chi-square statistical model on top of the HWE null mathematical model. In terms of model validation, it is important to note that when one compares the observed F_1 genotypes to the expected HWE values, one is checking to see if the *experimental population* is in HWE, but when one compares the expected HWE values to the observed population data, one is validating the assumptions of the *mathematical model*.

In the example above, the recognition that the genotype frequencies in the F_1 generation do not fit the HWE null model leads to the biological explanation that evolutionary forces may be at play. However, it should also be recognized that this is an evaluation of the assumptions of the mathematical model and that an appropriate revision of the model could improve its explanatory power. This calls for students to

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

cross the threshold into creating a new model through refinement of an existing model to meet new criteria or to apply it to a new situation. With a better understanding of the mathematical modeling process, students can thoughtfully explore the consequences of adding a third allele or a second locus, or explore the outcomes of various adaptive landscapes. These consequences could be reasoned through and tested with the PopGen simulator (http://www.radford.edu/~rsheehy/Gen_flash/popgen/) or the aforementioned Biological ESTEEM Project modules. We recognize that this may be beyond the scope of a typical introductory biology course, but include this example here to show where this learning trajectory is leading. In total, the modeling activities described in this essay form a learning progression from model description to model exploration to model development to model refinement (Box 2). In this process, students shift from seeing mathematical models as a "black box" to a "glass box", exploring the why and how of the observed behaviors. Finally, by developing and refining their own mathematical models, they are operating with "no box" (http://bioquest.org/esteem/Intro_to_ESTEEM.pdf), having learned skills that can be transferred to new biological problems.

Provide opportunities for repeated practice of modeling skills within courses and across the biology curriculum

While we have stayed with the HWE model throughout this essay to demonstrate how to implement the pedagogical strategies we suggest, we do not wish to imply that these pedagogical strategies only apply to this particular content domain. Instead, we suggest that instructors be explicit about the language of models and modeling throughout an introductory biology course. For example, have students consider the species concept as a collection of species models in different contexts. By exploring the assumptions and limitations of a species model, students will certainly gain more content knowledge about the biological relevance of the definition of species (What does "species" mean in the bacterial domain or for extinct species?). Moreover, by making a small shift and using the language of models and modeling in this different example, the activity reinforces what students learned about the modeling process in other situations such as the HWE model and contributes to their ability to apply modeling skills in another context.

It is important to provide students practice with modeling throughout their degree program, but it is not necessary for students to engage in the full mathematical modeling process from conception to dissemination (Figure 2) in every course. For example, it may not make sense to have students engage in all of the mathematical modeling activities we described, such as deriving the Hardy-Weinberg mathematical equation in introductory biology. In contrast, it may be an explicit goal of an upper division genetics or evolution course to not only do so, but to revise the equation for absolute selection against a homozygous recessive genotype or to move from a one locus-two allele model to more loci or alleles or both. These goals can be supported by using non-mathematical modeling language in introductory biology through discussions of how to measure evolution and emphasizing the assumptions of the HWE model. In addition, students will benefit from repeated experiences with modeling in different contexts.

Another reason to engage students in mathematical modeling is that students come to biology with quantitative knowledge and reasoning skills (AP Calculus or Statistics, college-level math courses; see Jungck, 2011), but need practice retrieving and applying them properly in novel contexts (Hester et al., 2014). Students have exhibited significant learning gains when applying their quantitative skills to biological problems in mathematics courses designed for biology majors (Eaton and Highlander, 2017), and we encourage biology faculty to provide opportunities for students to practice and apply these important skills in biology courses as well. If foundational mathematics and modeling knowledge is not practiced throughout the curriculum, proficiency in the skill will be lost.

An invitation to modeling: building a community with shared explicit practices

We hope this essay has shifted your thinking to see that models are ubiquitous in biology and to consider yourself a modeler. Our goal was to inspire you to incorporate model-based reasoning in your biology courses, adopting some of our suggested pedagogical strategies, starting with being explicit in

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

your language. Finally, we believe that as you gain experience using models and the language of modeling with your students, this will lay the foundation for incorporating more mathematical modeling into your courses. In short, we hope that you have revised your model of teaching modeling, and we invite you to join our community of modelers. There are many resources available to support your efforts in this area, some of which we have compiled in the Accessing Materials section. For mathematical modeling, we encourage you to read the GAIMME report (Garfunkel and Montgomery, 2016), explore the resources provided by the Society of Industrial and Applied Mathematics (SIAM, 2012; SIAM, 2014; https://m3challenge.siam.org/resources), and join the QUBES community (https://qubeshub.org; Donovan et al., 2015), which provides resources, tools, and professional development opportunities around quantitative biology, especially through the Modeling Hub (https://mmhub.qubeshub.org). While we have not addressed the assessment of modeling skills in this essay, sample rubrics are available (Garfunkel and Montgomery, 2016; Bryce et al., 2016). We extend to you an invitation to modeling; we hope you will join us!

Accessing Materials

- QUBES: https://qubeshub.org, Modeling Hub: https://mmhub.qubeshub.org
- Math Modeling Resources at SIAM: https://m3challenge.siam.org/resources
- PopGen: Population genetics simulation program http://www.radford.edu/~rsheehy/Gen_flash/popgen/
- Biological ESTEEM Project (http://bioquest.org/esteem/index.php)
    - Deme 2.0 (http://bioquest.org/esteem/esteem_details.php?product_id=193)
    - DeFinetti 1.0 (http://bioquest.org/esteem/esteem_details.php?product_id=204)
    - Evolution Through Natural Selection (http://bioquest.org/esteem/esteem_details.php?product_id=7080)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

Acknowledgments

This work was conducted as part of the "Unpacking the Black Box: Teaching Quantitative Biology" Working Group at the National Institute for Mathematical and Biological Synthesis (NIMBioS), sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. Co-authors Glenn Ledder and Richard C. Schugart were supported as NIMBioS Sabbatical Fellows. We want to acknowledge the intellectual contributions of working group members who were not able to work on this paper including Gregory D. Goins, Ben G. Fitzpatrick, and Edward F. (Joe) Redish and conversations with colleagues beyond the working group, especially Arietta E. Fleming-Davies and Anton E. Weisstein who gave valuable feedback on a draft of this essay. Many thanks to our brave students and colleagues who allowed us to refine our ideas with them. Organization of this co-authorship and working group between meetings would not have been possible without the QUBES Hub supported by NSF Awards #DBI 1346584, DUE 1446269, DUE 1446258, and DUE 1446284.

References

AAMC-HHMI Committee. (2009). *Scientific foundations for future physicians*. Washington, DC: Association of American Medical Colleges.

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, *33*, 131–152. http://doi.org/10.1016/S0360-1315(99)00029-9

Allen, L. J., Brauer, F., Van den Driessche, P., & Wu, J. (2008). *Mathematical Epidemiology (Volume 1945)*. Berlin: Springer.

American Association for the Advancement of Science (AAAS). (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*. Washington, DC: American Association for the Advancement of Science.

Barab, S. A., Hay, K. E., Barnett, M., & Keating, T. (2000). Virtual solar system project: Building understanding through model building. *Journal of Research in Science Teaching*, *37*(7), 719–756. http://doi.org/10.1002/1098-2736(200009)37:7<719::AID-TEA6>3.0.CO;2-V

Biological ESTEEM Project. (2015). DeFinetti 1.0. Retrieved August 24, 2017, from http://bioquest.org/esteem/esteem_details.php?product_id=204

Biological ESTEEM Project. (2015). Deme 2.0. Retrieved August 24, 2017, from http://bioquest.org/esteem/esteem_details.php?product_id=193

Biological ESTEEM Project. (2015). Evolution Through Natural Selection. Retrieved August 24, 2017, from http://bioquest.org/esteem/esteem_details.php?product_id=7080

Biological ESTEEM Project. (2015). An Introduction to ESTEEM. Retrieved August 24, 2017, from http://bioquest.org/esteem/Intro_to_ESTEEM.pdf

Brewe, E. (2008). Modeling theory applied: Modeling Instruction in introductory physics. *American Journal of Physics*, *76*(12), 1155–1160. http://doi.org/10.1119/1.2983148

Brewer, M. S., & Gardner, G. E. (2013). Teaching evolution through the Hardy-Weinberg Principle: A real-time, active-learning exercise using classroom response devices. *American Biology Teacher*, *75*(7), 476–479. http://doi.org/10.1525/abt.2013.75.7.6

Bryce, C. M., Baliga, V. B., De Nesnera, K. L., Fiack, D., Goetz, K., Tarjan, L. M., … Gilbert, G. S. (2016). Exploring models in the biology classroom. *American Biology Teacher*, *8*(1), 35–42. http://doi.org/10.1525/abt.2016.78.1.35

College Board. (2015). AP Biology Course and Exam Description. Retrieved July 7, 2017, from https://apcentral.collegeboard.org/pdf/ap-biology-course-and-exam-description.pdf?course=ap-biology

College Board. (2012). Investigation 2: Mathematical modeling: Hardy-Weinberg in AP Biology Investigative Labs: An Inquiry-Based Approach. Retrieved August 24, 2017, from http://www.collegeboard.com/html/apcourseaudit/courses/pdfs/cb-biology-lab-manual-1-24-12.pdf

Cooper, A. K., & Oliver-Hoyo, M. T. (2017). Creating 3D physical models to probe student understanding of macromolecular structure. *Biochemistry and Molecular Biology Education*, 1–10. http://doi.org/10.1002/bmb.21076

Cordell, S., Goldstein, G., Mueller-Dombois, D., Webb, D., & Vitousek, P. M. (1998). Physiological and morphological variation in Metrosideros polymorpha, a dominant Hawaiian tree species, along an altitudinal gradient: The role of phenotypic plasticity. *Oecologia*, *113*(2), 188–196. http://doi.org/10.1007/s004420050367

Dauer, J. T., Momsen, J. L., Speth, E. B., Makohon-Moore, S. C., & Long, T. M. (2013). Analyzing change in students' gene-to-evolution models in college-level introductory biology. *Journal of Research in Science Teaching*, *50*(6), 639–659. http://doi.org/10.1002/tea.21094

Dohn, N. B., Madsen, P. T., & Malte, H. (2009). The situational interest of undergraduate students in zoophysiology. *Advances in physiology education*, *33*(3), 196-201. http://doi.org/10.1152/advan.00038.2009

Donovan, S., Eaton, C. D., Gower, S. T., Jenkins, K. P., LaMar, M. D., Poli, D., … Wojdak, J. M. (2015). QUBES: A community focused on supporting teaching and learning in quantitative biology. *Letters in Biomathematics*, *2*(1), 46–55. http://doi.org/10.1080/23737867.2015.1049969

Eaton, C. D., Highlander, H. C., Dahlquist, K. D., LaMar, M. D., Ledder, G., & Schugart, R. C. (2017). A "Rule of Five" Framework for Models and Modeling to Unify Mathematicians and Biologists and Improve Student Learning. Revised version submitted to the journal *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies* on April 28, 2017, available from the *arXiv* preprint server, https://arxiv.org/abs/1607.02165v2

Eaton, C. D., & Highlander, H. C. (2017). The case for biocalculus: Design, retention, and student performance. *CBE Life Sciences Education*, *16*(2), 1–11. http://doi.org/10.1187/cbe.15-04-0096

Fretz, E. B., Wu, H. H.-K., Zhang, B., Davis, E. A., Krajcik, J. S., & Soloway, E. (2002). An investigation of software scaffolds supporting modeling practices. *Research in Science Education*, *32*(4), 567–589. http://doi.org/10.1023/A:1022400817926

Garfunkel, S., & Montgomery, M. (Eds.). (2016). *GAIMME Report: Guideline for Assessment and Instruction in Mathematical Modeling Education*. Bedford, MA: Consortium for Mathematics and Its Applications. ISBN: 978-1-611974-43-0

Gilbert, J. K. (2004). Models and modelling: Routes to more authentic science education. *International Journal of Science and Mathematics Education*, *2*, 115–130. http://doi.org/10.1007/s10763-004-3186-4

Gilbert, S. W. (1991). Model building and a definition of science. *Journal of Research in Science Teaching*, *28*(1), 73–79. http://doi.org/10.1002/tea.3660280107

Grimm, V., Augusiak, J., Focks, A., Frank, B. M., Gabsi, F., Johnston, A. S. A., … Railsback, S. F. (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecological Modelling*, *280*, 129–139. http://doi.org/10.1016/j.ecolmodel.2014.01.018

Hardy, G. H. (1908). Mendelian Proportions in a Mixed Population. *Science*, *28*(706), 49–50. http://doi.org/10.1126/science.28.706.49

Hester, S., Buxner, S., Elfring, L., & Nagy, L. (2014). Integrating quantitative thinking into an introductory biology course improves students' mathematical reasoning in biological contexts. *CBE - Life Sciences Education*, *13*, 54–64. http://doi.org/10.1187/cbe.13-07-0129

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of educational research*, *70*(2), 151-179. http://doi.org/10.3102/00346543070002151

Hoefnagel, M. H., Starrenburg, M. J., Martens, D. E., Hugenholtz, J., Kleerebezem, M., Van Swam, I. I., ... & Snoep, J. L. (2002). Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. *Microbiology*, *148*(4), 1003-1013. http://doi.org/10.1099/00221287-148-4-1003

Jungck, J. R. (1985). A problem posing approach to biology education. *The American Biology Teacher*, *47*(5), 264-266. http://doi.org/10.2307/4448046

Jungck, J. R. (2011). Mathematical biology education: modeling makes meaning. *Mathematical Modelling of Natural Phenomena*, *6*(6), 1-21. http://doi.org/10.1051/mmnp/20116601

Jungck, J. R., Gaff, H., & Weisstein, A. E. (2010). Mathematical manipulative models: In defense of "Beanbag Biology". *CBE-Life Sciences Education*, *9*(3), 201-211. http://doi.org/10.1187/cbe.10-03-0040

Just, W., Callender, H., LaMar, M. D., & Toporikova, N. (2015). Transmission of infectious diseases: Data, models, and simulations. In R. Robeva (Ed.), *Algebraic and Discrete Mathematical Methods for Modern Biology* (pp. 193–215). San Diego, CA: Academic Press.

Kain, K. H. (2009). Telomeres and tetrahymena: an interview with Elizabeth Blackburn. *Disease models & mechanisms*, *2*(11-12), 534-537. http://doi.org/10.1242/dmm.003418

Kutmon, M., Riutta, A., Nunes, N., Hanspers, K., Willighagen, E. L., Bohler, A., ... & Coort, S. L. (2015). WikiPathways: capturing the full diversity of pathway knowledge. *Nucleic acids research*, *44*(D1), D488-D494. http://doi.org/10.1093/nar/gkv1024

Lander, A. D. (2010). The edges of understanding. *BMC Biology*, *8*, 40. http://doi.org/10.1186/1741-7007-8-40

Lehrer, R., & Schauble, L. (2005). Developing modeling and argument in the elementary grades. In Romberg, T. A., Carpenter, T. P., & Dremock, F. (Eds.). *Understanding mathematics and science matters*. Routledge, 29–53.

Levins, R. (1966). The strategy of model building in population biology. *American scientist*, *54*(4), 421-431. http://www.jstor.org/stable/27836590

Long, T. M., Dauer, J. T., Kostelnik, K. M., Momsen, J. L., Wyse, S. A., Speth, E. B., & Ebert-May, D. (2014). Fostering ecoliteracy through model-based instruction. *Frontiers in Ecology and the Environment*, *12*(2), 138–139. http://doi.org/10.1890/1540-9295-12.2.138

Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, *64*(4), 471-492. http://doi.org/10.1080/00131911.2011.628748

Macal, C. M., & North, M. J. (2006). Tutorial on agent-based modeling and simulation part 2: How to model with agents. In L. F. Perrone, F. P. Wieland, J. Liu, B. J. Lawson, D. M. Nicol, & R. M. Fujimoto (Eds.), *Proceedings of the Winter 2006 Simulation Conference* (pp. 73–83).

Mayes, R. L., Peterson, F., & Bonilla, R. (2013). Quantitative Reasoning Learning Progressions for Environmental Science: Developing a Framework. *Numeracy*, *6*(1), Article 4. http://doi.org/10.5038/1936-4660.6.1.4

National Research Council (NRC). (2003). *BIO2010: Transforming undergraduate education for future research biologists*. National Academies Press. http://doi.org/10.17226/10497

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: National Academies Press.

Odenbaugh, J. A. Y. (2005). Idealized , inaccurate but successful : A pragmatic approach to evaluating models in theoretical ecology. *Biology and Philosophy*, *20*, 231–255. http://doi.org/10.1007/s10539-004-0478-6

P21 Partnership for 21st Century Learning. (2009). 21st Century Skills Map for Science. Retrieved August 24, 2017, from http://www.p21.org/storage/documents/21stcskillsmap_science.pdf

Papaevripidou, M., Constantinou, C. P., & Zacharia, Z. C. (2007). Modeling complex marine ecosystems: An investigation of two teaching approaches with fifth graders. *Journal of Computer Assisted Learning*, *23*(2), 145–157. http://doi.org/10.1111/j.1365-2729.2006.00217.x

Papaevripidou, M., & Zacharia, Z. C. (2015). Examining how students' knowledge of the subject domain affects their process of modeling in a computer programming environment. *Journal of Computers in Education*, *2*(3), 251–282. http://doi.org/10.1007/s40692-015-0034-1

Pearsall, N. R., Skipper, J. O. E. L. J., & Mintzes, J. J. (1997). Knowledge restructuring in the life sciences: A longitudinal study of conceptual change in biology. *Science Education*, *81*(2), 193–215. http://doi.org/10.1002/(SICI)1098-237X(199704)81:2<193::AID-SCE5>3.0.CO;2-A

Peterson, N. S., & Jungck, J. R. (1988). Problem-posing, problem-solving and persuasion in biology education. *Academic Computing*, *2*(6), 14–17 & 48–50.

QUBES. (2017). Math Modeling Hub. Retrieved August 24, 2017, from https://mmhub.qubeshub.org

QUBES. (2017). QUBES – Home. Retrieved August 24, 2017, from https://qubeshub.org/

Railsback, S. F., & Grimm, V. (2011). *Agent-based and Individual-based Modeling: A Practical Introduction.* Princeton, NJ: Princeton University Press.

Rothstein, D., & Santana, L. (2011). Teaching students to ask their own questions. *Harvard Education Letter*, *27*(5), 1–2.

Schuchardt, A. M., & Schunn, C. D. (2016). Modeling scientific processes with mathematics equations enhances student qualitative conceptual understanding and quantitative problem solving. *Science Education*, *100*(2), 290-320. http://doi.org/10.1002/sce.21198

Schwarz, C. V, Reiser, B. J., Davis, E. A., Kenyon, L., Fortus, D., Shwartz, Y., … Ache, A. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654. http://doi.org/10.1002/tea.20311

Schwarz, C. V., & White, B. Y. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and instruction*, *23*(2), 165-205. http://doi.org/10.1207/s1532690xci2302_1

Sheehy, R. (2017) Population genetics simulation program. Retrieved August 24, 2017, from http://www.radford.edu/~rsheehy/Gen_flash/popgen/

Sins, P. H., Savelsbergh, E. R., & van Joolingen, W. R. (2005). The Difficult Process of Scientific Modelling: An analysis of novices' reasoning during computer-based modelling. *International Journal of Science Education*, *27*(14), 1695-1721. http://doi.org/10.1080/09500690500206408

Society for Industrial and Applied Mathematics (SIAM). (2017). Math Modeling Challenge Resources. Retrieved August 24, 2017, from https://m3challenge.siam.org/resources

Society for Industrial and Applied Mathematics (SIAM). (2012). *Modeling across the curriculum I*. Retrieved August 24, 2017, from http://www.siam.org/reports/modeling_12.pdf

Society for Industrial and Applied Mathematics (SIAM). (2014). *Modeling across the curriculum II*. Retrieved August 24, 2017, from http://www.siam.org/reports/ModelingAcross%20Curr_2014.pdf

Soderberg, P., & Price, F. (2003). An examination of problem-based teaching and learning in population genetics and evolution using EVOLVE, a computer simulation. *International Journal of Science Education*, *25*(1), 35-55. http://doi.org/10.1080/09500690110095285

Speth, E. B., Shaw, N., Momsen, J., Reinagel, A., Le, P., Taqieddin, R., & Long, T. (2014). Introductory biology students' conceptual models and explanations of the origin of variation. *CBE Life Sciences Education*, *13*(3), 529–539. http://doi.org/10.1187/cbe.14-02-0020

Stewart, J., Passmore, C., & Cartier, J. (2010). Project MUSE: Involving high school students in evolutionary biology through realistic problems and causal models. *Biology International*, *47*, 78-90.

Svoboda, J., & Passmore, C. (2011). The Strategies of Modeling in Biology Education. *Science & Education*, *22*, 119–142. http://doi.org/10.1007/s11191-011-9425-5

Thompson, P. W. (2011). Quantitative reasoning and mathematical modeling. In L. L. Hatfield, S. Chamberlain, & S. Belbase (Eds.), *New Perspectives and Directions for Collaborative Research in Mathematics Education*. Laramie, WY: University of Wyoming.

Tomasi, J. (1988). Models and modeling in theoretical chemistry. *Journal of Molecular Structure: THEOCHEM*, *179*(1), 273–292. http://doi.org/10.1016/0166-1280(88)80128-3

Understanding Science. (2016). The Real Process of Science. Retrieved August 24, 2017, from http://undsci.berkeley.edu/article/howscienceworks_02

Weisstein, A. E. (2011). Building mathematical models and biological insight in an introductory biology course. *Mathematical Modelling of Natural Phenomena*, *6*(6), 198-214. http://doi.org/10.1051/mmnp/20116610

WikiPathways. (2017) WikiPathways. Retrieved August 24, 2017, from http://wikipathways.org

Williamson, B. (2015) AP Biology Artificial Selection Data Analysis with ExcelHD. Retrieved August 21, 2017, from https://www.youtube.com/watch?v=5ggSWuEzxeM

Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science Education*, *92*, 941–967. http://doi.org/10.1002/sce.20259

Figure 1

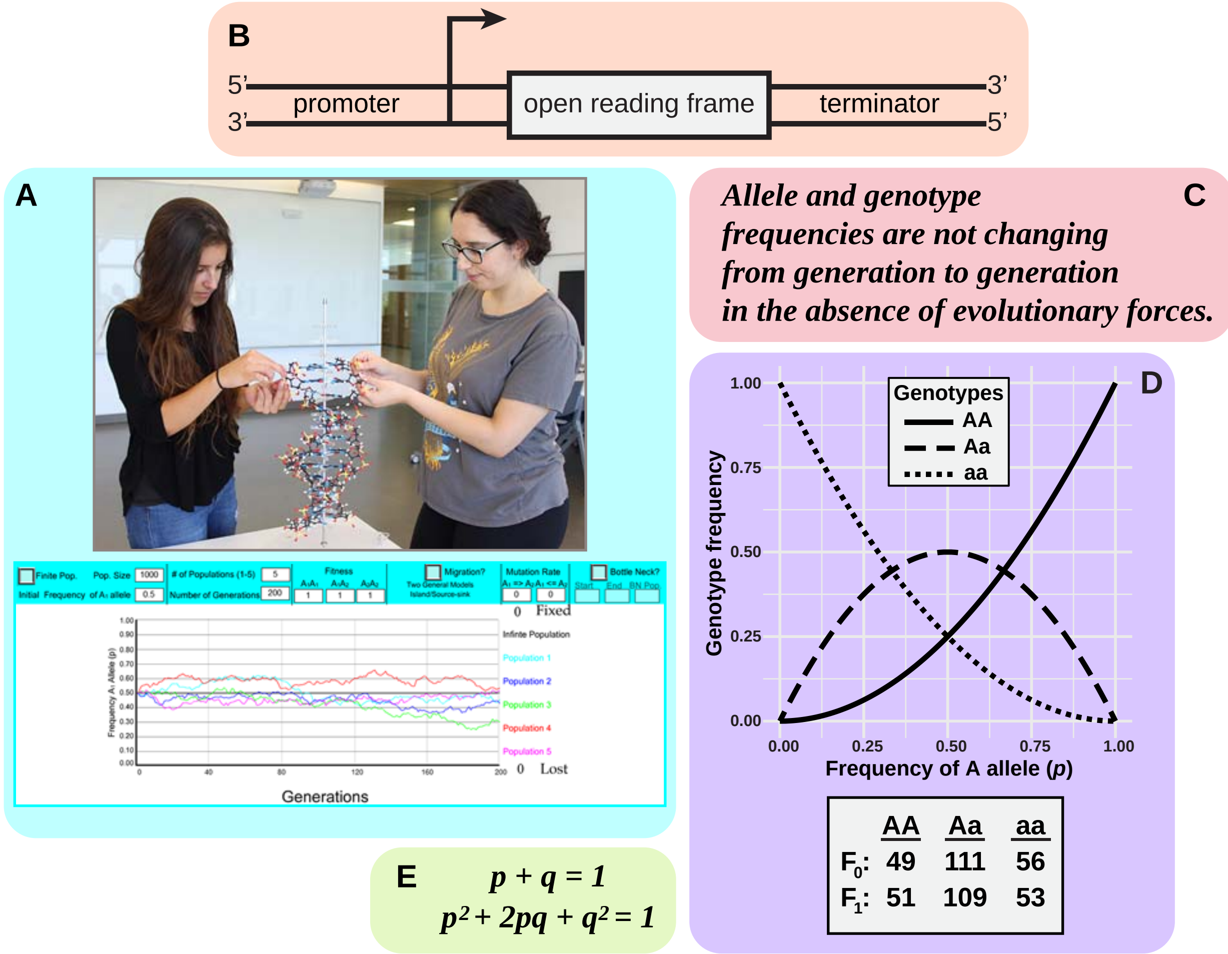

Photo credit: Tiffany Jonick

Figure 2

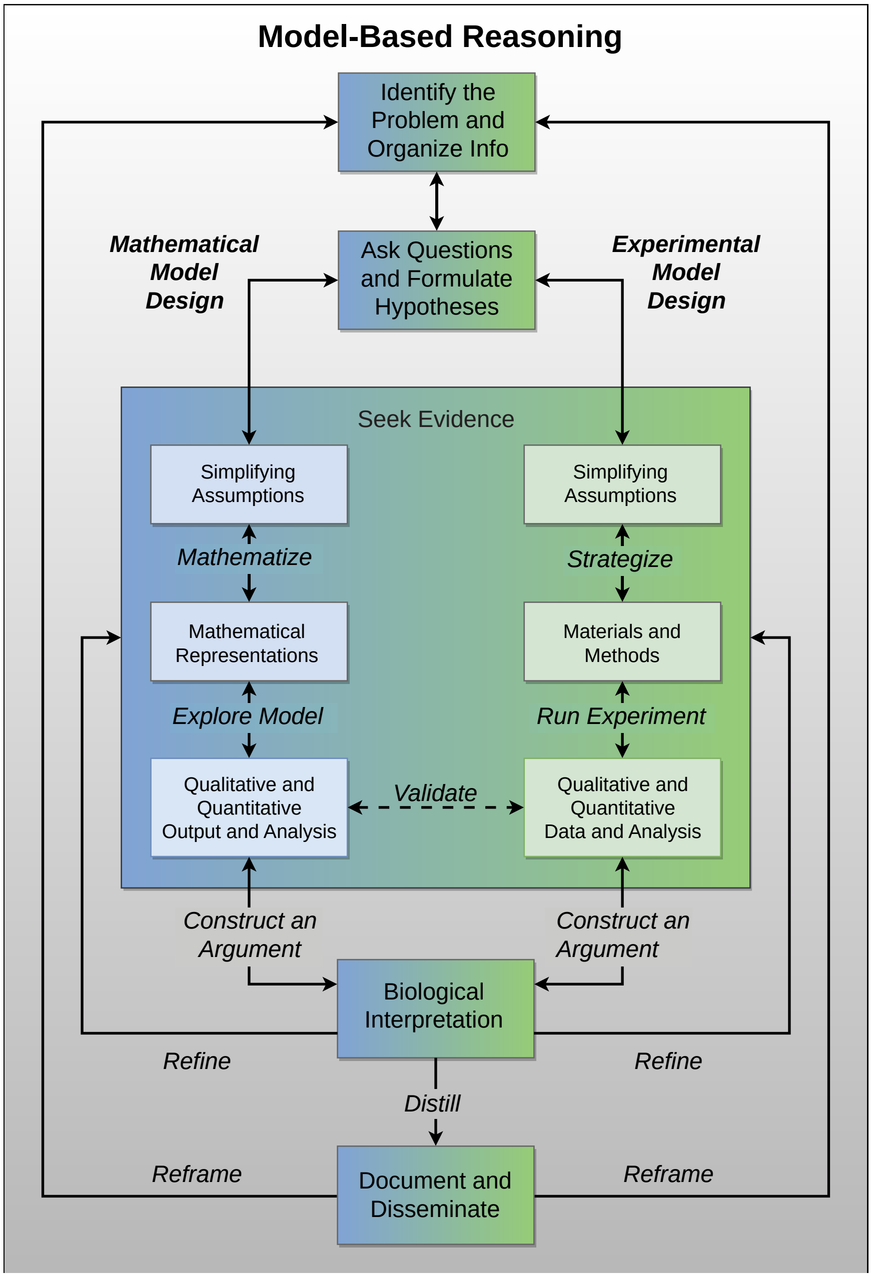
Model-Based Reasoning
Identify the Problem and Organize Info
Ask Questions and Formulate Hypotheses
Mathematical Model Design
Experimental Model Design
Seek Evidence
Simplifying Assumptions
Simplifying Assumptions
Mathematize
Strategize
Mathematical Representations
Materials and Methods
Explore Model
Run Experiment
Qualitative and Quantitative Output and Analysis
Validate
Qualitative and Quantitative Data and Analysis
Construct an Argument
Construct an Argument
Biological Interpretation
Refine
Refine
Distill
Reframe
Reframe
Document and Disseminate

Figure Legends

Figure 1. Different model representations. (A) a physical DNA model put together by students (see also, Cooper et al., 2017) and a screenshot of the PopGen simulator (http://www.radford.edu/~rsheehy/Gen_flash/popgen/) are examples of experiential model representations; (B) a schematic of a gene is an example of a visual model representation; (C) a statement of the null hypothesis of the Hardy-Weinberg model is an example of a verbal model representation; (D) a graph and data table of genotype frequencies are examples of numerical model representations; (E) the Hardy-Weinberg equations are an example of a symbolic model representation. Photo credit: Tiffany Jonick.

Figure 2. The parallel and iterative nature of the process of mathematical modeling and the process of experimental science as instances of model-based reasoning. When approaching a new problem, one begins by identifying the problem and organizing contextual information before proceeding to asking questions and formulating hypotheses, which then inform the seeking of evidence. Evidence can be obtained through either mathematical models or experiments (shown), or other avenues such as observational field studies, evolutionary reconstruction approaches, or meta-analyses. In each track (left and right), the steps of the mathematical and experimental model design have clear parallels. Validation (dashed arrow) occurs when experimental data are compared to model output or vice versa. The analysis of mathematical model and experimental model results are used to construct an argument for a particular biological interpretation, which is documented (Grimm et al., 2014) and disseminated to other scientists. This, in most cases, leads to even more questions. Each arrow is a modeling activity that can be performed with students. While this diagram was drawn in a top-to-bottom, linear fashion to facilitate easy viewing, we recognize that actual practice may be messier, requiring entering the diagram at different points, traversing the steps in a different order, and repeating steps (Eaton et al., 2017; Understanding Science)[1]

Footnote

[1] Interestingly, this figure, depicting the relationships and parallels between mathematical and experimental approaches to modeling, is itself a model. It has served an important purpose in the negotiation of our shared understanding of modeling over the course of collaboratively writing this paper. We have actively used this model as a point of focus during our attempts to articulate our claims about the modeling process. Questions like, "What exactly does this box represent?", "Why is this word used instead of another?", and "Why are there unidirectional arrows here but bidirectional arrows in another place?" have been asked by the biologists and mathematicians to each other when developing our ideas. This figure has been refined many times.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017

Table 1: Five types of model utility as described by Odenbaugh (2005) with example models.

Model Utility	Example Model
Simple, unrealistic models for exploring complex systems	• Using Tetrahymena is used as a model organism to study telomeres because it has tens of thousands of short linear chromosomes (Kain, 2009). • A model of an epidemic with different initial populations of susceptible, infected, and resistant individuals could be explored with different rules for transmission and recovery to provide insights into how different diseases spread through a population (Allen et al., 2008; Weisstein, 2011; Just et al., 2015).
Exploring unknown possibilities	• Building 3D models based on predicted protein structures could be used to understand drug-target interactions. • Agent-based models could be used to identify simple interaction rules that can lead to different emergent population level behaviors like flocking (Macal and North, 2006; Railsback and Grimm, 2011).
Developing conceptual frameworks	• A pathway diagram is a conceptual model summarizing experimental results (examples can be found at WikiPathways, http://wikipathways.org; Kutmon et al., 2016). • The Hardy-Weinberg null model can provide a starting point for explaining diverse evolutionary forces.
Making accurate predictions	• Data-driven population models of fish stocks inform sustainable harvests. • An enzyme kinetic model of pyruvate carbon distribution in lactic acid bacteria accurately predicted which genes to manipulate to increase flavor compound production (Hoefnagel et al., 2002).
Generating causal explanations	• A common garden experiment was used to determine whether differences in traits among populations of a plant species is due to genetic differences or phenotypic plasticity (Cordell et al., 1998). • The Hodgkin-Huxley symbolic model of ion flow across cell membranes helps to explain the all or none firing of action potentials.

Box 1: Definitions of Terms

Model: A simplified, abstract or concrete representation of relationships and/or processes in the real world, constructed for some purpose (Eaton et al., 2017).

Model utility: The purpose(s) for constructing and using the model, *e.g.* developing conceptual frameworks or making accurate predictions (see Table 1; Odenbaugh, 2005).

Modeling: An iterative process in which a model is proposed, explored, and refined (the arrows of Figure 2).

Model-based Reasoning: Forms of inquiry based on the process of modeling; using models to understand biological concepts.

Mathematization: The modeling process of going from a visual schematic or verbal description of the model and assumptions to a symbolic mathematical model representation.

Model refinement: Modifying aspects of the model, including changing the objects, processes and/or relationships.

Model exploration: Depending on the type of mathematical model, model exploration can consist of mathematical analyses or computer simulations to observe the behavior of the model as a function of its assumptions, inputs, and parameters.

Model validation: The process of assessing a model's output and assumptions with regard to its desired utility (is it addressing our goals?) and accuracy (is it consistent with other lines of evidence, *e.g.*, experimental data, observations and/or different models?).

Reframe: Incorporating the model and results into the broader set of scientific work, leading to new questions, hypotheses, or foci for scientific exploration.

Box 2: Examples of the range of ways that students can perform different modeling activities using the same HWE model.

Model description: Have the students discuss the utility of the null HWE model. Have students describe how the assumptions built into the HWE model limits the conditions under which it can usefully be applied. Have students identify the variables in the model, their attributes and measures. Link variables to assumptions by asking why these variables are included in the model. How do the attributes of the variables lead to relationships that result in the model?

Metamodeling: Have students keep a journal where they reflect upon the modeling process, answering questions such as: How did clarifying your assumptions help you develop a better model?, How did you determine if the results were biologically valid?, What was most surprising about your findings?, How has the modeling activity given you different insight into the biological problem?

Model use: Have students work with a small empirical dataset as an anchoring phenomenon that allows them to calculate both allele and genotype frequencies for a population and then test that data against the expectations generated by the HWE model.

Model exploration: Have students test the boundary conditions of the model or add alleles to the model to more deeply understand the quantitative relationships between allele frequencies and genotype frequencies under HWE model conditions.

Moving between multiple model representations: Have students ask their own questions about an anchoring phenomenon and then create a sketch or diagram (visual model). Have them then brainstorm what quantitative aspects could be added to the model.

Model development: Have students derive the HWE model after working with a physical "bean-bag" genetics simulation and collecting data on the relationships between allele and genotype frequency.

Model refinement: Have students write out (starting in English) changes to the physical bean-bag genetics simulation in order to account for one of the evolutionary forces that can influence allele frequencies. Then have them try to formalize it into the HWE model equation or have them build a spreadsheet simulation of the force acting on a population.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3215v1 | CC BY 4.0 Open Access | rec: 1 Sep 2017, publ: 1 Sep 2017