Strategies of Model Building in Population Genetics

Anya Plutynski[†‡]

In 1966, Richard Levins argued that there are different strategies in model building in population biology. In this paper, I reply to Orzack and Sober's (1993) critiques of Levins and argue that his views on modeling strategies apply also in the context of evolutionary genetics. In particular, I argue that there are different ways in which models are used to ask and answer questions about the dynamics of evolutionary change, prospectively and retrospectively, in classical versus molecular evolutionary genetics. Further, I argue that robustness analysis is a tool for, if not confirmation, then something near enough, in this discipline.

1. Introduction. There are two central aims in this paper. First, the aim is to defend a sensible middle ground between Levins, on the one hand, and Orzack and Sober, on the other, regarding the aims and strategies of model building. Second, this argument will be supported with examples from classical and molecular evolutionary genetics.

In 1966, Levins famously argued that population biologists must trade off generality, precision, and realism in model building. There is no "single, best all-purpose model" (1966, 7). There are, rather, three strategies of model building: modeling that sacrifices generality for the sake of realism and precision (what he calls type I), modeling that sacrifices realism for generality and precision (type II), and modeling that sacrifices precision for generality and realism (type III). Levins' argument is largely based on examples, and he does not define his terms precisely, so it is rather difficult to determine whether his claims are true. Orzack and Sober reply to Levins

in a 1993 paper. They define the terms "generality," "precision," and "realism" as follows:

(G) If one model applies to more real world systems than another, it is more general.

(R) If one model takes account of more independent variables known to have an effect than another model, it is more realistic.

(P) If a model generates point predictions for output parameters, it is precise. (1993, 534)

Orzack and Sober argue that Levins is mistaken. They claim that there are not necessarily trade-offs between these three virtues in model building. For instance, they show that one model may be both as general as and more realistic (in their senses of the term) than a second model if, for instance, the latter is a special case of the former. Thus, generality and realism can be maximized simultaneously. (Their example is density-independent versus density-dependent models of population growth. The former can be "nested" in the latter. Or it is an instance of the more general case: only the value of the variable describing the effects of density is zero.) Moreover, they show how we may increase realism and make no sacrifice in generality or precision to a model, simply by adding new independent variables. In other words, they show how in some cases, at least, it is possible to maximize all three virtues. The claim that these trade-offs are inevitable seems to be defeated. They conclude:

> Conflicts between models should be understood in terms of their substance, not their styles. Levins's thesis suggests that it is a matter of taste which two of the three desiderata a scientist chooses to pursue. If different models really did have different aspirations, it would be a mistake to attempt to decide which model is best. Models in different categories would then "go their own ways," answerable to their own standards. The perennial difficulties of understanding nature make this pluralism a temptation. It is, however, a temptation to be resisted. One way to do this is by separating models by their assumptions and by their fit to the data. (1993, 545)

I agree with Orzack and Sober that if it were an implication of Levins' view that models cannot be compared or evaluated or that which model one prefers is a matter of taste, then we ought to reject Levins' view. However, it's not clear that Levins' account has this implication.

I take it that the correct reading of Levins is that models serve different functions at different stages in the process of confirming theories. As Levins is at pains to emphasize in his reply, model building is a process, and the process can be developed along different lines to meet different

needs in the process of better understanding how the world works. (For excellent defenses of this interpretation, see Wimsatt [1981, 2002], Odenbaugh [2004], and Weisberg [n.d.].)

In a reply to their 1993 paper, Levins calls Orzack and Sober's approach to theory building and confirmation a "formalized" one. Here is a sketch of what I take this to mean. On Orzack and Sober's view, we construct models and then proceed to determine whether they fit the world. Better or worse fit can be engineered by adding or removing variables, or making models more or less realistic or precise. I don't believe that Levins has any conflict with Orzack and Sober over this characterization of model building as a regulative ideal. However, Levins is not concerned with ideals so much as what scientists actually do. In science such as ecology or evolution, which investigate complex systems, one cannot always generate a model that accurately represents a system in all its details. Indeed, doing so might be computationally impossible; rather, one can answer a number of limited questions about the dynamics of such a system under certain simplifying assumptions. Drawing on the results of several different models that have been constructed to address several such questions, one can arrive at a robust theorem. Such a theorem is not one that scientists would stake their careers on, but if one can discharge the assumptions of the different models, one can say that such theorems are indirectly or partially confirmed. Levins' claim that there is a trade-off between the three virtues does not amount to the logical claim that it is impossible to maximize all three, but a pragmatic one (Wimsatt 1981; Odenbaugh 2004). Sometimes population biologists sacrifice precision because they want a characterization of the key dynamics of the system, without predictive accuracy. This is one stage in indirect confirmation of a robust theorem, which is often the best one can get in sciences that are attempting to answer questions about complex, dynamic processes with multiple interacting causes (Weisberg n.d.).

My claim in this paper is that this strategy is one that is often used in the context of evolutionary genetics. In order to demonstrate this, I must first explicate two distinctions Orzack and Sober make early on in the paper, between models "as mathematical truths" and models "as empirical claims." Orzack and Sober introduce this distinction as follows:

> Consider any mathematical model said to describe or explain a particular phenomenon. The first distinction is the one between the model as a mathematical statement and the model as an empirical claim about part of the physical world. Within the second class, we can distinguish models whose parameter values are left unspecified and those whose values are specified. (1993, 535)

Orzack and Sober's example of a mathematical statement is the fol-

lowing. They ask us to consider a one-locus, two-allele model of viability selection in a diploid population. Three genotypes AA, Aa, and aa are assigned fitnesses wAA, wAa, and waa. "It is," they write, "a mathematical truth that if selection is the only force acting in the population, then a stable polymorphism will occur when wAa > wAA, waa."

Surely it is a true mathematical statement that any system so described will reach equilibrium. But this is not simply mathematical analysis; this model refers to empirical features of a system, such that this sort of analysis will serve any number of roles in confirming some empirical hypothesis at some later stage of the game. Analytical modeling of this sort may generate quite novel and surprising discoveries. For example, one might find that traits evolving under sexual selection are less likely to be sex-linked than autosomal (Kirkpatrick, personal communication). Theoretical models like this one (given that the assumptions are empirically well supported) may be used as evidence supporting assumptions in other contexts. And even when the assumptions are false, they can provide us with useful information about the constraints on various kinds of processes (Wimsatt 1987, 2002). This is why scientists are as interested (if not more interested) in discoveries of this sort as novel empirical observations. They want to understand what sorts of things are likely, possible, and necessary under certain assumptions. Understanding the dynamics of a system under some description, especially one that is enormously difficult to manipulate, requires the construction of an analytical model. When a system is complex, such as a population of interbreeding organisms, one useful means of discovering these properties of the system is to artificially "hold" other aspects of the world constant in an idealized mathematical model, allowing the model to count as a kind of artificial "bottle" experiment to test a hypothesis. This is a central strategy for posing and answering different kinds of questions in theoretical biology. It is in just such contexts that we might prefer a more general, but perhaps less precise, model. Moreover, depending on the question such a model is designed to answer, the standards for evaluation of analytical models such as this are rather different from the standards used to evaluate models built so as to generate very specific predictions for testing empirical claims.

According to Orzack and Sober, there are not distinct "types" of models with distinct "aspirations." However, Levins' point is not that these models per se are different, but that the construction of models exemplifying these different virtues enables scientists to answer different kinds of indirect questions about some system. These are stages in what Levins and, subsequently, Wimsatt will call "robustness" analysis. While I agree with Orzack and Sober that there are not different "types" of models per se, I think that the subtler Levins thesis is correct. There are different kinds

of questions we might pursue and thus different strategies to model building.

2. Classical versus Molecular Evolutionary Genetics. An examination of modeling in classical population genetics and molecular evolutionary genetics is illustrative of this contrast. Though the two fields are overlapping more and more, there are some important and interesting contrasts to be made between them. Classical population genetics models are primarily prospective; the assumptions are idealized and often, strictly speaking, false. Many of the analytical models constructed by population genetics test general claims about the dynamics of evolution under certain assumptions. They give general answers about evolutionary processes under certain idealized conditions.

Molecular evolutionary genetics models, for the most part, have been developed in the last 40 years to test very specific empirical hypotheses about rates of evolution, sequence divergence, and inter- and intraspecific variation. Many of these models are retrospective; beginning with a sequence, one makes inferences about time to the most recent common ancestor or compares rates of change across sequences. In other words, the premium is on generating a model so that an empirical hypothesis can be disambiguated from data. What are studied are the products of evolutionary processes rather than the processes themselves.

One important lesson to be gained from Levins is the way in which theory-driven and empirically based models can be used jointly to understand the causes of some pattern or process in nature. Especially when we are investigating a process that occurs over very long time scales, with multiple causes, we may not be able to do the sort of "turn the crank" kind of confirmation that Orzack and Sober seem to endorse. Rather, we take several different models, some more general and some more precise, and arrive at, if not the exact explanation of some state of affairs, the most likely one.

Analytic models can serve to provide mathematical derivations of simple, necessary generalizations about evolutionary dynamics. For example, here is an elegant argument made by Kimura (1968). Kimura found that:

- The rate of substitution or number of mutants reaching fixation per unit time (K) equals the rate of mutation, for neutral mutations.
- Substitution of neutral mutations is independent of population size.

How did he arrive at this result? Imagine that there are mutations that are entirely neutral in their effects. They appear at a certain rate, u, in a population of organisms of some size, N. The rate at which such neutral mutations arise per generation in a diploid population at some particular

locus will be $2Nu$ per generation. The probability of fixation of these mutations will be $P = 1/2N$. The reason is that for neutral mutations, the probability of fixation is equal to its initial frequency ($1/2N$); one neutral mutation has a chance of becoming fixed in a population equal to that of any other neutral mutation (assuming perfectly independent assortment and no linkage!). (Its probability of being lost is much higher [$(2N - 1)/2N$, approximately one]; but it does have a small chance of success.) Finally, the rate of substitution of new mutations is just the rate at which they arise times the probability of fixation, $K = 2NuP$. When we substitute $1/2N$ for P, we get $K = u$. Or the rate of substitution of neutral alleles is just equal to their rate of mutation and independent of population size.

What is the status of this argument and conclusion? First, the argument is a simple mathematical derivation. As such, the conclusion is a true mathematical statement, $K = u$, à la Orzack and Sober. The derivation is a true mathematical statement, but it entails a true (counterfactual) empirical claim. That is, if the assumptions of the model hold, we now know an empirical fact to also hold: that the rate of substitution of neutral mutations equals the mutation rate. And we can use this analytic result to answer other empirical questions.

Notice how many idealizing (or false) assumptions are in Kimura's conditional. Were there to be an entirely neutral mutation, then the chance of substitution equals the rate of mutation. We must assume perfect neutrality, and that change in one nucleotide site is independent of changes at other sites. This is almost never, strictly speaking, true. This is, as it were, a general and precise, but unrealistic, model.

According to Orzack and Sober, we should improve on such a model by adding in more parameters, perhaps. But, depending on the context of application of such a result, the assumptions can be discharged; in other words, the assumptions are harmless if they are irrelevant to the empirical hypothesis we're interested in testing. In this case, the important function of this model is as a null model. One assumes this equality to hold, and then a failure of this equality is taken to indicate nonneutral evolution.

The Kimura model permits construction of data-driven models in the context of molecular evolutionary genetics. For instance, consider the MacDonald-Kreitman test (1991). In 1991, MacDonald and Kreitman developed a test to determine whether most of the mutations that have contributed to genetic variation in populations were neutral or nearly neutral. The test compares degrees of polymorphism within and between species. The neutral theory predicts that the variation within populations and the differences between populations are mainly due to neutral or nearly neutral mutations. The rate of molecular evolution indicated by

comparisons between species should be positively correlated with within-population variation.

MacDonald and Kreitman originally applied their test to the alcohol dehydrogenase gene in two closely related species of *Drosophila*. For the two species, they measured the number of fixed differences for replacement and silent nucleotide changes and divided one by the other to get a ratio. They also measured polymorphism within each species for replacement and silent sites and expressed this as a ratio. The two ratios should be equal if all the changes are neutral. This is so because the rate of neutral evolution equals the mutation rate, and the ratio of species differences for replacement and silent sites should therefore equal the ratio of neutral replacement changes versus neutral silent sites.

They found a high ratio of fixed replacement site to synonymous or silent sites. The best explanation of this, they argued, was adaptive fixation.

3. Critique and Reply. Orzack and Sober reply that this argument confounded a number of issues (personal communication). First, they will not disagree that one might use models as proofs of principle. Indeed, Sober has drawn attention to several examples of this strategy in earlier work (Sober 1984). However, this use of modeling is properly part of the "context of discovery," not the "context of justification." Their critique, Sober alleges (personal communication), was of Levins' claim to distinctness of the variety of models in the context of justification.

However, I think that this is exactly why Orzack and Sober have missed the thrust of Levins' piece. In the actual practice of science, there are many cases in which a scientist's deployment of a model is not restricted to one context alone. The derivation of a mathematical statement may constitute a kind of "discovery" (if they like) about the constraints on processes, but the same mathematical statement may be used as one stage along the way to confirmation of some independent empirical hypothesis, as illustrated above.

Second, Levins was talking of strategies of model building, not models per se. That is, he was interested in why scientists build models, not just the models themselves, and how their choice of modeling strategy reflected the kind of question they wished to answer. Scientists are interested in answering questions about not only whether some model accurately reflects the world, but also whether or not some pattern or process is general, likely, possible, or necessary, given some initial conditions. Filing the answers under the distinct and exclusive contexts of "discovery" or "confirmation" is, I think, not an easy task.

Finally, Orzack and Sober might argue that the contrast I make between modeling in classical population genetics and in molecular evolutionary

genetics is an "apples and oranges" comparison. I take this to mean that both sorts of models address both theoretical problems and empirical questions. Of course, both do indeed do so. My claim was not that they exclusively do one or the other. Rather, my claim was that the former, "more often than not," aims to address questions about general prospective patterns and processes in evolution. This is so for a rather significant historical reason: when Haldane (1924a, 1924b), Wright (1931), and Fisher (1930) first developed these models, they did not know the structure or means of transmission of the hereditary materials. Thus, these models were necessarily "as if" models—making a number of assumptions about how evolutionary change can and should be represented at the genetic level. Many of the same modeling strategies that were initiated by these authors are still used today in theoretical population genetics. These kinds of models address very theoretical "as if" questions exactly because they are theory-laden models: they assume that traits may be represented by changes in allele frequency in our models, for instance, or that assortment is independent. We now have very good reason to be suspicious of these assumptions. Nonetheless, for the purposes of addressing broad general questions about the evolutionary process, such theoretical models are enormously useful.

In molecular evolutionary genetics, in contrast, I argued that such models are "for the most part" developed to test very specific empirical hypotheses about rates of evolution, sequence divergence, and so forth. The reason why this is so is that these models are constructed with the data, as it were, "in front of them." While there is significant overlap between the models in classical and molecular genetics, there are significant differences. As Ewens (2004, 288) remarks, "To some extent the purely mathematical theory [classical population genetic theory] carries through to the molecular level, with the nucleotide frequency replacing the gene frequency as the primary variable, but clearly, new models and viewpoints, as well as new 'typical' values of various fundamental parameters, are necessary at the molecular level." However, he also remarks on the contrasts:

> Molecular theory is "dynamic" as opposed to "static"—mutations lead to new allelic types rather than back to current or previously existent types.
>
> The assumption that different sites within one gene evolve independently is particularly questionable.
>
> The actual genetic material is known.
>
> Molecular considerations are retrospective as opposed to prospective. (Ewens 2004, 289)

Molecular evolutionary genetics is, in some sense, tracking a different kind of process—changes at the level of single base pairs—than that envisioned by the founders of classical population genetics. And they are working, not from "as if" assumptions, but from databases that allow them to compare actual genetic variation in different populations and retrodictively derive the causal story that leads to that pattern of variation. Thus their models are more often "data driven" in the sense that they are working from the data in construction of models rather than working from "as if" assumptions about heredity and evolution.

4. Conclusion. On the one hand, the aim of model building in classical population genetics is more often than not the production of models that provide proof of principles. On the other hand, the aim of model building in molecular evolutionary genetics is to find a way to evaluate data concerning very specific empirical hypotheses.

Orzack and Sober are correct that there are not distinct "types" of models in the sense that the same model can exhibit type II or type III characteristics, depending on the mode of application. However, the subtler reading of Levins—namely, that there are distinct functions that modeling serves at different stages of inquiry—seems more plausible.

By deemphasizing the merely "mathematical" part of the work of theoretical modeling, Orzack and Sober shift the focus to confirmation and testing. Thus, the stages of model building, where one builds a "scaffold" or derives mathematical relationships that enable one to make counterfactual claims about the dynamics of evolution, are merely preliminary. However, I wish to suggest that this part of the modeling process satisfies a different set of aims that are significant to both theory testing and determining what will count as adequate tests of the theory. Thus they are not "distinct from" the context of justification, insofar as they determine the very conditions on the possibility of justification and frequently do themselves justify assumptions made in testing empirical claims in other contexts. As Wimsatt (1987, 2002) has argued, less realistic models often function to provide information about constraints on processes. Or use of several such "analytic" models can allow the derivation of robust theorems. And when analyzing systems with many moving parts, robustness analysis can give us, if not confirmation, then something near enough.

REFERENCES

Ewens, W. (2004), *Mathematical Population Genetics*. Interdisciplinary Applied Mathematics, vol. 27. 2nd ed. New York: Springer.
Fisher, R. A. (1930), *The Genetical Theory of Natural Selection*. Oxford: Clarendon. 2nd ed., New York: Dover, 1958.

Haldane, J. B. S. (1924a), "A Mathematical Theory of Natural and Artificial Selection, Part 1," *Proceedings of the Cambridge Philosophical Society* 23: 19–41.

——— (1924b), "A Mathematical Theory of Natural and Artificial Selection, Part II. The Influence of Partial Self-Fertilization, Inbreeding, Assortative Mating and Selective Fertilization on the Composition of Mendelian Populations, and on Natural Selection," *Proceedings of the Cambridge Philosophical Society* 1: 158–163.

Kimura, M. (1968), "Evolutionary Rate at the Molecular Level," *Nature* 217: 624–626.

Levins, R. (1966), "The Strategy of Model Building in Population Biology," *American Scientist* 54: 421–431.

——— (1993), "A Response to Orzack and Sober: Formal Analysis and the Fluidity of Science," *Quarterly Review of Biology* 68: 547–555.

MacDonald, J., and M. Kreitman (1991), "Adaptive Protein Evolution at the Adh Locus in *Drosophila*," *Nature* 351: 652–654.

Odenbaugh, J. (2004), "Complex Systems, Trade-offs and Mathematical Modeling: A Response to Sober and Orzack," *Philosophy of Science* 70 (December): 1496–1507.

Orzack, S., and E. Sober (1993), "A Critical Assessment of Levins' 'The Strategy of Model Building (1966),'" *Quarterly Review of Biology* 68: 534–546.

Sober, E. (1984), *The Nature of Selection*. Cambridge, MA: MIT Press.

Weisberg, M. (n.d.), "Robustness."

Wimsatt, W. (1981), "Robustness, Reliability, and Overdetermination," in M. Brewer and B. Collins (eds.), *Scientific Inquiry and the Social Sciences*. San Francisco: Jossey-Bass, 124–163.

——— (1987), "False Models as Means to Truer Theories," in M. Nitecki and A. Hoffman (eds.), *Neutral Models in Biology*. London: Oxford University Press, 23–55.

——— (2002), "Using False Models to Elaborate Constraints on Processes: Blending Inheritance in Organic and Cultural Evolution," *Philosophy of Science* 69 (Proceedings): S12–S24.

Wright, S. (1931), "Evolution in Mendelian Populations," *Genetics* 16: 97–159. Reprinted in William Provine (ed.), *Evolution: Selected Papers by Sewall Wright*. Chicago: University of Chicago Press, 1986.