

Laboratory 1: Forces of evolution
Handed out: September 19/20
Due: October 3

Tools: Population genetics & evolution simulation - *Populus* java computer program, version 5.3 & PopG computer program

Populus: Download site and installation instructions at the University of Minnesota located at <http://www.cbs.umn.edu/populus/Download/download.html>. (.jar files available for PC, Mac, Linux-Unix)

Make sure you have checked for java compatibility according to the instructions on that page. A PDF file of instructions for all modules is available in the labs section. Check with us if you need to run the program on the MIT server workstations rather than your own computer.

PopG: Download site and installation instructions at the University of Washington located here.

This lab will allow you to explore the impact of various *forces* on allele dynamics. Here we will present an overview of the laboratory, along with some introductory questions. This is followed by a link to a pdf file for the full laboratory itself. We also include a 'review' section on the Hardy-Weinberg equilibrium theory, just to make sure that is understood before you tackle evolutionary forces.

Overview: some methodological preaching - Studying evolutionary processes using models

Just about any process can be approached with a mathematical model. One needs to be able to identify objects (such as genotypes), have a way to quantify (such as counts), and have reasonable guesses about factors that influence changes in the objects (such as mortality, reproductive success). If we want to understand how the relative abundance of different genotypes changes over generations, we can construct equations that represent the abundance of each genotype and the fitness of each genotype.

What biological phenomena can be studied with models?

Of course, a mathematical model must simplify the complexity of the real world. The basic question is, can we capture the essence of biological process with our simpler model? The model provides predictions about how organisms might react to differences in conditions (such as different fitnesses), but these predictions must be tested against the real world. If the models are too simple to accurately predict experimental outcomes or observations in the real world, then we can increase reality (?) to the simple equations by adding more complex interactions (e.g., allow fitness to change with the abundance of each genotype).

Part 0: Hardy-Weinberg warmup (no questions or computer use involved)

In order to get started, we have to make sure you have the basic "Newton's First Law" of population genetics well in hand – the Hardy-Weinberg equilibrium theory. Calculation of allele and genotype frequencies is central to what follows.

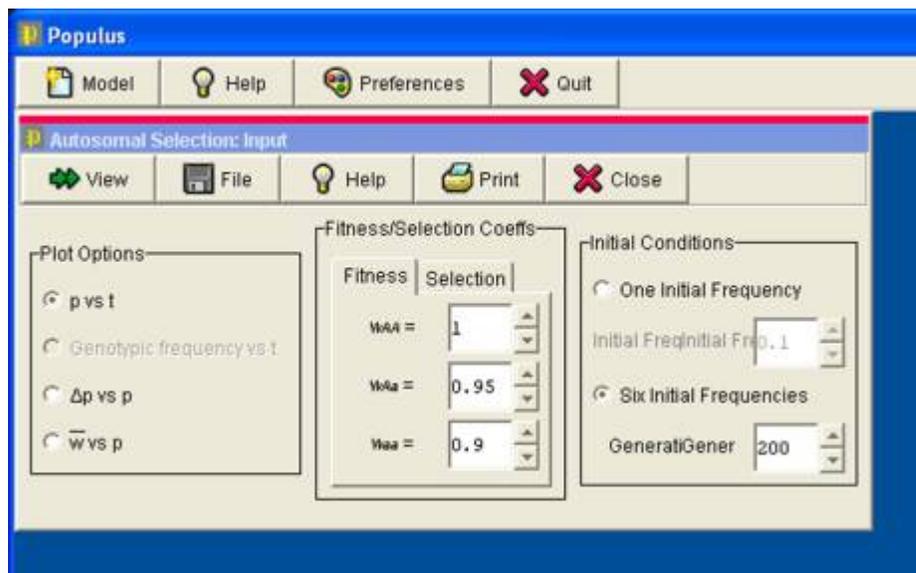
Population geneticists study frequencies of genotypes and alleles within populations rather than the ratios of phenotypes (external forms) that Mendelian geneticists use. By comparing these frequencies with those predicted by null models that assume no evolutionary

mechanisms are acting within populations, they draw conclusions regarding the evolutionary forces in operation. In a constant environment, genes will continue to sort similarly for generations upon generations. The observation of this constancy led two researchers, G. Hardy and W. Weinberg, to express an important relationship in evolution. The law that describes this relationship bears their names. The Hardy-Weinberg Equilibrium Theory serves as the basic null model for population genetics.

Every individual has *alleles* (=variants of the same gene) that were passed on from their parents. If we take all of the alleles of a group of individuals of the same species (that is, a population) we have what is called the *gene pool*. The frequency, or proportion, of individuals in that population that possess a certain allele is called the *allele frequency*. Populations can have allele frequencies, but individuals cannot. This makes populations a reasonable hierarchical unit, or level, to study evolution, *on the assumption* that evolution is basically the study of the change in allele frequencies over time. (You should be prepared to defend or attack this assumption! How is it true? How is it not true?)

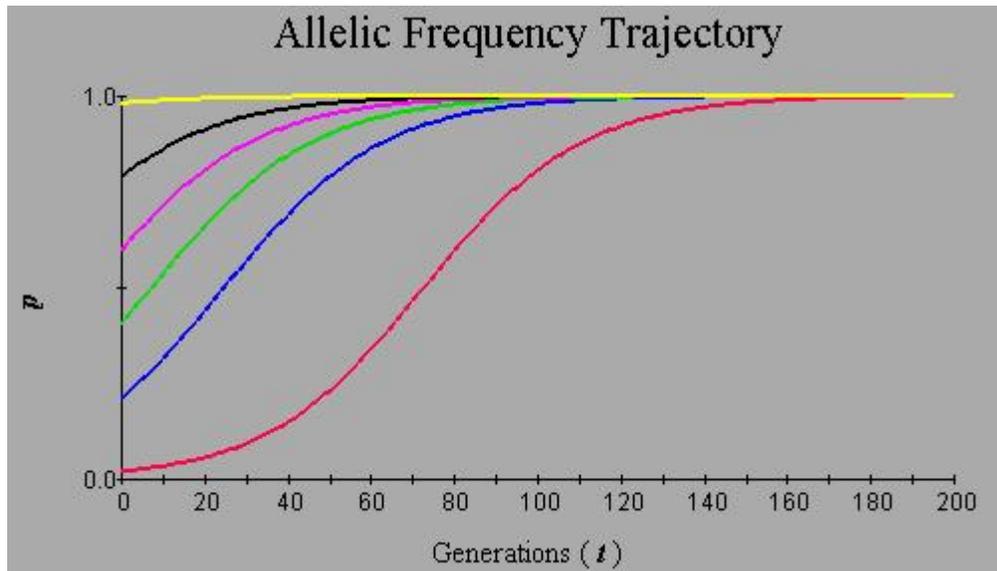
The rest of the lab: using *Populus* (parts 1 and 2a) and *PopG* (part 2b)

1. Start *Populus* by double-clicking on the *Populus* icon on the computer desktop.
2. Click in the Model box in the left side of the menu bar; select autosomal selection.
3. In Plot Options, keep p vs t ; later, you can change this to examine shifts in genotype frequencies. p is, of course, the frequency of the model A allele.
4. In the "Fitness/Selection Coefficients" box, make sure the "fitness" tab is selected. Later, you can select the "selection" tab, in which you examine the effects of different selection coefficients s .
5. Set genotype fitnesses according to the hypothesis you wish to test (see below).
6. Under "Initial Conditions" highlight the button for "Six Initial Frequencies."



Courtesy of Don N. Alstad, University of Minnesota.

7. Now you are ready to run the simulation: Click on the "View" button (green arrows). The following graph will appear:



Courtesy of Don N. Alstad, University of Minnesota.

Each color represents the predicted change in the frequency of "A" over 200 generations, given different starting frequencies. Note that the model predicts that A will become the only allele eventually (at equilibrium) no matter what the starting frequency. Using options at the top of the graph screen, you can save your results to a new file. You can save the graph as a picture, or you can save a text file with the data used to generate the graph. (Tip: On a Macintosh, it appears that is easiest to save a graph file as a pdf by using the Print menu, rather than saving it as a jpeg file. Saving jpegs appears to work fine on Windows and Linux.) Be careful to note where you save files.

Also, you should take notes on the important predictions of each run of the model. For example, if you reduce the heterozygote Aa and aa homozygote fitnesses, how does the time to reach equilibrium change?

Using PopG

PopG limits itself to a one-locus, 2 allele model, but lets one investigate all the interactions between selection, migration, mutation, and drift (by specifying the population size). Please follow the instructions on the same web page as the installation instructions. You will either execute the popg file in a terminal window (on Linux); or click or double-click the PopG icon (on a Macintosh or Windows). This will pull up a menu where where you can start a new run of 100 generations at a time, after specifying fitnesses, mutation rates, migration rates, and initial frequency of the A allele. We will not be using PopG until Part 2b of this lab.

Part 1: Populus warmup exercises

One of the tasks facing newcomers to the evolutionary biology simulation game is how to express and present results. The key question to ask yourself here is: What is the question that you are exploring? To gain some facility with the *Populus* program and this issue, here are two questions you should start with:

Question 1. *If an allele is lethal but recessive, how fast can selection remove it from a population?*

- For this question, you would set aa fitness to zero, AA and Aa to 1.0.
- Run the program, and determine the number of generations to fixation.
 - You may want to use "Options" on the graph menu to set up a grid.
- Write the number of generations as a function of starting frequency.
 - (If you can find an analytic relationship using the equations from the text/notes, so much the better)
- Report your results in the form of a short table.

Question 2. *How is the rate of elimination of a deleterious allele affected by the selection coefficient s ?*

- For this question, you will need to vary *aa* fitness (fitness = $1 - s$).
- Choose a single starting frequency.
- Run the simulation and note the time to fixation.
- Reset the simulation for a different fitness, and note the new time to fixation.
- You will need to set up a table with a column for *aa* fitness and a column for the corresponding time to fixation.

An Excel graph would be a good way to present this relationship.

In the last part of this lab, you will be asked to analyze real data and apply the *Populus* models to this data.

How are selection coefficients computed from real data? Here is a simple example, and a third question for Part I.

Question 3. *How can we calculate fitness values and selection coefficients from real data?*

Remember that selection does not act on a gene in isolation. Locus *A* has two alleles *A* and *a* and three genotypes.

We can calculate fitness and the force of selection on these genotypes- the selection coefficients.

Step 1. The most fit genotype is always set as 1. Suppose *AA* is most fit. The less fit genotypes are calculated as proportions of the most fit genotype. The s values are "selection coefficients" but often we shall just use the ratios directly, these are the *relative fitnesses*.

| Genotype | AA | Aa | aa |
|-------------|-------------|-----------------|-----------------|
| Fitness | W11 | W12 | W22 |
| Proportions | $W11/W11=1$ | $W11/W12=1-S_1$ | $W11/W22=1-S_2$ |

Step 2. You superimpose this calculation on the Hardy-Weinberg frequencies to compute the frequency *after* selection.

| Genotype | AA | Aa | aa |
|--|--------------|--------------|--------------|
| Frequency before selection | p^2 | $2pq$ | q^2 |
| Proportional contribution to next generation | p^2 W11 | $2pq$ W12 | q^2 W22 |

Step 3. Calculating selection coefficients from empirical data. You assume that the frequency before selection is in H-W equilibrium. The "proportional contribution to next generation" is what is *observed*.

| Genotype | AA | Aa | aa |
|--|-------------------|------------------|------------------|
| Frequency before selection | 0.25 | 0.50 | 0.25 |
| Proportional contribution to next generation | 0.35 | 0.48 | 0.17 |
| Relative survival value | $0.35/0.25 = 1.4$ | $0.48/0.50=0.96$ | $0.17/0.25=0.68$ |
| Relative fitness value | $1.4/1.4 = 1.0$ | $0.96/1.4=0.70$ | $0.68/1.4=0.40$ |
| Selection coefficient | $1-1 = 0$ | $1-0.7 = 0.3$ | $1- 0.4 = 0.6$ |

Step 4. Calculating fitness values & selection coefficients from real data: Applying Step 3 to actual numbers. *Biston betulia*, the peppered moth, and its response to industrial melanism (aka 'smoke from British factories') was studied by Kettlewell (1959). To test his hypothesis that there was selection favoring darker colored moths that could more readily hide on the pollution-darkened bark of trees, he did mark-recapture experiments in a polluted wood. He released 154 *carbonaria* (the melanic form) and 65 *typica* (lighter) moths for a total of 218. He recaptured 82 *carbonaria* and 16 *typica*. From this data you can calculate the selection coefficients, as follows. We give you the first 5 rows of the data table. **You** are to calculate the numbers that go in the last two rows: first relative fitness, and then the selection coefficients. Relative fitness, w , is just the survival ratios amongst all the types (so we normalize the value to 1.0 for the highest surviving type). The selection coefficient, s , is computed as in the table above.

| Genotype | CC (<i>carbonaria</i>) | Cc (<i>carbonaria</i>) | cc (<i>typica</i>) | Total |
|--|-----------------------------|-----------------------------|----------------------|-------|
| Number before selection | 77 | 77 | 65 | 219 |
| Number after selection next generation | 41 | 41 | 16 | 98 |
| Frequency before selection | $77/219=0.35$ | $77/219=0.35$ | $65/219=0.297$ | 1 |
| Frequency after selection | $41/98 = 0.42$ | $41/98 = 0.42$ | $16/98 = 0.16$ | |
| Relative survival | $0.42/0.35=1.20$ | $0.42/0.35=1.20$ | $0.16/0.3=0.53$ | |
| Relative fitness | | | | |
| Selection coefficients, s | | | | |

Question 3 (finally): With the fitness values that you just computed that presumably favor the *carbonaria* form, how long, in terms of # of moth generations, would it take to change the frequency of *carbonaria* observed in an unpolluted wood, 2%, to the 87% frequency Kettlewell observed in a polluted wood? *Note* that since the Populus model is continuous, but, obviously the moth populations are discrete, you'll have to reconcile this distinction.

Reference: Kettlewell, H. B. D. 1959. Darwin's missing evidence. *Scientific American* 200 (8): 48-53.

**Part 2: The forces of evolution - selection, mutation, migration, and drift
- *Populus* and *Popgen* exercises.**

OK, with your computer muscles all warmed up, please now proceed to Part 2 of the laboratory, a more extensive exploration of the interaction between different evolutionary forces.