

# Population Genetics Simulations

*Heath Blackmon and Emma E. Goldberg*

*last updated: 2016-04-02*

## Contents

|   |          |
|---|----------|
| <b>Introduction</b>                                     | <b>1</b> |
| Evolution Across Generations . . . . .                  | 1        |
| Launching the Population Genetics Application . . . . . | 2        |
| <b>Selection</b>  | <b>3</b> |
| Directional selection . . . . .                         | 3        |
| Overdominance . . . . .                                 | 4        |
| Underdominance . . . . .                                | 4        |
| <b>Genetic Drift</b>                                    | <b>5</b> |
| Drift Alone . . . . .                                   | 5        |
| Drift and Selection . . . . .                           | 5        |
| <b>Mutation</b>   | <b>6</b> |

## Introduction

Once an allele enters a population (by mutation or migration), its evolutionary fate is determined by selection and drift. In this lab, you will use a computer program to explore how the prevalence of an allele in a population (its frequency) is affected by a variety of factors—its initial frequency, whether it is dominant or recessive, the relative fitnesses of individuals bearing different alleles, and population size.

Throughout this lab, we will focus on a single locus with two alleles, A and a. We will assume that mating is random with respect to this locus, and that there is no migration from other population.

## Evolution Across Generations

The core of a population genetics model is to describe how allele frequencies change over time. A computer can quickly simulate many generations, rather than doing a ton of calculations by hand. We will use an R computer program in this lab. You don't need to worry about the actual code that it contains, however, it is important to understand what goes on inside such a program.

Let's focus on the scenario in which there is selection, but not drift, mutation, or migration. The basic idea is to start with some initial frequency of allele A, compute how it changes within the first generation, and subsequent generations. Here is how a computer program does this:

1. Have some initial frequency of allele A.
2. Compute frequencies of genotypes AA, Aa, and aa, based on the frequency of A in step 1.

3. Adjust the genotype frequencies based on the relative fitnesses of each genotype.
4. Compute the frequency of A from the genotype frequencies in step 3.
5. Return to step 1, but now using the frequency of A determined in step 4.

## Launching the Population Genetics Application

The population genetics application you will be using is a computer program written in R. There is an R package called [shiny](#) that makes it easy to write programs with graphical user interfaces. So, you'll have sliders to drag rather than lines of code to execute. (The Shiny application we'll be using is in the package [poggensim](#))

If you are working on your own computer follow these installation instructions:

First [install R](#)

Then install and load the devtools package that allows you use packages on github:

```
install.packages("devtools")
```

```
library(devtools)
```

Now you can install and load poggensim:

```
install_github('coleoguy/poggensim', build_vignettes = T)
```

The lab computers already have poggensim installed.

Start R or RStudio.

In your R window, type (or copy-paste) the following commands:

```
library(poggensim)
# Launch the population genetics application.
simulate(x="popgen")
```

If you can't install software on your computer you can also use the [web version](#) of the simulator.

You should now see a new window that looks like the screenshot below. Click on the "open in browser" button and maximize the new window.

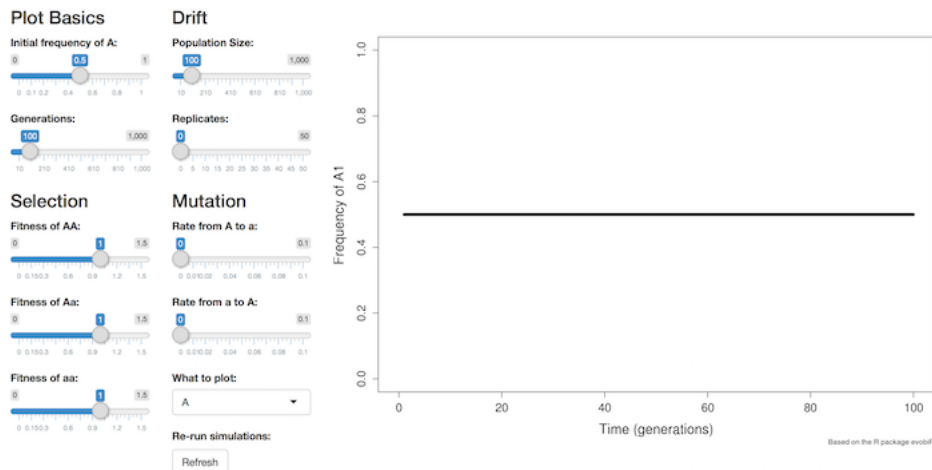


Figure 1: our popgen R app

For each topic below, the goal is for you to develop a sense of how the various parameters affect the frequency of allele A. Each topic begins with some suggested scenarios to investigate, but you are encouraged to experiment further. Each topic concludes with a question for you to answer and turn in with your lab report.

## Selection

The default settings are that all three genotypes have the same fitness. You can see the relative fitness values on the sliders in the “Selection” section of the application. So you have already obtained the first population genetics result in this lab: when fitness does not depend on genotype, allele frequency remains constant over time, regardless of what its initial value is.

In this first set of explorations, you will be adjusting the fitnesses of the three genotypes to learn how that affects the allele frequencies over time.

Note: The fitness of the AA genotype will sometimes be called  $w_{11}$ . Similarly,  $w_{12}$  is the fitness of the Aa genotype, and  $w_{22}$  is the fitness of the other aa genotype.

## Directional selection

Here are some scenarios to explore, to get a sense for how directional selection alone affects allele frequencies.

1. Keeping  $w_{11} = w_{12} = 1$ , reduce  $w_{22}$ . (Which allele is dominant now?) See how this affects the eventual (equilibrium) allele frequency, and how fast that equilibrium is reached. You may need to increase the number of generations plotted.
2. Keeping  $w_{12} = w_{22} = 1$ , increase  $w_{11}$ . (Which allele is dominant now?)
3. Compare the results from the previous two scenarios, starting with a very small initial frequency of A (like 0.01). To make this comparison “fair” (i.e., focused only on dominance), be sure that you either decrease  $w_{22}$  or increase  $w_{11}$  by the same amount.
4. While you are exploring #3, switch the plot to show the frequency of Aa over time. With strong enough selection, you should see it increase, peak, and then decrease. (Why?)
5. Set selection to any values that favor A. Adjust the initial frequency of A through the entire possible range, from 0 to 1. Observe the equilibrium frequency of A that is reached.
6. Switch between plotting A and a. Note that always,  $\text{freq}(A) + \text{freq}(a) = 1$ , so as one goes up, the other goes down symmetrically.

You should be able to understand intuitively why directional selection causes either allele A or a to fix (depending on the fitness values chosen). (Discuss with your lab partner or TA if necessary.)

When you feel you understand the above explorations, answer the following question.

### Question 1

- (a) When allele A is deleterious and rare, is it eliminated by selection more quickly when it is dominant or when it is recessive? Explain your intuition, with particular reference to the frequency and fitness of heterozygotes.
- (b) Show numerical examples that support your conclusion. Each example (one for A dominant, one for A recessive) should include the three fitness values you chose and the corresponding plot of  $\text{freq}(A)$ .

## Overdominance

Here are some scenarios to explore, to get a sense for how overdominance (also called balancing selection) affects allele frequencies.

1. Set the fitnesses so the heterozygote has highest fitness. Adjust the initial frequency of A through the entire possible range. Observe the equilibrium frequency of A that is reached.
2. Set  $w_{11} = w_{22} = 1$ , and  $w_{12} > 1$ . Now increase  $w_{12}$  further. What happens to the equilibrium frequency of A?
3. Set  $w_{11} = w_{22} = 1$ , and  $w_{12}$  to some value greater than 1. Now decrease  $w_{22}$ . What happens to the equilibrium frequency of A?

You should be able to understand intuitively why overdominant selection causes both alleles to be maintained indefinitely. (Discuss with your lab partner or TA if necessary.)

When you feel you understand the above explorations, answer the following question.

### Question 2

Set  $w_{11} = 1$  and  $w_{22} = 0.85$ .

- (a) What value of  $w_{12}$  yields an equilibrium frequency for allele A of approximately 0.7? (The slider does not allow ultra-precise adjustment, so just find the value that seems closest.)
- (b) Show two plots of your results to illustrate how the same equilibrium frequency of approximately 0.7 is reached from very-different initial frequencies.
- (c) What is the precise equilibrium value for these three fitnesses? Use the formula we discussed in class. Show your work.

## Underdominance

Here is the main scenario to explore, to get a sense for how underdominance (also called disruptive selection) affects allele frequencies.

1. Set the fitnesses so the heterozygote has lowest fitness. Adjust the initial frequency of A through the entire possible range. Observe the equilibrium frequency that is reached. (How does this result differ qualitatively from what you saw for overdominance?)

You should be able to understand intuitively why underdominant selection causes either allele A or a to fix, depending on the initial value. (Discuss with your lab partner or TA if necessary.)

When you feel you understand the above exploration, answer the following question.

### Question 3

Set  $w_{11} = 1$ ,  $w_{12} = 0.7$ , and  $w_{22} = 0.9$ .

- (a) Continuously adjust the initial frequency of A until you find the value at which its equilibrium frequency jumps from 0 to 1. This is the unstable equilibrium allele frequency. Approximately what is this frequency? (As best as you can tell given the resolution of the slider.)
- (b) For these values of fitness, compute the precise equilibrium frequency of A. Show your work.

## Genetic Drift

All our investigations so far have assumed there are an infinite number of individuals in the population. This is a reasonable approximation for populations of large size. When there are few individuals, however, exactly which alleles happen to get passed from one generation to the next is influenced by sampling. Evolution by this kind of probabilistic sampling is called genetic drift.

### Drift Alone

Here are some scenarios to explore, to get a sense for how drift alone affects allele frequencies.

Set all three fitnesses to the same value, so that selection is not influencing the results. The expected outcome is the horizontal black line: no change in allele frequency.

1. Increase the “Replicates” slider from 0 (the default) to 10. Each of the new colored lines is the trajectory for a single simulated population. They all start at the same value in generation 0, but then they diverge and wander. Click the “Refresh” button a few times to get a sense of the range of variation.
2. Change the population size. What happens to the range of variation? What happens to the number of populations in which A is fixed, or is lost? What happens to the speed with which fixation or loss occurs?
3. Set the initial frequency of A to 0.5. Across all the timesteps and all the populations, is the allele frequency more likely to go up, down, or neither? Is it more likely to fix at  $\text{freq}(A) = 1$  or  $\text{freq}(A) = 0$ ?
4. How do your answers to #3 change if you adjust the initial frequency of A to a more extreme value, like 0.2 or 0.8?

When you feel you understand the above explorations, answer the following question.

#### Question 4

Set all fitnesses at the same value, the initial frequency of A to 0.5, the number of generations to 100, and the number of replicate populations to 50. Show plots of results for a small population with 50 individuals, and for a large population with 500 individuals. Discuss the differences you see. (If you’re not sure what to discuss, refer to the numbered explorations above.)

### Drift and Selection

Here are some scenarios to explore, to get a sense for how drift and selection interact to affect allele frequencies.

1. Set  $w_{11} = w_{12} = w_{22} = 1$ . Gradually increase the strength of selection against a (i.e., decrease  $w_{22}$ ). Try this for different population sizes and different initial frequencies of A.
2. How are the results affected if A is dominant versus recessive?
3. Set population size to 100, the frequency of A to 0.01, fitness of AA to 1.5, and replicates to 50. This represents a new beneficial mutation.
4. Are beneficial mutations lost often? why?

When you feel you understand the above explorations, answer the following question.

### Question 5

As the strength of selection increases, what happens to the role that drift plays in determining how allele frequencies change over time? Your answer should discuss both the range of variation experience by a single population, and by the overall behavior across populations. Include plots that support your answer, for both small and large population size.

## Mutation

Here are some scenarios to explore, to get a sense for how mutation interacts with other evolutionary processes to affect allele frequencies.

1. Remove the influence of selection and drift. (That is, set the fitnesses equal and the number of replicates to 0). Gradually increase the rate of mutation from A to a. What happens to the equilibrium allele frequency? (Note that the values on the slider are far, far greater than is realistic for most loci.)
2. Now include selection against a, as well as mutation to a. What happens to the equilibrium allele frequency?
3. Now add drift, by increasing the number of replicate populations and setting the population size to be very small. Can the deleterious allele a ever fix?

When you feel you understand the above explorations, answer the following question.

### Question 6

Set the rate of mutation from A to a to be 0.002. Set allele a to be deleterious, so  $w_{11} = 1$  and  $w_{22} = 0.9$ .

- (a) Set  $w_{12}$  so that the mutant allele a is completely recessive. As best as you can read off the graph, what is the equilibrium frequency of A?
- (b) Set  $w_{12}$  so that the alleles are co-dominant (i.e.,  $w_{12}$  is midway between  $w_{11}$  and  $w_{22}$ ). As best as you can read off the graph, what is the equilibrium frequency of A?
- (c) Explain why the answers in (a) and (b) differ.