Bi/Ge 105: Evolution
Homework 3
Due Date: Wednesday, February 4th, 2026

> "An ounce of algebra is worth a ton of verbal argument."
>
> JBS Haldane

1. Simulating genetic drift with coin flips

In class we learned about the mathematical formalism behind population genetics, one of the centerpieces of evolutionary theory. Those ideas will provide a quantitative foundation for understanding the different evolutionary forces that shape life on our planet. It is both profound and amusing how much we can learn about evolution by thinking about coin flips and similar games of chance. The broad reach of the mathematics of coin flips calls to mind what Harvard professor (and former Caltech undergrad!) Joe Blitzstein likes to say: "The nouns change, but the verbs remain the same."

In this problem, we'll explore evolutionary forces using stochastic simulations. This approach will lead us to insights about the interplay between genetic drift and mutation rates, while sidestepping the more advanced stochastic differential equations needed to model these processes analytically.

The Buri genetic drift experiment

In 1956, Peter Buri published a classic paper in which he experimentally demonstrated the concept of genetic drift in fruit flies.[1] An overview of his approach is depicted in Figure 1. Briefly, Buri began with eight male and eight female flies, all of which were heterozygous[2] at the *bw* locus. There are two possible alleles of *bw*: one associated with white eyes, and one associated with red eyes. A fly homozygous for the white allele will have white eyes; a fly homozygous for the red allele will have red eyes. But being *bw* heterozygotes,

[1] Buri 1956.

[2] Like humans, fruit flies are diploid: they have one maternal and one paternal allele of each gene. Where both alleles are the same, that individual is a homozygote; where they differ, a heterozygote.

each of Buri's flies had one copy of the white allele and one copy of the red allele. This combination leads to a third possible phenotype: that of orange eyes.

Thus starting with his eight male and eight female flies, all of which had orange eyes, Buri let them reproduce. This yielded many larvae. From these offspring, he randomly chose eight new males and females without attention to their eye color. These sixteen flies were transferred to a new flask, and the procedure was repeated for nineteen successive generations.

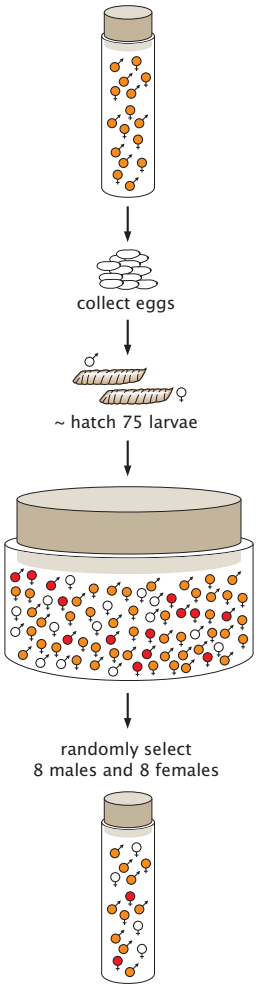

Figure 1: **Buri's experimental setup.** Each generation, eight heterozygous females and eight heterozygous males were allowed to reproduce. From their offspring, eight new males and eight new females were chosen at random and transferred into a new flask.

Question 1a. Work out the expected genotype frequency of red-eyed flies (f_{rr}), white-eyed flies (f_{ww}), and orange-eyed flies (f_{rw} and f_{wr}) after the first generation. Recall that each allele is drawn from the parent's pool at random with replacement.

Since both the mating process and the selection of offspring were random, Buri knew that the outcome would be different if he repeated an identical experiment in different vials. As a result, and for statistical power, he simultaneously tracked 107 flasks as shown in Figure 2. At each generation, he recorded the number of red-eyed, white-eyed and orange-eyed flies he had randomly chosen. After 19 generations, many vials contained only white-eyed or red-eyed flies, though some vials still contained a mixture of eye colors.

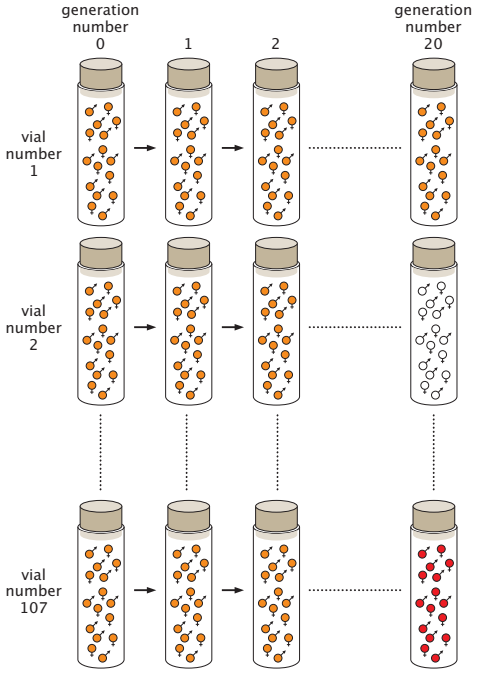

Figure 2: **Multiple replicates of the Buri experiment.** Buri repeated his experiment in 107 separate vials, with the evolutionary trajectory different each time as a result of genetic drift. In the long time limit, many of the vials became fixed: all flies had either white or red eyes, meaning that the opposite allele had been completely eradicated from the population.

Question 1b. Write down a formula for fly genotype frequencies in terms of eye color count. That is, given the counts of the number of red-, white-, and

orange-eyed flies in a single vial, what is the frequency of red alleles (f_r) and white alleles (f_w)? Use N_{red}, N_{white}, and N_{orange} to denote the number of red-, white-, and orange-eyed flies in a vial.

Figure 3 summarizes the results of the experiment. Buri witnessed evolution in real time driven entirely by genetic drift! He saw that in some populations, an allele could go extinct, all due to nothing more than the random fluctuations inherent in small breeding populations. To see this for ourselves, we'll now develop a computational simulation of his experiment.

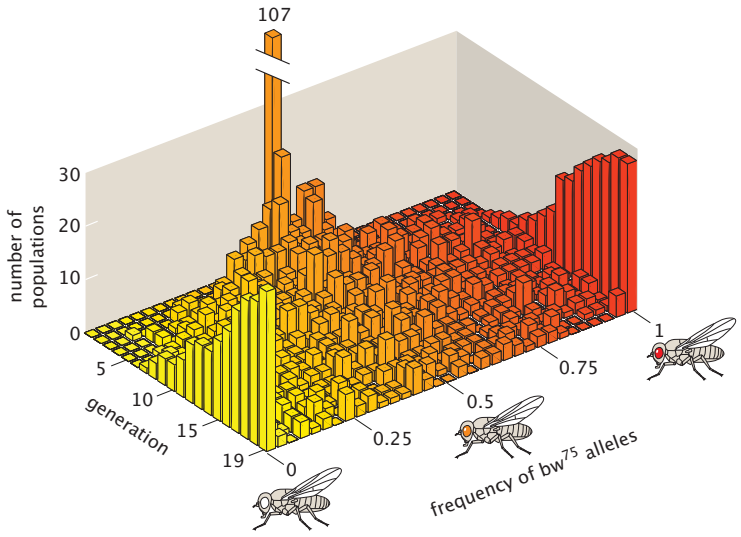

Figure 3: **Results of the Buri experiment.** By tracking the eye color of flies, Buri was able to infer allele frequencies for many populations. Changes in these frequencies arose due to genetic drift. After 19 generations, many of the vials resulted in the fixation of alleles: a tangible evolutionary milestone.

Question 1c. Perform stochastic simulations of genetic drift using the same parameters as Buri: 107 populations over 19 generations, where each population consists of 16 flies (hence 32 *bw* alleles). Plot histograms of the allele frequency for generations 0, 1, 10, and 19.

Question 1d. Repeat the simulations from the previous problem, this time continuing for 10^3 generations. Quantify how long it takes for each population to become fixed. In other words, record how many generations it takes for each population to reach an allele frequency of zero or one. Do this for populations of size $N = 4$, 8, 16, 32, and 64. Plot the mean time to fixation as a function

of population size. Hence, discuss the effect of population size on genetic drift.

Let's now explore the effect of another evolutionary force — mutation. In our toy model, rather than thinking about tracking the complexity of single base pair mutations, we will think of a "reaction" of the form

$$A \underset{\mu_2}{\overset{\mu_1}{\rightleftharpoons}} a \tag{1}$$

where A and a are the two versions of the allele (e.g. red and white), and μ_1 and μ_2 are the mutation rates that take you from one allele to the other. To simplify things even further, assume that $\mu_1 = \mu_2 \equiv \mu$.

Question 1e. Implement mutations into your simulation, at a rate of $\mu = 0.001$. Plot the allele frequency for a single population of 16 flies over 109 generations. What do you see? (Hint: one way to think of this is that mating still happens at random, but now before being selected for the next generation, each allele must flip a *second*, very biased coin to decide if it will mutate or not.)

Question 1f. Simulate 100 populations with the same mutation rate as before. Plot 10 of these trajectories. Also plot histograms of allele frequencies at $t =$ 0, 5, 10, 50, 100, and 500 generations. Compare these histograms to the cases in which $\mu = 0$. What do you notice?

Question 1g. Simulate a single population of 16 flies over 109 generations for $\mu = 0$, 0.001, 0.01, and 0.1. In each case, plot the histogram of allele frequencies from the final time point. How does μ affect evolution and drift in this case?

2. Experimental evolution in the era of DNA sequencing

Over the course of your lifetime, sequencing an organism's entire genome has gone from a pipe dream to a routine procedure. This has opened up an exciting new field, where the technology of next-generation sequencing meets the experimental benefits of studying microbial populations, making it possible to quantitatively test evolutionary theories.

One particularly interesting long-term experiment in this space comes from the lab of Richard Lenski at Michigan State University. On February 24, 1988, he began growing twelve *E. coli* cultures in parallel, passaging their offspring much like Buri did. But unlike Buri's flies, these bacteria are haploid, asexual, and divide extremely quickly, with generation times on the order of an hour. Three decades and nearly 100,000 generations later, it's the longest-running continuous evolution experiment in the world.

These bacterial cultures have been adapting to a very simple environment with a fixed media (liquid food source) composition. The advantage of working with these microbes is that every few generations, a sample can be frozen indefinitely and brought back to life at will. In this sense, Lenski's $-80°$C freezers act as an evolutionary time machine, allowing him to recover organisms from a "fossil record"!

One surprising outcome of this experiment has been the appearance of a bacterial strain capable of metabolizing a new carbon source: citric acid, or citrate. For historical reasons (most likely to avoid phage infection), the cultures have always been grown in the presence of citrate. But even though citrate contains carbon, typical *E. coli* cells cannot eat this molecule.[3]

In fact, things get weirder: although *E. coli* has the machinery to ferment citrate encoded in its genome, those genes are only expressed under anaerobic conditions. Lenski discovered that the bacteria in one of his replicates had evolved to metabolize citrate even under aerobic conditions. In practice, this meant that once the glucose in their media ran out, these cells could continue growing off of citrate until they received fresh media in the morning — resulting in a clear fitness advantage over their competitors!

With this story in mind, we'll work out a very simple model to analyze how long it might take for such a mutant to overtake a culture.

[3]In fact, before the advent of DNA sequencing, a common way microbiologists differentiated bacterial species was by what they could and couldn't eat. Among other things, *E. coli* was known for its ability to ferment arabinose, lactose, and mannitol, and *inability* to ferment citrate.

A toy model for two competing bacteria strains

Consider the case in which two alleles, A_1 and A_2, are present in a population with initial frequency p and $q = 1-p$, respectively. Assume that cells harboring allele A_1 have a growth rate m_1, while those harboring A_2 have a growth rate m_2. Let us further assume that A_1 represents the allele that allows bacteria to metabolize citrate, and as a consequence $m_1 > m_2$. In particular, we will say that $m_2 = m_1(1-s)$, where s is a small parameter $s \ll 1$. If N_1 represents the number of cells with allele A_1, and N_2 the number of cells with allele A_2, the equation that describes the growth curve is given by

$$\frac{dN_i}{dt} = m_i N_i \tag{2}$$

for $i \in \{1, 2\}$. The solution to this differential equation results in an exponential growth profile, namely

$$N_i(t) = N_i(0)e^{m_i t}, \tag{3}$$

where $N_i(0)$ is the initial number of cells with allele A_i.

Question 2a. Write an expression for the total number of cells $N_{tot}(t)$ as a function of time. Remember we have two competing cell types and we are assuming they don't interfere with each other.

Having this expression for $N_{tot}(t)$ is interesting, but what we really care about is the frequency of alleles in the population given that one of the alleles has a fitness advantage over the other.

Question 2b. Write an expression for $p(t)$, the frequency of the mutant allele A_1 over time. Similarly, find $q(t)$, the frequency of the wild-type allele A_2. These should be a function of the initial cell counts $N_1(0)$ and $N_2(0)$, as well as the selection coefficient s.

Hopefully you ended up with a nice, compact expression that looks like a logistic function. Let's now explore the consequences of this expression.

Question 2c. Let $p(0) = 10^{-9}$ denote the initial frequency of A_1. Plot $p(t)$ and $q(t)$ for the cases where $s = 10^{-4}$, 10^{-3}, and 10^{-2}. Assume the doubling time of the mutant is 1 hour. Comment precisely: how does the time that it takes for the higher-fitness allele to overtake the population vary with the selection parameter s? When plotting, use x-axis units of years (rather than

hours) to give a better sense of the timescales needed for mutants to overtake the population.