

Genetics of Evolution - Mechanisms

Preliminaries

If you are not already familiar with the structure of these exercises, read the Introduction first.

Note

Reminder: Save your work regularly.

Important

If you are using a Mac, we recommend that you use either Chrome or Firefox to complete these exercises. Some of the default settings in Safari prevent these exercises from running.

Contact information

If you have questions about these exercises, please contact Dr. Kevin Middleton (middletonk@missouri.edu) or drop by Tucker 224.

Learning objectives

The learning objectives for this exercise are:

- Describe and identify the mechanisms by which variation arises and is fixed (or lost) in a population over time.
- Explain how the processes of drift, natural selection, migration, and mutation can affect the elimination, maintenance or increase in frequency of alleles in a population.

Exploring the assumptions of Hardy-Weinberg Equilibrium

Recall that the requirements for a population to be in Hardy-Weinberg Equilibrium are:

1. Infinite population size
2. All mating is random
3. No migration
4. No selection

5. No mutation

In reality, these conditions are never strictly met. For example, it isn't possible to have a truly infinite population size. For each of the assumptions above, there is a mechanism which can cause the deviation from the predictions of Hardy-Weinberg Equilibrium: **evolution**:

1. Drift
2. Non-random mating (e.g., inbreeding, sexual selection)
3. Addition of new alleles via migration (gene flow)
4. Differential survival of individuals with certain combination of alleles (selection)
5. Changing alleles via mutation

Because evolution is the *change in allele frequencies over time*, we will be tracking allele frequencies in simulated populations across many generations. Simulations like the ones in this exercise are one of the primary ways that evolutionary biologists make predictions about the course of evolution. These predictions are often then tested experimentally in the laboratory or in natural populations in nature.

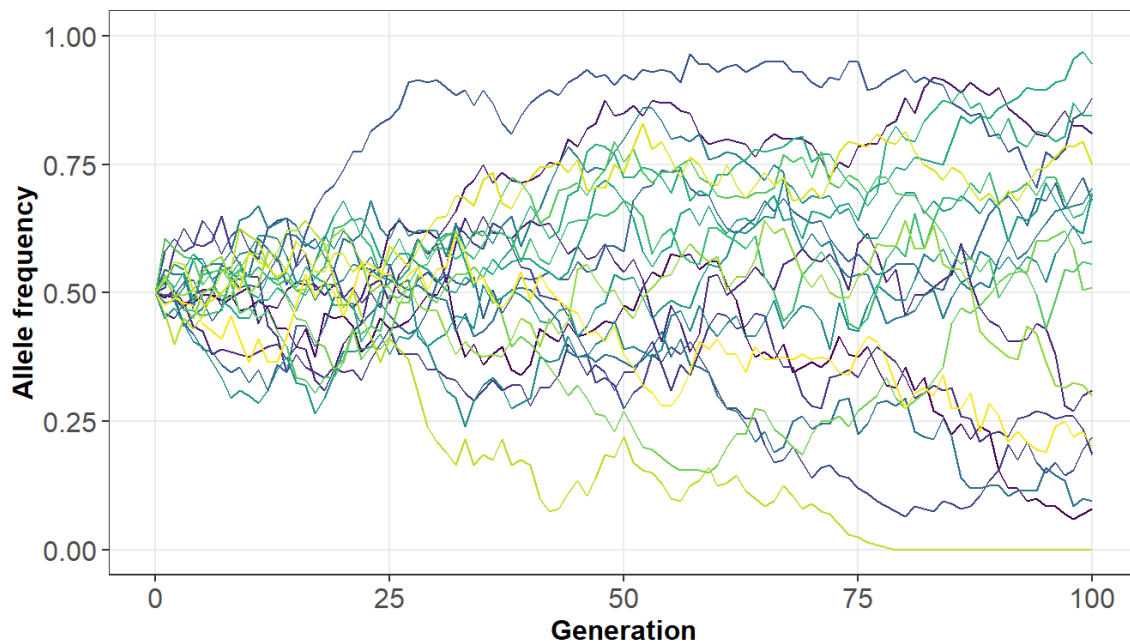
Simulating evolution

The figure below shows a very simple but very powerful model of the process of evolution. We provide the simulation a set of parameters, and these parameters are fed into the equations that govern population biology. This is called a *stochastic simulation* because the randomness ("stochasticity") that is an inherent part of nature is modeled as part of the process.

Given an initial allele frequency (in this case 0.5), 20 independent populations are allowed to evolve for 100 generations. Among the possible parameters we can set for these populations are:

- Number of generations (*n_Generations*): 100
- Number of individuals in the population (*Population_Size*): 100
- Initial allele frequency (*Initial_Frequency*): 0.5
- Fitness of the *AA* genotype (*Fitness_AA*): 1
- Fitness of the *AB* genotype (*Fitness_AB*): 1
- Fitness of the *BB* genotype (*Fitness_BB*): 1
- Rate of (*Migration*): 0
- Rate of mutation from *A* to *B* (*Mutation_AB*): 0
- Rate of mutation from *B* to *A* (*Mutation_BA*): 0

In this first simulation, which uses the parameter values above, fitness is equal among the genotypes, and there is no migration or mutation.



Each line represents a single population that starts at generation 0. For each population the population begins with a set number of individuals and allele frequency. Each generation, some individuals reproduce and, depending on their fitness, some do not. We also can simulate migration and/or mutation. At each generation the frequency of one allele is calculated. This process is repeated for each population.

There are a few things that we can note from just this one set of populations:

- We only need to plot the frequency of one allele (*A*), because the frequency of the *B* allele is just $1 - A$.
- Evolution is occurring in the populations. We know this because the allele frequencies are changing across generations.
- In this case, evolution is occurring without migration or mutation, and with equal fitness across all genotypes.
- In some populations, the frequency goes up and in some it goes down.
- In one population, the *A* allele is lost completely at about generation 80. After this generation, all individuals will have the *B* allele.

We will use this simulation framework to explore the effects of different factors on the process of evolution. There are many different combinations of each of these parameters, so we will only change one or two at a time, holding everything else constant.

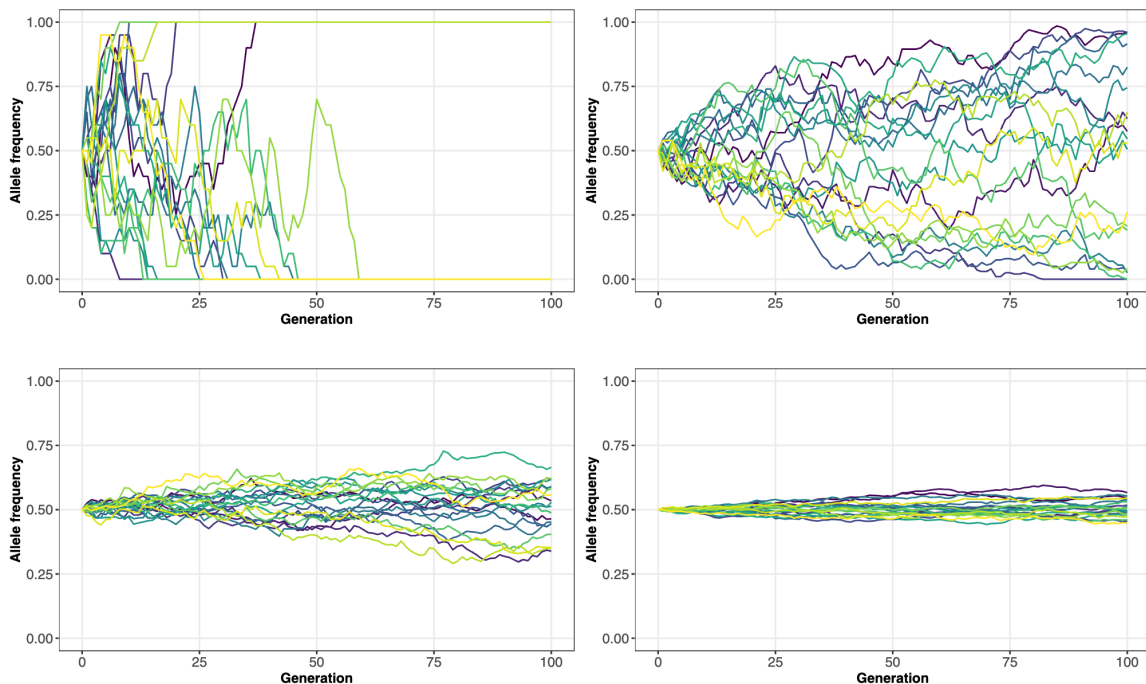
Effect of Population Size

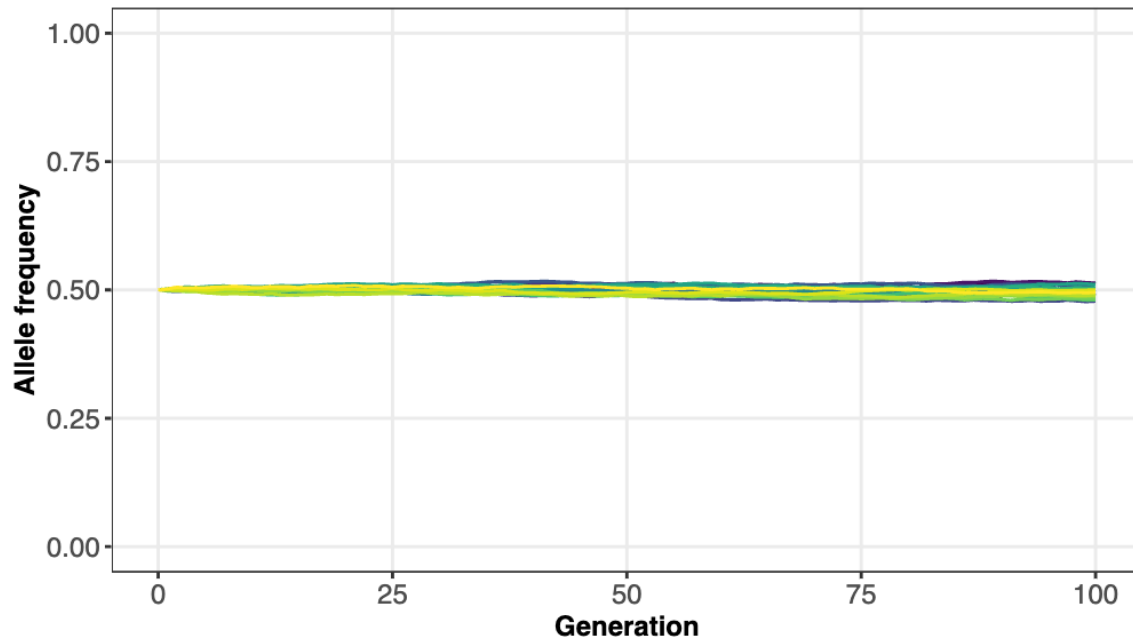
For the first set of simulations, we will explore the role of population size in allele frequency change. Below, the population size is set to 10 (`Population_Size = 10`). Run the simulation a few times and observe the patterns.

Now gradually increase the population size, rerunning the simulation a few times for each population size. Try to keep the population size under 100,000 or the simulations will run very slowly.

Remember that you can't use commas in the code, so 100,000 is `100000`.

```
Simulate_Population(  
  Population_Size = 10  
)
```





What do you observe about allele frequencies in the populations as you increase population size from 10 to 100, 500, 1,000, ..., etc. up to 100,000? How is a population size of 10 different from 100,000?

At a population size of 10, allele frequencies change very quickly. As population size increases, the frequencies get much more stable and change slowly. At 100,000, they barely change at all from the original frequencies of 0.5.

How do you think the allele frequencies would change if you kept increasing population size to infinity?

At infinity, the allele frequencies will be in Hardy-Weinberg Equilibrium and would not change at all. Recall that "infinite" population size is a requirement for HWE.

How do you predict the allele frequencies would change if you allowed the simulation to run for more generations at very small (~10-20), intermediate (~1,000-2,000), and large (50,000-100,000) population sizes?

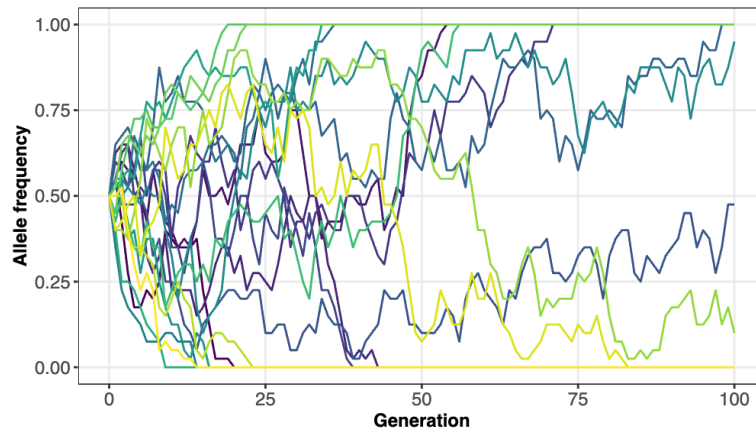
Running more generations though longer simulations will allow more time for change. For small populations, more populations will go to fixation. For intermediate population sizes, there will be more variation and possibly fixation. For very large population sizes, there will still be more variation, but likely not enough for fixation (as we saw above).

Run these simulations using the code block below:

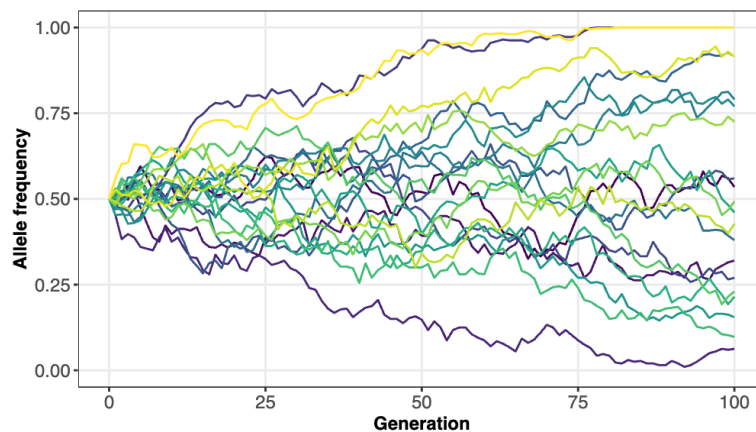
- Use one small, one intermediate, and one large population size that you choose.
- For each gradually increase the number of generations from 100 to 1,000. Avoid going over 1,000 generations or the simulation will run very slowly.

```
Simulate_Population(  
  Population_Size = 20,  
  n_Generations = 100  
)
```

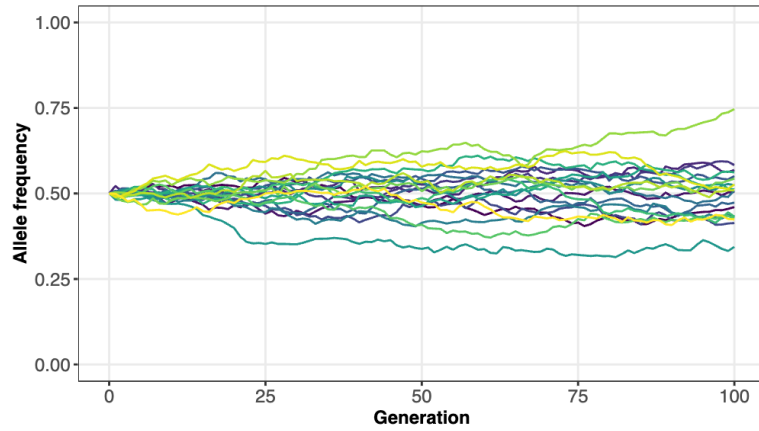
Population_Size = 20, n_Generations = 100



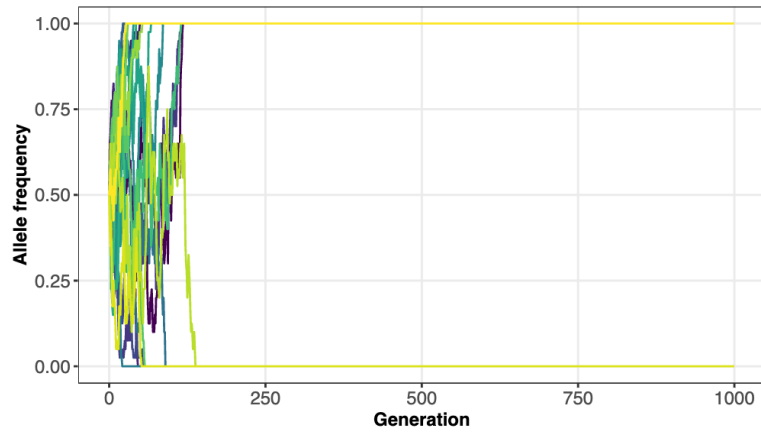
Population_Size = 200, n_Generations = 100



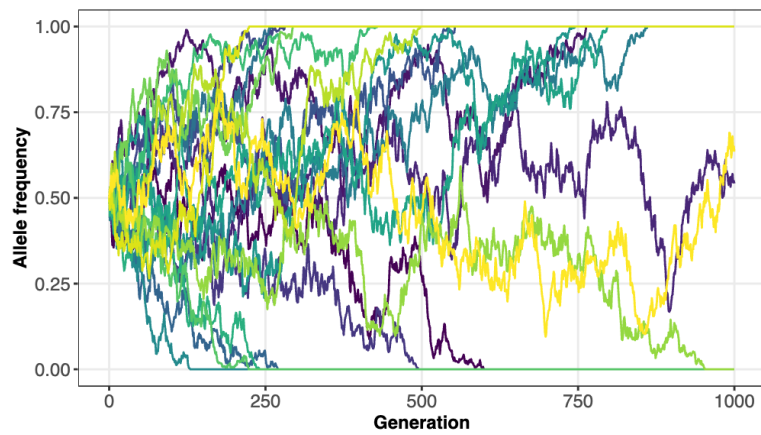
Population_Size = 2000, n_Generations = 100



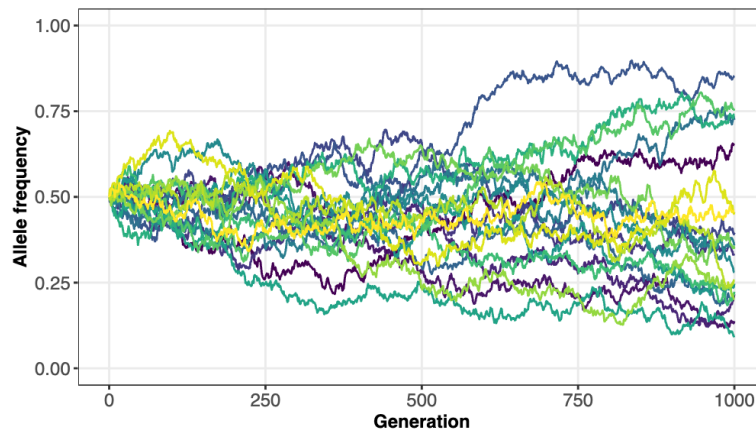
Population_Size = 20, n_Generations = 1000



Population_Size = 200, n_Generations = 1000



Population_Size = 2000, n_Generations = 1000



Were your predictions correct? If not, how did your predictions differ from what you observed in the simulations?

Small: In under 200 generations, all alleles are fixed.

Intermediate: By 1000 generation, there are alleles at about 0.1 and 0.95, but none have gone to fixation in these 20 populations.

Large: By 1000 generations, frequencies are only between about 0.45 and 0.55.

Longer numbers of generations allow more time to drift to occur.

Note that because each set of simulations is slightly different, your results may differ slightly. However, the overall patterns should remain.

What conclusions can you draw about the role of population size and time (number of generations) on the process of evolution?

Increasing the number of generations allows more time for drift and thus alleles can change more from the starting value. Small population sizes magnify the effects of drift, leading to larger changes in allele frequencies. Large population size reduces the effects of drift.

Effect of Migration (gene flow)

Migration is the rate of emigration: percentage of individuals that move to a new population each generation. To begin, we will use 100 generations and a population size of 100.

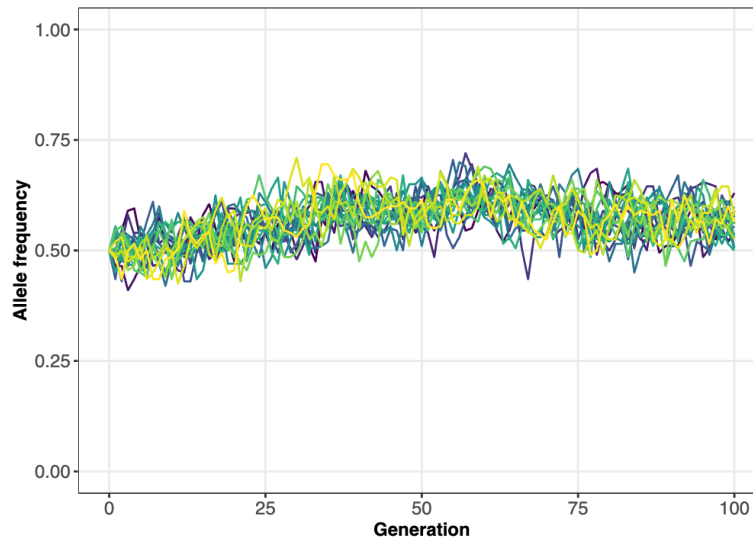
What are your predictions for the effect(s) of migration on allele frequencies? As migration increases, will you observe more or less allele frequency divergence from the initial frequency? Briefly explain your predictions.

Migration will reduce diversity of allele frequencies overall, because individuals with certain alleles will be leaving the gene pool.

To begin, we will use 100 generations and a population size of 100.

- Gradually increase the *Migration* parameter from 0 by units of 0.01 until you get to 0.1.
- Also feel free to change the *Initial_Frequency* if you want to see what effect that has.

```
Simulate_Population(  
  Population_Size = 100,  
  n_Generations = 100,  
  Initial_Frequency = 0.5,  
  Migration = 0.5  
)
```



What conclusions can you draw about the role of migration in changing allele frequencies across generations?

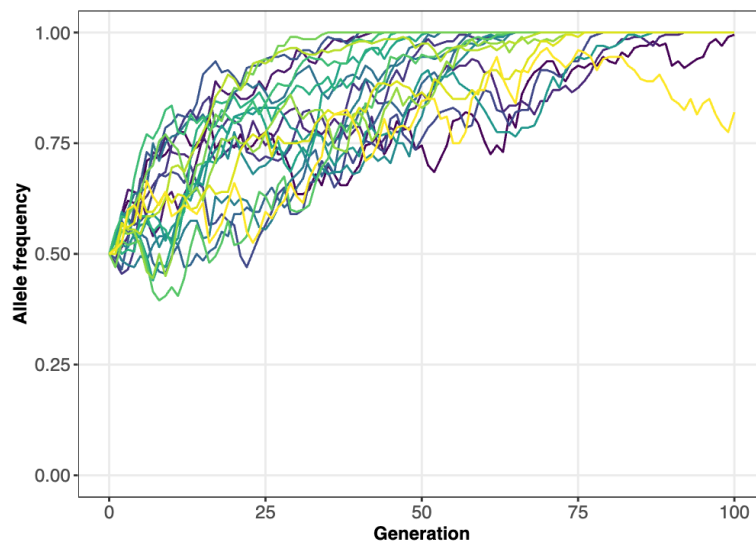
As migration increases from 0 to 0.1, the range of allele frequencies appears to be increasingly constrained. It stays at about 0.25 across all populations. So, migration appears to reduce genetic diversity in this alternate allele system. Migration might bring in different alleles for other genes.

Effect of Selection (differential fitness)

The parameters *Fitness_AA*, *Fitness_AB*, and *Fitness_BB* control the relative fitness of the *AA*, *AB*, and *BB* genotypes. It works best to keep them restricted to values between 0 and 1.

Below we change the fitness of *BB* to 0.9 and of *AB* to 0.95 (imagine the additive effects of A and B, such that *AB* has an intermediate fitness).

```
Simulate_Population(  
  Population_Size = 100,  
  n_Generations = 100,  
  Initial_Frequency = 0.5,  
  Fitness_AA = 1,  
  Fitness_AB = 0.95,  
  Fitness_BB = 0.9  
)
```

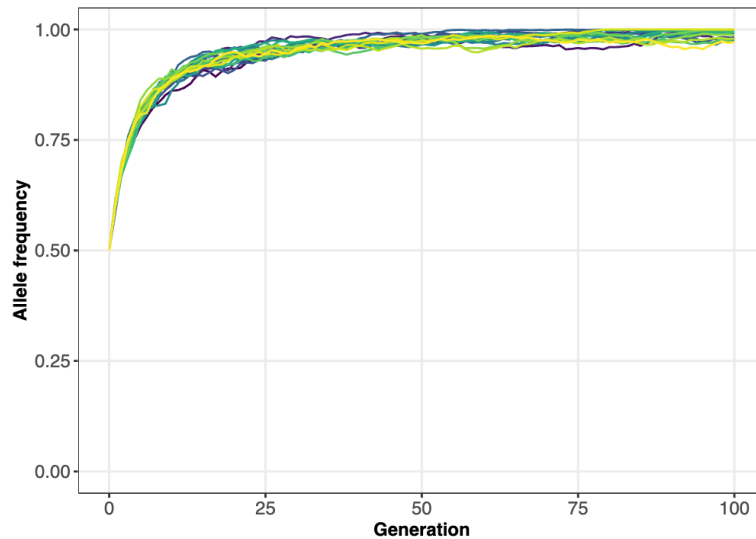


What pattern of allele frequency change do you observe in the populations?

Lower fitness in AB and BB genotypes leads to fixation of the A allele.

Now we change the population size to 1000 and the fitness of the *AB* genotype to 1, and the *BB* genotype to 0.3.

```
Simulate_Population(  
  Population_Size = 1000,  
  n_Generations = 100,  
  Initial_Frequency = 0.5,  
  Fitness_AA = 1,  
  Fitness_AB = 1,  
  Fitness_BB = 0.3  
)
```



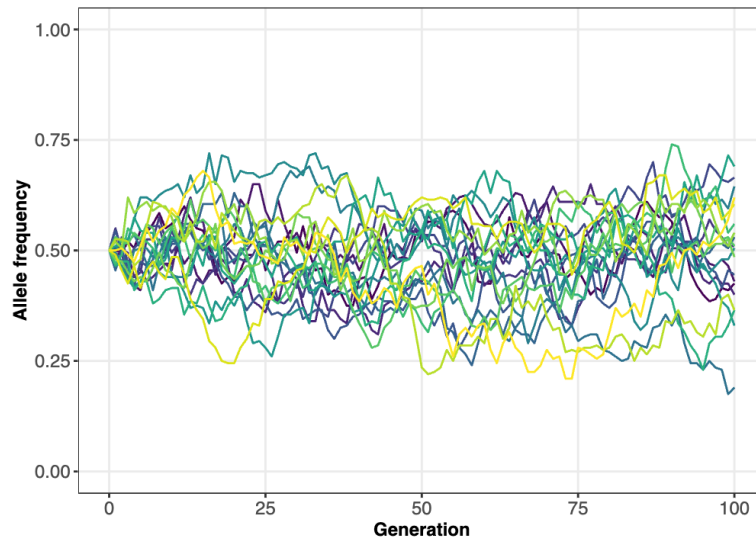
What pattern of allele frequency change do you observe in the populations? How does this pattern differ from the one above? What do these results tell you about the ability of deleterious alleles to persist in populations?

The A allele nearly goes to fixation, but it does not completely in most of the population. Above, many more populations went to fixation.

Even quite deleterious alleles are able to persist in populations as long as the heterozygote fitness is high enough.

Finally, we change the population size back to 100, the fitness of the *AB* genotype back to 1, and both the *AA* and the *BB* genotype to 0.9.

```
Simulate_Population(
  Population_Size = 100,
  n_Generations = 100,
  Initial_Frequency = 0.5,
  Fitness_AA = 0.9,
  Fitness_AB = 1,
  Fitness_BB = 0.9
)
```



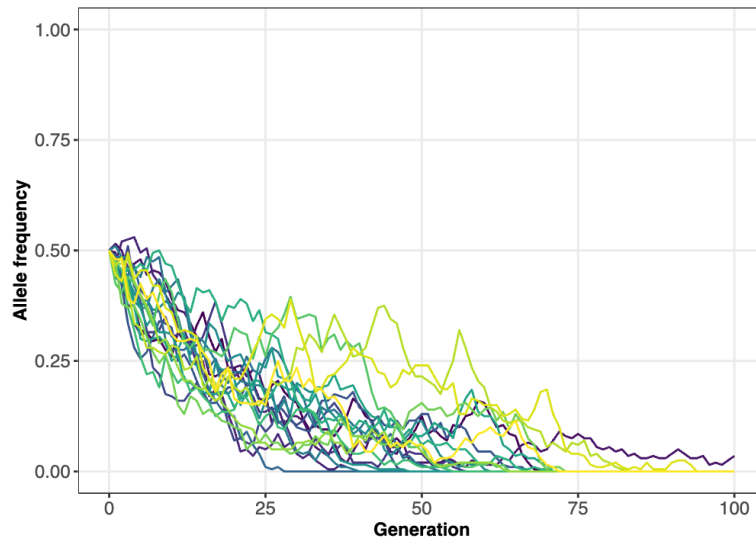
What pattern of allele frequency change do you observe in the populations? Can you think of a biological explanation for this set of parameter values?

The allele frequencies drift away from 0.5 somewhat but do not go to fixation for either allele. This process represents a heterozygote fitness advantage.

Effect of Mutation

The *Mutation_AB* parameter controls the rate of change from *A* to *B* and the *Mutation_BA* parameter the reverse. In the code block below we start with the *A* to *B* rate at 0.05.

```
Simulate_Population(
  Population_Size = 100,
  n_Generations = 100,
  Initial_Frequency = 0.5,
  Mutation_AB = 0.05,
  Mutation_BA = 0
)
```

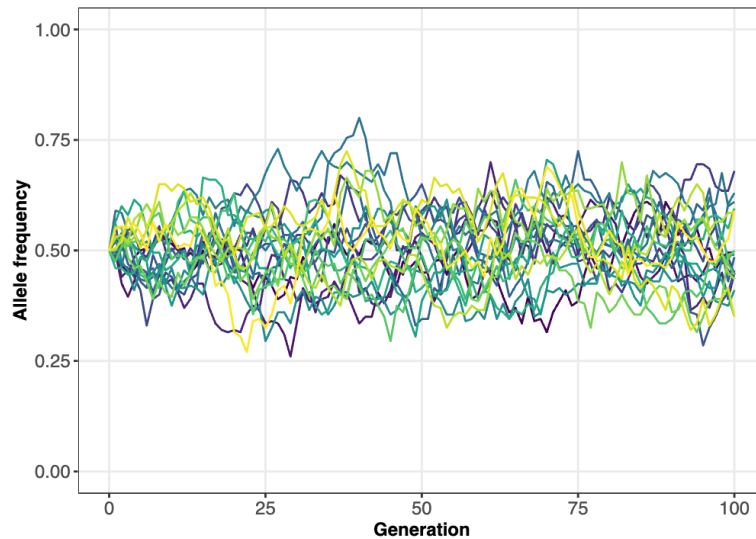


*What pattern of allele frequency change do you observe in the populations?
What do you hypothesize is going on in the populations?*

The A allele is gradually lost (B goes to fixation). Without some balancing B to A mutation, A will be lost eventually in all lineages.

Use the code block below to try to find a set of values for the mutation parameters that will lead to allele frequencies that do not go to fixation (either 0 or 1).

```
Simulate_Population(
  Population_Size = 100,
  n_Generations = 100,
  Initial_Frequency = 0.5,
  Mutation_AB = 0.05,
  Mutation_BA = 0.05
)
```



What values for the mutation parameters stabilize the variation in allele frequencies? Does this allow you to draw any conclusions about the role of mutation in natural populations (Hint: Compare the mutation probabilities you decided upon to the case where there is no mutation)?

If the rates of mutation of A \rightarrow B and B \rightarrow A are roughly equal, then the allele frequencies will be stable in the population. At 0.05 and 0.05, the range is about 0.25 to 0.75. When the mutation rates are higher (but still equal), the average allele frequency in the populations remains closer to 0.5. When the mutation rates are lower, allele frequencies will vary more.

Summation

This exercise has allowed you explore four of the factors that are involved in the process of evolution:

1. Drift
2. Gene flow (migration)
3. Differential survival (selection)
4. Mutation

Which of these factors do you think is most important in driving evolution in populations? Note that there is not a single correct answer. We are just interested in hearing your thoughts.

Your answer will vary for this question.

Feedback

We would appreciate your anonymous feedback on this exercise. If you choose to, please fill out this optional 4-question survey to help us improve.