

Workshop on Modelling in Population Genetics

Julien Barrere and Tanush Jagdish

Activity A) - Effect of population size:

Background: Let's say that we want to study an animal population. A question one could ask is how this population will react to the unpredictable environment due to climate change. One piece of information that will be very useful for us is the genetic diversity of the population: that is, for each gene, how many versions of this gene exist in the population. This is of fundamental importance because the more genetically diverse a population, the more likely it is to adjust to a new environment.

Evolution is fuelled by mutations, random mistakes that occur when DNA is copied, creating different versions of genes. Over time, these versions of genes accumulate or disappear, contributes to increasing or decreasing genetic diversity respectively.

In this activity, we will try to understand the mechanisms that will influence genetic diversity in a population, especially the mechanisms that lead to the loss of genetic diversity.

Part One: A Theoretical Scenario

Let's start with a theoretical framework with the following parameters:

- The population is of **infinite size**
- There is **no selection**
- Individuals **randomly mate with one another**

These are assumptions and they are not true in nature, but they can be a good start to understand population genetics.

To convince yourself of the statement above, open the following web page (<http://popgensimulator.pitt.edu/graphs/allele>). This page should already be open on your computers.

- a. Under the "Base Simulation Model" tab, make sure that the number of generations is 500.
- b. Make sure that the starting "Starting Allele Frequency" is 0.5.
- c. Then go to the tab below, make sure it is not checked, but choose the number of simulations to be 10 (see image below).
- d. You can then press "Generate Graph" below the graph.

Simulation Parameters

Open All
Show Help
⚙️ 🔗 📄

Generate Graph
+ Add Line

Base Simulation Model ^

Generations: ? 500

Starting Allele Frequency: ? 0.5000

Finite Population ⌵

Population Size: ? 500

Number of Simulations: ? 10

e. What do you see about how the allele frequency is changing over time?

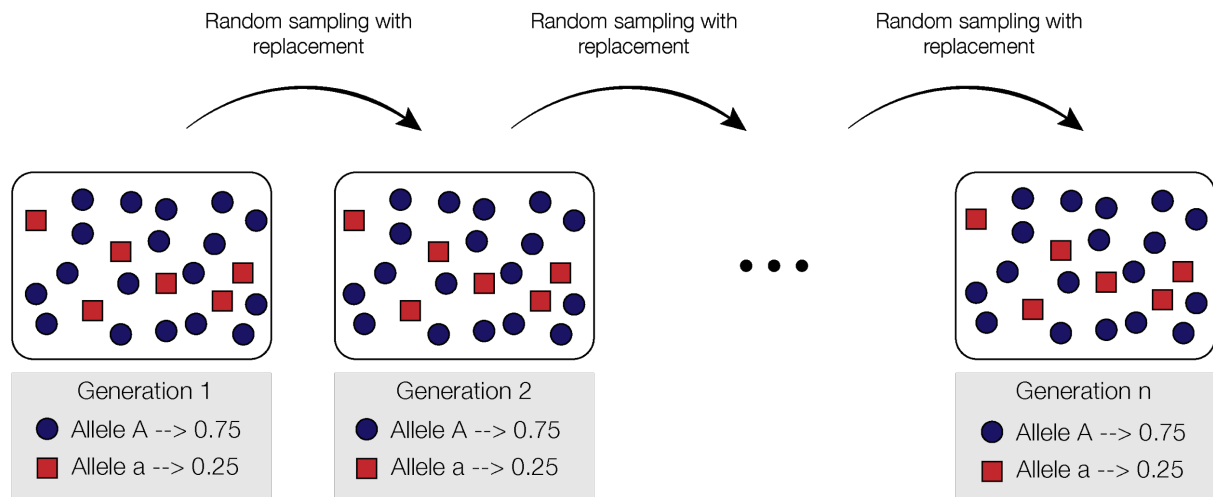
f. Does what you see reflect natural scenarios? How is it different?

g. What would happen if you change the starting allele frequency?

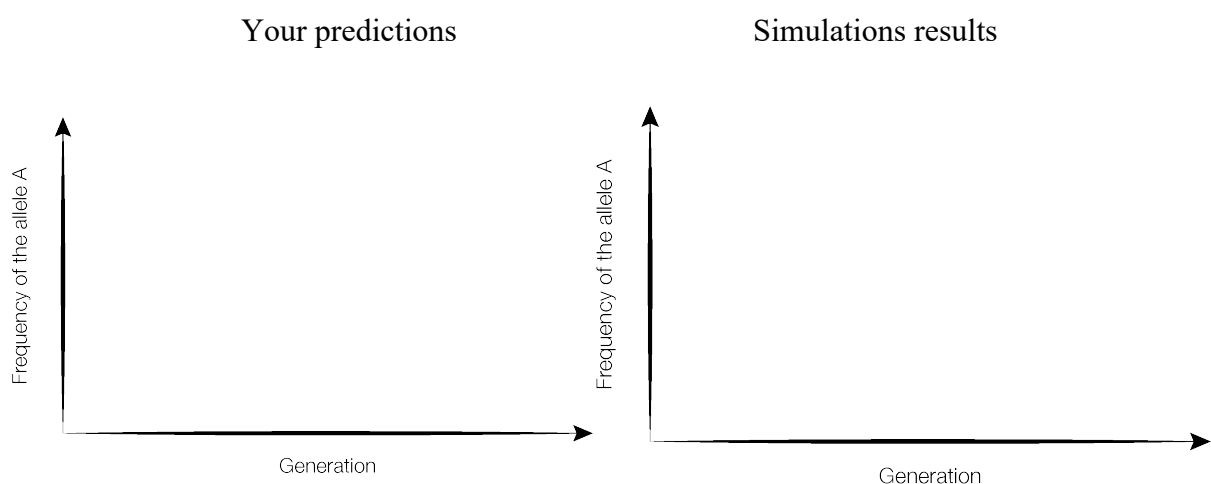
*Building from the theoretical framework of allele fixation in a population (assuming infinite population size, no selection and random mating) we will explore how the **relaxation of some of these assumptions will influence fixation of alleles** (and therefore the loss of genetic diversity) in our population.*

Part Two: Relaxing Infinite Population Size

We will start by relaxing (here, relaxing means that we will “remove” the assumption) the non-reasonable assumption of infinite population size. To explore how infinite population size influence the frequency of an allele in a population we need to understand what is behind the simulations. The simulations reflect a scenario in which at each generation, the new population is created by randomly drawing alleles with replacement in the previous generation. (see figure below: hypothetical scenario of infinite population size).



- To begin, think about what you would expect the result of 10 simulations to be if you were to keep the same parameters in a population of 10,000 individuals instead of an infinite population. Draw the results you expect in the “Your Predictions” empty graph on the left below.



- b. Now run the simulations! Just keep the same parameters as before, change only one thing: check the box for “Finite Population” and chose a population of 10,000 individuals.

- c. What do you see? Draw the simulation results next to your prediction, in the “Simulation results” empty graph.

- d. Why are you seeing these specific results?

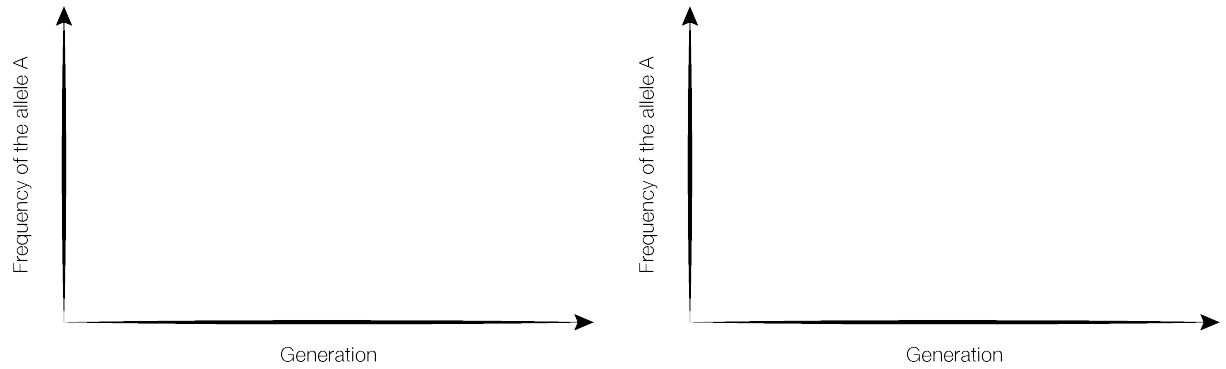
- e. Is this expected? You can increase the number of generation under the “Base Simulation Model” tab and push the number of generations to its maximum (10 000), you will see that even after 10 000 generations the alleles won’t have fixed.

- f. Let’s now see what happens when we **decrease the population size**.

- To recap, we started with an infinite population size, then a population of 10,000, and now we will decrease it even further.
- Before playing around with the simulations, discuss with people around to predict what will happen.
- You can draw what you expect in the graph below on the left panel.
- Make sure you agree with each other on what will be the outcome of the simulations as well as the reasons for such a behavior.

Your predictions

Simulations results



g. You can now decrease the population, to 1000, 100 and even 10. Graph the simulation results of one of your population sizes.

h. What is happening?

i. Why does fixation occur faster in a smaller population size?

Activity B) - Fitness Effects of Mutations

We have seen how the size of a population impacts its diversity of mutations over time. However, we were only thinking about how random occurrences were impacting these frequencies. That is, the mutations were neutral and provided no benefit. In fact, this is another key assumption in the Hardy-Weinberg Equilibrium. In addition to infinite population size, the HWE requires that no new mutations with positive or negative effects crop up in the population.

Part One: A Theoretical Scenario

- a. But, say, for instance, a very beneficial mutation arises in an individual. This means that the mutation provides a ‘fitness benefit’ to the individual. Discuss the consequences of this with others in your group. What does increased fitness mean in this case? After many generations, what might one expect of the mutation and descendants of the individual that first got this mutation? Write down one or two predictions.

- b. Since the mutation is passed consistently from parent to offspring, in a world with no other factors influencing evolution, over a long enough period of time, the entire human population will contain this mutation (that is, all individuals will be direct descendants of the person who first got the mutation). This is called a **selective sweep** of the mutation. But reality is more complicated and various factors influence whether a mutation will sweep through a population. What are these factors? Discuss with your group and write about one or two possible mechanisms that would sway the theoretical outcome.

Part Two: Adding Different Fitness Effects

We will now use simulations to think more deeply about the predictions you made. Return to the population genetics simulator (<http://popgensimulator.pitt.edu/graphs/allele>).

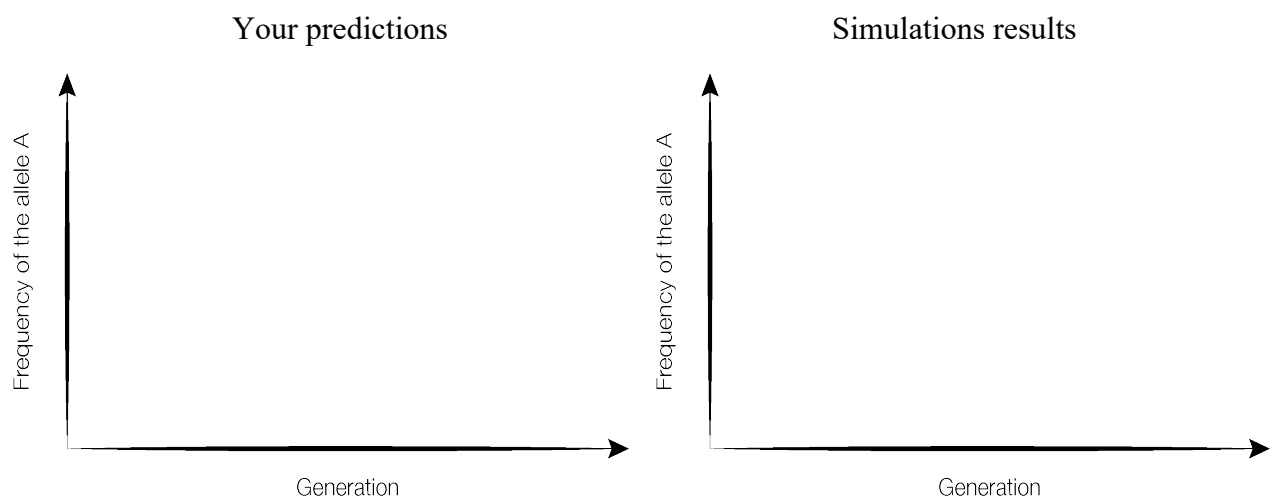
- a. Increase generation time to 500 generations (how long is 500 generations in human beings in terms of years?) Change the starting allele frequency to 0.05.

Here we are presuming that initially the mutation is at about 5% frequency in the human population. This is not really the frequency at which a new mutation arises. It is usually much lower (what do you think the true starting frequency of a new mutation in a human population is?) But we want to be able to see where the line is on the graph, so we will keep it higher.

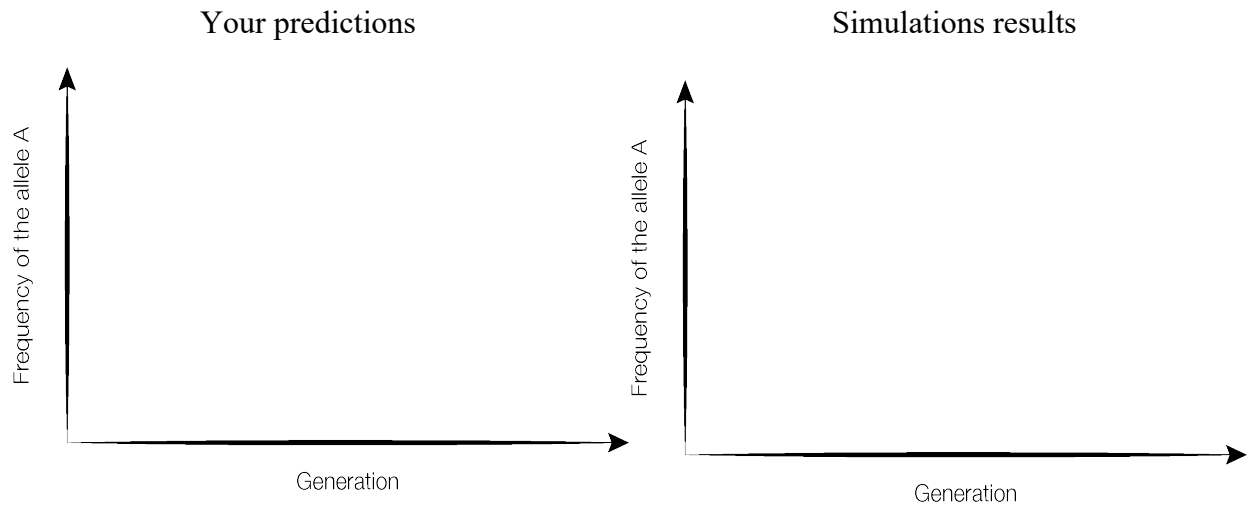
Now, keeping the population size as infinite, we will change the fitness effect of the different alleles. Suppose the beneficiality is additive. That is, if both your chromosomes have the same allele, you will have exactly twice as much fitness than you would if you had just one of the two alleles. Let's say the fitness differences are small. That is the fitness of AA is 1, the fitness of Aa is 0.99, and the fitness of aa is 0.98.

First, discuss what these fitness differences mean for the individuals with each of the three possible pairs of chromosomes.

- b. Now graph out a prediction for how the frequency of allele A should change over time. Does your simulation match your predictions? Why or why not?



- c. Now, let's consider increasing the fitness of Aa to 1 as well. Essentially, we are saying that having the beneficial mutation on just one chromosome is sufficient and adding an extra one doesn't make a difference. Graph your predictions before running the simulation.



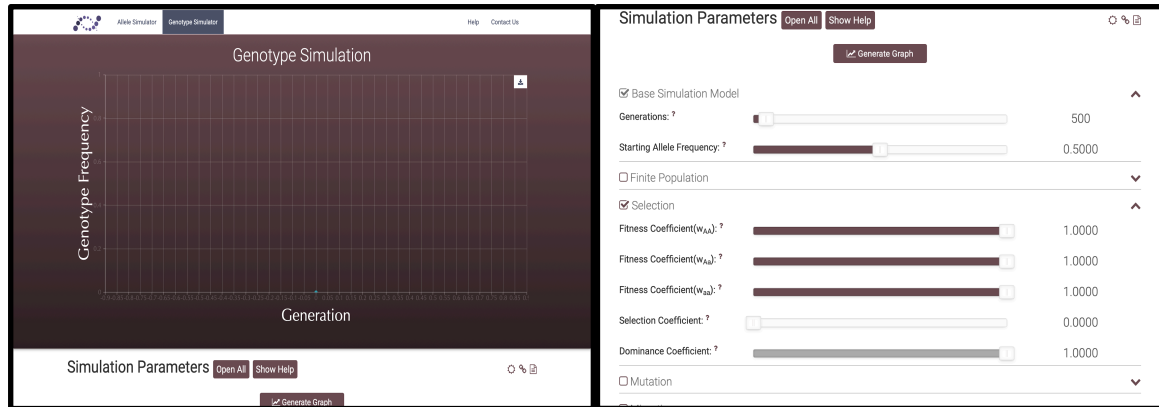
- d. Would you expect A to fix faster or slower than it did in the case of problem 1?

- e. What does the simulation tell you?

- f. Discuss possible mechanisms that might be responsible for these dynamics?

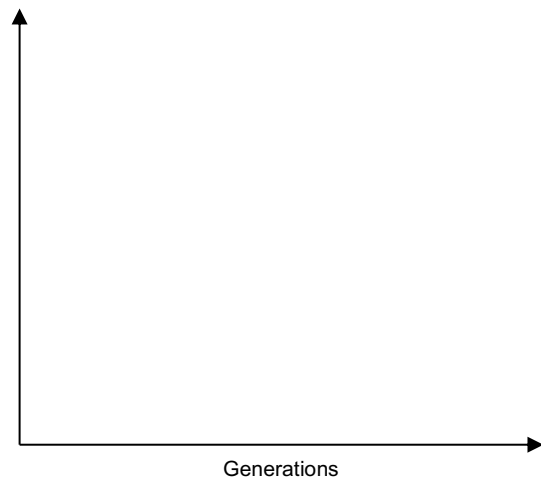
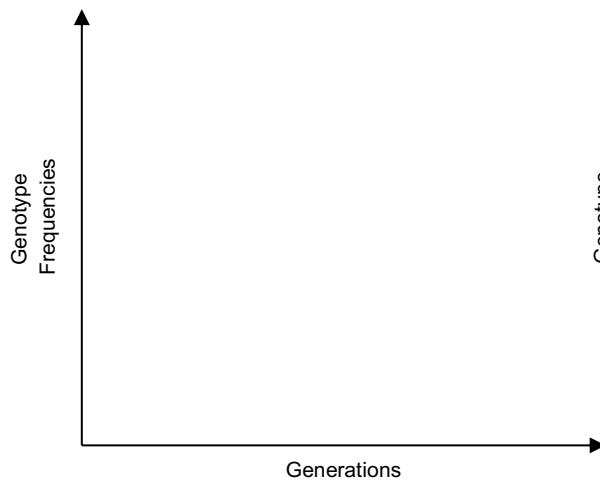
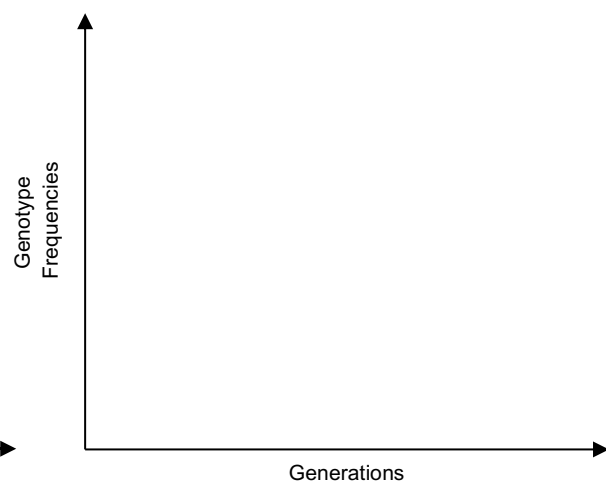
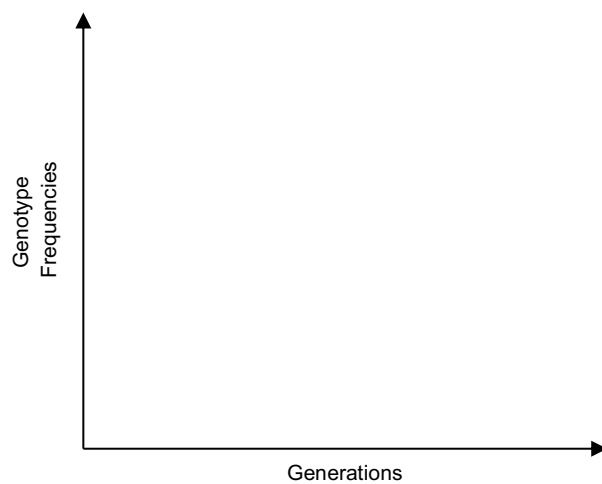
Part Three: Dynamic of Genotypes

- In a new tab, open a genotype simulator. This will show us the frequencies of individuals with each of the three allele combinations (AA , Aa , aa) instead of just showing us the frequencies of alleles over time. Conduct simulations from part two with genotypes instead of alleles.



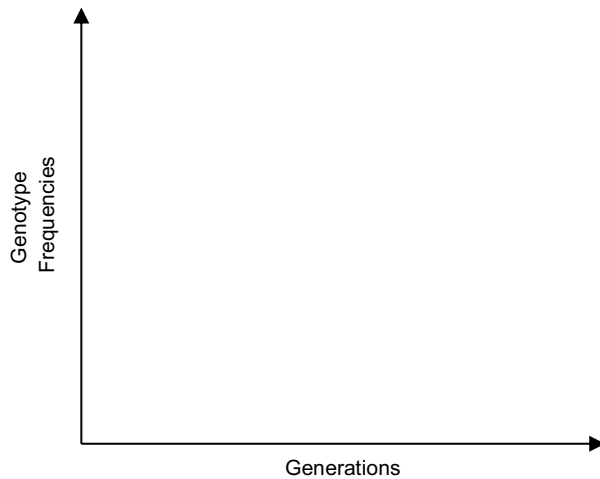
Your predictions

Simulations results

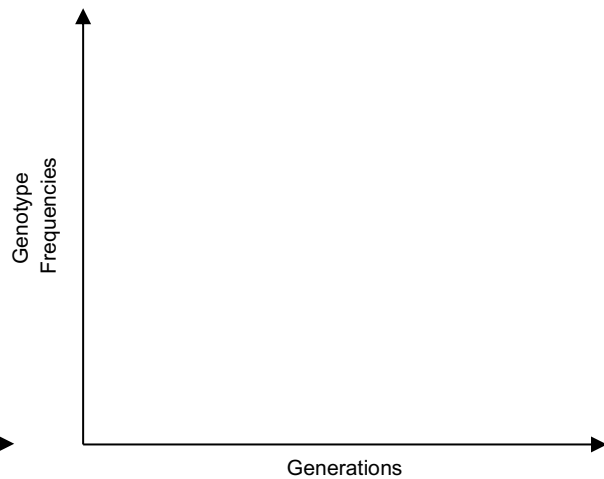


- c. Finally, we will try an example where a mutation is beneficial only when it is one of the two chromosomes but not both. This is called heterozygous advantage. Change the fitness of AA to 0.96, Aa to 1, and aa to 0.98. So when a mutation arises it is beneficial in a certain context, but having both copies of the allele is deleterious. What do you see, both in the genotype and allele simulators?

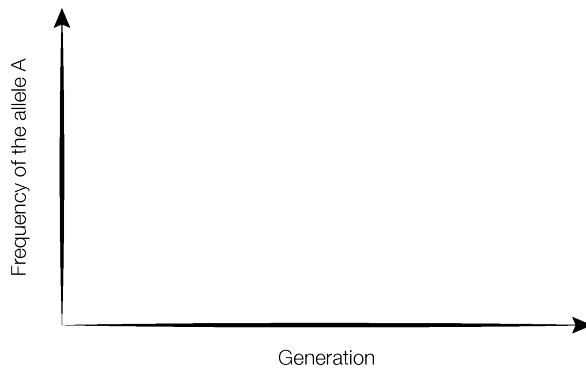
Your predictions (Genotype)



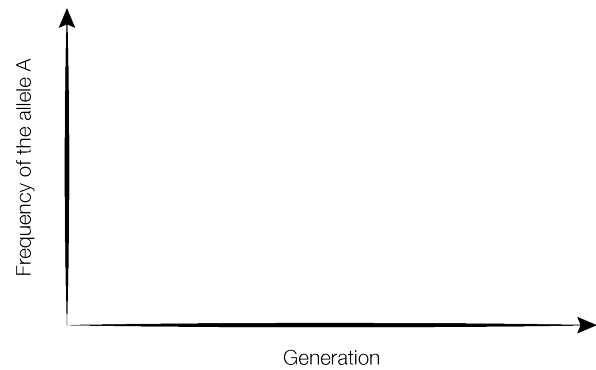
Simulations results (Genotype)



Your predictions (Alleles)



Simulations results (Alleles)




Activity C) - Interaction of population size and selection effect.


In this last part of the activity, we will explore how both population size and selection influence the fixation of alleles in a population.

- a. If an allele has a fitness advantage, will it always fix in a population?

- b. Let's explore this question using the simulations, enter the following parameters:

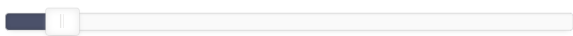
 **Graphs 1-30** [Hide Legend]


Generations: $t = 1000$	Population Size: $N = 50$	Starting Allele Frequency: $p = 0.5$
Fitness Coefficient: $W_{AA} = 1$	Fitness Coefficient: $W_{Aa} = 0.99$	Fitness Coefficient: $W_{aa} = 0.98$
Batch Runs: $= 30$		

 **Batch Graph Stats**


Avg. Gens to A fixation, $P(A)=1$: 124.85	No. of simulations reaching fixation: 20
Avg. Gens to A elimination, $P(A)=0$: 128.70	No. of simulations reaching elimination: 10


Base Simulation Model ^

Generations: ?  1,000


Starting Allele Frequency: ?  0.5000


Finite Population ^


Population Size: ?  50

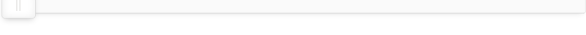
Number of Simulations: ?  30


Selection ^

Fitness Coefficient(w_{AA}): ?  1.0000

Fitness Coefficient(w_{Aa}): ?  0.9900

Fitness Coefficient(w_{aa}): ?  0.9800

Selection Coefficient: ?  0.0000

Dominance Coefficient: ?  1.0000

c. What do you see? Is this what you expected?

Activity D) - Summary concept map:

Using what you learned during the activity, try to come up with a concept map for population genetics. You should work on this as a group. You can use the board to write draw the concept maps so that we can discuss them and so that you can share your understanding with the other groups. You may use terms from the glossary to begin.