POPULATION GENETICS SIMULATIONS—HW#1 **DUE Sept 26, at the beginning of class**

How to get the simulation program (AlleleA1)

Go to the following web site:
http://faculty.washington.edu/~herronjc/SoftwareFolder/AlleleA1.html

Download the version of the program that is appropriate for your operating system.

Software Notes

The program is mostly self-explanatory. It can be used to simulate the effects on a single di-allelic locus of selection, mutation, migration, and genetic drift. You can set parameter values (fitness of genotypes, mutation & migration rates, population size, etc.) and watch allele frequencies change through time in a series of simulated populations.

These exercises should solidify your understanding of how selection, mutation, migration, and drift affect the evolutionary process. Additionally, you should come away with an understanding of how beneficial or deleterious recessive alleles may or may not persist in a population depending on the evolutionary forces at work. You should also understand the effects of interacting evolutionary forces on the evolutionary process.

How to set values in the program:

With the mouse, click on the box neighboring the value you would like to set (or use the TAB button).
Delete the quantity in there and type in the new value.

Please write your name on every page of the assignment.

When the Instructions indicate that you should include a printout of your simulations, be sure the printout includes the entire window shown in the AlleleA1 program, which includes the graph and the values of the parameters used to produce it. This is what you will get if you simply choose “Print” from the File menu.

Good luck and have fun!

**Name:** ___________________________________

## Part 1. MUTATION

Consider a simple case of mutation alone. Things will happen extremely slowly if we use realistic mutation rates, so pretend our study population is at Love Canal, Chernobyl, or Three Mile Island. Set mutation rate ($\mu$) from A1 to A2 = 0.1, and mutation rate from A2 to A1 to 0.01. Starting freq A1=0, All fitnesses = 1, and no migration, no inbreeding, and infinite population size. Choose "Auto" for the "Graph lines" option. Leave the "Generations" set to 500. Answer the following questions before running any simulations.

a) Is mutation from allele A1 to allele A2 deleterious, beneficial or neutral?

b) Is mutation from allele A2 to allele A1 deleterious, beneficial or neutral?

c) Will allele A1 get more or less common through time?

d) Will allele A1 be fixed (p = 1), lost (p = 0), or maintained at some intermediate frequency?

Make sure all parameter values are set as described above. Click on "Run". Include a printout of your simulation with your HW, label this graph "Mutation Graph 1".

e) Does mutation increase or decrease genetic variation within populations?

f) Try some more simulations with lower (more realistic) mutation rates. For example, try mutation rate of *A1* to *A2* = 0.01 and mutation rate of *A2* to *A1* = 0.001. How do decreased mutation rates affect the time it takes for an allele to reach an intermediate frequency? Include a printout of your simulations with your answers, labeled "Mutation Graph 2". Select the "Auto" button under graph lines, so you can have the results of multiple simulation runs on the same graph.

g) Do you think mutation alone is a weak or a strong evolutionary force?

## PART 2. GENETIC DRIFT

From now on, I will use the shorthand that p = freq of *A1* and q = freq of *A2*. I will also use the shorthand $w_{A1A1}$ for fitness of the *A1A1* genotype, etc.

Click on "Clear" to erase the previous graphs. Set p = 0.5, $w_{A1A1} = w_{A1A2} = w_{A2A2} = 1$, mutation rates = 0, migration = 0, number of generations = 100.

a) Before running any simulations, predict what will happen if the population size = 1000 (will allele frequencies change from their initial values)?

Set the population size to 1000 and run the simulation 5 times. Write down the final *A1* frequency after each run (this value appears to the right of the graph). If you have selected "Auto" graph lines, five different lines should appear on your graph. Include a printout of your simulations with your answers, labeled "Drift Graph 1".

Prediction:

Final value of freq of *A1* (p) for each run:

b) Does each line have the same trajectory? ___Does each run end with the same final frequency of *A1*? ___ Why or why not (answer in the space below)?

c) What is the mean of the five different final values of p? What is the variance?
Mean= Variance=

d) Next, we'll reduce the population size by to 100. Predict how these simulations will differ from one another. Specifically, will the mean of the final value of p change? _____. Will the variance change? _____.

e) Clear the previous graphs, run the simulation 5 times, and calculate the mean and the variance of the final p values. Include a printout of your simulations with your answers, labeled "Drift Graph 2". Do these values match your predictions?

Why or why not?

Mean= Variance=

Why do these values differ (or not differ) from your predictions?

f) Does genetic drift tend to increase or to decrease genetic variation *within* populations (Hint: calculate the Heterozygosity (H) at the beginning and the end of the simulation for a couple of your runs.)

Increase or Decrease (circle one)

g) Does genetic drift tend to increase or to decrease genetic variation *between* populations? (Hint: do the populations represented by each line tend to get more similar or more different with time?)

Increase or Decrease (circle one)

**PART 3. MIGRATION and DRIFT**

Migration tends to homogenize allele frequencies between two populations. There are many ways migration could happen between populations, and population geneticists have devised several different migration models.

Continent-Island or Source-Sink: One-way gene flow from a continent with a large population to an island with a smaller population. This is the model of gene flow discussed in chapter 6 of the text and is the only one considered in this exercise.

Island model: Gene flow occurs among several islands. Each island has an equal probability of receiving or sending migrants, thus the rate of gene flow is equal among all islands.

Stepping-Stone model: Each island is more likely to receive migrants from adjacent islands than from islands that are further away.

This program only simulates the continent-island model, but this is sufficient to illustrate the basic principles of migration. In order to make the parameters of the simulation match the formulae in your notes, let q represent the frequency of the *A1* allele for this set of questions. Consider the following situation. On an island, let the initial frequency of *A1* be q = 0.5, $w_{A1A1} = 1$; $w_{A1A2} = 1$; $w_{A2A2} = 1$. Let mutation rates = 0, population size = 1000, and generations= 50.

a) What do you predict the mean final frequency of *A1* will be if you run this simulation many times? ________. What is the mean final frequency of A1 after running the simulation 5 times? ________.

b) Now let's introduce a small amount of migration. Suppose there are only *A1A1* genotypes on the mainland near our island and every generation 10 migrants from the continent arrive on the island. Set the value for "Frequency of A1 in the source population" to 1.0, and set "Fraction migrants each generation" to 0.01. Use the formula for one-way migration (see lecture notes) to calculate the expected frequency of the A1 allele on the island after 50 generations of migration. Using the notation in your lecture notes, this would be the value of $q_{50}$, where the subscript indicates the number of generations of migration. Predicted $q_{50}$ = __________.

c) What is the final frequency of *A1* ($q_{50}$) in your simulation? _______. Does this match your prediction? _________ Why or why not?

d) Run the simulation 4 more times, each time noting the $q_{50}$. Include a printout of your simulation with your answers, labeled "Migration Graph 1". What is the mean value of $q_{50}$? _________. What is the variance of $q_{50}$? __________ . Is the mean closer to the predicted value than was the value you got for the first run in part (c)? _________. Why or Why not?

e) If you change the population size to 100, but leave the migration rate at m=0.01, how will this affect the expected mean value of $q_{50}$? _______. How do you think it will change the variance? ________.

Why?

f) Run the simulation 5 times with the population size=100. What is the mean value of $q_{50}$? _________. What is the variance? _________.

g) Increase the migration rate to m=0.1 and change the number of generations to 15. Include a printout of your simulation with your answers, labeled "Migration Graph 2". What is the expected value of $q_{50}$ now? _______. What is the mean value of $q_{50}$ over 5 simulation runs? _______.

h) Compared to mutation and genetic drift, do you think migration is a strong or a weak evolutionary force? _______. Why?

**PART 4. INBREEDING**

Now consider what happens to allele frequencies and genotype frequencies when there is inbreeding. Say you want to create an inbred line to use in a genetic mapping project. In most animals, the most extreme form of inbreeding you can accomplish is brother-sister mating. The offspring of brother-sister mating have alleles that are "identical by descent" at 25% of their genes, because that is the probability that the two alleles at a locus are descended from exactly the same allele in a grandparent. The inbreeding coefficient of the offspring of brother-sister mating is F=0.25, reflecting the "identity by descent". After 15 generations of brother-sister mating, the inbreeding coefficient is approximately F=0.96.

a) Assume that you have a population of fruit flies in which the allele frequency at a given di-allelic locus is p=q=0.5. Using the formula from your lecture notes for the effect of inbreeding on heterozygosity, what is the expected heterozygosity after 15 generations of brother-sister mating (use F=0.96)? Show your calculations.

Answer: ____________.

b) Run a simulation in which the allele frequencies are p=q=0.5, all genotypes have equal fitness, no mutation, no migration, a population size of 100, and an inbreeding coefficient of 0.96. Set the number of generations to 15. From the pull-down menu at the far left of the graph, select "Frequency of genotype A1A2". This sets the program to plot this genotypic frequency, instead of the A1 allele frequency. Run the simulation five times. Include a printout of your simulation with your answers, labeled "Inbreeding Graph 1". What is the mean final heterozygosity from your five simulation runs? Mean: ____________. Is this close to value you predicted in part a? _________.

## PART 5. SELECTION

Consider a simple case of overdominance alone (no other evolutionary forces) using the parameters: $p = 0.1$, $w_{A1A1} = 0.9$, $w_{A1A2} = 1.0$, and $w_{A2A2} = 0.9$.

a) Why is this a case of overdominance?

b) Will allele *A1* get more or less common through time? _______________.

c) What do you predict will be the final frequency of A1 after many generations? Show calculations.

Answer: ________________.

d) Set the simulation program to: $p = 0.1$, $w_{A1A1} = 0.9$, $w_{A1A2} = 1$, $w_{A2A2} = 0.9$, population size = 1000, migration and mutation rates = 0, number of generations = 100. Also set the y-axis back to "Frequency of A1" using the pull-down menu at the far left of the graph. Include a printout of your simulation with your answers, labeled "Selection Graph 1". In the space below, summarize how overdominance alone affects the evolutionary process.

e) Now consider a case where allele *A1* is recessive to allele *A2*. Set $p = 0.5$ and $w_{A2A2} = 1.0$. Set the number of generations = 500. All other parameters remain the same as above. How do the three fitness values ($w_{A1A1}$, $w_{A1A2}$, $w_{A2A2}$) show that allele *A1* is recessive?

f) Is *A1* a beneficial or a deleterious allele? _____________.

g) Will allele *A1* get more or less common through time? _______________.

h) Will allele *A1* be fixed ($p = 1$), lost ($p = 0$), or maintained at some intermediate frequency? ________________.

i) In the space below, summarize how recessiveness alone affects the spread or loss of an allele in the population (compared to what would happen if the allele had the same fitness, but was dominant).

j) Run a simulation in which $W_{A1A1}$, remains the same, but now allele *A1* is dominant. Include a printout of this simulation with your assignment, labeled "Selection Graph 2".

k) How does the dominance of the *A1* allele affect its rate of spread or loss (relative to when it was recessive)?

e) How would you increase the strength of selection against allele *A1*? Try simulating this situation. How does the strength of selection affect the rate of spread or loss of a recessive allele?

**PART 6. SELECTION and MUTATION**

Even though mutation alone is a weak evolutionary force in the short term, in the long term mutation is an extremely important evolutionary force. These next two examples will illustrate the effects on the evolutionary process of mutation and selection combined.

Consider a case of mutation – selection balance and the loss of deleterious recessive alleles. Set p = 0.5, $w_{A1A1}$ = 1, $w_{A1A2}$ = 1, $w_{A2A2}$ = 0.5, $\mu$ (A1 to A2) = 0.001, $\mu$ (A2 to A1) = 0. All other parameters remain the same as before.

a) Will allele *A1* be fixed (p = 1), lost (p = 0), or maintained at some intermediate frequency? ________________.

b) What is the predicted equilibrium frequency of the allele A1 after many generations (use equation from text or lecture notes, and show your work below)?

Answer: ________________.

c) What is the mean frequency of allele A1 in your simulations (do five simulations)? Include a printout of this simulation with your assignment, labeled "Selection-Mutation Graph 1".
Mean: ________________.

There are two possible reasons why allele *A* was not fixed in this population (note: these are not mutually exclusive). First, mutation from *A1* to *A2* may be recreating allele *A2* faster than selection takes it out. Second, allele *A2* is 'completely recessive'. This means that heterozygotes do not suffer a fitness cost because they have one copy of allele *A2*. These next two simulations will test two hypotheses for the maintenance of allele *A2* in this population.

Hypothesis I: Mutation keeps allele *A2* in the population.

d) Set the mutation rates to zero, leaving all other parameters the same. Re-run the simulation and compare your results to the previous simulation (with mutation rate *A1* to *A2* = 0.001). Does mutation alone explain the maintenance of allele *a* in the population?

Hypothesis II: Allele A2 is kept in the population because it is completely recessive.

e) Leaving the mutation rates at zero, change $w_{A1A2}$ to 0.9. This makes allele *A2* 'partially dominant', meaning that one copy of allele *A2* has a small deleterious effect on the heterozygote. Re-run the simulation and compare your results to the above two simulations. Does *A2* decline in frequency more quickly when it is partially dominant?

f) Now add mutation back in by setting $\mu$ (A1 to A2) = 0.001 again. Is *A2* maintained in the population now? ______. Include a printout of this simulation with your assignment, labeled "Selection-Mutation Graph 2".

g) Is the frequency of the deleterious allele *A2* maintained at a higher level when it is recessive, or when it is partially dominant? ____________________________.

h) Reset $w_{A1A1} = 1$, $w_{A1A2} = 1$, and $w_{A2A2} = 0.5$ leaving all other parameters the same (note: allele *A2* is now completely recessive). Re-run several simulations altering the mutation rate from A1 to A2. (leave mutation (*A2* to *A1*) = 0). How does the mutation rate affect the equilibrium frequency of allele *A*?

Finally, consider another case of mutation – selection balance and the spread of beneficial recessive alleles. Set the following parameters: p = 1.0, $w_{A1A1} = 0.5$, $w_{A1A2} = 0.5$, $w_{A2A2} = 1.0$, $\mu$ (*A1* to *A2*) = 0.0001, $\mu$ (*A2* to *A1*) = 0, number generations = 1000, population size = 1000. Include a printout of this simulation with your assignment, labeled "Selection-Mutation Graph 3".

i) How do the fitness values show that allele *A2* is a beneficial recessive?

j) Will allele *A1* be fixed (p = 1), lost (p = 0), or maintained at some intermediate frequency? ________________________.

k) Why does it take allele *A2* so long to make it into the population even though it is an extremely beneficial allele?

l) Set $w_{A1A2}$ = 0.6. This makes allele a 'partially dominant' because one copy of allele *a* has a small beneficial effect on the heterozygote. All other parameters remain the same. Re-run the simulation. What is the effect of 'partial dominance' compared to 'complete recessiveness' on the evolutionary dynamics when a beneficial recessive allele is considered? (how is this simulation different from the previous one).

**There are three important things to notice from these Mutation – Selection exercises:**

1) Mutation can counter-balance selection. Equilibrium allele frequencies may change depending on the mutation rate and the strength of selection.

2) Mutation provides the raw genetic variation that selection can act on.

3) The degree to which an allele is dominant or recessive in a population (from completely dominant to completely recessive and everything in between) affects evolutionary dynamics. The 'effectiveness' of selection in eliminating or fixing alleles in a population depends not only on the strength of selection but also on how alleles are expressed in the heterozygote (degree of dominance).