We can study many evolutionary processes over a longer time period using computer simulations. Computer simulations offer an excellent opportunity to model some of the processes we will discuss in lecture. Computer models can show how selection and genetic drift affect the frequency of alleles over time. The program we are going to use is freeware developed by ecologists at the University of Minnesota. The program is available on all the computers in the computer labs, and is also available free from the University of Minnesota web site (http://www.cbs.umn.edu/software/populus.html).

The Populus software contains a set of simulation models that all share a common format, as follows: After a model is chosen from the menu, the program displays (optionally) several screens of background material which introduce the theory and mathematics, and end with basic references. You should see a window listing all of the input parameters; you can change initial defaults to values specified below or of your own choosing. The program sets permissible maxima and minima for each parameter and filters input values accordingly. Usually there are several possible outputs (e.g., allele frequency, p, vs. generation) which can also be selected from the parameter input screen and appear in a separate window. Alternatively, you can view the different outputs in sequence, by clicking on the appropriate button in turn. A “help” pdf document is available from the input and output screens of every model.

Instructions for Using POPULUS: Population Biology Simulations
• Open Populus by double-clicking the icon.

Model Drop-Down Menu (the ones in bold with sub-menus will be the ones we’ll use):
• Single-Species Dynamics
• Multi-Species Dynamics
• Mendelian Genetics
  - Genetic Drift
  - Drift and Selection
• Natural Selection
  - Woozleology
  - Selection on a Diallelic Autosomal Locus
• Quantitative Genetics Models
• Spatial Models
• Interaction Engine
• Load Model From File
Woozleology
For over a century now, those who contest evolution have argued that the evolution of complex structures by natural selection is simply too mathematically unlikely. To explore this, we will examine how aspects of brood size, mutation rate, and recombination rate affect the time it takes to “evolve” the phrase “METHINKS IT IS LIKE A WOOZLE”.

Open POPULUS. From the Model menu, go to “Natural Selection” and select "Woozleology".

Be sure to read the “help” pdf document (pg 78) and make sure you know how the program works. For an explanation of the Woozle demonstration, go to the "Help" menu and scroll to page 78 in the pdf document. Note that in the real world, unlike in Woozleology, natural selection has no long-term goal. Note: you can run more than 1 model at once.

• BROOD SIZE this is the number of descendent copies from the original phrase. The one phrase that is most similar to the target phrase will be selected to be the parent phrase for the next generation.
  ◦ Set mutation rate = 0.01;
  ◦ model a diploid, sexual process – leave unchecked
  ◦ Run 3 simulations at each of the following brood sizes (15 total simulations):
    2,10,50,100,500
  ◦ Calculate the mean number of generations it takes to “evolve” the phrase at each brood size.

<table>
<thead>
<tr>
<th>Run #</th>
<th>2</th>
<th>10</th>
<th>50</th>
<th>100</th>
<th>500</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• MUTATION RATE this is the probability that a letter will change or “mutate” to a different letter in the progeny of the parental phrase.
  ◦ Set brood size = 50; model a diploid, sexual process – leave unchecked
  ◦ Run 3 simulations at each of the following mutation rates (12 total simulations): 0.001, 0.01, 0.1, 0.2
  ◦ Calculate the mean number of generations it takes to “evolve” the phrase at each brood size.

<table>
<thead>
<tr>
<th>Run #</th>
<th>0.001</th>
<th>0.01</th>
<th>0.1</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Be prepared to discuss in class the how and why changing your parameter affects the time it takes to evolve the phrase.
(This Woozleology exercise was adapted from one used in the Evolution Laboratory course at the University of Virginia.)
**SELECTION and DRIFT**

For all simulations and problems below make the following assumptions. Assume that coat color in a certain strain of mice is controlled by one gene with 2 alleles. One allele codes for black coats (A allele), and the other codes for white coats (a allele). In the population you find 3 coat phenotypes: black (AA), gray (heterozygotes – Aa), and white (aa). Now, assume we have a stable population of mice living on an island with no owls. For convenience, let’s assume that there are just as many “A” alleles in the population as “a” alleles (unless otherwise noted), and the population starts out in Hardy-Weinberg equilibrium. Find helpful information on page 80 in the “help” pdf.

**Simulation A:** Here we will assume we have a very large isolated mouse population with no appreciable mutations in coat color alleles and random mating. When owls find their way to the island, it suddenly becomes somewhat more dangerous to be a white mouse. We want to know how the mouse population evolves in response to this selection pressure. How strong does selection have to be in order for there to be a response to it?

1. Open up Populus and go to the “Natural Selection” models. Choose Selection on a Diallelic Autosomal Locus (by the way, what is a diallelic autosomal locus?).
2. Set plot options to “genotypic frequencies vs. t.”
3. Choose “Fitness” (rather than “Selection”). Fitness is expressed relative to other genotypes.
   a. For the fitness of AA, enter 1.0.
   b. For the fitness of Aa, enter 1.0.
   c. For the fitness of aa, enter 0.7
4. For initial conditions, choose one initial frequency and enter 0.5. Set number of generations at 130.
5. Hit “view.”
6. If you select “6 Initial Frequencies” the plot shows p vs. t for 6 computer-generated initial frequencies of the A allele. However, you can’t plot genotypic frequencies vs. time for this selection; if you want to examine genotype frequencies for different initial conditions, you must enter them one at a time (see question f below).

7. Answer the following questions.
   a. Identify the lines representing the 3 genotypes. What happens to each one?
   b. If AA and Aa have equal fitness, why does the frequency of AA go up and the frequency of Aa go down?
   c. If aa is bad, why doesn’t that genotype disappear entirely? Why doesn’t the a allele disappear? Now, go back to the Plot Options box and check “p vs. t”. This shows how the allele frequency (p = frequency of allele A) changes over time. What do you see?
   d. What does this simulation tell us about the relationship between fitness and genotypic frequency?
   e. Natural selection is very good at driving deleterious recessives into rarity, but it’s not so good at eliminating them entirely. What does this say about rare genetic diseases?
f. Change the initial frequency of the \( A \) allele to 0.1 (leave everything else the same). In other words, we’re assuming that for whatever reason, white mice outnumber dark mice on the island prior to the arrival of owls. So, why does the \( aa \) line start so high and drop so fast? Why does \( Aa \) increase, then decrease?

g. Plot “p vs. t”. What does this tell you about how selection can work?

Simulation B: Here we will simulate the same large, isolated population of mice with no appreciable mutation in coat color alleles, random mating, and where individuals with white coats are spotted most frequently by predators, individuals with black coats are the next most frequently spotted, and gray individuals are rarely spotted by predators. This situation, where heterozygotes have the highest fitness, is called heterozygote advantage.

1. In the same model, Selection on a Diallelic Autosomal Locus, set everything up as before (\( p = .5 \)), except that this time, set the fitness of the \( AA \) allele at 0.9 (with \( Aa \) at 1.0 and \( aa \) at 0.7).

2. What is the equilibrium condition (and what does equilibrium condition mean)? What are the major differences between this simulation and the previous one? Why is the low fitness genotype not eliminated? Why is the high fitness genotype not pushed to fixation?

3. What human disease is the classic example of heterozygote advantage? What causes the low fitness of \( AA \) and \( aa \) homozygotes in this disease?

4. How would the fitness of the \( AA \) genotype vary between areas of sub-Saharan Africa with a serious vs. not so serious malaria problem? Compare genotypic frequencies when \( w_{AA} = .9 \) vs. \( w_{AA} = .8 \).

5. What is the fitness of the \( AA \) genotype in the US today? What effect will this have on the prevalence of sickle-cell disease (model it)? What is the fitness of the \( aa \) genotype in the US today? What effect will this have on the prevalence of sickle-cell disease here (model it)?

Simulation C: Now, we’ll simulate the same isolated population of mice, but with a small population size. We will assume random mating, no appreciable mutation in coat color alleles, and no differential survival among coat color phenotypes. Take a look at page 71 in the “help” pdf.

1. Open up Populus, then go to the Genetic Drift model under Mendelian Genetics. Choose the Monte Carlo tab. Make sure the default settings read:

   a. Runtime = 100 generations
   b. Loci = 6
   c. Initial frequency = 0.5
   d. Population size = 500

2. Hit view. Each color follows the trajectory through time of the frequency of a particular allele (think of them as six randomly chosen independent loci within the genome). Note that
these are neutral alleles (i.e., there is no selection acting on them, and they confer no survival or reproductive advantage relative to other alleles at that locus).

3. Run 18 trials, six at $N = 500$, six at $N = 50$, and six at $N = 5$.
4. For each trial, record the following information:
   a. Trial #
   b. Population size ($N$)
   c. Generations to first fixation (of any allele)
   d. Color of first to fixation
   e. Number alleles fixed in trial (note whether to 1.0 or 0.0).

<table>
<thead>
<tr>
<th>Run#</th>
<th>$N = 500$</th>
<th>$N = 50$</th>
<th>$N = 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. Answer the following questions:
   a. Within each trial, did each of the 6 loci behave similarly? Why or why not?
   b. Did each color loci behave similarly across the iterations? Why or why not?
   c. Were particular colors most likely to be the first to go to fixation? Why or why not?
   d. Within a given population size, how much did time to first fixation vary?
   e. How does population size affect time to first fixation?
   f. If these loci are neutral with respect to selection, why are they changing in frequency over time? Why are some alleles winners and others losers?
   g. Many people confuse small population size effects with drift. Genetic drift is one effect of small population size (see also founder effects and bottleneck effects). One easy way to remember drift is that the colored lines were drifting randomly around on the plot. That random drifting is genetic drift. Note that drift occurs even in large populations, but is more dramatic and consequential in small populations.
   h. These simulations show changes in gene frequency over time. Isn’t that the definition of evolution? Were we watching evolution?
   i. What is the probable genetic fate of endangered species? Does a species have to be critically endangered to suffer loss of genetic variation? What is the 50:500 rule?
   j. For a given population, can you precisely predict when loss of genetic variation (fixation) will occur?
Simulation D: Drift and Selection - In the real world, drift and selection often operate simultaneously. In fact, drift and selection are probably the two most important agents of evolutionary change. But do they necessarily work hand in hand? Consider again our island mice. With the arrival of owls, the selective regime is against those very common white mice (but note that it’s not lethal to be a white mouse – they do 90% as well as darker mice since they hide well in dense island vegetation…). Information on the drift and selection models is found on page 77 of the “help” pdf.

1. Go to Genetic Drift Models under Mendelian Genetics. Choose Drift and Selection. Alter the default settings to read:
   a. N = 500, p = 0.1, Generations = 500
   b. AA = 1, Aa = 1, aa = .9

2. Before you run the simulation, consider: in the absence of drift, do you expect the a allele to go extinct in such a population? Explain your answer.

3. Now hit view and see what happens over 500 generations. Hit view 5 more times, each time seeing what happens. What did happen? Was it the same every time? Why did the A allele go to fixation in this exercise but not in Simulation 1?

4. So … selection is pushing the frequency of the A allele upwards. But unlike what we saw in Simulation 1, it is not a smooth monotonic increase. The increase is jerky. That drifting line IS genetic drift in action.

5. Now change the population size to 50 and hit view. What happens in this smaller population?

6. Hit view 5 more times. Are your results similar to what you saw in the population of 500?

7. When the population size was 500, you undoubtedly saw the frequency of A make its way upward (jerkily) until it eventually (at least often) became fixed in the population. Remember in our first simulation how the frequency of A went close to fixation but never quite became fixed? How did we explain the persistence of deleterious recessive alleles in the population in that exercise? So what happened here to purge these deleterious alleles from the population? Drift! Natural selection can’t weed out rare recessives, but drift can! So drift can take a population to places that natural selection alone can’t!! Of course, this is a double-edged sword. Rare recessives aren’t necessarily deleterious … and may be a reservoir of genetic variation within a population. Drift is very effective at eliminating rare alleles, regardless of their “value”.

8. Of course, drift doesn’t just remove deleterious and neutral alleles. If you keep running the program at population = 50, you’ll probably eventually run into a situation where the frequency of A declined and hit zero (that is, a became fixed in the population), EVENTHOUGH “a” WAS BEING SELECTED AGAINST! We saw in earlier trials that natural selection can easily push the frequency of a beneficial allele from 0.1 to 0.99. So why was selection not able to do that in this case? A rare allele is vulnerable to drift, even when beneficial.
9. Were the mice on the island evolving? If so, what mechanism was responsible?

10. What do these results say about the power of selection and drift in small and large populations?

Assignment
Your task is to consider ways of presenting these raw data (from *Woozleology* and *Simulation C*) in graphical fashion to illustrate two major points regarding the data. Processing of these data is critical, because a report that included just the raw data would be difficult to read and understand. In addition, you will want to practice producing publication quality graphs for future reports and assignments. Use the information I handed out last week regarding construction of Excel graphs to guide your efforts. Cut and paste the graphs into Word and add proper figure and table legends (see Appendix C in handout). Work in pairs and turn in via electronic submission. That is, email the Word document (NOT Excel files) to me. The file should be named using the following convention: lastnames_graphs_ex1.doc.
Calculation of Allele and Genotype Frequencies & Hardy-Weinberg Equilibrium Theory

Written By Dr. Patricia Peroni

INTRODUCTION
Population geneticists study frequencies of genotypes and alleles within populations. By comparing these frequencies with those predicted by null models that assume no evolutionary mechanisms are acting on populations, they draw conclusions regarding the evolutionary forces in operation. In a constant environment, genes will continue to sort similarly for generations upon generations. The observation of this constancy led two researchers, G. Hardy and W. Weinberg, to express an important relationship in evolution. The Hardy-Weinberg Equilibrium Theory serves as the basic null model for population genetics.

If we take all of the alleles of a group of individuals of the same species (that is, a population) we have what is called the gene pool. The frequency, or proportion, of individuals in that population that possess a certain allele is called the allele frequency. Populations can have allele frequencies, but individuals cannot. This obviously makes populations the best hierarchical level to study evolution, as evolution is basically the study of the change in allele frequencies over time.

Allele Frequencies
Consider an individual locus and a population of diploid individuals where two different alleles, A and a, can be found at that locus. If your population consists of 100 individuals, then that group possesses 200 alleles for this locus (100 individuals x 2 alleles at that locus per individual). The number of A alleles present in that population expressed as a fraction of all the alleles (A or a) at that locus represents the frequency of the A allele in the population.

1. To calculate allele frequencies for populations of diploid organisms, first multiply the number of individuals in the population by 2 to obtain the total number of alleles at that locus.

2. Select one of the alleles for your first set of calculations. Let’s first choose the A allele from the example provided above.

a. Individuals homozygous for the A allele will each possess 2 A alleles. Multiply the number of AA homozygotes by 2 to calculate the number of A alleles.

b. Heterozygotes will each possess only one A allele.

c. The total number of A alleles in the population = [(the number of Aa heterozygotes) + (2 x the number of AA homozygotes)]

3. The frequency of the A allele = [(total number of A alleles in the population) / (total number of alleles in population for that locus)]

4. The frequency of the a allele = (1 - frequency of the A allele)

Genotype Frequencies
Consider the same population, locus, and alleles described above. Genotype frequencies represent the abundance of each genotype within a population as a fraction of the population
size. In other words, the frequency of the AA genotype represents the fraction of the population homozygous for the A allele.

1. To calculate genotype frequencies for populations of diploid organisms, first determine the number of individuals with each genotype in the population. In the example above, count the number of individuals with each of the following genotypes: AA, Aa, and aa.

2. To determine the frequency of each genotype, divide the number of individuals with that genotype by the total number of individuals in the population. For example, frequency of AA genotype = # AA individuals / population size.

**IMPORTANT NOTE:**

Unless you know that a population meets Hardy-Weinberg equilibrium assumptions, you must use the above procedure to calculate genotype frequencies. If you know that a population meets Hardy-Weinberg expectations, then you can calculate genotype frequencies using allele frequencies and the Hardy-Weinberg equations (see below).

**Assertions of the Hardy-Weinberg Equilibrium Theory**

The Hardy-Weinberg Equilibrium Theory refers to loci within populations that experience no evolutionary mechanisms (i.e., selective forces). For such populations the theory asserts that:

1. Allele and genotype frequencies should remain constant from one generation to the next (i.e., no evolution has occurred). If, at a certain gene locus, there are only two alleles each will have a frequency such that the frequency of one allele plus the other equals one. Remember, we are discussing the frequency in a population, not in an individual. Formally, we can state the allelic frequency in a population as follows:

\[
p = \text{Frequency of allele A} = \text{freq}(A)
\]
\[
q = \text{Frequency of allele } a = \text{freq}(a) \text{ and } p + q = 1
\]

2. Given a certain set of allele frequencies, genotype frequencies should conform to those calculated using basic probability. In a one locus/two allele system such as the one described above, the genotype frequencies should be as follows:

a. Frequency of AA genotype = (frequency of A allele)²

b. Frequency of aa genotype = (frequency of a allele)²

c. Frequency of Aa genotype = 2 x (frequency of A allele) x (frequency of a allele)

Within a population, the frequency of the possible combinations of a pair of alleles at one locus is related to the expansion of the binomial \((p + q)²\). The expansion is

\[
(p + q) x (p + q) = p² + 2pq + q² = 1, \text{ where}
\]

\[
p² = \text{Frequency of genotype } A/A
\]
\[
2pq = \text{Frequency of genotype } A/a
\]
\[
q² = \text{Frequency of genotype } a/a
\]
3. If the genotype frequencies obtained from a real population do not agree with those predicted by the Hardy-Weinberg Theory, then we know that some evolutionary mechanism or mechanisms must operate on the locus of interest. Knowledge of the theory can help narrow down possible mechanisms. Then we can use experiments to determine which potential mechanism or mechanisms operate on the locus. Thus, the Hardy-Weinberg Equilibrium Theory serves as an important tool for population geneticists.

**Assumptions of the Hardy-Weinberg Equilibrium Theory (Evolutionary Mechanisms)**

The assumptions that populations must meet in order for the H-W assertions to hold are:

1. Large population size (i.e., no genetic drift). Random chance can alter allele frequencies through mating processes and death within small populations.

2. Random mating, which means that the choice of mates by individuals in the population is determined by chance, and not influenced by the genotypes of the individuals in question.

3. No difference in the mutation rates between alleles at the same locus.

4. Reproductive isolation from other populations (i.e., no gene flow or migration).

5. No differential survival or reproduction among phenotypes (i.e., no natural selection).

**Example**

Consider a population of 1000 individuals and the locus and alleles described above. Assume that you have no information on the presence or absence of evolutionary mechanism in this population. You find that the population consists of:

- 90 individuals homozygous for the A allele (AA genotype)
- 490 individuals homozygous for the a allele (aa genotype)
- 420 heterozygotes (Aa genotype)

1. Calculate the genotype and allele frequencies for this locus.

2. Determine if this population meets Hardy Weinberg Assumptions (in other words determine if evolutionary mechanisms operate in this population).

**Calculation of Allele and Genotype Frequencies**

Since you do not know if this population meets Hardy Weinberg Assumptions, you must calculate both the allele and genotype frequencies using the raw data.

1. **Allele Frequencies:**
   - The frequency of the A allele will equal: \( \frac{(\text{total number of A alleles in the population})}{(\text{total number of alleles in population for locus})} = \frac{(90*2) + 420}{(1000*2)} = 0.30 \)
   - The frequency of the a allele will equal: \( 1 - 0.30 \) or \( \frac{(\text{total number of a alleles in the population})}{(\text{total number of alleles in population})} = \frac{(490*2) + 420}{(1000*2)} = 0.70 \)

2. **Genotype frequencies:**
• Frequency of AA genotype = # AA individuals / population size = 90/1000 = 0.09

• Frequency of Aa genotype = # Aa individuals / population size = 420/1000 = 0.42

• Frequency of aa genotype = # aa individuals / population size = 490/1000 = 0.49

**Hardy-Weinberg Predictions**

If no evolutionary mechanisms operate on this locus, then the Hardy-Weinberg Equilibrium Theory predicts that the genotype frequencies should be as follows:

• Frequency of AA = (frequency of A allele)² = (0.3)² = 0.09

• Frequency of Aa = 2 x (frequency of A allele) x (frequency of a allele) = 2*0.3*0.7 = 0.42

• Frequency of aa = (frequency of a allele)² = (0.7)² = 0.49

**Conclusion**

Since the observed genotype frequencies equal those predicted by the Hardy-Weinberg Equilibrium Theory, we conclude that no evolutionary mechanisms operate on this locus in this population (i.e., the population meets the assumptions of the Hardy Weinberg Theory).