ECOLOGY and EVOLUTION

Week 2: September 6 – September 9, 2011

LAB II – Part 2. Evolutionary mechanisms

Before coming to lab, read this lab manual and be sure to know the following terms:

- Gene locus
- Haploid
- Diploid
- Allele
- Homozygote
- Heterozygote
- Allele Frequency
- Genotype Frequency

Last week we focused on the presentation format for formally reporting statistical results.

This week we will focus on biology. Use your lab notebook to take notes as you go. Print outs of graphs and statistical results will go in your Bio Binders. You should write notes on the graphs to remind you what was being studied, and answer the questions that you find in the lab handout. This is not to turn in, but for your use in studying, so the format requirements are relaxed this time.

Among the mechanisms underlying the evolution of species are those that bring about changes in allele frequencies in populations. For such changes to take place, there must be genetic variation among individuals. This variation results from gene mutation, migration, or genetic recombination. These genotypic variants may be exposed to random sampling error (stochastic processes), natural selection, and nonrandom mating, all of which can result in long-term changes in the genetic structure of populations. In addition to being important to an understanding of evolutionary biology, these factors are also at the heart of conservation biology. Does a particular population have a genetic structure that will allow it to survive in the face of environmental degradation? How can we enhance conditions that would promote a favorable genetic structure, and what can we avoid that would threaten it?

Allele Frequency Changes in a Population:

The one gene locus, two allele model

Evolutionary processes are difficult to study because they occur over fairly long time periods. Even with an organism like a fruit fly that has a short generation time, it would take up to several weeks to complete a single generation. As a starting point for understanding processes that cause genetic changes in populations, we can begin by building both qualitative and quantitative (mathematical) models. (Be sure to read the

Bio Binder Quantitative Experimental Design Section F 1- 5.) Once the principles exemplified by our models are understood, we will be in a position to ask more refined questions that justify the investment in time and effort required to study a real population. Today in lab we are going to use **simulation techniques** to learn the basic principles of **population genetics**. Population genetics provides an excellent opportunity for exploring the interplay between biology and mathematics – it has one of the richest bodies of mathematical theory in biology and has been developing continuously for more than 90 years.

The following experiments will first be performed on **physical models** that will illustrate several processes that, because they are based on genetic events, are strongly influenced by **random processes**. We will then turn to the computers in lab and repeat these experiments and extend our analysis of the model with **computer simulations** that have the same properties. For the physical model, we will use colored beads to represent mating individuals and gametes, as specified below.

The Founder Principle

This term was introduced by Ernst Mayr, who pointed out that the establishment of a new colony of a species involves a sampling effect because the limited number of founders contains only a small fraction of the genetic variation in the parental population. The ultimate success or extinction of the new colony will depend in part on the quality of its genes in its new environment. Different colonies of the same size stemming from the same parental population are likely to contain different genetic compositions (different allele frequencies) and may therefore have different fates.

In this model, the large class container represents a parental population and individual student samples taken from it will represent colonies of diploid individuals. We shall consider only one pair of alleles. The genotypes are represented by colored beads as follows: *red AA; white Aa; green aa*. The genotype frequencies in the parental population are unknown.

Procedure - This exercise is to be done individually**.**

First, take a random sample of 20 beads to your desk in a bowl. This represents your **founder population**. Really think of these beads as individual animals. You therefore just behaved like a hurricane blowing 20 animals out of their parental population in South America onto the Galapagos Islands 600 miles away.

On the next page and in your lab notebook, record the genotypes of your founder population by counting the different colors in your sample and **compute** the genotypic and allele frequencies (see next page for equations).

 $(f = proportion of individuals of each genotype; i.e., # of a color divided by 20)$

Allele frequencies: Let the frequency of $A = f(A) = p$, and frequency of $a = f(a) = q$.

Compare the above allele frequencies with others in your class.

Hardy-Weinberg Equilibrium

A population which is said to be in Hardy-Weinberg equilibrium is one in which no change occurs in the genetic structure from one generation to the next. There are five specific assumptions that must be met for this situation to occur (no mutation, no migration, no natural selection, random mating, and very large population size). Under these special conditions, we can develop a specific, numerical expectation of what the genotype frequencies of the next generation will be. These genotype frequencies are determined solely by the Mendelian ratios after random mating, and the allele frequencies themselves will not change. This special condition of population genetic equilibrium was first pointed out by the English mathematician Hardy in 1908 and independently by Weinberg the same year. For a model like the one we are using that has one gene locus and two alleles, the sum of the two allele frequencies must always be 1.0, and the sum of the three genotype frequencies must also be 1.0.

Alleles: $p + q = 1$ (where p is the frequency of allele "A" and q is the frequency of allele "a")

If you know the two allele frequencies **and** if the five assumptions hold, then the following genotype frequencies must occur and the sum of the three must also equal 1.0:

Genotypes: $p^2 + 2pq + q^2 = 1$ (where $p^2 = f(AA)$, $2pq = f(Aa)$, and $q^2 = f(aa)$ only when the population is in Hardy-Weinberg equilibrium).

Calculate the expected genotype frequencies for a population in Hardy-Weinberg equilibrium that has the same allele frequencies as your **founder population**. Use your calculated p and q from **page 3** to calculate p^2 , 2pq, and q^2 . **Compare** these three values to your f(AA), f(Aa), and f(aa) to see if your original sample (Founder Population) was by chance in Hardy-Weinberg equilibrium.

Please note: checking that **f(AA)**, **f(Aa)**, and **f(aa)** sum to 1.0 is merely a check of your math and is not an indication of equilibrium. This is a common mistake on exam questions of this nature.

Actual deviations from equilibrium either in a model or in the real world indicate that the genetic structure of the population is changing, i.e., it is evolving and one or more of the five assumptions is being violated.

Genetic Drift (violating the assumption of very large population)

The Hardy-Weinberg equilibrium gives the theoretical expectation of genotypic distributions in a very large population. We shall now use our physical model to investigate how statistical sampling affects allele frequencies in a small population.

- Start with the founder group of 20 individuals.
- **Set up** a small beaker of gametes with one amber (A) and one yellow (a) bead to determine mating outcomes for heterozygous individuals.
- **Draw** two beads randomly from the 20 beads in the bowl. This is considered a mating.
- **Record** the mating type on the worksheet (page 6).
- **Return** the beads to the bowl.
- **Repeat and record for a total of 10 matings.**

We will assume that the population size remains constant. This means that for each pair of parents, two and only two progeny survive to reproduce. Note that for (AA) x (AA), (AA) x (aa), and (aa) x (aa) matings no gamete drawing is necessary because the progeny will be of a single genotype. The yellow and amber beads will be used to determine the progeny genotypes for matings that involve heterozygotes (whites).

So for (Aa) x (AA) or (Aa) x (aa) one parent contributes an uncertain gamete that needs to be determined by chance. For (Aa) x (Aa) each parent contributes an uncertain gamete that needs to be determined by chance. So,

- Draw gametes as required.
- Record on the worksheet (page 6) the genotypes of each of the two progeny obtained from each of the ten matings.
- Compute the totals for each progeny column this yields the number of offspring of each genotype and should sum to 20.
- Reconstitute the population by making the number of beads you are about to mate match the new offspring numbers. (Your adult beads are about to die and their progeny get to mate in the next generation). To accomplish this you will return the Red, Green and White beads to the main population and replace them with offspring of the proper genotype.

Calculate the genotype and allele frequencies for this first post-founder generation. Then continue your matings (using the new offspring from the post-founder generation) for one more generation.

Calculate the new genotype and allele frequencies of the second generation after the founding event.

Return to the class container and adjust your population of 20 to match your Generation 2 genotype frequencies.

In your lab notebook, roughly **plot** the frequency of **one** of the alleles on the y-axis and on the x-axis plot time in generations (time 0 for the founder) and then for the next two generations $(i.e., t = 0, 1, 2)$. **What will happen if you keep allowing your population to breed in this manner?** To further explore the unfolding of this "**random walk" (genetic drift),** we will continue our simulation on computers.

Genetic Drift Simulation: WORKSHEET

Starting allele frequency: A: $____ +$ a: $____ = 1.0$

Matings for Generation 1:

Be sure to change beads to reflect the **progeny** genotypes from the first generation. The parents have died, and these 20 progeny are now the mating population for this next round of the simulation.

Matings for Generation 2:

A Computer Simulation study of Genetic Drift

Part 1. Getting to know the simulation software

Working with **PopBio** software and combining it with last week's analytical work with **JMP** and **Word**.

For the remainder of the exercises, work in pairs.

- **Connect** to the **Home Server** so that you will have access to your folder for saving files.
- **Open JMP**.

• **Open PopBio** from the dock $\left[\frac{\mathbf{E}[\mathbf{x}]}{\mathbf{x}}\right]$ and select **Population Genetics**.

PopBio has been developed by students and faculty of Reed's Biology Dept. to provide a computer simulation that has properties similar to the bead experiment that you just completed. During your computer simulations, you will not be able to watch changes in the actual genotype frequencies. The output from **PopBio** is only expressed in terms of allele frequencies.

• **Set up** the parameters as shown below, **except** for **A allele frequency**. You should enter the **A allele frequency** you (or your partner) recorded on page 6 for the second generation of your bead simulation. You will therefore now begin the simulation at generation 3.

• Click **Calculate**

What was the outcome of your simulation? Spend a moment to consider how long this would have taken you if you had continued using your bead model for 500 more generations. Then consider how long it would take in a real population. Does the time it takes for the process to operate in the real world diminish the value of the computer simulation model?

What is the effect of genetic drift on the amount of genetic variation in populations?

- 1. **Click** on the **Calculate** button repeatedly.
- 2. **Run** simulations until you get an idea as to the answer.
- 3. **Run** the same simulation 100 times by setting the **Number of Trials** to 100.

The data window will automatically record the amount of time for fixation or loss of the allele for each run. Notice the outcome is rarely ever the same. This is a stochastic model in which chance plays an important role in determining the outcome.

- 4. Using the **Save Options** menu at the bottom of the dataset window, **Save table as text**. **Save** your **PopBio** data set to your **Desktop** with a clear file name.
- 5. From within **JMP**, **Open** from the **File** menu, and select the **.txt** file from the **Desktop**. **Open As: Data (Best Guess). Open.**
- 6. Construct a **frequency distribution** (see Bio Binder JMP Instructions G-6) for the **Time to Fix/Loss** Y variable.
- 7. **Select All, Copy** and **Paste** to a **Word** file labeled as **Genetic Drift.**
- 8. **Save** to the **Desktop.**
- 9. **Add** the following figure legend to the **Word** file below the graph:

Figure 1. Frequency distribution of time in generations to the completion of a random walk of a population with 20 individuals and a starting A allele frequency of .

- 10. **Add** your name at the top of the page.
- 11. **Save, Print** two copies**,** and place in your Bio Binders with this lab manual for your records**.**

We can't predict exactly how long it will take for a particular random walk to be completed. Can we derive any general principles from this model?

* Close the data window in PopBio before starting the next experiment.

Part 2. Experiments with the simulation software

Experiment 1. What is the effect of initial allele frequency on the time to fixation or loss of an allele from a population analyzing your results with ANOVA?

Experiment 2. What is the effect of population size on time to fixation or loss of an allele from a population analyzing your results with Bivariate Regression Analysis?

Experiment 1: Effect of initial allele frequency

In order to answer the first question (What is the effect of initial allele frequency on the time to fixation or loss of an allele from a population using ANOVA?), we will design an experiment where the null hypothesis is that there is no significant effect of initial allele frequency on the amount of time to fixation or loss of an allele from a population. Our experiment will be designed so that the independent variable is initial allele frequency and our response variable is time to fixation or loss. All other parameters of the model will be kept constant.

- 1. Set up **PopBio Population genetics** with a **population size** of **20**.
- 2. In one data window, run the following 3 initial allele frequencies.
- 3. **Run** the simulation with **50 multiple trials** with an **initial allele frequency** of **0.1**. Do not close the data window.
- 4. Change the **initial allele frequency** to **0.25** and repeat the **50 multiple trials**.
- 5. Change the **initial allele frequency** to **0.50** and repeat the **50 multiple trials**.
- 6. **Save** the data table to the **Desktop** as **text** as before.
- 7. **Open** the **.txt** file in **JMP**. Double click on the column name for **Allele A freq** and set **Data Type** to **Character**. **Modeling Type** will then automatically change to **Nominal**. Click **OK**.
- 8. Run an **ANOVA** (Bio Binder JMP Instructions G-9+10) on **Time to fixation/loss** as the **Continuous** (Y) response variable and the discrete (nominal) **initial allele frequency** as the **Construct Model Effect**.
- 9. Double click on the **Y-axis** of your **JMP LS Means Plot**, and change the scale to make the differences more visible. In the same window, add **gridlines** to the major and minor tickmarks for the Y-axis. Print out two copies of this, punch holes, and put it in your Bio Binders for study purposes.
- 10. What is the effect of changing allele frequency on the time to fixation or loss of an allele from a population of a constant size of 20 individuals?
- 11. The null hypothesis is that there is no significant effect of initial allele frequency on the time to fixation or loss of an allele from a population.

Experiment 2: Effect of population size

In order to answer the second question (What is the effect of population size on time to fixation or loss of an allele from a population using Bivariate Regression Analysis?), we will design an experiment where the null hypothesis is that there is no significant relationship between **population size** and the amount of time to fixation or loss of an allele from a population. Our experiment will be designed so that the independent variable is population size and our response variable (as before) is time to fixation or loss of an allele. All other parameters of the model will be kept constant.

*** Close the data window in PopBio before starting the next experiment.**

- 1. Set up **PopBio Population genetics** with **allele frequency** at **0.5**.
- 2. In one data window, run the following 5 population sizes. Do not close the data window until you have run all 5 population sizes.
- 3. **Run** the simulation with **15 multiple trials** with a **population size** of **10**.
- 4. Change the **population size** to **15, 20, 25, and 30** repeating the **15 multiple trials** for each population size.
- 5. **Save** the data table to the **Desktop** as **text**.
- 6. **Open** the **.txt** file in **JMP** and **run** a **bivariate regression analysis** (like last week) on **Time to fixation/loss** as the **continuous** (Y) response variable and **population size** as the **continuous** (X) independent variable. (See Bio Binder JMP Instructions G-12+13.) Don't forget **Fit Line**.
- 7. Double click on the **Y-axis** of your **JMP Plot**, and change the scale to make the differences more visible. In the same window, add **Gridlines** to the **Major** and **Minor Tickmarks** for the Y-axis. Print out two copies of this, punch holes, and put it in your Bio Binders for study purposes
- 8. Based on the **P value**, what is the effect of changing population size from 10 individuals to 30 individuals?
- 9. The null hypothesis is that there is no significant effect of population size on the time to fixation or loss of an allele from a population.

The important point of simulation experiments 1 and 2 is to understand that even in stochastic models where the outcome is unpredictable, there are general laws operating that can tell us much about how the genetic structure of populations will change over time. An understanding of the results of these experiments is crucial to an understanding of the genetic factors that are important to both evolution and conservation.

Using PopBio to Explore Natural Selection

As you recall, the Hardy-Weinberg equilibrium maintains that as long as populations are very large and there is no mutation, migration, or natural selection, and mating is random, then the allele frequencies will not change and the genotypes will be distributed **binomially** according to the formula $p^2 + 2pq + q^2 = 1$ where p is the frequency of one allele and q is the frequency of the other allele in a one gene locus, 2 allele system. We just saw that when we violate one of the assumptions, namely large population size, **allele frequency is not constant but randomly walks to ultimate extinction or fixation**.

Now we're going to make our model **deterministic** as opposed to **stochastic** by restoring population size to infinity and violate the assumption of no natural selection and see what happens to the genetic equilibrium. **De-select** the **Simulate stochastic drift** option box and **Select** the **Simulate natural selection** box.

This causes the population size parameter to be bypassed and the population is set at infinity automatically. **The model is now deterministic and not stochastic**, meaning that the outcome of each simulation is identical as long as the values of the parameters don't change. The deterministic model can be written as follows:

$$
P_{t+1} = \frac{(P_t W_{AA} + q_t W_{Aa}) P_t}{P_t^2 W_{AA} + 2 P_t q_t W_{Aa} + q_t^2 W_{aa}}
$$

Where p_{t+1} is the allele frequency in the next generation and "W" denotes the **genotype fitness** (i.e., adaptive value of a particular genotype).

Let's explore **three** basic types of natural selection. (Note: the characterization of natural selection as a single force always leading to improved phenotypes is a misleading oversimplification at best.) We will not use ANOVA or regression techniques but will use a new feature of the software called an **Accumulator Plot** to make a visual record of patterns of allele frequency changes that result from the three basic types of natural selection.

12 Lab II Evolutionary Mechanisms Reed College Bio 101/102: Kaplan

First type of Natural Selection: selection against a recessive allele - If "A" is a recessive allele, it only impacts the phenotype when it is in the homozygous state (i.e., genotype AA). Since "A" allele is going to decline in frequency because we're selecting against its homozygote genotype, let's set its allele frequency to 0.95 in the parameter box.

The starting population of diploid individuals is set up by the program to consist of genotypes with Hardy-Weinberg proportions based on this allele frequency (i.e., 90.25% AA, 9.50% Aa, and 0.25% aa).

a. **Recessive lethal** – A recessive lethal condition occurs when the homozygote for an allele results in zero Darwinian fitness but in heterozygous form it does not. Heterozygotes are carriers.

- **Quit Popbio** and **restart**.
- The **Simulate stochastic drift** box should not be checked.
- **Select** the **Simulate natural selection** box.
- **Set** the **A allele frequency** to 0.95.
- **Set** the **Genotype AA fitness** to 0. This is also called the **adaptive value** and is equivalent to a **selection coefficient (s)** of 1, where **s** is a proportional selection against a genotype. Therefore, the *adaptive value or relative fitness* $(= 1 - s)$ is 0 for the AA genotype. Each time an AA individual is born, it is genetically dead (i.e., sterile at least).
- **Set** the number of **Generations** to 250.
- **Calculate** to run the simulation.
- **Observe** the shape of the graph. Is the A allele ever eliminated from the population? Why? This phenomenon is called a protected polymorphism - the last A recessive allele is always hidden in the last heterozygote.
- **Click** the **A+** symbol at the bottom of the graph.
- **Assign** the name **recessive lethal** to the **accumulator plot** when prompted.
- **Add Plot**.

b. **Reduced selection against a recessive allele** – Selection against a recessive allele does not have to be lethal in the homozygous condition. Continuing to work with the **Accumulator Plot** above:

- **Set** the **Genotype AA fitness** to 0.90, and click **Calculate**.
- **Click** the **A+** symbol at the bottom of the graph.
- **Assign** the name **recessive fitness 0.9** to the **accumulator plot** when prompted.
- **Add Plot**. Note the accumulator plot shows both **recessive lethal** from above and **recessive fitness 0.9**.
- Under **Save Options…, Save Plot as TIFF. Save** the graph (accumulated **recessive lethal** and **recessive fitness 0.9**) to the **Desktop** calling it **Selection against a recessive allele**.
- **Print** two copies.
- Use this to help you to label the curves by hand on your printout:

Show Curves List -> Select Plot Name-> red line is that plot.

Thought question: How and why do the two accumulator plots from a and b differ?

Second type of Natural Selection: selection against a dominant allele. - In this case the allele that is selected against is selected against equally in both the homozygote and heterozygote condition.

- **Quit Popbio** and **restart**.
- The **Simulate stochastic drift** box should not be checked.
- **Select** the **Simulate natural selection** box.
- **Set** the **A allele frequency** to 0.95.
- **Set** the **Genotype AA fitness** to 0.5.
- **Set** the **Genotype Aa fitness** to 0.5.
- **Set** the number of **Generations** to 50.
- **Click Calculate** to run the simulation.
- **Observe** the shape of the graph. Is the A allele ever eliminated from the population? Why?
- **Click** the **A+** symbol at the bottom of the graph.
- **Assign** the name **Dominant fitness 0.5** to the **accumulator plot** when prompted.
- **Add Plot**.
- **Adjust** the intensity of selection (fitness) up to 0.75 using the same value of selection against both AA and Aa.
- **Click** the **A+** symbol at the bottom of the graph.
- **Assign** the name **Dominant fitness 0.75** to the **accumulator plot** when prompted.
- **Add Plot**. Note the accumulator plot shows both graphs.
- Under **Save Options…, Save Plot as TIFF. Save** the graph to the **Desktop** calling it **Selection against a dominant allele**.
- **Print** two copies.

Use **Show Curves List** to help you to label the curves by hand on your printout.

Thought question: How and why does selection against a recessive differ from selection against a dominant?

Third type of Natural Selection: selection favoring heterozygotes -

- **Quit Popbio** and **restart**.
- The **Simulate stochastic drift** box should not be checked.
- **Select** the **Simulate natural selection** box.
- **Set** the **A allele frequency** to 0.95.
- **Set** the **Genotype AA fitness** to 0.8.
- **Set** the **Genotype Aa fitness** to 1.0.
- **Set** the **Genotype aa fitness** to 0.8.
- **Set** the number of **Generations** to 50.
- **Click Calculate** to run the simulation.
- **Observe** the shape of the graph. What happened? Why?
- **Click** the **A+** symbol at the bottom of the graph.
- **Assign** name **fA=0.95; fitness AA, aa=0.8** to the **accumulator plot.**
- **Add Plot**.
- **Set** the **A allele frequency** to 0.05. Leave all other settings.
- **Click** the **A+** symbol at the bottom of the graph.
- **Assign** name **fA =0.05; AA, aa = 0.8** to the **accumulator plot** when prompted.
- **Add Plot**. Note the **accumulator plot** shows both graphs.
- Under **Save Options…, Save Plot as TIFF.** Save the graph to **Desktop** calling it **Selection favoring heterozygotes**.
- **Print** two copies and label by hand.

Thought question: How and why does selection against a dominant differ from selection favoring a heterozygote?

Optional runs:

- **Change** the intensity of selection from 0.8 to 0.9. Accumulate the plots.
- **Put** some asymmetry into the system. Make the fitness value for $AA = 0.8$ and the fitness value of aa $= 0.3$. Can you guess what will happen? Run the program to find out. Try some different asymmetrical values until you get a feel for the outcome.

This phenomenon of heterozygote superiority has great evolutionary importance as a mechanism by which genetic variability is conserved in populations. It is a type of balancing selection. It is also the explanation for why some genetically based diseases such as sickle cell anemia are not eliminated via selection. Try a run where the heterozygote is superior to both homozygotes, but one of the homozygotes is lethal to see how this system works.

You should have printed two copies of the three saved TIFF files.

1) Selection against a recessive allele.

- 2) Selection against a dominant allele.
- 3) Selection favoring heterozygotes.
- Place these printed documents in your Bio Binders.
- **Save** (**Drag**) the files from the **Desktop** to your **Home Server**.
- **Log Out**. (Do not Shut Down or close the computer lid.)