POPULATION GENETICS SIMULATIONS

The goal of this lab is to investigate the mechanisms of evolution at the population level. Specifically, we will quantify how finite population size, non-random mating, and natural selection cause the allele and genotype frequencies of populations to change. The three mechanisms just listed should sound familiar because they are based on the assumptions of Hardy-Weinberg Equilibrium (HWE). In other words, we are going to use HWE as a null model and explore the consequences of breaking the assumptions of HWE.

To do this, we are going to use simulations that I built in the R Statistical Program. There are five different R functions that you will use to carry out the simulations. You must download the package that contains the functions and the code (script) to run the functions from Canvas. The package is a folder called pop.gen.sims2 (you may need to click on the downloaded .zip file in your internet browser to unzip it). The script with the code is called "Population genetics simulations.r". After you download the script open it in Rstudio. Rstudio is a graphical user interface for the R program.

The script starts with a set of code needed to install and upload the package into R. Run those lines of code first (further detail is given in the script). After the package is uploaded you will be able to run the five functions. Again, further detail for using the functions is included in the script.

You will run the functions and answer the questions given below. You may answer them as a group in a single document that you upload to Canvas when you complete the lab.

Finite population size

The first function you will use is called lp. Run the code ?lp to learn more details about the function and how to input and manipulate simulation parameters. Use the lp function to simulate a large population (N=1000) using initial allele frequencies of p1 = 0.3 and p2 = 0.7 (these should be the default values).

The output includes many things. Some description is given in the function help documentation. H_o is the observed heterozygosity (frequency of heterozygous individuals). H_e is the expected heterozygosity assuming HWE. Wright's F is an index used to compare the observed to expected heterozygosities and is described by the equation:

$$F = 1 - \frac{H_o}{H_e},$$

where a value close to zero indicates that the observed heterozygosity matches the expected, which implies that the assumptions of HWE are met. The allele frequencies are provided along with the results of Fisher's Exact Tests (see documentation in R package for more details) that determine if the genotype counts are equal to those dictated by the initial allele frequencies and those expected given the allele frequencies after the simulation is completed.

1. Do the observed allele frequencies differ significantly from there initial values? How about H_o as compared to H_e? How do the Fisher's Exact Tests help you answer these questions.

2. Do your answers to the questions given in number one change when you decrease the population size to 10? If yes (they may not!), describe how.

3. Give a biological (OK, it is really a mathematical) explanation for why Wright's F value is not zero.

The next function is called finite_pop. It is very similar to the lp function except it simulates a population over many generations. Run the simulation (after using ?finite_pop) using the default parameter values. Note that they are all the same as the initial values for the lp function, except there is an additional parameter: generations.

This function outputs three plots that are described in the help documentation and Fisher's Exact Tests similar to those described above. You may need to resize the plot window to see all of the figure legends.

4. Are there any patterns in the plot of allele frequency? Give an explanation for why or why not.

5. How does H_o compare to H_e? They should change in concert with each other through time (from generation to generation). Why? How is that reflected in Wright's F and Fisher's Exact Test results?

6. A relatively recent study (Cameron et al. 2011) indicates that due to habitat fragmentation and destruction many native bumble bee species have experienced reductions in abundance of up to 96% (Figure 1). Reduce the population size from 1000 to 40. Run the function several times and note how the allele frequencies change through time. Explain why reducing the population size changes the allele frequency dynamics. Use the Fisher's Exact Test results in your explanation. Speculate how this could be detrimental to the native bumble bee populations.

Figure 1. Image of *Bombus impatiens* (Hymenoptera: Apidae). One of the many bumble bee species declining in number due to anthropogenic effects.

Non-random mating

The third function you will run is called non_rand_mating. This function simulates a population produced as a result of completely random mating (outcrossing rate = 1), completely self mating (outcrossing rate = 0), or any value in between. Run the function with the default values. The default outcrossing rate is 1.

7. Is there any notable difference between your results and the ones you got when you first ran the lp function? Focus your answer on Wright's F and the Fisher's Exact Tests.

8. Run the function again, but with an outcrossing rate = 0. Explain how and why your results change.

The fourth function you will use is called finite_non_rand. This function combines the effects of both finite population size and non-random mating over many generations. It also runs many replicates meaning that it runs the simulation many times (default = 100) and averages values over those many runs. These simulations may take awhile to run, so be patient. While they are running, they print which replicate was just completed.

Run the simulation with the default settings (outcrossing rate = 0.5). The plots produced are similar to those described in previous functions, but the heterozygosity plot also includes expected H (H_e) assuming that genetic drift is the only evolutionary force (look in your lecture notes to review the equation).

9. How does the allele frequency change through time (note the y-axis scale!)? How about Wright's F. Use the Heterozygosity plot to explain your answer.

10. The aphid (Figure 2) is a very interesting organism because its environment significantly influences its mode of reproduction. If mates are not available, some individuals will reproduce asexually. Decrease the outcrossing rate to 0 and run the function. Use the results from the plots produced to predict how aphid genotype and allele frequencies will change as a result of asexual reproduction.

11. Return the outcrossing rate to 0.5 and decrease the population size (ND_nr) to 10. How do the three plots change, as compared to the simulations you ran for numbers 9? Note that in the top left corner of the plot window are backward and forward arrows that you can use to toggle between simulation plots.

Figure 2. *Aphis nerii* (Hemiptera: Aphidae) reproduce both sexually and asexually (parthenogenesis), which can significantly alter genotype frequencies through time.

<u>Natural selection</u>

The final function you will use is called finite_selection. This function combines the effects of both finite population size and natural selection over many generations. It also runs many replicates meaning that it runs the simulation many times (default = 100) and averages values over those many runs. These simulations may take awhile to run, so be patient. While they are running, they print which replicate was just completed.

12. Run the simulation with the default settings. Explain why the allele frequency changes the way it does and why the H_o is so different than the H_e - drift. More specifically why does H_o increase at first and then decrease?

13. Look at Figure 3 below. In the research depicted in the figure, scientists raised experimental populations of fruit flies on food spiked with ethanol, and monitored the frequency of the *Adh*^{F} allele over 50 generations. *Adh*^{F} encodes a version of the alcohol dehydrogenase enzyme that breaks down ethanol at twice the rate of the version encoded by *Adh*^{S}. The starting frequency of *Adh*^{F} was about 0.35 in both experimental populations; the ending frequency was about 0.9 in one population and about 1 in the other. Use the finite_selection function to estimate the strength of selection for the *Adh*^{F} allele during this experiment. Let p1g2_nr represent the *Adh*^{F} allele. Set the starting frequency of p1g2_nr to 0.35. Set the number of generations to 50, population size to 50, dominance coefficient (h) to 0, and number of replicates to 1. Try different selection coefficient values (between 0 and 1) until you reproduce the pattern of change over time for the flies exposed to ethanol in Figure 3. What selection coefficient works best? What does this selection coefficient (and dominance coefficient) tell you about the relative survival (or reproductive success) of the three genotypes?

14. Change the population size to 10 and run the simulation again several times. Does the *Adh*^{S} gene ever reach fixation? How does decreasing the population size change the probability of fixation? Use the function $N_e s$ to explain your answer.

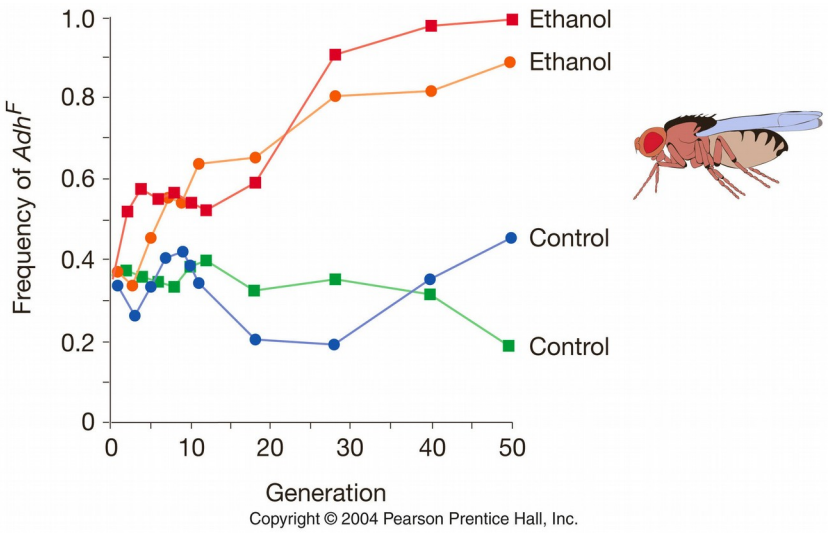

Figure 3. Frequency of *Adh*^{F} gene over time in *Drosophila melanogaster* (Diptera: Drosophilidae) exposed to ethanol and controls (not exposed to ethanol).

Literature Cited

Cameron, S.A., J.D. Lozier, J.P. Strange, J.B. Koch, N. Cordes, L.F. Solter, T.L. Griswold. 2011. Patterns of widespread decline in North American bumble bees. PNAS 108:662–667.