

Exercise 1: Exploring Evolutionary Mechanisms with Theoretical Computer Simulations, and Calculation of Allele and Genotype Frequencies & Hardy-Weinberg Equilibrium Theory

INTRODUCTION

Evolution is defined as a change in the genetic composition of a population of organisms that occurs over time. More precisely, evolution is a change that occurs over time in the proportions of organisms in a population that differ genetically in one or more characteristics. Just as the life of an individual organism is dynamic, so is that of a species. In the study of biological evolution we can ask what factors are capable of causing change within a population.

Due to time constraints (thousands of generations may be required for one species to evolve) it is difficult to perform evolution experiments on populations in one semester. Our approach will thus be to first investigate these questions using a hypothetical population. We will conduct simulations to determine the factors that can facilitate or inhibit genetic change at one locus within a population. Our hypothetical population will be very simple, and we'll focus on how different factors affect the genotype and allele frequencies at one locus.

The simulation approach we will use represents a type of theoretical investigation. Theoretical inquiry serves as a guide for empirical research (i.e., research that involves taking measurements on real organisms). Real systems are complex, and experimental research with these systems is often time consuming and expensive. Theoretical research allows us to ask "what if" questions using very simple systems that we refer to as models. In a general sense, a model should be thought of as a formalized working hypothesis. That is, in the context of a computer simulation model, the program is written to test a prediction about the mechanisms responsible for a given event (in this case, parameters affecting populations). If the hypothesis is correct, then the program will accurately simulate an event.

The results of such investigations can suggest questions that can be tested with experiments. If the predictions of theoretical investigations and empirical studies are at odds with each other, then we must refine our theoretical models to account for factors omitted from our initial inquiries. The operation of testing a model, and changing it as required, is part of the scientific process. All active areas of research involve this type of interplay between theoretical and empirical research, and our understanding of how the world operates depends upon both types of investigations.

COMPUTER SIMULATIONS AND PROBLEMS

We can study many evolutionary processes that normally occur over a longer time period using computer simulations. Computer simulations offer an excellent opportunity to model some of the processes, such as natural selection and genetic drift, we will discuss in class. The program we are going to use is freeware developed by ecologists at the University of Minnesota. The program is available on the computers in the labs, and is also available from the University of Minn.: <http://www.cbs.umn.edu/software/populus.html>.

Populus 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 that 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. Context-sensitive help screens are 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 (not all shown; ones in bold will be the ones you may use):

- **Genetic Drift Models:**
 - **Genetic Drift**
 - **Inbreeding**
 - Population Structure
 - **Drift and Selection**
- **Selection Models:**
 - Wozleology
 - **Selection on a Diallelic Autosomal Locus**
 - **Selection on a Sex-Linked Locus**
 - **Selection on a Multi-Allelic Locus**
 - **Two Locus Selection**
 - **Selection and Mutation**
- Quantitative Genetics Models:

Brief Synopsis of Assumptions and Questions:

Here you will explore the simulations, and design experiments to test hypotheses about how genetic drift, mutation, and natural selection all affect populations of organisms. For most simulations and problems, 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.

Make the following predictions prior to each simulation, either as a group or as a class:

1. Predict whether the observed genotype frequencies will change substantially as the simulation proceeds (i.e., deviate substantially from those predicted by Hardy-Weinberg equilibrium theory).
2. What will be the nature of the change you expect (e.g., excess of both homozygotes and a deficiency of heterozygotes)?
3. What will be the ultimate outcome of the situation for the population (e.g., loss of the black allele from the population)?

Simulation: 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 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 show 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 8f below).
7. Print out or save copies of the **most relevant** graphs.
8. 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? In fact, 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 ” for this scenario. What does this tell you about how selection can work?

Other Simulations: Investigate the evolutionary mechanisms of genetic drift, mutation, and natural selection using other simulations available. Perform 3 other simulations, with the goal of investigating strength of selection (relative fitnesses), the strength of genetic drift vs. natural selection, and how mutation and natural selection may work relative to the other. In the real world, mutation, 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? More complicated models could be used to investigate selection on two loci, or a single locus with more than 2 alleles. For either of those, you could investigate the strength of selection on different phenotypes.

1. For each simulation you perform, read the background material, consider what you learned in the text and in class, and brainstorm with your partners and instructor to consider assumptions and predictions, using the questions above to guide your efforts.
2. At any point where you and your partners are even a little confused about the simulations and how to use them, **consult your instructor!**
3. Questions to consider for each simulation:
 - a. What is the equilibrium condition?
 - b. What are the major differences between this simulation and others you've performed?
 - c. Is any mechanism of evolution stronger than any other, and if so, under what conditions?
 - d. Most simulations will show changes in gene frequency over time. Isn't that the definition of evolution? Were we watching evolution? Explain your answer.
 - e. For a given population, can you precisely predict when loss of genetic variation (fixation) will occur in a particular simulation?
 - f. Was the population in question evolving? If so, what mechanism was responsible?
 - g. What do your results say about the power of mutation, selection or drift in small and large populations?

Assignment

You and your partners will generate data from the Populus simulations. Your task is to learn more about how the Hardy-Weinberg equilibrium and the mechanisms of evolution work, both by themselves and in conjunction with other mechanisms. Also consider ways of presenting raw data to illustrate major points regarding the data. You will want to practice producing publication quality graphs for future assignments. Use your experience and course resources to guide your efforts in construction of Excel graphs. **See the additional handout on preparing graphs and tables.**

Calculation of Allele and Genotype Frequencies & Hardy-Weinberg Equilibrium Theory

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, and the information below will help you to understand the simulations you are performing on Populus.

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

$$(p + q) \times (p + q) = p^2 + 2pq + q^2 = 1, \text{ where}$$

$$p^2 = \text{Frequency of genotype } A/A$$

$$2pq = \text{Frequency of genotype } A/a$$

$$q^2 = \text{Frequency of genotype } a/a$$

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: (total number of A alleles in the population) / (total number of alleles in population for locus) = $[(90 \times 2) + 420] / (1000 \times 2) = 0.30$
 - The frequency of the a allele will equal: $(1 - 0.30)$ or (total number of a alleles in the population) / (total number of alleles in population) = $[(490 \times 2) + 420] / (1000 \times 2) = 0.70$
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 $AA = (\text{frequency of } A \text{ allele})^2 = (0.3)^2 = 0.09$
- Frequency of $Aa = 2 \times (\text{frequency of } A \text{ allele}) \times (\text{frequency of } a \text{ allele}) = 2 \times 0.3 \times 0.7 = 0.42$
- Frequency of $aa = (\text{frequency of } a \text{ allele})^2 = (0.7)^2 = 0.49$

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).

ACKNOWLEDGEMENTS

The synopsis of Hardy-Weinberg Equilibrium Theory was written by Dr. Patricia Peroni and the Populus Instructions were written by Dr. Chris Paradise.