22 GENE FLOW AND POPULATION STRUCTURE

Objectives

- Model two subpopulations that exchange individuals through gene flow.
- Determine equilibrium allele frequencies as a result of gene flow.
- Calculate H (heterozygosity) statistics for the population.
- Calculate F statistics for the population.
- Determine how H, F, and allele frequencies change over time as a result of gene flow.

Suggested Preliminary Exercise: Hardy-Weinberg Equilibrium

INTRODUCTION

Think about a favorite plant or animal species, and consider how it is distributed across the earth. Are the individuals all in one place, or are individuals scattered in their distribution? Most of the earth's species have distributions that are "patchy" in some way. In other words, the greater population is subdivided into smaller units or subpopulations. For example, a species of fish may have a subdivided distribution if individuals inhabit a number of different lakes. Similarly, maple forests may be patchily distributed within a mosaic of farm land, resulting in a number of subpopulations. Even dandelions in a lawn may have distinct patches to which individuals belong. But does this "subdivision" in distribution suggest that the species is made up of several "subpopulations," each with an independent evolutionary trajectory? Or does the species "behave" as a single, panmictic population, where individuals can mix freely in spite of the patchiness? Or perhaps the population is somewhat subdivided, where individuals from one location can mix (breed) with individuals from other locations, but not as freely as a single panmictic population because they are spatially separated from each other.

These questions concerning **gene flow** and **population structure** are important from the perspectives of evolution, ecology, and conservation. A population is "structured" if the individuals that make up the greater, overall population are subdivided spatially, and hence random mating among individuals in the greater population is limited. The degree to which populations are structured depends in large part on the amount of gene flow— the migration of individuals between subpopulations, with subsequent breeding—that takes place between the subdivided populations (or subpopulations). If there is little or no gene flow, then each subpopulation evolves independently of the other. In contrast, if there is substan-

tial gene flow, the structure in the population breaks down because sufficient genetic mixing has occurred. Gene flow is therefore a homogenizing force that causes allele frequencies in subdivided populations to converge (Wilson and Bossert 1971).

Allele Frequencies in Subpopulations

Let's consider gene locus A in two subpopulations. To keep things simple, we'll assume locus A exists in two forms, or **alleles**, A_1 and A_2. Let's assume that subpopulation 1 has an A_1 allele frequency, p_1, of 0.7, while subpopulation 2 has an A_1 allele frequency of $p_2 = 0.2$. Let's now let the two subpopulations exchange individuals through migration, where m is the migration rate of individuals *into* a subpopulation. The individuals that make up the population that did not migrate in are called **residents**, and the resident population is designated as $1 - m$. If $m > 0$, then after a single generation of mixing, p_1 in subpopulation 1 will be changed; subpopulation 1 now consists of some portion of individuals that remained within subpopulation 1, plus some portion of individuals that migrated from subpopulation 2 into subpopulation 1. Mathematically, the new frequency of allele A_1 is designated as p_1', and

$$p_1' = (1 - m)p_1 + mp_2 \qquad \text{Equation 1}$$

Equation 1 says that the new frequency of allele A_1 will have two components: $(1 - m)p_1$, which represents the proportion of subpopulation 1 that does not emigrate times the frequency of A_1 in subpopulation 1 before migration, and mp_2, which represents the proportion of immigrants from subpopulation 2 times the frequency of A_1 in subpopulation 2.

$$\Delta p = p_1' - p_1 \qquad \text{Equation 2}$$

Substituting p_1' from Equation 1 into the Equation 2, we get

$$\Delta p = (1 - m)p_1 + mp_2 - p_1 = p_1 - mp_1 + mp_2 - p_1$$

The p_1s drop out of the equation, and we can factor out $-m$ from the remaining terms to get

$$\Delta p = -m(p_1 - p_2) \qquad \text{Equation 3}$$

Equation 3 says that a change in allele frequency of a recipient population (subpopulation 1) due to migration is a function of the migration rate, as well as of the difference in the allele frequency between the migrants and the recipient population. If the migration rates remain constant over time, eventually the two subpopulations will have exactly the same allele frequencies (Figure 1; Wilson and Bossert 1971).

H and F Statistics

When two populations have reached the same allele frequencies, the larger population will appear to be unstructured. Or is it? Structure depends not only on allele frequencies but also how the A_1 and A_2 alleles are distributed among individuals. Therefore, we must also consider genotype frequencies in the subpopulations.

In many species, especially animals, individuals carry two copies of most genes, one from each parent. Let's assume that subpopulation 1 consists of 5 individuals with genotypes A_1A_1, A_1A_1, A_1A_2, A_2A_2, A_2A_2, and that subpopulation 2 consists of 5 individuals with genotypes A_1A_2, A_1A_2, A_1A_2, A_1A_2, A_1A_2. The subpopulations have identical frequencies of the A_1 allele, $p = 0.5$, but the two subpopulations have quite different levels of **heterozygosity**. Most of the individuals in subpopulation 1 are homozygotes—they carry either two copies of A_1 or two copies of A_2; but the individuals in subpopulation 2 are heterozygotes and each of them carries one copy each of allele A_1 and A_2. So allele frequency alone does not tell us everything about a population's structure. The level of structure depends on levels of heterozygosity in the subpopulations, as well as the level of heterozygosity in the greater population.
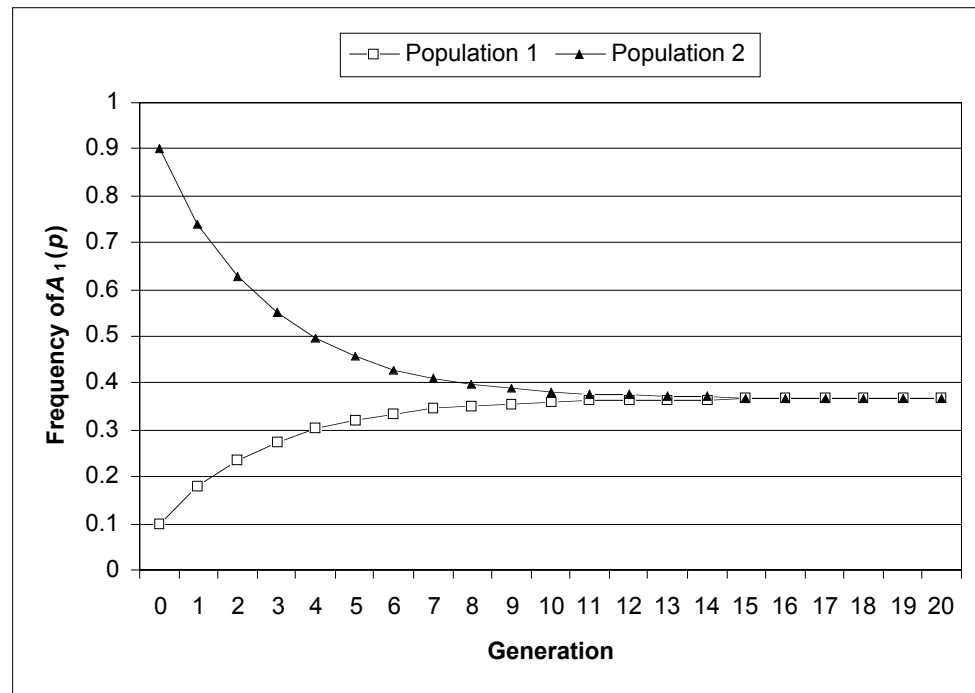
Figure 1   Two subpopulations with different initial frequencies of allele A_1 exchange individuals at two different rates (the migration rate, *m*). As individuals move between the two populations, the frequency of A_1 in subpopulation 1 approaches that in subpopulation 2, and they eventually become equal.

Why is heterozygosity used to estimate structure? And how is the degree of structuring measured through heterozygosity statistics? Two measures are commonly used, *H* and *F* (Hartl 2000).

H is a measure of heterozygosity; it is used to measure structure because individuals within subdivided populations are likely to inbreed due to small population sizes, which typically results in decreased heterozygosity (see Exercise 41/24, "Inbreeding and Outbreeding"). Thus, if there is no gene flow between subpopulations, each subpopulation will (theoretically) have more homozygotes (A_1A_1 or A_2A_2) than predicted by Hardy-Weinberg.

The statistic H_i measures the *observed* level of heterozygosity in a subpopulation For example, 1 of 5 individuals in subpopulation 1 from our previous example were heterozygotes while 5/5 individuals in subpopulation 2 were heterozygotes. This measure is averaged across subpopulations, and can be interpreted as the average heterozygosity of an individual in a subpopulation, or the proportion of the genome that is heterozygous within an individual. For example, *H* for subpopulation 1 equals $1/5 = 0.2$. *H* for subpopulation 2 equals $5/5 = 1.0$. The average of the two *H* scores $= 0.6 = H_i$.

The observed levels of heterozygosity in subpopulations are compared to two other measures of heterozygosity, H_s and H_t. H_s is the *expected* level of heterozygosity in a subpopulation if the subpopulation is randomly mating as predicted by Hardy-Weinberg. This measure is also averaged across subpopulations. Returning to our example, both subpopulations have allele frequencies $p = 0.5$ and $q = 0.5$. If each subpopulation were in Hardy-Weinberg equilibrium, we would expect the genotype frequency of heterozygotes to be $2 \times 0.5 \times 0.5 = 0.5$. This number is averaged for the two subpopulations to give us H_s: $(0.5 + 0.5)/2 – 0.5$. Thus, in our example, $H_i = 0.6$ and $H_s = 0.5$. This means that the observed levels of heterozygotes are, on average, higher than what is expected for a population in Hardy-Weinberg equilibrium.H_t is the expected level of heterozygosity that should be observed in the subpopulations if the greater population (subpopulation 1 and subpopulation 2) were really a single, randomly mating, pan-

mictic population. If our subpopulations were really a single, panmictic population, the expected genotype frequency of heterozygotes would be $2 \times p \times q$, where p and q are the *averages* of the subpopulation allele frequencies (Hartl, 2000). In out example, $p = q = 0.5$ for both subpopulations, so the equation is $2 \times 0.5 \times 0.5 = 0.5$.

The three H statistics are used to calculate F statistics, which are common measures of population subdivision and inbreeding; F is sometimes referred to as the **inbreeding coefficient**. The F statistics use the different H statistics to reveal different things about population subdivision. F_{is} compares observed and expected heterozygosities *within* a subpopulation. It is calculated as

$$F_{is} = \frac{H_s - H_i}{H_s} \qquad \text{Equation 4}$$

and suggests the level of inbreeding at the subpopulation level. Thus, F_{is} is often called the *inbreeding coefficient within subpopulations*. The numerator reveals how much the heterozygosity observed in the subpopulations differs, on average, from what is expected from Hardy-Weinberg. For mathematical reasons, this difference is then "adjusted" by the expected level.

When H_i is approximately the same as H_s, the deviation from Hardy-Weinberg is small, and F_{is} is close to 0, suggesting that observed and expected levels of heterozygosity within subpopulations are close in value. When H_i is much different than H_s, F_{is} deviates from 0. When F_{is} is positive, fewer heterozygotes are observed in subpopulations than predicted by Hardy-Weinberg. When F_{is} is negative, more heterozygotes are observed in the subpopulation than predicted by Hardy-Weinberg. F_{is} is usually large in self-fertilizing (inbred) species.

F_{it} also measures inbreeding, but is concerned with how individuals (H_i) deviate, on average, from the heterozygosity of the larger population (H_t). It is calculated as

$$F_{it} = \frac{H_t - H_i}{H_t} \qquad \text{Equation 5}$$

Thus, it calculates a level of *inbreeding at the total population level*. When H_i is similar to H_t, the observed heterozygosities in subpopulations are close to what is predicted as if the population were really a single large, panmictic population, and F_{it} is 0. When H_i is much different than H_t, F_{it} deviates from 0. When F_{it} is positive, fewer heterozygotes are observed in subpopulations than predicted by Hardy-Weinberg. When F_{it} is negative, more heterozygotes are observed in the subpopulation than predicted by Hardy-Weinberg. These differences can be caused by both inbreeding and by genetic drift, both of which reduce heterozygosity in a subpopulation. Thus, F_{it} measures the amount of inbreeding due to the combined effects of nonrandom mating within subpopulations and to random genetic drift among subpopulations.

F_{st} is a measure of *nonrandom mating among or between subpopulations* relative to the total population, and hence this statistic is often used to indirectly measure the amount of population subdivision. It is calculated as

$$F_{st} = \frac{H_t - H_s}{H_t} \qquad \text{Equation 6}$$

F_{st} is a measure of the genetic differentiation of subpopulations and is always positive. The formula "compares" two expected values from Hardy-Weinberg calculations. The numerator in the formula measures the difference in H_t (the average of the expected heterozygosity in the total population) and H_s (H_s is the average *expected* heterozygosity within the subpopulations). F_{st} is not concerned with individual subpopulations, so it measures the reduction in heterozygosity due to factors other than inbreeding (such as genetic drift). When population subdivision is great, the difference between the values in the numerator increases, F_{st} takes on a high value.

PROCEDURES

The *H* and *F* statistics can be confusing until you sit down and work through the math. The purpose of this exercise is to set up a model of two subpopulations of equal size that interact through migration. You'll enter observed genotype frequencies, then calculate gene frequencies and how these frequencies change over time. You'll also calculate and interpret the *H* and *F* statistics as gene flow occurs between the two populations. As the simulation progresses, you'll be able to see how the *H* and *F* statistics change as the two subpopulations become homogenized, and you'll interpret what the statistics mean.

    As always, save your work frequently to disk.

INSTRUCTIONS	ANNOTATION

A. Set up the spreadsheet.

1. Open a new spreadsheet and set up headings as shown in Figure 2.

	A	B	C	D	E	F	G	H
1	*Gene Flow and Population Structure*							
2								
3				Parameters			Genotype frequencies	
4			*N*	*m*	*r*	A1A1	A1A2	A2A2
5		Subpopulation 1:	100	0	1	0.36	0.48	0.16
6		Subpopulation 2:	100	0	1	0.04	0.32	0.64

Figure 2

2. Enter *N* and *m* subpopulation parameters as shown.

We'll consider a general model of gene flow and population structure that focuses on a single locus, the *A* locus. We'll start with two subpopulations, 1 and 2, that each consist of *N* individuals; we designate *N* as 100 in cells C5 and C6. In this exercise, *N* will be the same for both populations.

The migration rate, *m*, ranges between 0 and 1 and is the proportion of the population that migrates from one subpopulation to the other. The value in cell D5 gives the migration rate *into* subpopulation 1 (from subpopulation 2). The value in cell D6 gives the migration rate *into* subpopulation 2 (from subpopulation 1). To begin the exercise, we'll consider two subpopulations where the migration rate between them is 0. We'll modify *m* later in the exercise.

3. Enter a formula to calculate the value of *r* (the proportion of each subpopulation that are residents as opposed to migrants).

Enter **=1-D5** in cell E5 and **=1-D6** in cell E6.
The total subpopulation consists of migrants that move into the population plus the residents that remain in the population, so the sum of *m* (the migration rate) and *r* (resident population proportion) is equal to 1.

4. Enter the observed genotype frequencies for each subpopulation in cells F5–H6 as shown in Figure 2.

For the purpose of this exercise, we'll assume that you have the ability to determine the genotype of each individual in the subpopulations, and can then calculate the proportion of A_1A_1, A_1A_2, and A_2A_2 genotypes. The current values in cells F5–H6 indicate that both subpopulations are in Hardy-Weinberg equilibrium. (Prove this to yourself before you continue). You will be able to manipulate the observed genotype proportions later in the exercise (i.e., you can model populations that are not in Hardy-Weinberg equilibrium).

5. Sum the genotype frequencies for each subpopulation in cells I5 and I6.

Enter the formula **=SUM(F5:H5)** in cell I5 and **=SUM(F6:H6)** for subpopulation 2. These equations are used to ensure that the genotype frequencies for each subpopulation sum to 1. If the frequencies don't sum to 1, change the observed genotype frequencies so that they sum to 1.

6. Save your work.

B. Set up the general model of gene flow.

1. Set up new headings as shown in Figure 3.

	A	B	C	D	E	F	G
10				**Observed allele frequencies**			
11			**Subpop 1**			**Subpop 2**	
12	**Generation**	**A1**	**A2**	**Delta A2**	**A1**	**A2**	**Delta A2**

Figure 3

2. Set up a linear series from 0 to 50 in cells A13–A63.

We'll calculate the allele frequencies in our two subpopulations over a 50-generation period. Year 0 will represent the initial conditions in terms of allele frequencies.

3. In cell B13 and C13, enter formulae to calculate the initial frequencies of the A_1 and A_2 alleles in subpopulation 1, respectively.

Remember that a population of 100 individuals has 200 "gene copies" or "total alleles" present. (Each individual has 2 copies). We just need to know how many of those are A_1 alleles, and how many are A_2 alleles. Homozygote A_1A_1 individuals carry two of the A_1 alleles, and heterozygotes carry 1 A_1 allele.
Enter the formula **=(2*F5*C5+G5*C5)/(2*C5)** in cell B13.
Enter the formula **=1-B13** in cell C13.

4. In cells E13 and F13, enter formulae to calculate the starting frequencies of the A_1 and A_2 alleles in subpopulation 2.

Enter the formula **=(2*F6*C6+G6*C6)/(2*C6)** in cell E13.
Enter the formula **=1-E13** in cell F13.

5. Enter formulae in cells B14 and C14 to calculate the allele frequencies of subpopulation 1, given the migration and resident parameters.

Remember that the frequencies in the next time step can be computed as

$$p_{1,t+1} = (1-m)p_1 + mp_2$$

We used the formula **=$E$5*C13+$D$5*F13** in cell C14 to calculate the frequency of the A_2 allele, and then calculated A_1 as $1 - q$ in cell B14 **(=1-C14)**.
Make sure you understand the C14 formula. It says that the frequency of the A_2 allele in subpopulation 1 in year 1 depends on two factors: (1) the frequency of the A_2 allele in the resident population (**$E$5*C13**), and (2) frequency of the A_2 allele in the immigrants (**$D$5*F13**).

6. Calculate the change in the frequency of the A_2 allele (ΔA_2) in cell D14.

We used the formula **=C14-C13**. (You can make a delta symbol, Δ, by typing in a capital D, and then changing the font to Symbol.)

7. Calculate the allele frequencies and change in the A_2 allele frequency in subpopulation 2 for year 1.

Enter the following formulae:
- E14 **=1-F14**
- F14 **=$E$6*F13+$D$6*C13**
- G14 **=F14-F13**

8. Select cells B14–G14 and copy their formulae down to row 63.

9. Save your work.

C. Make graphs.

1. Graph the frequency of the A_1 allele over time.

Use the line graph option and label your axes fully. Your graph should look something like Figure 4. (We have graphed only the first 15 generations for clarity.)
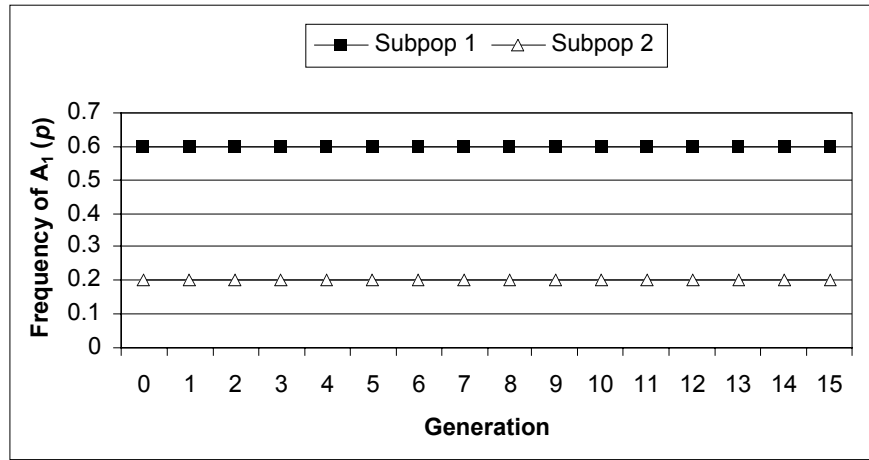

Figure 4

2. Change the migration rate for your two populations (choose any rate between 0 and 1), and construct a new graph of allele frequencies over time.

We generated the graph in Figure 5 by changing the migration rate for subpopulation 1 from 0 to 0.2.
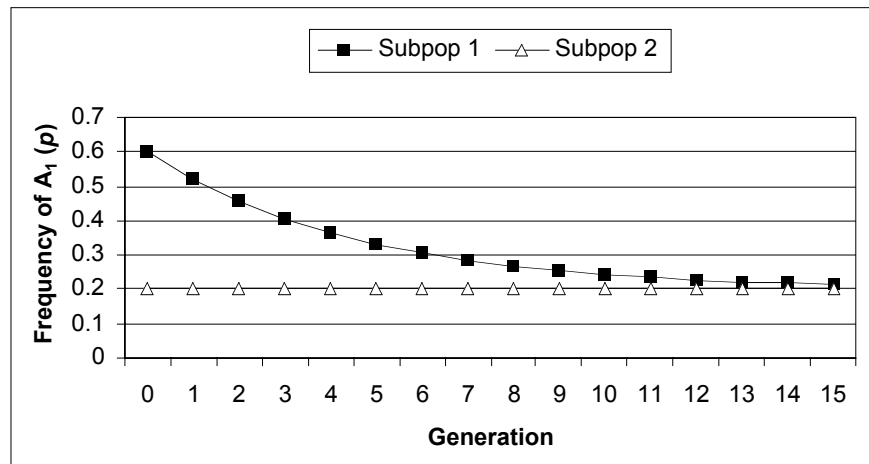

Figure 5

3. Save your work, and answer questions 1–3 at the end of the exercise.

D. Calculate H and F statistics.

1. Set up new headings as shown in Figure 6.

	H	I	J	K	L	M
11	*H* **Statistics**			*F* **Statistics**		
12	H_i	H_s	H_t	F_{is}	F_{it}	F_{st}

Figure 6

2. In cell H13, enter a formula to calculate H_i.

Enter the formula **=AVERAGE($G$5:$G$6)** in cell H13.

H_i is the average observed heterozygosity within a total population. Thus, we take the average of cells G5 and G6, which are the frequencies of heterozygotes in subpoplation 1 and subpopulation 2. Keep in mind that by making cells G5–G6 absolute references, you are forcing the heterozygote proportions to remain constant over time—this will affect the calculation of F statistics later in the exercise.

3. In cell I13, enter a formula to calculate H_s.

Enter the formula **=AVERAGE(2*B13*C13,2*E13*F13)** in cell I13.

H_s is the average *expected* heterozygosity within the subpopulations. Cell B13 and C13 give the frequency of the A_1 (p) and A_2 (q) allele in subpopulation 1. Cells E13 and F13 give the frequency of the A_1 (p) and A_2 (q) allele for subpopulation 2. The Hardy-Weinberg principle tells us that, for each subpopulation, the expected heterozygote frequency is $2 \times p \times q$. The formula in I13 tells Excel to multiply $2 \times p \times q$ for subpopulation 1, then multiply $2 \times p \times q$ for subpopulation 2, and finally to average these two values together.

4. In cell J13, enter a formula to calculate H_t.

Enter the formula **=2*AVERAGE(B13,E13)*AVERAGE(C13,F13)** in cell J13.

H_t is the average of the expected heterozygosity in the total population. H_t is similar to H_s, but it's the average expected heterozygosity for the population at large. Therefore, first we calculate an overall p, then an overall q, and then multiply by 2. The result tells us what heterozygosity should be if the two subpopulations were one panmictic population.

5. In cell K13, enter a formula to calculate F_{is}.

Enter the formula **=(I13-H13)/I13** in cell K13.

Now that we have the H statistics calculated, the F statistics are fairly straightforward. The F statistics compare the different levels of heterozygosities to reveal how the population is structured. All three F statistics (F_{is}, F_{it}, F_{st}) have H_t or H_s as the denominator, which "adjusts" for the expected level of heterozygosity if the population were a single randomly mating, panmictic population (H_t) or randomly mating subdivided populations (H_s).

F_{is} measures of the deviation from Hardy-Weinberg heterozygote proportions within subpopulations (or the deviation of H_i from H_s). Remember that F_{is} also called the inbreeding coefficient because it measures the decrease in heterozygosity *within* a subpopulation (due to inbreeding). The numerator in the equation $F_{is} = (H_s - H_i) / H_s$ thus reveals the difference between the actual, observed heterozygosities in the subpopulations (H_i) and the expected heterozygosities if the subpopulations were in Hardy-Weinberg equilibrium (H_s). When H_i is approximately the same as H_s, the deviation from Hardy-Weinberg is small, and F_{is} is close to 0. When H_i is much different than H_s, F_{is} deviates from 0. When F_{is} is positive, fewer heterozygotes are observed in subpopulations than predicted by Hardy-Weinberg. When F_{is} is negative, more heterozygotes are observed in the subpopulation than predicted by Hardy-Weinberg.

6. In cell L13, enter a formula to calculate F_{it}.

Enter the formula **=(J13-H13)/J13** in cell L13.

F_{it} measures the total inbreeding coefficient. It measures the deviations of observed heterozygosities within subpopulations from Hardy-Weinberg proportions of the total population (or the deviation of H_i from H_t). The equation for calculating F_{it} is $F_{it} = (H_t - H_i)/H_t$. When H_i is similar to H_t, the observed heterozygosities in subpopulations are close to what is predicted as if the population were really one large, panmictic population, and F_{it} is 0. Thus, F_{it} measures the amount of inbreeding due to the combined effects of nonrandom mating within subpopulation and to random genetic drift among subpopulations. When H_i is much different than H_t, F_{it} deviates from 0. When F_{it} is positive, fewer heterozygotes are observed in subpopulations than predicted by Hardy-Weinberg. When F_{it} is negative, more heterozygotes are observed in the subpopulation than predicted by Hardy-Weinberg.

7. In cell M13, enter a formula to calculate F_{st}.

Enter the formula **=(J13-I13)/J13** in cell M13.

F_{st} is a measure of the genetic differentiation of subpopulations and is always positive. The formula "compares" two expected values from Hardy-Weinberg calculations. The numerator in the formula $F_{st} = (H_t - H_s)/H_t$ measures the difference in H_t (the average of the expected heterozygosity in the total population) and H_s (H_s is the average *expected* heterozygosity within the subpopulations). Thus, F_{st} is the amount of "inbreeding" due solely to population subdivision (i.e., due to genetic drift). When inbreeding due to subdivision is great, the difference between the values in the numerator increases, and F_{st} takes on a high value.

8. Select cells H13–M13, and copy their formulae down to row 63.

At this time, you might want to play around with your model parameters and contemplate the meaning of the H and F statistics in Generation 0. Then consider the statistics as gene flow occurs in subsequent generations.

9. Save your work.

E. Create graphs.

1. Set the migration rate to 0, and graph the H statistics and allele frequencies as a function of time. Use the line graph option and label your axes fully.

Interpret your graph. Your graph should resemble Figure 7.

Figure 7

2. Graph the F statistics and allele frequencies as a function of time.
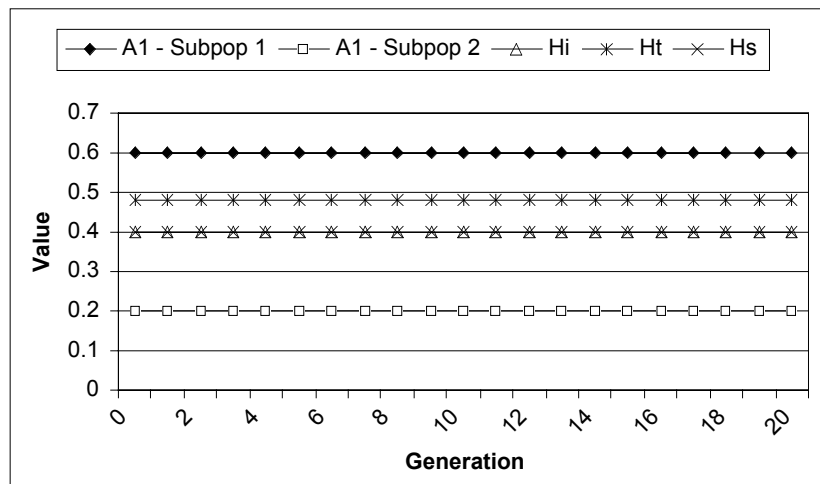
Your graph should resemble Figure 8. Interpret your graph.

3. Save your work.
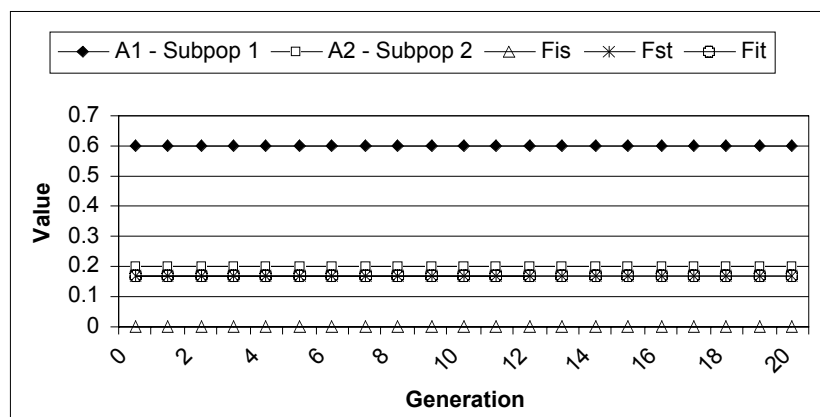
Figure 8

QUESTIONS

1.Enter the following values in your spreadsheet:

	A	B	C	D	E	F	G	H
3			**Parameters**			**Genotype frequencies**		
4			**N**	**m**	**r**	**A1A1**	**A1A2**	**A2A2**
5	**Subpopulation 1:**		100	0	1	0.25	0.5	0.25
6	**Subpopulation 2:**		100	0	1	0.09	0.42	0.49

Change cell D5 by increments of 0.1. What is the equilibrium allele frequencies for subdivided populations with gene flow? How does changing *m* determine the point in time is equilibrium reached?

2. How do allele frequencies change in the two populations in an island model (gene flow is uni-directional) compared to a general model in which gene flow is bi-directional? Set *m* for subpopulation 1 to 0 to indicate that subpopulation 1 is a mainland that sends out emigrants but does not receive immigrants. Set *m* = 0.5 for subpopulation 2 to indicate that subpopulation 2 is an island that receives immigrants from subpopulation 1. Graph your results. Then change *m* for subpopulation 1 from 0 to 1 in increments of 0.1. How do the two models compare? How do your results change if *m* for subpopulation 2 is changed?

3. What determines the amount of time to reach equilibrium frequencies in subdivided populations that have gene flow? Set up population genotypes as shown.

	A	B	C	D	E	F	G	H
3			**Parameters**			**Genotype frequencies**		
4			**N**	**m**	**r**	**A1A1**	**A1A2**	**A2A2**
5	**Subpopulation 1:**		100	0.1	0.9	0.83	0.16	0.01
6	**Subpopulation 2:**		100	0.1	0.9	0.01	0.16	0.83

The allele frequencies for the subpopulations are $p = 0.91$ for subpopulation 1 and $p = 0.09$ for subpopulation 2. Keeping *m* fixed at 0.1 for both subpopulations, change the intial genotype frequencies (the allele frequencies will also be altered). How does change in initial genotype frequency (and allele frequency) affect the amount of time until equilibrium is achieved?

Return your spreadsheet to its initial settings (Figure 2) and continue to Part D in the exercise.

4. Set *m* to 0 in both subpopulations, and enter genotype frequencies in cells F5–H6 so that both subpopulations are in Hardy-Weinberg equilibrium, and have identical allele frequencies. (In the exercise both subpopulations were in Hardy-Weinberg equilbrium and had different allele frequencies within them.) How does this change affect the *H* and *F* statistics? Graph the results and fully interpret the meaning of the *H* and *F* statistics.

5. Set *m* as 0 values for both subpopulations, then enter genotype frequencies in cells F5–H6 so that at least one subpopulation is out of Hardy-Weinberg equilibrium. For example, you might enter values as shown:

	A	B	C	D	E	F	G	H
3			**Parameters**			**Genotype frequencies**		
4			*N*	*m*	*r*	**A1A1**	**A1A2**	**A2A2**
5	**Subpopulation 1:**		100	0	1	0.5	0	0.5
6	**Subpopulation 2:**		100	0	1	0.04	0.32	0.64

How do *H* and *F* statistics reflect structure? How did F_{is} change? Is it positive or negative? Is it large or small? Explain why you obtained the F_{is} value that you did. What does this tell you about the populations? (Remember that the genotype frequencies will remain out of Hardy-Weinberg equilibrium over time because of the formula entered in cell H13.)

6. For this question, you will ignore the genotype frequencies given in rows 5 and 6, and directly enter the initial allele frequencies for subpopulations in cells B13–F13. (We'll assume the genotypes are in Hardy-Weinberg proportions.) Start with $p = 0.6$ for subpopulation 1 and $p = 0.5$ for subpopulation 2. Record the *F* statistics for that generation. Then let $p = 0.8$ in supopulation 1 and $p = 0.2$ in subpopulation 2, and record the *F* statistics. Then let $p = 0.9$ in subpopulation 1 and subpopulation 2, and record the *F* statistics. How did the *F* statistics change as the two subpopulations became more differentiated (allele frequencies diverged)? Which *F* statistic changed the most? Why?

LITERATURE CITED

Hartl, D. 2000. *A Primer of Population Genetics*, Third Edition. Sinauer Associates, Sunderland, MA.

Wilson, E. O., and W. H. Bossert. 1971. *A Primer of Population Biology*. Sinauer Associates, Sunderland, MA.