25 | *GENETIC DRIFT*

Objectives

- Set up a spreadsheet model of genetic drift.
- Determine the likelihood of allele fixation in a population of 10 individuals.
- Evaluate how initial allele frequencies in a population of 10 individuals affect probability of fixation.
- Compare the effects of genetic drift on small versus large populations.

Suggested Preliminary Exercise: Hardy-Weinberg Equilibrium

INTRODUCTION

Random events play a strong role in evolution, especially in small populations. Genetic drift is a random process; it is the chance fluctuations in allele frequencies within a populations as a result of random sampling among gametes (Hartl 2000). To understand what genetic drift is, we start with a very brief refresher in population genetics.

For diploid organisms such as vertebrates, each *individual* carries two alleles in their genetic makeup at each locus (one allele was inherited from the mother and one allele was inherited from the father). Let's assume that there are two types of allele, A_1 and A_2, for a given gene in a population. If the two alleles in an individual are of the same type, the individual is said to be *homozygous* (A_1A_1 or A_2A_2). If the alleles are of different types, the individual is said to be *heterozygous* (A_1A_2). Although *individuals* are either homozygous or heterozygous at a particular gene, *populations* are described by their **genotype frequencies** and **allele frequencies**. The word "frequency" in this case means the proportion of occurrence in a population. To obtain the genotype frequencies of a population, simply count up the number of each kind of genotype and divide by the total number of individuals in the population. For example, if we study a population of 55 individuals, and 8 individuals are A_1A_1, 35 are A_1A_2, and 12 are A_2A_2, the genotype frequencies (f) are

$$f(A_1A_1) = 8/55 \quad = 0.146$$
$$f(A_1A_2) = 35/55 \quad = 0.636$$
$$f(A_2A_2) = 12/55 \quad = 0.218$$
$$\text{Total} \quad = 1.00$$

The sum of the genotype frequencies of a population always equals 1.

Allele frequencies, in contrast, describe the proportion of all alleles in the population that are of a specific type (Hartl 2000). For our population of 55 individuals, there are a total of 110 gene copies present in the population (each of 55 individuals has 2 copies, so $55 \times 2 = 110$). To calculate the allele frequencies of the population, we need to calculate how many of those allele copies are of type A_1 and how many are of type A_2. To calculate how many copies are A_1, count the number of A_1A_1 homozygotes and multiply that number by 2 (each homozygote has two A_1 copies), then add to it the number of A_1A_2 heterozygotes (each heterozygote has one A_1 copy). The number of A_1 alleles in the population is then divided by the total number of gene copies in the population to generate an allele frequency. Thus, the total number of A_1 alleles in the population is $(2 \times 8) + (1 \times 35) = 51$. The frequency of A_1 is calculated as $51/(2 \times 55) = 51/110 = 0.464$. Similarly, the total number of A_2 alleles in the population is $(2 \times 12) + (1 \times 35) = 59$. The frequency of A_2 is calculated as $59/(2 \times 55) = 59/110 = 0.536$. As with genotype frequencies, the total of the allele frequencies of a population always equals 1. By convention, frequencies are designated by letters. If there are only two alleles in the population, these letters are conventionally p and q, where p is the frequency of one kind of allele and q is the frequency of the other. For genes that have only two alleles,

$$p + q = 1 \qquad \text{Equation 1}$$

If there were more than two kinds of alleles for a particular gene, we would calculate allele frequencies for the other kinds of alleles in the same way. For example, if three alleles were present, A_1, A_2, and A_3, the frequencies would be p (the frequency of the A_1 allele), q (the frequency of the A_2 allele) and r (the frequency of the A_3 allele). No matter how many alleles are present in the population, the frequencies should always add to 1. Note that when we describe a population in terms of its allele frequencies, we don't necessarily know the genetic makeup of *individuals* in the population. For instance, all individuals can be homozygous (A_1A_1, A_1A_1, A_2A_2, A_2A_2, A_2A_2) or individuals can be a mix of homozygous and heterozygous genotypes (A_1A_2, A_1A_2, A_1A_1, A_2A_2, A_2A_2); the allele frequencies are the same in both situations.

In summary, for a population of N individuals, suppose the number of A_1A_1, A_1A_2, and A_2A_2 genotypes are n_{A1A1}, n_{A1A2}, and n_{A2A2}, respectively. If p represents the frequency of the A_1 allele, and q represents the frequency of the A_2 allele, the estimates of the allele frequencies in the population are

$$f(A_1) = p = (2n_{A1A1} + n_{A1A2}) / 2n \qquad \text{Equation 2}$$

$$f(A_2) = q = (2n_{A2A2} + n_{A1A2}) / 2n \qquad \text{Equation 3}$$

Genetic Drift and Evolution

Evolution is often described as a change in allele frequencies in a population over time (Hartl 2000). For example, we may notice that the frequency of the A_1 allele in our population changed from a value of 0.4 at time t to a value of 0.5 at time $t + 1$. There are several evolutionary forces that could have produced this change, such as natural selection, mutation, and gene flow. **Genetic drift**, the change in allele frequencies in populations that occurs by chance, without direction, is another kind of evolutionary force that can alter allele frequencies over time. Its impact is often greatest in small populations, and results in a loss of genetic diversity for a given (single) population.

Suppose, for example, that a population of 5 individuals has two alleles, A_1 and A_2, at a given locus, with frequencies p and q, respectively. Suppose further that in a certain generation, $p = q = 0.5$ (in other words, the frequency of allele A_1 is equal to the frequency of allele A_2). We will let this population mix and breed randomly to produce 5 new offspring that make up the next generation. Thus, the birth rates will remain low in this population. We can simulate random breeding by using a random number gen-

erator, where the random numbers 0, 1, 2, 3, and 4 represent the passing down of the A_1 allele to the next generation, and random numbers 5, 6, 7, 8, and 9 represent the passing down of the A_2 allele to the next generation.

Note that the two alleles are each represented by five numbers because the allele frequencies are initially equal. By drawing 10 random numbers to represent the 10 alleles making up the "new" generation, we can assign genotypes to the 5 new offspring and then calculate the new gene frequencies. For example, if the random numbers 0, 1, 5, 3, 9, 8, 3, 4, 8, and 2 are drawn, the 10 alleles in the next generation are A_1, A_1, A_2, A_1, A_2, A_2, A_1, A_1, A_2, and A_1, with genotypes taken in the order A_1A_1, A_2A_1, A_2A_2, A_1A_1, and A_2A_1. If you count how many alleles in this new population are A_1 and how many are A_2 (out of 10 total alleles), you find that this new generation has allele frequencies of p = 0.6 and q = 0.4. The population has evolved due to genetic drift.

We can continue this process for several generations to examine how the allele frequencies will continue to fluctuate over time. We used this method to track the frequency of the A_1 allele in 5 different populations, each consisting of 10 individuals, as shown below (Figure 1). In all populations, the frequency of A_1 was 0.5 to begin with. Inspection of Figure 1 shows that the frequency of the A_1 allele is 1 after 20 generations in two populations (populations 2 and 5). This means, by definition, that the frequency of the A_2 allele is 0. In contrast, the frequency of the A_1 allele is 0 after 20 generations in two other populations (populations 1 and 3).

In the first situation, we say that the A_1 allele has become **fixed** in the population, so that its frequency is 1. In the second situation, the A_2 allele has become fixed and the A_1 allele has been **lost** from the population. In both cases, allelic diversity has been lost from the population because there is now only one allele where previously there were two. Population 4 was also subjected to drift, but both the A_1 and A_2 alleles remained present in the population for at least 20 generations.

The important point is that when populations are very small, and are kept small over time, genetic drift tends to eliminate alleles from within a population, ultimately fixing the population at a frequency of either p = 1 or q = 1. You'll see how this happens as you work through the exercise. We can also think of the effects of drift *across* all five populations. Taking this larger view, genetic drift results in different populations becoming genetically different from each other because by chance, different alleles will become
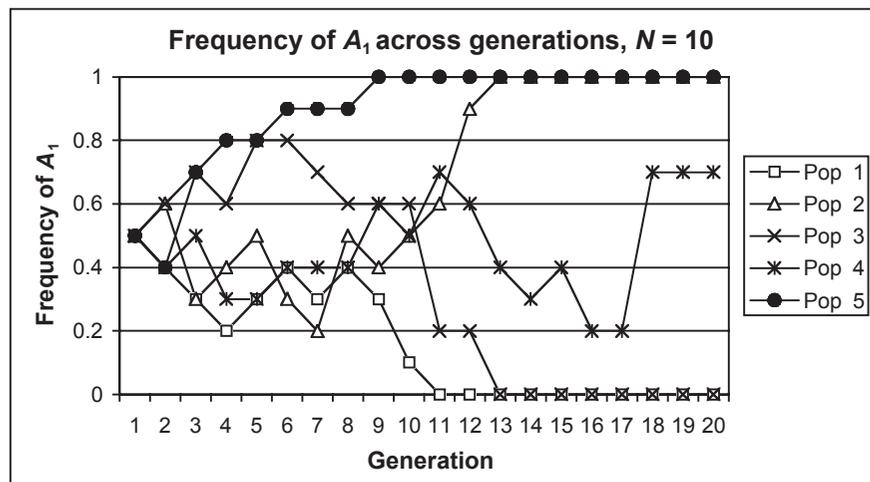

Figure 1 In five populations of size N = 10, the initial frequencies of A_1 and A_2 were each 0.5. After 20 generations, allele A_1 has become fixed in populations 2 and 5 and lost in populations 1 and 3. Only in population 4 do both alleles still exist.
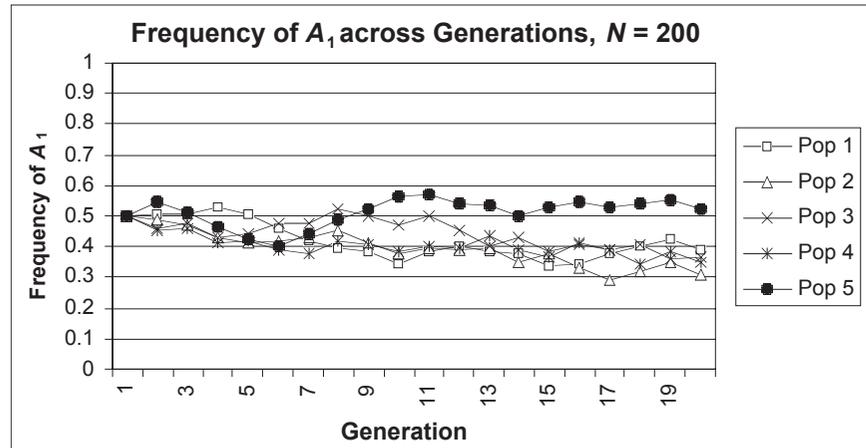
Figure 2 The effects of genetic drift are less dramatic when N is large (population size is large).

fixed in different populations. Some populations will be fixed at one allele, while another population will be fixed at a different allele.

The effects of drift become less important as population size increases. Figure 2 shows five populations, each consisting of 200 individuals and with an initial frequency of 0.5 for the A_1 allele. For the larger population sizes, drift is still apparent, but in no case did the A_1 or A_2 allele become fixed.

PROCEDURES

In this exercise, you'll set up a spreadsheet model to explore the effects of genetic drift. In doing so, you should learn why drift occurs and how it affects genetic diversity.

As always, save your work frequently to disk.

INSTRUCTIONS ANNOTATION

A. Set up the population parameters.

1. Open a new spreadsheet and set up headings as shown in Figure 3.

	A	B	C	D	E	F	G	H	I
1	*Genetic Drift*								
2	Examine the effect of genetic drift on two alleles in a population of 5 individuals.								
3	**Initial frequencies**				**New frequencies for subsequent generations**				
4	A1 => *p* =					*p*	*q*		
5	A2=> *q* =				G(2) =>				
6					G(*N*) =>				
7	**Random mating and genotypes**								
8	**of offspring in next generation**								
9	Allele #	Random #	Allele ID		Allele #	Random #	Allele ID		
10	1								
11	2								
12	3								
13	4								
14	5								
15	6								
16	7								
17	8								
18	9								
19	10				10				
20									
21						**New A1 =**	**New A2 =**		
22									

Figure 3

2. Enter 0.5 in cells B4 and B5.	Set the initial population's allele frequencies in cell B4 (frequency of the A_1 allele) and B5 (frequency of the A_2 allele). These frequencies are the gene frequencies of your initial generation, or generation 1.
3. Set up a linear series from 1 to 10 in cells A10–A19.	Enter 1 in cell A10. Enter **=1+A10** in cell A11. Copy your formula down to cell A19.
4. In cells B10–B19, use the **RAND** function to generate a random number between 0 and 1.	In cell B10 enter the formula **=RAND()**. Copy the formula down to cell B19. Press F9, the calculate key, to generate new random numbers. These random numbers will be used to assign an allele that is inherited by the next generation in the next step.
5. In cells C10–C19, use the **IF** function to simulate which alleles are passed down from the parental generation as a result of random mating.	In cell C10, enter the formula **=IF(B10<B4,"A1","A2")**. Copy the formula down to cell C19. The initial population of 5 individuals mates randomly and produce 5 new offspring that will make up generation 2. Each offspring in the population will inherit 2 alleles at the locus. The first offspring in generation 2 will inherit alleles given in cells A10–A11. The second offspring in generation 2 will inherit alleles given in cells A12–A13, and so on. The formula in C10 uses an **IF** function to determine whether the random number is associated with the A_1 allele or the A_2 allele. The formula tells the spreadsheet to evaluate cell B10; if the random number in cell B10 is less than the frequency of the A_1 allele designated in cell B4, then allele number 1 in the next generation will be an A_1 allele. Otherwise, allele number 1 in the next generation will be an A_2 allele.
6. Note the genotypes of the 5 offspring in generation 2.	The genotypes of our 5 offspring (Figure 4) were A_2A_2, A_1A_1, A_2A_1, A_2A_2, and A_1A_2. Your genotypes will likely be different than ours. Press F9, the calculate key, to generate new random numbers, and hence new offspring genotypes.

	A	B	C	D
3	**Initial frequencies**			
4	A1 => p =	0.5		
5	A2=> q =	0.5		
6				
7	**Random mating and genotypes**			
8	**of offspring in next generation**			
9	Allele #	Random #	Allele ID	
10	1	0.97142	A2	
11	2	0.572015	A2	
12	3	0.300821	A1	
13	4	0.438254	A1	
14	5	0.658612	A2	
15	6	0.447047	A1	
16	7	0.6244	A2	
17	8	0.699563	A2	
18	9	0.180683	A1	
19	10	0.505648	A2	
20				

Figure 4

7. In cell F5, use the **COUNTIF** function to count the number of A_1 alleles in the second generation (labeled G2), and calculate the new frequency of the A_1 allele (p).

Enter the formula **=COUNTIF(C10:C19,"A1")/10** in cell F5.
The **COUNTIF** function counts the number of cells within a range that meet the given criteria. It has the syntax **COUNTIF**(**range**,**criteria**), where range is the range of cells you want to examine, and criteria is the item that will be counted. Since you entered **=COUNTIF(C10:C19,"A1")**, the program will examine cells C10–C19 and count the number of times A_1 appears. This number, when divided by the total alleles in the population, **/10**, gives the new A_1 allele frequency p for generation 2.

8. In cell G5, enter a **COUNTIF** formula to calculate the new frequency of the A_2 allele (q).

In cell G5 enter the formula **=COUNTIF(C10:C19,"A2")/10**. This equation is analogous to the one in Step 7.
After these formulas have been entered, each time you press F9 the spreadsheet will generate a new set of random numbers and will automatically compute the new allele frequencies in cells F5 and G5. We obtained allele frequencies of $p = 0.4$ and $q = 0.6$ for generation 2 (see Figure 4; you probably obtained different results; that's fine).

9. Manually type whatever frequencies you obtained in cells F5 and G5 into cells F6 and G6. Save your work.

Your frequencies will change each time the spreadsheet is calculated. By entering the frequencies in cells F6 and G6 by hand, you are "fixing" the frequencies for future generations.

B. Project allele frequencies to generation 3.

Now we'll repeat the entire process over time by letting generation 2 grow and reproduce 5 new individuals that will make up generation 3. To simulate the third generation, set up a new set of alleles, random numbers, and allele identifications in columns E, F, and G, as you did for generation 2.

1. Set up a linear series from 1 to 10 in cells E10–E19.

Enter the number 1 in cell E10.
Enter the formula **=1+E10** in cell E11. Copy the formula down to cell E19.

2. In cells F10–F19, generate a random number between 0 and 1.

Enter **=RAND()** in cells F10–F19 to assign a random number to each allele in generation 3.

3. In cells G10–G19, use an **IF** formula to determine whether the first allele in generation 3 is A_1 or A_2.

In cell G10 enter the formula **=IF(F10<F6,"A1","A2")**. Copy your formula down to cell G19.
This **IF** formula tells the spreadsheet to examine the random number in cell F10 and assign it a value of A_1 if it is less than the allele frequency designated in F6. If the random number is greater than the allele frequency designated in F6, the program assigns it an A_2 allele.

We determine the results of random mating in generation 2 by assigning an allele (A_1 or A_2) to each new random number in generation 3. Remember that the assignment of random numbers now depends on the allele frequencies in the *second generation* (listed in F6 and G6), and no longer depend on the initial population.

4. In cells F22 and G22, calculate the new allele frequencies inherited by the third generation.

In cell F22 enter the formula **=COUNTIF(G10:G19,"A1")/10** to compute the frequency of the A_1 allele.
In cell G22 enter the formul **=COUNTIF(G10:G19,"A2")/10** to compute the frequency of the A_2 allele.
As before, we use the **COUNTIF** formula to count the total number of A_1 and A_2 alleles..

5. Examine the change in allele frequencies over generations 1–3.

In our version of the exercise, generation 1 had initial allele frequencies of $p = 0.5$ and $q = 0.5$; generation 2 had allele frequencies of $p = 0.4$ and $q = 0.6$; and generation 3 had 0.4 and 0.6 (given in cells B4 and B5). You will almost certainly obtain different results from your own spreadsheet, and that's fine!

6. Obtain allele frequencies for generation 4.

You can quickly obtain the allele frequencies for generation 4 by copying the frequencies of generation 3 in cells F22 and G22 and pasting these values into cells F6 and G6, replacing the frequencies you used for generation 2. (This is why in Figure 3, this cell is labeled generation N, or G(N) for short.) So you can:

- Copy cells F22–G22.
- Select cells F6–G6.
- Open Edit | Paste Special. Select Paste Values and OK.
- Press F9 to automatically calculate new allele frequencies for generation 4 in cells F22 and G22.

What happened? Because of the way you typed in formulas for designating allele types in cells G10-G19, your assignment of alleles to the next generation depends on the parental generation that preceded it. Now the frequencies from generation 4 have been automatically counted in cells F22 and G22.

7. Save your work.

C. Track allele frequencies over time.

1. Set up some new headings as shown in Figure 5, but extend your generations to 20.

	J	K	L	M	N
1	**Change in frequencies over generations**				
2	**Generation**	**A1**	**A2**		
3	1	0.5	0.5		
4	2				
5	3				
6	4				
7	5				
8	6				
9	7				
10	8				
11	9				
12	10				

Figure 5

2. Enter 0.5 in cells K3 and K4.

Ultimately, you will track the fate of the frequencies of the A_1 and A_2 alleles over 20 generations. We will start again with generation 1, which has allele frequencies of $p = 0.5$ and $q = 0.5$.

3. Enter 0.5 in cells F6 and G6.

The values in cells F6 and G6 now represent the allele frequencies for G(1), or generation 1. We can now track how these frequencies change over 20 generations.

4. Write a macro to track allele frequencies for 20 generations.

From the menu, open Tools | Options | Calculations and select Manual Calculation. Then open the Macro function (see Exercise 2) to Record and assign a shortcut key. Perform the following steps:

- Press F9 to generate a new set of random numbers.
- Highlight cells F22 and G22, the new gene frequencies for the second generation.
- Go to Edit | Copy.
- Select cell K2, then go to Edit | Find | Find What. Leave the Find What cell *completely* blank, but make sure the Search by Columns option is selected.

- Select Find Next. The first blank cell in column K should be highlighted. Close the Find box.
- Open Edit | Paste Special, and paste in Values.
- Select cells F6 and G6
- Open Edit | Repeat Paste Special. This action will paste the new frequencies into cells F6 and G6, and will ensure that the spreadsheet uses these new frequencies to assign allele types to the offspring that make up the next, new generation.

Stop recording. Press your shortcut key until you have obtained allele frequencies for 20 generations.

5. Save your work.

D. Create graphs.

1. Graph the frequencies of the A_1 and A_2 allele over time.

Use the Line Graph option and make sure your axes are clearly labeled. Your graph should resemble Figure 6.

Figure 6

2. Save your work. Answer Question 1 at the end of the exercise before proceeding.

E. Run 100 trials.
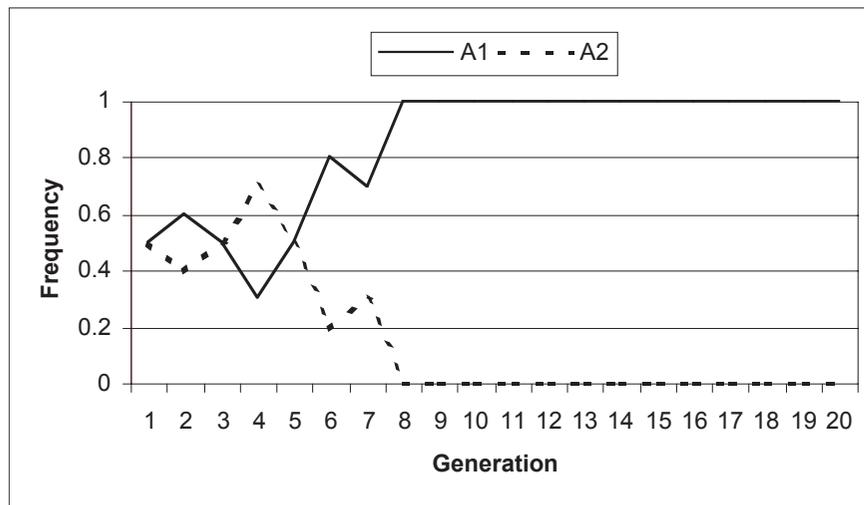
How likely is it that a given allele would become fixed in your population? To know the probability of fixation in your population of 5 with initial gene frequencies of $p = 0.5$ and $q = 0.5$, you will need to repeat your entire simulation a minimum of 100 times (more would be better) and examine the outcomes of a variety of different simulations.

1. Make sure you are in the automatic calculation mode.

Open Tools | Options | Calculation and select Automatic.

2. Set up your spreadsheet as shown in Figure 7, but allow for 20 generations and 100 trials (extend your generations to cell U28 and your trials to cell A128).

	A	B	C	D	E	F	G
27	*Trial*	*Generation*					
28		*1*	*2*	*3*	*4*	*5*	*6*
29	*1*						
30	*2*						
31	*3*						
32	*4*						
33	*5*						

Figure 7

3. Copy results from Trial 1 (cells K3–K22, the frequency of A_1 over 20 generations) into cells B29–U29.

You've run 1 "trial" so far, with the results listed in cells K3–K22. We need to put those values into cells B29–U29.

Highlight cells K3–K22.
Open Edit | Copy. Select cell B29.
Open Edit | Paste Special | Paste Values, and select Paste Transpose. The transpose option will paste in the allele frequencies in row 29, automatically filling in the frequency of the A_1 allele across generations.

4. Switch to manual calculation.

Open Tools | Options | Calculation, and select Manual.

5. Develop a new macro to run 100 trials.

Try writing this macro on your own. If you get stuck, here are the steps we recorded to perform the task:
- Enter 0.5 in cells F6–G6. This will re-set the initial allele frequencies to 0.5 in generation 1.
- Press F9 to generate a new set of random numbers. The spreadsheet automatically calculates the new frequencies listed in cells F22 and G22.
- Use your mouse to highlight cells K4–L22.
- Press the delete key. The results of generation 2–20 from your first trial will be wiped out.
- Press the shortcut key (usually <Control> + some key) that runs your *first* macro (Step 4 in Section C), until you have generated allele frequencies for 20 generations.
- Select cells K3 to K22, and open Edit | Copy.
- Select cell B28, and go to Edit | Find. At the Find What prompt, leave the cell completely blank.
- Select the Search by Columns option. Select Find Next and then Close. This action will move your cursor to the next open cell in Column B.
- Open Edit | Paste Special. Select the Paste Values and Transpose options.

Stop recording. Press your shortcut key until you have run *a minimum* of 100 trials.

F. Calculate probability of fixation.

Now that you have run a number of trials, you can determine how likely it is that an allele would become fixed in the population after 20 generations. First, we'll count the number of times the A_1 allele went "extinct" (the frequency of the A_1 allele = 0, and the A_2 allele was fixed at 1). Then we'll count the number of times the A_1 allele was fixed at 1 (the A_2 allele went extinct).

1. Return to automatic calculation.

Open Tools | Options | Calculation and select Automatic.

2. Set up column labels as shown in Figure 8.

	V	W
27	**A1 extinct?**	**A1 fixed?**
28	**1 = yes**	**1 = yes**

Figure 8

3. In cells V29–V128, use the IF function to calculate how many times the A_1 allele went extinct.

Enter the formula **=IF(U29=0,1,0)** in cell V29. Copy this formula down to cell V128. The **IF** statement in V29 tells the spreadsheet to examine the contents of cell U29 (the allele frequency of the twentieth generation in trial 1). If cell U29 = 0, then assign it a value of 1; otherwise assign it a value of 0. Thus, if the A_1 allele went extinct in the course of 20 generations for a particular trial, the value in column V is scored as 1.

4. In cells W29–W128, use the **IF** function to calculate how many times the A_2 allele went extinct.

Enter the formula **=IF(U29=1,1,0)** in cell W29. Copy this formula down to cell W128. The spreadsheet will return a "1" if the A_1 allele became fixed at 1 (and thus the A_2 allele went extinct).

5. Sum the number of times the A_1 allele went extinct in cell V129. Sum the number of times the A_1 allele was fixed in cell W129.

Enter the formula **=SUM(V29:V128)** in cell V129.
Enter the formula **=SUM(W29:W128)** in cell W129.
In this step you simply add the number of times the A_1 allele went extinct ($p = 0$) and the number of times the A_1 allele became fixed at $p = 1$ for your trials.

6. In cell V130, enter a formula to calculate the probability of fixation as the probability that either the A_1 or A_2 allele will be fixed in the population. Label this value in U130.

We entered the formula **=(V129+W129)/100** in cell V130.
Now you can estimate the probability of fixation of an allele for a population of size 5 with initial gene frequencies of $p = 0.5$ and $q = 0.5$. These probabilities are simply the total number of times the A_1 allele went extinct *or* became fixed at 1, divided by the total number of trials you ran.

7. Save your work.

QUESTIONS

1. Trace the fate of the frequency of the A_1 allele over time. Did it vary dramatically? What was its frequency in the 20th generation? Was the frequency of the A_1 allele ever 1 or 0 at any time during your simulation? If so, did it bounce back to a new frequency, or did it remain fixed at a given level over time? Why?

2. How do the initial frequencies in the population affect the probability of extinction or of fixation? Change your initial allele frequencies to $p = 0.8$, $q = 0.2$. Set cell K3 to 0.8, and cell L3 to 0.2. Open Tools | Macro | Macros, then edit your Trials macro. You should see the Visual Basic for Applications Code that Excel "wrote" as you recorded your macro. Modify the values from 0.5 to 0.8 and 0.2. Close out of the edit box and return to your spreadsheet. Clear the results of your 100 trials, then run your 100 trials again. Graph and explain your results.

3. (*Advanced) What are the effects of genetic drift in a much larger population (say $N = 50$ or $N = 100$), where the initial allele frequencies are $p = 0.5$ and $q = 0.5$? Expand your model to compare the results of the effects of drift on small versus large populations. Copy the entire spreadsheet to a new page, and make your modifications on the new sheet.

4. What are some possible consequences of drift in populations, particularly if drift leads to fixation of alleles? Should this be of concern to wildlife managers? Could you use your model to estimate the minimum population size required to minimize the effects of drift?

LITERATURE CITED

Hartl, D. 2000. *A Primer of Population Genetics*, 3rd Edition. Sinauer Associates, Sunderland, MA.