23 *MEASURES OF GENETIC DIVERSITY*

Objectives

- Estimate allele frequencies from a sample of individuals using the maximum likelihood formulation.
- Determine polymorphism for a population, *P*.
- Determine heterozygosity for a population, *H*.
- Evaluate how sample size affects estimates of allele frequency, polymorphism, and heterozygosity.

Suggested Preliminary Exercise: Hardy-Weinberg Equilibrium

INTRODUCTION

The amount of genetic variation on earth is astounding. Think of the genetic programming that creates first a larva, then a caterpillar, then a cocoon, and finally an adult butterfly. Or think of the programming that created a single Sequoia tree, and then the different kinds of programming that created an entire forest of Sequoias. Or marvel at the programming required to create you inside your mother's womb. Who would have thought that a mere four molecules—adenine, thymine, cytosine, and guanine, the bases of the genetic code—could be arranged in such a multitude of ways to produce the astonishing variation found among the organisms, living and extinct, that have called the earth home.

The total genetic variation existing on earth today can be "partitioned" or "organized" into four different levels: variation among species; variation among populations of a species; variation among individuals within a population; and variation within a single individual (Hunter 1996). The genetic differences among species such as Sequoias, butterflies, and humans clearly accounts for a large chunk of the total genetic diversity. But populations and individuals of the same species differ in their genetic makeup too. For example, a population of garter snakes living near Lake Ontario may have a very different genetic make-up than a population of the same species of snake living in the Ozark Mountains. Even within a single population, individuals can be quite variable, although they can also be genetically very similar to one another. And within an individual—you, for instance—some portion of the total genome is heterozygous (two different alleles of a gene are present at a locus), and some portion of the genome is homozygous (the two alleles at a locus are both the same). The diversity within any individual can be great or small, depending on how many gene loci are heterozygous. It is important to realize that diversity is measured as a continuum from little or no diversity to very high levels of diversity.

How is genetic diversity measured in populations? Typically, a sample of individuals is obtained from the population and the genotype of each individual is determined using one of several methods (e.g., protein electrophoresis or DNA sequencing). From there, allele frequencies can be estimated, and two other measures of genetic diversity—polymorphism and heterozygosity—can be measured (Hartl 2000).

Let's illustrate these measures with an example. Suppose you sample five individuals of mice from a nearby farm field. For two loci, you obtain the genotypes shown in the table.

Individual	Locus A Genotype	Locus B Genotype
1	A1A1	B1B1
2	A1A2	B1B2
3	A1A2	B1B3
4	A1A1	B1B1
5	A2A2	B1B1

Based on your sample, there are two "alleles" present at the A locus (A_1 and A_2) and three alleles present at the B locus (B_1, B_2, B_3). For the A locus, the frequency of the A_1 allele is 0.6 because 6 of the 10 total alleles (5 individuals, each with two alleles) at this locus are A_1. Likewise, the frequency of the A_2 allele is 0.4. For the B locus, the frequency of the B_1 allele is 0.8, the frequency of the B_2 allele is 0.1, and the frequency of the B_3 allele is 0.1. Note that the sum of the frequencies for any locus must equal 1. By sampling five individuals from the population and deriving allele frequency estimates, you are hoping that the five individuals sampled reflect the greater population of mice that live in the field but were not sampled. But does the greater population of field mice really have these frequencies? If we sampled five additional mice, our allele frequency estimates might change. And they might continue to change until every single mouse in the field population is sampled; at that point we could calculate (as opposed to estimate) the true allele frequency of the mouse population.

Estimating Polymorphism and Heterozygosity

Sampling, by nature, involves some error. But we can estimate what the *most likely* allele frequencies are in the greater population, given the size of our sample. The procedure to estimate the frequencies is called **maximum likelihood formulation**. And we can make a statement about how accurate our estimates are by calculating the variance of the estimates themselves.

If we assume that the genetic system of the A and B alleles is one of co-dominance, the **maximum likelihood estimate** of p (the frequency of the A_1 allele) is

$$\hat{p} = \frac{0.5 \times N_{A_1A_2} + N_{A_1A_1}}{N} \qquad \text{Equation 1}$$

and the variance in $\hat{p}$ is

$$V(\hat{p}) = \frac{p(1-p)}{2N} \qquad \text{Equation 2}$$

Equation 1 should look familiar to you. Using these formulae, the maximum likelihood estimate of the A_1 allele is 0.6, and the variance is .024. The frequency of the A_2 allele, q, can be similarly calculated.

Once we have estimated the allele frequencies in the population, we can estimate another useful measure of genetic diversity, **polymorphism**, P. The word "polymorphism" literally means "many forms." It follows that P measures whether a locus contains many different forms of a gene (i.e., alleles), or whether a locus contains few forms or even just one allele. In our example above, the A locus has two alleles (A_1 and A_2), while the B locus has three (B_1, B_2, and B_3). Both loci are polymorphic. Since 2 loci out of 2 loci sampled (A locus and B locus) each have different kinds alleles, $P = 2/2 = 1$. On the other hand, if all five individuals were B_1B_1 genotypes at locus B, the B locus would be monomorphic (literally, "one form"), and so 1 of 2 loci examined would be polymorphic, and P would equal 0.5. Thus, P can be defined as

$$P = \text{Number polymorphic loci/Total number loci evaluated} \qquad \text{Equation 3}$$

In a large population, almost all loci will have more than one allele (Hartl 2000), so if we consider a polymorphism to be any locus that has more than one allele, the value of P will never be very far from 1. To make P more meaningful, a locus is usually considered to be polymorphic only if the *frequency of the most common allele* is less than some arbitrary threshold, usually 0.95 (Ayala 1982). Sample size is therefore a key issue in estimating P. Suppose, for example, that we are examining the C locus in a population and the first four individuals all have the genotype C_1C_1, but the fifth has the genotype C_1C_2. Of the ten alleles we have sampled so far, all but one are C_1, so our estimate of the frequency of C_1 is 9/10, or 0.9. On the basis of this very small sample we would conclude that the C locus is polymorphic. If we continue to sample and find that the next 45 individuals are all C_1C_1, however, we need to reconsider—now we've sampled 100 alleles (from 50 individuals in all), and 99 of them are C_1, so our new estimate of the frequency of C_1 is 0.99. It's beginning to look as if the C_2 allele is less common than our initial sample of five individuals suggested, and the C locus may actually not be polymorphic (if we use a frequency of 0.95 as the cutoff in our definition). A larger sample size yet would give us greater confidence in our results.

Another useful measure of genetic diversity is **heterozygosity**, H, which measures the percentage of genes at which the *average individual* is heterozygous. In our example, individual 1 is homozygous at both the A and B locus, so its heterozygosity is 0 out of 2 loci = 0. Individual 2 is heterozygous at both the A and B locus, so its heterozygosity is 2 out of 2 loci examined = 1. The average individual heterozygosity for these two individuals is then the average of individual 1 and individual 2, so $H = 0.5$. In mathematical terms, **average heterozygosity** is calculated as

$$\hat{H} = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{j=1}^{m} H_{ij} \qquad \text{Equation 4}$$

and the variance in $\hat{H}$ is

$$V(\hat{H}) = \frac{\hat{H}(1-\hat{H})}{Nm} \qquad \text{Equation 5}$$

where N is the sample size and m is the number of loci examined. You'll see clearly how these formulae function as you work through the exercise.

PROCEDURES

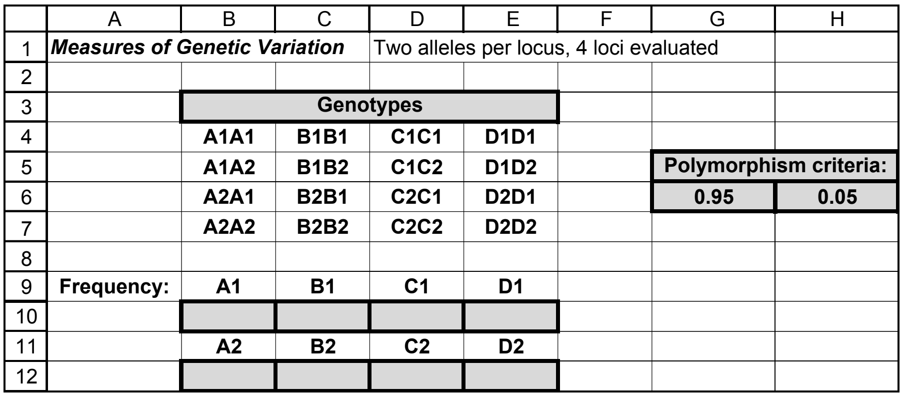
In this exercise, you'll learn how to estimate allele frequencies using the maximum likelihood formulation, and you will learn how to calculate P, H and $\hat{H}$. We'll examine only four loci (A, B, C, and D) and we will assume that each locus has only two alleles present in the population. We'll also assume that you are sampling individuals one at a time from a very large population and can identify the genotypes of each individual at the different loci. You'll examine how your estimates of allele frequencies, P, and $\hat{H}$ change as new individuals are sampled and sample size increases. As always, save your work frequently to disk.

INSTRUCTIONS

ANNOTATION

A. Set up the hypothetical population.

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



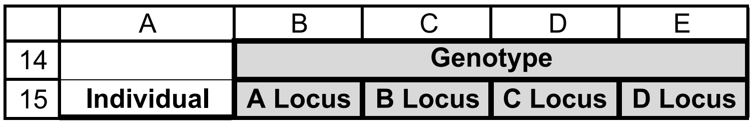
	A	B	C	D	E	F	G	H
1	***Measures of Genetic Variation***			Two alleles per locus, 4 loci evaluated				
2								
3		**Genotypes**						
4		**A1A1**	**B1B1**	**C1C1**	**D1D1**			
5		**A1A2**	**B1B2**	**C1C2**	**D1D2**		**Polymorphism criteria:**	
6		**A2A1**	**B2B1**	**C2C1**	**D2D1**		**0.95**	**0.05**
7		**A2A2**	**B2B2**	**C2C2**	**D2D2**			
8								
9	**Frequency:**	**A1**	**B1**	**C1**	**D1**			
10								
11		**A2**	**B2**	**C2**	**D2**			
12								

Figure 1

2. In rows 10 and 12, assign true allele frequences to a very large hypothetical population. We will try to estimate these frequencies by sampling individuals from the population.

Enter 0.5 in cells B10–E10.
Enter **=1-B10** in cell B12, and copy this formula across to cell E12.
The values in cells B10–E10 and B12–E12 represent the true allele frequencies of a very large (infinite) population from which we'll sample individuals and estimate allele frequencies, *P*, and *H*. To begin, we'll let the true frequencies of each allele for each locus be 0.5. Remember, the sum of the allele frequencies for a given locus must equal 1. The values in cells B10–E10 can be modified directly as you go through the exercise (cells B12–E12 will automatically be updated).

3. Set up spreadsheet headings as shown in Figure 2.



	A	B	C	D	E
14		**Genotype**			
15	**Individual**	**A Locus**	**B Locus**	**C Locus**	**D Locus**

Figure 2

4. Set up a linear series from 1 to 100 in cells A16–A115.

Enter 1 in cell A16.
In cell A17, enter the formula **=1+A16**.
Copy the formula down to cell A115.
We will sample 100 individuals from this large population and determine the genotypes of each individual. We will then assume that individuals are sampled in order (from 1 to 100), and will then estimate the allele frequencies, polymorphism, and heterozygosity as new individuals are included in the total sample.

5. Assign genotypes at the *A* locus to each individual in the population, based on the allele frequencies designated in Step 2.

In cell B16, enter the formula **=IF(RAND()<$B$10,$B$9,$B$11)&IF(RAND()<$B$10,$B$9,$B$11)**.
This formula will assign genotypes based on the allele frequencies that we designated in cells B10 and B12. The **IF** formula in cell B16 is used to determine the genotype of individual 1. The first part of the formula in cell B16 tells the spreadsheet to choose a random number between 0 and 1 (the **RAND()** portion of the formula), and if that random number is less than the value designated in cell B10, then return the value in cell B9 (A_1); otherwise, return the value in cell B11 (A_2). All individuals have two alleles for

a given locus, so you need to repeat the formula again, and then join the two alleles obtained from the two **IF** formulas by using the **&** symbol.

Once you've obtained genotypes for individual 1, copy this formula down to cell B115 to obtain genotypes for all 100 individuals in the population. Note that when you press F9, the calculate key, the spreadsheet generates a new random number, and hence a new genotype.

6. Enter formulae in cells C16–E16 to generate genotypes for individual 1 at the *B*, *C*, and *D* loci.

Enter the formulae:

- C16 **=IF(RAND()<$C$10,$C$9,$C$11)&IF(RAND()<$C$10,$C$9,$C$11)**
- D16 **=IF(RAND()<$D$10,$D$9,$D$11)&IF(RAND()<$D$10,$D$9,$D$11)**
- E16 **=IF(RAND()<$E$10,$E$9,$E$11)&IF(RAND()<$E$10,$E$9,$E$11)**

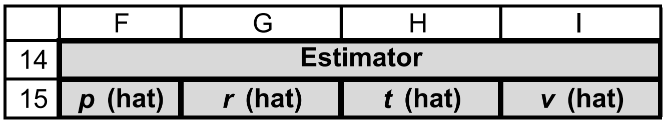
7. Copy cells B16–E16 down to row 115.

When you copy the formula down, note that the genotypes are assigned based on the random numbers and the allele frequencies in row 10, and the allele designations in rows 9 and 11. These formulae require absolute cell references (with row and columns preceded by **$** signs) so that when the formulae are copied down to individual 100, the spreadsheet will go back to the appropriate, fixed, cells in assigning genotypes to individuals.

8. Save your work.

B. Calculate likelihood estimators.

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



	F	G	H	I
14	**Estimator**			
15	***p* (hat)**	***r* (hat)**	***t* (hat)**	***v* (hat)**

Figure 3

2. In cell F16, enter a formula to estimate the frequency of the A_1 allele of our population (this will be a maximum likelihood formula based on Equation 1).

We'll let $\hat{p}$ estimate the frequency of the A_1 allele, $\hat{r}$ be the estimate of the B_1 allele frequency, $\hat{t}$ be the estimate of the C_1 allele frequency, and $\hat{v}$ be the estimate of the D_1 allele frequency. Enter the formula **=(COUNTIF($B$16:B16,$B$4)+COUNTIF($B$16:B16,$B$5) *0.5+COUNTIF($B$16:B16,$B$6)*0.5)/$A16** in cell F16. This represents Equation 1, the formula for estimating the frequency of an allele in a population:

$$\hat{p} = \frac{0.5 \times N_{A_1A_2} + N_{A_1A_1}}{N}$$

The first step is to tally the number of A_1A_1 homozygotes and the number of A_1A_2 heterozygotes. The tally of heterozygotes is then multiplied by 0.5. The sum is divided by the number of individuals sampled, *N*. The formula in cell F16 does this with the **COUNTIF** function. The formula in cell F16 counts the number of A_1A_1 homozygotes (cell $B$4) in the range of cells $B$16–B16, then counts the number of A_1A_2 heterozygotes in the same range and multiplies this number by 0.5, then counts the number of A_2A_1 heterozygotes and multiplies this number by 0.5. (Remember that a heterozygote can be either A_1A_2 or A_1A_2 in your spreadsheet.) The sum of these numbers is bracketed by parentheses so that the total is divided by *N*, the sample size. In this case, the sample size is 1, given in cell A16. Note the use of absolute and relative references. This will allow you to copy your formula down to cell F115 while updating *N* and the range of cells to be counted.

3. Enter formulae in cells G16–I16 to compute the estimated allele frequencies of the B_1, C_1, and D_1 alleles.

In cell G16, enter the formula **=(COUNTIF($C$16:C16,$C$4)+COUNTIF($C$16:C16,$C$5)* 0.5+COUNTIF($C$16:C16,$C$6)*0.5)/$A16.**
In cell H16, enter the formula **=(COUNTIF($D$16:D16,$D$4)+COUNTIF($D$16:D16,$D$5)* 0.5+ COUNTIF($D$16:D16,$D$6)*0.5)/$A16.**

In cell I16, enter the formula**=(COUNTIF($E$16:E16,$E$4)+COUNTIF($E$16:E16,$E$5)*0.5+ COUNTIF($E$16:E16,$E$6)*0.5)/$*A*16.**

4. Select cells F16–I16 and copy their formulae down to row 115.

5. Graph the estimated allele frequencies as a function of sample size. Set the *y*-axis scale to range between 0 and 1.

Use the line graph option and label your axes fully. Your graph should resemble Figure 4.

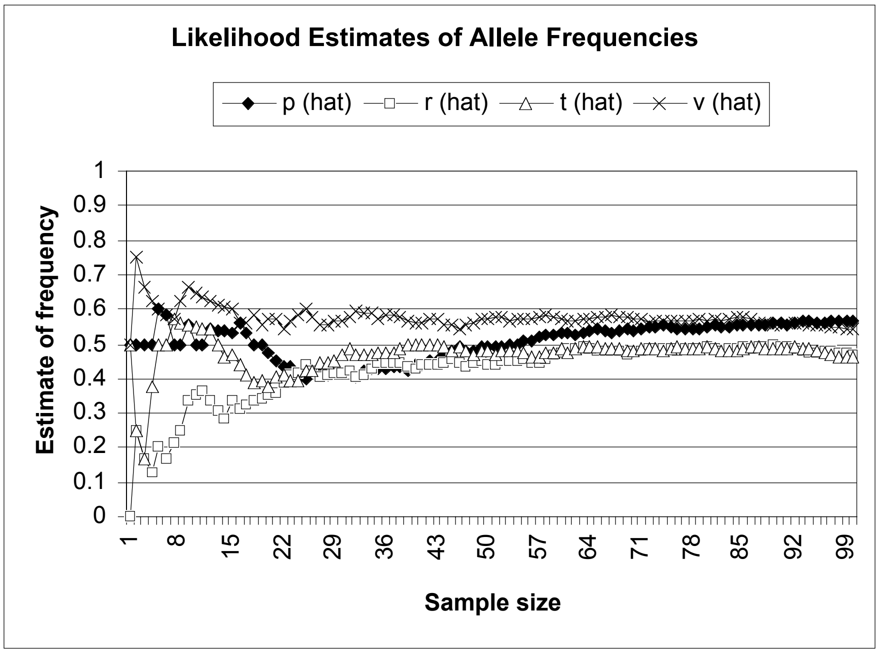


Figure 4

6. Press F9 to generate new random numbers, and hence new genotypes.

How closely do your samples reflect the allele frequencies given in rows 10 and 12? Examine your graph carefully and write a one or two sentence summary of the major results.

7. Save your work.

C. Estimate polymorphism, P.

1. Enter new spreadsheet headings as shown in Figure 5.

	J	K	L	M	N
14	**Polymorphism**				
15	**A Locus?**	**B Locus?**	**C Locus?**	**D Locus?**	**P**

Figure 5

2. In cell G6, enter the criterion parameter for polymorphism.

Now we are ready to estimate polymorphism. To begin, our criterion will be 0.95, so enter 0.95 in cell G6.

3. Enter **=1-G6** in cell H6.

Remember, a gene locus is polymorphic if the frequency of the *most common allele* is less than the criterion. Another way of saying this is that a locus is considered monomorphic if *any* of the alleles at that locus has a frequency >0.95. Thus, if either the A_1 or A_2 allele has a frequency of greater than 0.95, the locus is monomorphic. Concentrating on just the A_1 allele, the *A* locus is polymorphic if the A_1 allele has a frequency of <0.95 (which means that the A_1 allele frequency is <.95) or >0.05 (which means that the A_2 allele frequency is <.95). Otherwise, it is monomorphic.

4. Determine whether the locus is polymorphic (1) or monomorphic (0).

In cell J16, enter the formula **=IF(OR(F16>G6,(F16<H6)),0,1**).
We have already calculated the estimated allele frequencies for our population. We'll examine these estimates to determine whether or not the locus is polymorphic. The formula in cell J16 evaluates individual 1. Based on this single sample, if the value in cell F16 is either greater than the criterion in cell G6 or less than the criterion in cell H6, we will consider the locus to be monomorphic (0). Otherwise, it is considered to be polymorphic (1). The **OR** part of this formula—**OR(F16>G6,(F16<H6)—**allows us to evaluate both conditions; if either one is true the spreadsheet will return the number 0. If both criteria are false, the spreadsheet will return the number 1.

5. Enter formulae in cells K16–M16 to determine the polymorphism at the *B*, *C*, and *D* loci.

Select cell J16, and copy its formula across to cell M16, or enter the following:
In cell K16, enter the formula **=IF(OR(G16>G6,(G16<H6)),0,1).**
In cell L16, enter the formula **=IF(OR(H16>G6,(H16<H6)),0,1).**
In cell M16, enter the formula **=IF(OR(I16>G6,(I16<H6)),0,1).**

6. In cell N16, compute the average *P* for individual 1.

Enter the formula **=AVERAGE(J16:M16).**

7. Select cells J16–N16, and copy their formulae down to row 115.

Keep in mind that although the average polymorphism appears to be calculated for each individual, column A really gives the sample size from the population. The allele frequency estimates are based on all of the samples up to and including the individual sampled, so the *P* estimates are really estimates that change as individuals are added to the sample. Also keep in mind that since only four loci have been evaluated, *P* can take on only five values: 0, 0.25, 0.5, 0.75, and 1, where 0/4, 1/4, 2/4, 3/4, or 4/4 loci are polymorphic.

8. Graph *P* as a function of sample size. Set the *y*-axis scale to range between 0 and 1.

Use the line graph option and label your axes fully. Your graph should resemble Figure 5.

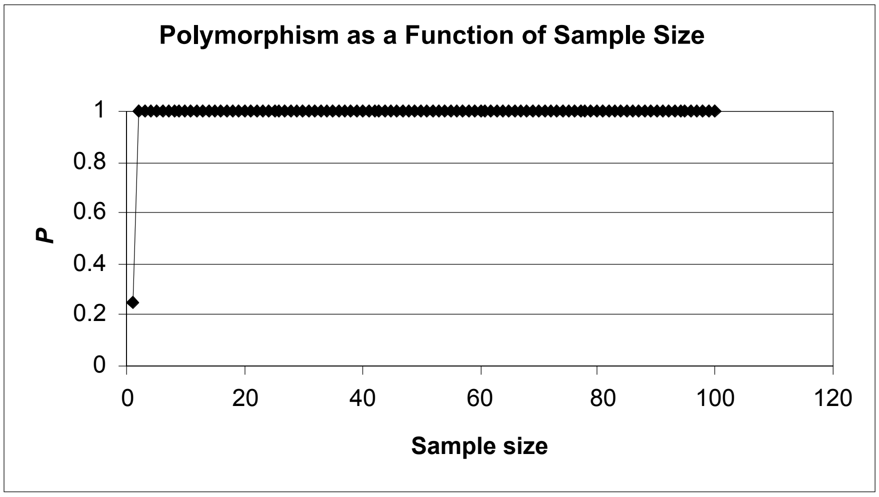


Figure 5

9. Press F9 several times and examine how *P* changes as the sampled individuals change in genotypes.

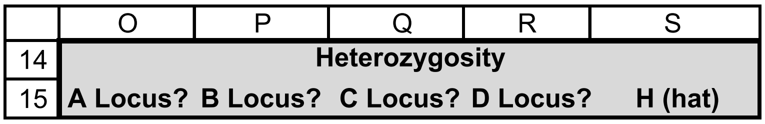
Since all loci have allele frequencies around 0.5 (for large enough sample sizes), *P* should equal 1, indicating that all four loci are polymorphic.

10. Save your work.

D. Estimate heterozygosity, H.

1. Enter new spreadsheet headings as shown in Figure 6.

Remember that heterozygosity has two components: within individuals (*H*) and among (or across) individuals ($\hat{H}$). Columns O through R tackle the within individual component. Column S uses that information to calculate the among-individuals component.



	O	P	Q	R	S
14	Heterozygosity				
15	A Locus?	B Locus?	C Locus?	D Locus?	H (hat)

Figure 6

2. Determine the heterozygosity of locus *A* for each individual.

In cell O16, enter the formula **=IF(OR(B16=$B$5,B16=$B$6),1,0).**
Within an individual, heterozygosity is the proportion of loci that are heterozygous. The O16 formula examines the *A* locus for individual 1 and returns a 1 if the individual is heterozygous at that locus, and a 0 if it is homozygous at that loci. An **OR** formula is used because either A_1A_2 or A_2A_1 heterozygotes should be counted. Copy this formula down to row 115 to determine the heterozygosity of the *A* locus for each individual in the sample.

3. Enter formulae in cells P16–R16 to compute heterozygosity for each individual in the sample at the *B*, *C*, and *D* loci.

In cell P16, enter the formula **=IF(OR(C16=$C$5,C16=$C$6),1,0)**
In cell Q16, enter the formula **=IF(OR(D16=$D$5,D16=$D$6),1,0)**
In cell R16, enter the formula **=IF(OR(E16=$E$5,E16=$E$6),1,0)**

4. Determine $\hat{H}$, the average heterozygosity across all individuals.

Now we are ready to calculate $\hat{H}$, which is calculated with Equation 4:

$$\hat{H} = \frac{1}{Nm}\sum_{i=1}^{N}\sum_{j=1}^{m} H_{ij}$$

5. Select cells O16–S16, and copy their formulae down to row 115.

In cell S16, enter the formula **=1/(4*A16)*SUM($O$16:R16)**. The formula **=AVERAGE($0$16:R16)** gives the same result.
In row 16, we are considering $\hat{H}$ when the sample size consists of a single individual. Our sample size, *N*, is 1 in this row, designated by cell A16. The number of loci evaluated, *m*, is 4. So the first part of the formula is easy to take care of. For the second part of the equation (the summation signs, $\sum$), we simply need to sum the 0's and 1's for individual 1, then multiply this sum by 1/*Nm*, or 1/4. As you copy this formula down to row 115, $\hat{H}$ will be automatically updated as a running estimate as sample size changes.

6. Graph $\hat{H}$ as a function of sample size.

Use the line graph option and label your axes fully. Your graph should resemble Figure 7.

7. Press F9 and evaluate how changes in sampling affect your estimates.

8. Save your work.

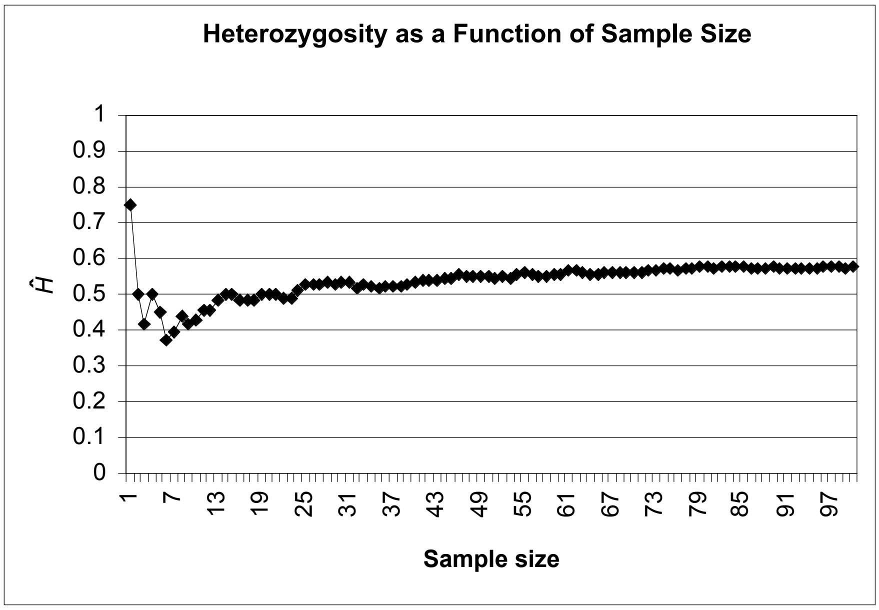
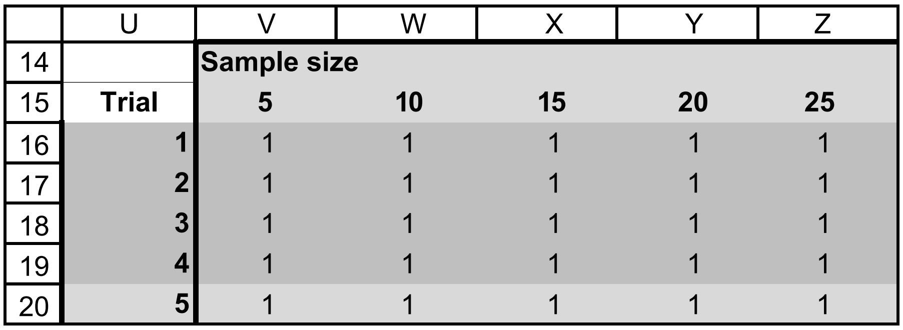


Figure 7

E. Generate 100 estimates of P as a function of sample size.

By now you should have noticed that when you press F9, the calculate key, all of your results, including your graphs, change. This is because the genotypes of individuals change when a new random number is generated. Although you can get a "feel" for how estimates change as sample size increases by pressing F9 a number of times and examining the graphs, quantitative approaches are usually used. How can you therefore assess how sample size affects your estimates of *P*, $\hat{H}$, and $\hat{p}$, when your results keep changing? In order to determine how sample size affects these estimates, we need to press F9 many times (say 100), and compute the average estimate. This is called a **Monte Carlo simulation**. We will do this in the next step for *P*; you may wish to evaluate other metrics as well.

1. Set up new spreadsheet headings as shown, in Figure 8, but extend the trials down to 100 (cell U115), and extend the sample size out to 100 (in increments of 5, cell AO15).



	U	V	W	X	Y	Z
14		**Sample size**				
15	**Trial**	**5**	**10**	**15**	**20**	**25**
16	**1**	1	1	1	1	1
17	**2**	1	1	1	1	1
18	**3**	1	1	1	1	1
19	**4**	1	1	1	1	1
20	**5**	1	1	1	1	1

Figure 8

2. Write a macro to record estimates of *P* for different sample sizes, tracking your results for 100 trials.

See Exercise 2, "Spreadsheet Functions and Macros," for information on how to record a macro. When you are in the Record Macro mode, assign a name (e.g., Trials) and a shortcut key (e.g., <Control>+t) to your macro. Then record the following steps:

- Press F9, the Calculate key, to generate new genotypes for the population.
- Highlight cell N20 (the *P* estimate for a sample size of 5).
- Press down the <Control> key, and *P* estimates for sample sizes 10 (N25), 15 (N30), up to N(115).

Courier CERTIFIED created with Synapse

- Open Edit | Copy.
- Select cell V15.
- Open Edit | Find. A dialog box will appear. Leave the Find What box blank, search by columns and values. Select Find Next, and then Close.
- Open Edit | Paste Special, and select the Paste Values and Transpose options. Click OK. Your results should be pasted into row 16.
- Open Tools | Macro | Stop Recording.

Now when you press your shortcut key 100 times, your estimates of *P* under different sample sizes will automatically be recorded.

3. Compute the average *P* in row 116.

Enter the formula **=AVERAGE(V16:V115)** in cell V116. Copy this formula over to cell AO116.

4. Graph your results, the average *P* as a function of sample size.

Use the line graph option and under the Series tab, select cells V15–AO15 as Category (x) axis labels. Your graph should look like Figure 9. Perhaps this figure is a bit boring, but it suggests that when the frequencies at all four loci are 0.5 for each allele (set in cells B10–E10), the estimate of *P* is insensitive to sample size. You will see that this is not the case when there are rare alleles at a locus.

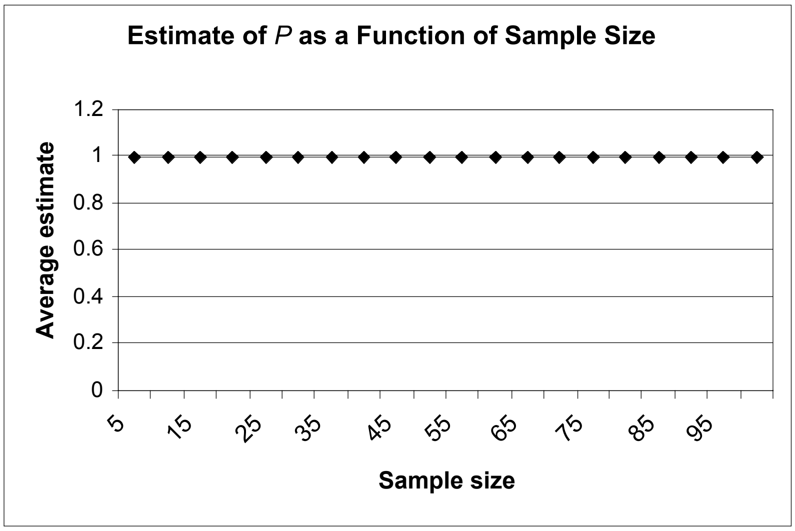


Figure 9

5. Examine the visual basic for application code to learn how to modify your macro for other metrics.

You can edit your macro to examine other metrics ($\hat{p}$, *H*) by making some slight modifications. (You can also just record a brand new macro if the idea of editing the code of a current macro does not appeal to you).

Open Tools | Macro, then select the macro Trials and Edit. You should now see the Visual Basic for Applications code that the spreadsheet "wrote" when you went through your keystrokes. Read through the code. It should make some sense to you, since it is simply a record of which cells you selected, copied, and pasted. We added two sentences to our code: `For counter = 1 to 100` was added after the fourth line (a keyboard shortcut) and the word `Next` was typed into the second to the last line of the code (before the last line, `ENDSUB`) so that when the macro is run, all 100 trials are completed. In this macro, estimates of polymorphism (*P*) are given in column N. If you manually replace the letter N with the letter F in all of the appropriate places, your macro can be used to evaluate how $\hat{p}$ or other estimates change as a function of sample size.

QUESTIONS

1. Examine your estimates of *P* as a function of sample size (last step). How do the allele frequencies affect your result? Set cell B5 to 0.05 (cell B6 should be updated to 0.95). Erase your macro results (cells V16–AO115), and then run your macro again. Your graphs should automatically be updated. Interpret your results.

2. Change polymorphism criteria from 0.95 to some other value, such as 0.9. How does the criteria affect the polymorphism estimate?

3. Which measure is a better indicator of genetic diversity for your population, *P* or *H*? Why is it useful to have multiple measures of diversity?

4. Add a fifth and sixth allele to your spreadsheet model. How does increasing the number of alleles affect polymorphism and heterozygosity estimates? If you were given additional funds to evaluate additional loci, would these dollars be well spent? Use graphs to illustrate your answer.

*5. (Advanced) Our model is based on a co-dominant allele system, but several other kinds of genetic systems are possible. Modify your model to estimate allele frequencies in a system where one allele is dominant over the other. Compare your results in terms of maximum likelihood estimators, polymorphism, and heterozygosity.

LITERATURE CITED

Ayala, F. 1982. *Population and Evolutionary Genetics*. Benjamin Cummings, Menlo Park, CA.

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

Hunter, M. L. Jr. 1996. *Fundamentals of Conservation Biology*. Blackwell Science, Inc., Cambridge, MA.

Courier CERTIFIED created with Synapse