

32

NATURAL SELECTION AND FITNESS

Objectives

- Mathematically define absolute fitness, relative fitness, and the selection coefficient.
- Predict the course of evolution by natural selection from any given initial allele condition, using the formula

$$p_{t+1} = \frac{W_{11}p_t^2 + W_{12}p_tq_t}{(W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2)}$$

- Predict the change in population size over time, using the formula

$$N_{t+1} = (W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2) \times N_t$$

- Develop a spreadsheet model of a population of 100 individuals that undergo natural selection and track genotypes through time.

Suggested Preliminary Exercises: Geometric and Exponential Population Models; Hardy-Weinberg Equilibrium

INTRODUCTION

Evolutionary biologists are interested in how genotypes and allele frequencies will change over time. **Natural selection** takes place in a population when different genotypes have different probabilities of survival or different abilities to reproduce (Roughgarden 1998). That is, genotypes themselves have growth rates, where “fit” genotypes increase in the population relative to “less fit” or “unfit” genotypes. Stated more succinctly, dN/dt varies among genotypes when natural selection is acting on a population (Wilson and Bossert 1971). Because natural selection affects the growth rates of genotypes, it can profoundly affect how allele frequencies change from one generation to the next. One of the assumptions of the Hardy-Weinberg principle is that natural selection does not act on the population. In this exercise, you’ll explore how violating this assumption affects the evolution of a population.

Let’s start with a quick review of the Hardy-Weinberg principle. You might recall that if there are only two alleles at a given locus, A_1 and A_2 , the frequencies of the alleles are symbolized by p and q , where p is the frequency of the first allele (A_1)

and q is the frequency of the second allele (A_2). Recall further that, for genes with only two alleles,

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

Assume that the A locus has allele frequencies of $p = 0.6$ and $q = 0.4$. Given these frequencies, the Hardy-Weinberg principle allows us to predict the genotype frequencies of a population, assuming that the population is large, that mating occurs at random, and that there is no gene flow, natural selection, or mutation acting on the population. The predicted genotypes of a population in Hardy-Weinberg equilibrium are $p^2:2pq:q^2$, where p^2 is the frequency of the A_1A_1 genotype, $2pq$ is the frequency of the heterozygous genotype (A_1A_2 and A_2A_1), and q^2 is the frequency of the A_2A_2 genotype. The sum of the genotype frequencies will be 1. In this example, a population in Hardy-Weinberg equilibrium will have roughly the following genotype frequencies:

- frequency (A_1A_1) = $p^2 = p \times p = 0.6 \times 0.6 = 0.36$, or 36% of the population will be A_1A_1 .
- frequency (A_1A_2) = $2 \times p \times q = 2 \times 0.6 \times 0.4 = 0.48$, or 48% of the population will be A_1A_2 .
- frequency (A_2A_2) = $q^2 = 0.4 \times 0.4 = 0.16$, or 16% of the population will be A_2A_2 .

Note that the genotype frequencies add to 1:

$$p^2 + 2pq + q^2 = 1 \quad \text{Equation 2}$$

The numbers of *individuals* of each genotype that are expected in the population can be calculated by multiplying the genotype frequencies by the population size, N .

$$\text{Number of } A_1A_1 \text{ individuals} = p^2 \times N$$

$$\text{Number of } A_1A_2 \text{ individuals} = 2pq \times N \quad \text{Equation 3}$$

$$\text{Number of } A_2A_2 \text{ individuals} = q^2 \times N$$

If our population consists of 400 individuals, for example, $0.36 \times 400 = 144$ individuals are expected to be A_1A_1 , $0.48 \times 400 = 192$ individuals are expected to be A_1A_2 , and $0.16 \times 400 = 64$ individuals are expected to be A_2A_2 .

Natural Selection

When natural selection is at work on a population, the genotype frequencies may not match the frequencies predicted by Hardy-Weinberg. If some genotypes are more likely to survive than others, the genotype frequencies in the population will be altered. In turn, the allele frequencies of the population may also change.

Consider a population of 100 individuals that consists of 25 A_1A_1 individuals, 50 A_1A_2 individuals, and 25 A_2A_2 individuals. Given the numbers of individuals of each genotype, the allele frequencies can be calculated and are $p = 0.5$ and $q = 0.5$. With these frequencies, $p^2 \times N = 0.5 \times 0.5 \times 100 = 25$ individuals are expected to be A_1A_1 , $2pq \times N = 2 \times 0.5 \times 0.5 \times 100 = 50$ individuals are expected to be A_1A_2 , and $q^2 \times N = 0.5 \times 0.5 \times 100 = 25$ individuals are expected to be A_2A_2 . Because the observed genotype frequencies equal the expected genotype frequencies, the population is in Hardy-Weinberg equilibrium.

Now let's consider what happens to the population when natural selection acts on it. In this exercise we will assume that our population has discrete, nonoverlapping generations, in which individuals start out as zygotes, reach sexual maturity, reproduce, and then immediately die. The probability of surviving to sexual maturity (adulthood) is given by the letter l . Given that individuals survive to reproductive age, the number of gametes that an adult contributes to the next generation's gene pool is given by $2m$. (The reason m is multiplied by 2 will become clear later on.) The life cycle of such an organism is depicted in Figure 1.

Let's assume that the A_2A_2 genotype has a low probability—say, 0.2—of surviving to reproductive age. If in fact only 20% of the A_2A_2 genotypes survive and *all* of the A_1A_1

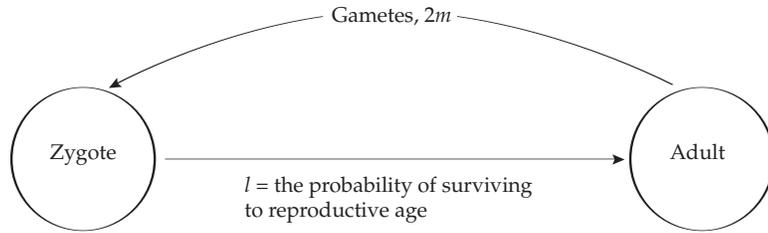


Figure 1

and A_1A_2 genotypes survive, the genotype frequencies of the adult population will be 25 A_1A_1 , 50 A_1A_2 , and 5 A_2A_2 (because 20 of the A_2A_2 individuals died). A graph of the genotype numbers before and after selection is shown in Figure 2.

Not only has natural selection altered the genotype frequencies, but the allele frequencies have consequently been altered as well. After selection, $p = 0.625$ and $q = 0.375$. Because p and q have changed, it might be tempting to conclude that the population has evolved. However, evolution is a change in allele frequencies *across* generations; so far we have examined the effects of selection *within* a generation. In order to determine the effects of natural selection on evolution, we must calculate p and q in the *next* generation, which depends on both the survival and the reproduction of the different genotypes.

To determine what p and q will be in the next generation, we will utilize the notation outlined by Roughgarden (1998) to follow the progress of a set of individuals from the zygote stage until death, keeping track of how many individuals of each genotype survive to sexual maturity (adulthood) and how many of the total gametes produced by each genotype make it into the next generation's gene pool (Table 1). The starting number of individuals (zygotes) of various genotypes in the population is shown in row 1 of Table 1. This is the Hardy-Weinberg genotype frequency multiplied by the total number of individuals in the population (Equation 3). The probability that a zygote of a given genotype will survive to sexual maturity (adulthood) is denoted by the letter l . The subscript after the letter l indicates the survival probability for a specific genotype; thus, l_{12} is the probability that an A_1A_2 genotype will survive to adulthood. The number of adults of a particular genotype can then be computed as the probability of surviving to adulthood multiplied by the number of zygotes of that genotype. This value appears in row 2 of Table 1.

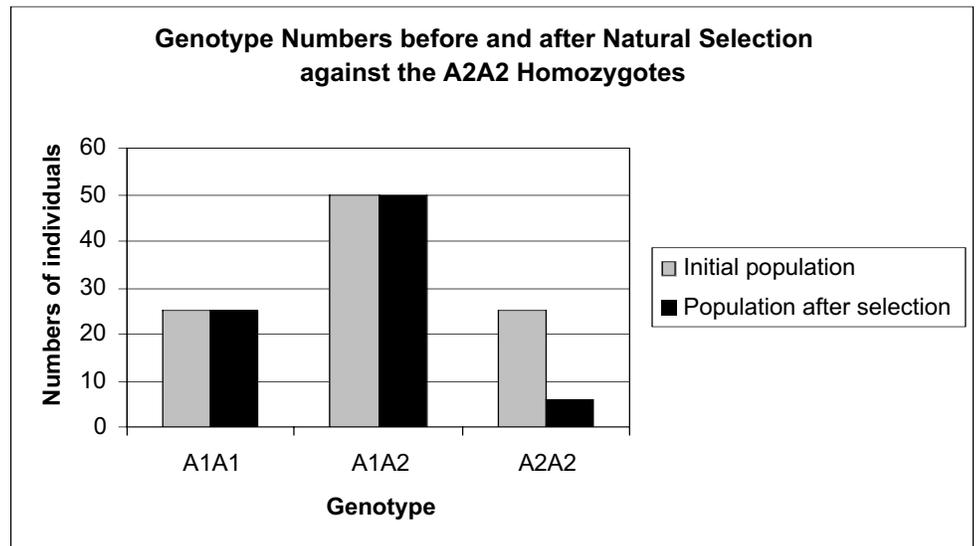


Figure 2 Genotype numbers before and after natural selection against the A_2A_2 homozygote.

TABLE 1.			
	A_1A_2	A_1A_2	A_2A_2
1	p^2N	$2pqN$	q^2N
2	$l_{11}p^2N$	$l_{12}2pqN$	$l_{22}q^2N$
3	$2m_{11}l_{11}p^2N$	$2m_{12}l_{12}2qpN$	$2m_{22}l_{22}q^2N$

The number of gametes that are produced per individual of a specified genotype that actually become incorporated into the next generation’s gene pool is $2m$: m represents one-half the gametes produced per individual. The total number of gametes from a single genotype in next year’s gene pool is $2m$ multiplied by the probability of survival and by the number of individuals of that genotype in the population. This value appears in row 3 of Table 1.

To be clear, let’s walk through an example. If $p = 0.5$ and there are 100 individuals in the population, then there would be $p^2N = 25$ A_1A_1 , $2pqN = 50$ A_1A_2 , and $q^2N = 25$ A_2A_2 zygotes in the population. If $l_{11} = 1$, $l_{12} = 1$, and $l_{22} = 0.2$, all of the A_1A_1 and A_1A_2 zygotes would reach adulthood, but only $0.2 \times 25 = 5$ A_2A_2 zygotes would reach adulthood. If $2m = 3$ for all genotypes, then the number of gametes contributed to the next generation is $3 \times 25 = 75$ gametes for A_1A_1 individuals, $3 \times 50 = 150$ gametes for A_1A_2 individuals, and $3 \times 5 = 15$ gametes for A_2A_2 adults. Thus, in total the next generation consists of $75 + 150 + 15$ gametes (240 total), which translates to 120 zygotes in the next generation. Thus, given information in Table 1, you can compute directly how each genotype will impact the gene pool in the next generation.

Absolute and Relative Fitness

We can also be more general in our computations. The frequency of the A_1 allele, p , at time $t + 1$ is

$$p_{t+1} = \frac{A_1 \text{ alleles}_{t+1}}{\text{Total alleles}_{t+1}} \tag{Equation 4}$$

The denominator is the total number of alleles or “gene copies” at the A locus for the offspring population. Obviously, these copies are from the parent’s gametes, so you can compute the denominator of Equation 3 as the sum of the bottom row in Table 1:

$$\text{Total alleles}_{t+1} = 2m_{11}l_{11}p_t^2N + 2m_{12}l_{12}2p_tq_tN + 2m_{22}l_{22}q_t^2N \tag{Equation 5}$$

We use the subscript t with p_t and q_t to indicate that these represent the frequencies at time t . To compute the numerator of Equation 4, we need to count up the gametes contributed by the A_1A_1 individuals (all of the gametes from this genotype will be A_1), plus one-half the gametes contributed by A_1A_2 individuals (only half of the gametes from this genotype will be A_1 ; the other half of the gametes will be A_2). Thus, the numerator can be rewritten as

$$2m_{11}l_{11}p_t^2N_t + (1/2)(2m_{12}l_{12}2p_tq_tN_t) \tag{Equation 6}$$

Thus, we can now rewrite Equation 4 as Equation 6 divided by Equation 5:

$$p_{t+1} = \frac{2m_{11}l_{11}p_t^2N_t + (1/2)2m_{12}l_{12}2p_tq_tN_t}{2m_{11}l_{11}p_t^2N_t + 2m_{12}l_{12}2p_tq_tN_t + 2m_{22}l_{22}q_t^2N_t} \tag{Equation 7}$$

You will notice that you can factor out both a 2 and an N from both the numerator and the denominator, which cancel out and give

$$p_{t+1} = \frac{m_{11}l_{11}p_t^2 + m_{12}l_{12}p_tq_t}{m_{11}l_{11}p_t^2 + m_{12}l_{12}2p_tq_t + m_{22}l_{22}q_t^2} \quad \text{Equation 8}$$

Thus, although $2m$ is the number of gametes contributed by an individual to the next generation's gene pool, our computations, once simplified, express each individual's contribution to the next generation's gene pool as m . To simplify things even further, we can combine both the survival probabilities and gamete contributions of a genotype into a single value, capital W :

- $W_{11} = m_{11}l_{11}$
- $W_{12} = m_{12}l_{12}$
- $W_{22} = m_{22}l_{22}$

and by substitution

$$p_{t+1} = \frac{W_{11}p_t^2 + W_{12}p_tq_t}{W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2} \quad \text{Equation 9}$$

Equation 9 is a fundamental formula in evolutionary biology. It was derived in the 1920s by R. A. Fisher, J. B. S. Haldane, and S. Wright. The value W is the **absolute fitness** of a genotype. Knowing W provides information on a genotype's survival probability and its reproductive contribution to the next generation's gene pool (Roughgarden 1998). Accordingly, "fitness" has both survival and reproductive components. Absolute fitness is sometimes designated as λ because it is the finite rate of increase for a particular genotype. Thus, when $W > 1$, the genotype is increasing over time; when $W < 1$, the genotype decreases over time; and when $W = 1$, the genotype remains stable over time. In a broad sense, absolute fitness can be formally defined as *the average per capita lifetime contribution of individuals of that genotype to the population after one or more generations* (Futuyma 1998).

By convention, W is "scaled" such that the genotype with the largest W has the value 1; this scaled value is its **relative fitness**. Relative fitness is designated by a lowercase w , and is computed by

$$w_{ij} = W_{ij} / W_{\max} \quad \text{Equation 10}$$

For instance, assume that the following W 's depict the absolute fitnesses of genotypes in the population:

- $W_{11} = 2$
- $W_{12} = 1$
- $W_{22} = 0.4$

The A_1A_1 genotype has the largest absolute fitness, and so we establish the **relative fitness** of this genotype as the standard genotype (the denominator of Equation 10) with which other genotype fitnesses will be compared:

- $w_{11} = W_{11} / W_{11} = 1$
- $w_{12} = W_{12} / W_{11} = 1 / 2 = 0.5$
- $w_{22} = W_{22} / W_{11} = 0.4 / 2 = 0.2$

The relative fitness values can be interpreted as the growth rate of a genotype relative to the fastest growing genotype. Thus, the A_1A_2 genotype grows at one-half the rate as the A_1A_1 genotype, and the A_2A_2 genotype is growing at 1/5 the rate of the A_1A_1 genotype.

The expression $1 - w$ is called the **selection coefficient** and indicates the degree to which natural selection selects "against" a genotype. Evolutionary modelers often use the relative fitness calculation and selection coefficients rather than the absolute fitnesses, because then the exact numbers of individuals of each genotype in the population do not need to be known. However, in this exercise you will track the fates of 100 individuals over time and will therefore be able to compute absolute fitnesses without difficulty.

The use of absolute fitness over relative fitness has another advantage: Because the number of gametes that each genotype contributes to the next generation is known, the population size of the next generation can also be determined. Refer again to Equation 4:

$$p_{t+1} = \frac{A_1 \text{ alleles}_{t+1}}{\text{Total alleles}_{t+1}}$$

The denominator gives the total number of alleles that will be incorporated into the next generation’s gene pool. Since we are talking about a diploid organism, the total number of individuals in the next generation is simply the total number of alleles in $t + 1$, multiplied by 0.5.

$$N_{t+1} = 0.5 \times \text{total alleles at } t + 1 \tag{Equation 11}$$

Remember that the total number of alleles at $t + 1$ is the sum of the bottom row in the table, given in Equation 5:

$$\text{Total alleles}_{t+1} = 2m_{11}l_{11} p_t^2 N + 2m_{12}l_{12} 2p_t q_t N + 2m_{22}l_{22} q_t^2 N$$

Multiply Equation 5 by 0.5, then replace the m_{ij} ’s and l_{ij} ’s with W_{ij} ’s, and we are left with the formula

$$N_{t+1} = (W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2) \times N_t \tag{Equation 12}$$

Hopefully, Equation 12 has a form that is familiar to you.

In Exercise 7, “Geometric and Exponential Population Models,” we developed a model with the form

$$N_{t+1} = \lambda \times N_t \tag{Equation 13}$$

Thus, the term $W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2$ in Equation 12 is the same thing as λ in Equation 13, the finite rate of increase for the population. This should not be too surprising, since fitness is the growth rate of the various genotypes over time. It is computed by summing the W ’s for each genotype, weighting each W by the frequency of each genotype (given by Hardy-Weinberg) in the population.

PROCEDURES

In this exercise, you’ll set up a spreadsheet model of a population of 100 individuals and subject the population to various selective forces. Your population will consist of individuals that reproduce sexually during a discrete time period and then die (think of an annual plant whose seeds are viable only until the following year). The ultimate goal of the model is to predict the allele frequencies p and q at time $t + 1$ given their initial state at time t , and to predict the new population size as well. As always, save your work frequently to disk.

INSTRUCTIONS

A. Set up the model parameters.

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

ANNOTATION

We’ll consider a population of 100 zygotes of varying genotypes and track their fates to adulthood.

	A	B	C	D	E	F
1	<i>Natural Selection and Fitness</i>					
2			Tally			
3	Genotypes	# of individuals (zygotes)	0			
4	A1A1	25				
5	A1A2	50				
6	A2A2	25		<== this number MUST total 100		

Figure 3

2. Enter numbers in cells B4–B6 as shown.
3. Enter 0 in cell C3.
4. Enter =SUM(\$B\$4:B4) in cell C4. Copy this formula down to cell C6.
5. Set up new headings as shown in Figure 4.

6. Calculate the initial genotype frequencies in cells C9–E9.

7. Enter values in cells C10–E10 as shown in Figure 4.

8. Enter values in cells C11–E11 as shown in Figure 4.

9. Compute the absolute fitness, W , in cells C12–E12.

10. Save your work.

B. Simulate the survival and reproduction of the 100 individuals in the population.

1. Set up column headings as shown in Figure 5.

To begin, we will have the population consist of 25 A_1A_1 homozygotes, 50 A_1A_2 heterozygotes, and 25 A_2A_2 homozygotes.

Cells C3–C6 will keep track of the total number of individuals by “tallying” the numbers in cells B4–B6. This tally will be used to assign genotypes to individuals in a few steps.

The spreadsheet should return the number 25 in cell C4. Your result in cell C6 should be 100, indicating that the population consists of 100 individuals. Later in the model, you will be free to change the genotype composition of the 100 individuals in cells B4–B6, but you’ll want to make sure that cell C6 totals 100.

	A	B	C	D	E
8			A1A1	A1A2	A2A2
9	Initial genotype frequencies =		0.25	0.5	0.25
10	Probability of genotype survival = l =		1	1	0.2
11	Half the # of gametes in next gen. = m =		2	2	2
12	Absolute fitness = $W = l * m$ =		2	2	0.4

Figure 4

In cell C9, enter the formula =B4/C6.

In cell D9, enter the formula =B5/C6.

In cell E9, enter the formula =B6/C6.

The frequency of the various genotypes is simply the number of individuals of a given genotype divided by the total number of individuals in the population.

Cells C10–E10 give the viability (survival) fitness component, or the probability of surviving to reproduction. (Make up a hypothetical situation in which the A_2A_2 genotypes are selected against; perhaps their phenotype is more susceptible to being eaten by an introduced herbivore.) A survival probability for A_2A_2 genotypes of 0.2 means that each individual has a 20% probability of surviving to reproductive maturity.

For now, let’s assume that each genotype that reaches sexual maturity will contribute an equal number of gametes to the next generation (that is, fitness is not affected by reproductive potential). Let m be one-half the number of gametes that a sexually reproducing individual will contribute to the next generation. Since m is the same for all genotypes, each individual (regardless of its genotype) will contribute roughly the same number of gametes per adult to the next generation as any other individual (given that individuals reach adulthood). Note that we don’t care *how* these gametes recombine in the next population, only that they *are present* and available for counting when we calculate the p ’s and q ’s in the next generation.

In cell C12, enter the formula =C10*C11.

In cell D12, enter the formula =D10*D11.

In cell E12, enter the formula =E10*E11.

Recall that the absolute fitness, w , is equal to $l \times m$.

	A	B	C	D	E	F	G
14			Fitness components		Surviving	Next generation gametes	
15	Individual	Zygote genotype	Survival	Reproduction	genotypes	A1 gametes	A2 gametes

Figure 5

2. Set up a linear series from 0 to 99 in cells A16–A115.

3. Enter a **LOOKUP** formula to assign a genotype to each of the 100 individuals in the population. Copy the formula down to cell B115.

4. In cell C16, enter an **HLOOKUP** formula to calculate the survival probability of each zygote in the population and list the survival probability of its genotype in column C. Copy the formula down to cell C115.

5. In cell D16, use the **HLOOKUP** function to return the gamete contributions for each individual in the population. Copy your formula down to cell D115.

Enter the value 0 in cell A16.
Enter = 1+A16 in cell A17. Copy this formula down to cell A115.

In cell B16, enter the formula =**LOOKUP**(A16,\$C\$3:\$C\$6,\$A\$4:\$A\$6). The **LOOKUP** function will allow us to assign genotypes according to the numbers you entered in cells B4–B6. The vector form of the **LOOKUP** function looks in a one-row or one-column range (known as a vector) for a value and returns a value from the same position in another one-row or one-column range. For instance, the formula in cell B16 tells the spreadsheet to look up the individual’s number given in cell A16 in the vector C3–C6 (the genotype tallies), and return the appropriate genotype in cells A4–A6. If **LOOKUP** can’t find an exact match (for instance, individual 37 cannot be found because the number 37 is not part of the “tally”), **LOOKUP** returns the genotype that is associated with a number in the tally less than 37. Thus, individuals 0–24 are assigned the genotype listed in cell A4, individuals 25–74 are assigned the genotype listed in cell A5, and individuals 75–99 are assigned the genotype listed in cell A6. The result is that the genotypes are assigned exactly the way you specified in cells B4–B6.

In cell C16, enter the formula =**HLOOKUP**(B16,\$C\$8:\$E\$12,3,TRUE). We need the spreadsheet to examine individual 0’s genotype in cell B16, look up its survival probability in the table in cells C8–E12, and return that probability to cell C16. The **HLOOKUP** function can be used for this purpose. The **HLOOKUP** formula searches for a value in the top row of a table, and then returns a value in the same column from a row you specify in the table. The **HLOOKUP** formula has the form **HLOOKUP**(lookup_value,table_array,row_index_num,range_lookup), where **lookup_value** is the value to be found in the first row of the table (in our case, we want to look up the individual’s genotype in cell B16); **table_array** is a table of information in which data is looked up (in our case, we want to look up information in the table consisting of cells C8–E12); **row_index_num** is the row number in **table_array** from which the matching value will be returned (in our case, we want to return the value associated with survival probabilities, which is the third row in the table). The word **TRUE** tells the program that you require an exact match in the table.

Select cell C15, select the **HLOOKUP** function, and follow the prompts to create your formula. Copy your formula down to record survivorship probabilities for the remaining 99 individuals in the population.

We used the formula =**HLOOKUP**(B16,\$C\$8:\$E\$11,4,TRUE). Your spreadsheet should now look something like Figure 6.

	A	B	C	D	E
8			A1A1	A1A2	A2A2
9	Initial genotype frequencies =		0.25	0.5	0.25
10	Probability of genotype survival = <i>l</i> =		1	1	0.2
11	Half the # of gametes in next gen. = <i>m</i> =		2	2	2
12	Absolute fitness = <i>W</i> = <i>l</i> * <i>m</i> =		2	2	0.4
13					
14			Fitness components		Surviving
15	Individual	Zygote genotype	Survival	Reproduction	genotypes
16	0	A1A1	1	2	
17	1	A1A1	1	2	
18	2	A1A1	1	2	

Figure 6

6. In cell E16, enter a formula to determine which zygotes survive to adulthood. Copy your formula down to cell E115.

7. In cell F16, enter a formula to count how many gametes the surviving individuals actually contribute to the next generation.

8. In cell G16, enter a formula to compute the number of A_2 gametes contributed to the next generation by each surviving individual.

9. Copy the formulae in cells F16–G16 down to cells F115–G115.

10. Save your work.

In cell E16 enter the formula `=IF(RAND()<C16,B16,"")`.

Remember, the survival probabilities indicate the probability that an individual will survive to reproductive maturity (adulthood). In cell E16, we need a formula that will randomly determine whether individual 0 will survive to adulthood or not, based on the survival probability given in cell C16. The formula in cell E16 uses an **IF** formula to accomplish this task. The formula draws a random number between 0 and 1 (the **RAND()** portion of the formula). If the random number is less than the survival probability given in cell C16, the spreadsheet returns the value in cell B16 (the genotype of the zygote, or shall we now say the genotype of the adult). If the random number is greater than the survival probability, however, the individual died and a period (which designates a missing value) is returned instead. So far, we know which individuals survived to reproduce.

In cell F16, enter the formula `=IF(E16="A1A1",D16*2,IF(E16="A1A2",D16,""))`.

We will keep track of the A_1 gametes in column F and A_2 gametes in column G. The formula in cell F16 is *two* nested IF functions. The first part of the formula, `IF(E16="A1A1",D16*2`, tells the spreadsheet to examine cell E16, and if cell E16 is an A_1A_1 genotype, to multiply cell D16 by 2 (remember that cell D16 is one-half the gametes contributed, so when this number is multiplied by 2 it is the total number of gametes that an individual of genotype A_1A_1 contributes to the next generation).

However, if cell E16 is *not* genotype A_1A_1 , the spreadsheet walks through the second **IF** statement, `IF(E16="A1A2",D16,"")`. This states that if the genotype is A_1A_2 , then return the value in cell D16; otherwise return a missing value. Remember that the gametes produced by A_1A_2 genotypes include both A_1 gametes and A_2 gametes in approximately equal numbers, so that half of an individual's gametes are A_1 and half are A_2 . Therefore, to count the A_1 gametes from heterozygotes, only half the gamete contribution can be tallied in column F, which is simply m .

We entered the formula `=IF(E16="A2A2",D16*2,IF(E16="A1A2",D16,""))`.

Your spreadsheet should look something like Figure 7.

	F	G
14	Next generation gametes	
15	A1 gametes	A2 gametes
16	4	.
17	4	.
18	4	.
19	4	.
20	4	.
21	4	.
22	4	.
23	4	.

Figure 7

C. Calculate selection statistics.

1. Set up new column headings as shown in Figure 8.

2. Enter formulae to count the initial number of zygotes in the population in cells J9–L9 and the number of adults in cells J10–L10.

3. Enter formulae to recompute the absolute fitnesses of each genotype in cells J11–L11.

4. Use the **LARGE** formula in cell N11 to determine the largest absolute fitness of the three genotypes.

5. In cells J12–L12, enter formula to compute the relative fitness, symbolized with a lowercase *w*, for each genotype.

6. Calculate the selection coefficient, *S*, as $1 - w$ for each of the genotypes in cells J13–L13.

7. Save your work.

D. Make graphs of the selection statistics.

1. Graph the numbers of zygotes and breeding adults for each genotype (cells I8–L10).

	H	I	J	K	L	M	N
8			A1A1	A1A2	A2A2		
9	Number of zygotes =>						
10	Number of adults =>						
11	Absolute fitness = <i>W</i> =>					Largest <i>W</i> =	
12	Relative fitness = <i>w</i> =>					Average <i>w</i> =	
13	Selection coefficient = <i>S</i> =>						

Figure 8

We entered the following formulae, although you may have come up with other methods for counting individuals:

- J9 =B4
- K9 =B5
- L9 =B6
- J10 =COUNTIF(\$E\$16:\$E\$115,J8)
- K10 =COUNTIF(\$E\$16:\$E\$115,K8)
- L10 =COUNTIF(\$E\$16:\$E\$115,L8)

In cell J11, enter the formula =C12.
 In cell K11, enter the formula =D12.
 In cell L11, enter the formula =E12.

The **LARGE** function returns the largest (or second largest, or third largest, etc.) value in a data set. We entered =**LARGE**(J11:L11,1), where cells J11 and L11 give the data set, and the number 1 at the end of the formula indicates that we want the largest value returned (as opposed to the second largest or third largest value).

In cell J12, enter the formula =J11/\$N\$11.
 In cell K12, enter the formula =K11/\$N\$11.
 In cell L12, enter the formula =L11/\$N\$11.

The *relative fitness* of each genotype is the fitness of each genotype relative to the fittest genotype in the population. In cell N11, you’ve calculated the largest of the *W* values in the population. This represents the “fittest” genotype in the population, and all other genotypes will be assigned fitness values relative to this genotype. Relative fitness, *w*, can be obtained for each genotype by dividing the genotype’s absolute fitness (*W*) by the largest absolute fitness (W_{\max}).

In cell J13, enter the formula =1-J12.
 In cell K13, enter the formula =1-K12.
 In cell L13, enter the formula =1-L12.

Another useful characterization of the strength of natural selection against a genotype is the selection coefficient, *S*. *S* is simply $1 - w$, and indicates the relative *decrease* of a genotype due to selection. A high *S* indicates that a genotype was selected against, while a low *S* indicates that it was not selected against.

Use a column graph and label your axes fully. Your graph should resemble Figure 9 (although the number of A2A2 adults may differ from our graph).

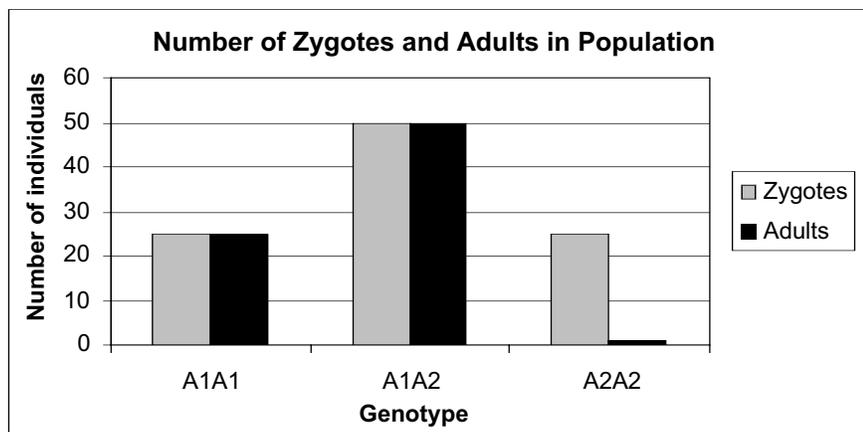


Figure 9

2. Graph W (absolute fitness), w (relative fitness), and S (the selection coefficient) for each genotype (cells I11–L13). Select the Series tab as you make your chart, and select cells J8–L8 as the Category (x) axis labels.

Use a column graph and label your axes fully. Your graph should resemble Figure 10.

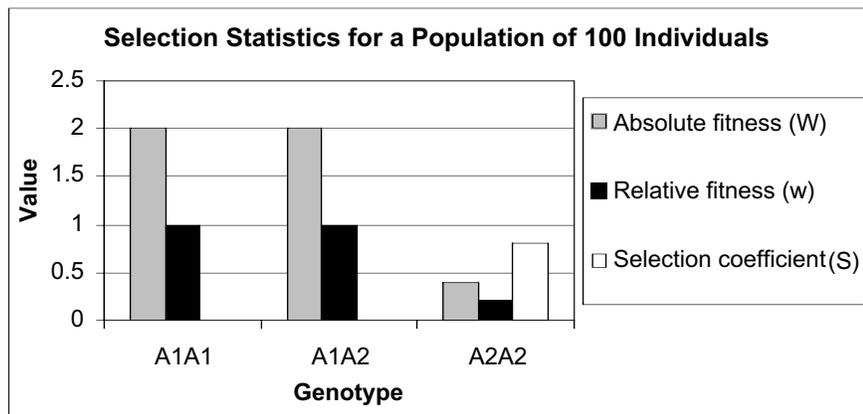


Figure 10

3. Answer Question 1 at the end of this exercise before proceeding.

E. Project allele frequencies and population numbers to next generation.

1. Set up new column headings as shown in Figure 11.

	I	J	K	L
16		A1 allele	A2 allele	
17	Time Step	p	q	N
18	1			
19	2			
20	3			
21	4			
22	5			
23	6			
24	7			
25	8			
26	9			
27	10			

Figure 11

2. In cells J18 and K18, enter formulae to compute p and q for the initial zygote population.

3. In cell L18, enter the formula =C6.

4. In cell J19, enter a formula to compute the new frequency of the A_1 allele, p . Copy your formula down to cell J27. Refer to Equation 9 in the Introduction.

5. In cell K19, compute the new frequency of q as =1-J19. Copy your formula down to cell K27.

6. For comparison, in cell H19 compute the frequency of the A_1 allele by counting the A_1 gametes in cells F16–F115, and divide that number by the total gametes (in cells F16–G115).

7. In cell L19, enter a formula to compute N_{t+1} . Copy your formula down to cell L27. Refer to Equation 12 in the Introduction.

In cell J18, enter the formula =`(COUNTIF(B16:B115,"A1A1"))*2+COUNTIF(B16:B115,"A1A2")`/(2*C6).

In cell K18, enter the formula =`1-J18`.

Refer to the exercise on Hardy-Weinberg equilibrium if you are rusty on the computations.

This represents the total initial population, tallied in cell C6.

We are now ready to write an equation to predict the change in allele frequencies from one time step to the next as a result of natural selection. Remember that selection happens *within* generations, but in this step we will now consider how natural selection may alter allele frequencies *between* generations. That is, how populations *evolve* as a result of natural selection.

In cell J19, enter the formula =`((J$11*J18^2)+(K$11*J18*K18))/((J$11*J18^2)+(K$11*2*J18*K18)+(L$11*K18^2))`.

This corresponds to Equation 9,

$$p_{t+1} = \frac{W_{11}p_t^2 + W_{12}p_tq_t}{W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2}$$

Follow the equation outlined in Step 4 above.

We used the formula =`SUM(F16:F115)/SUM(F16:G115)`.

Your results might not exactly match cell J19. Why? Press F9, the calculate key, and you will generate a new set of random numbers, and hence a new set of adults in column E. Only when the number of surviving adults exactly equals the survival probability times the number of zygotes will your answer match cell J19. This is because you selected which zygotes would reach adulthood with a random number function.

We will now use the p 's, q 's, and absolute fitnesses to calculate the new population size in cell L19.

We used the formula =`(J$11*J18^2+K$11*2*J18*K18+L$11*K18^2)*L18`. This corresponds to Equation 12:

$$N_{t+1} = (W_{11}p_t^2 + W_{12}2p_tq_t + W_{22}q_t^2) \times N_t$$

Your spreadsheet should now look something like Figure 12.

	I	J	K	L
16		A1 allele	A2 allele	
17	Time Step	p	q	N
18	1	0.5	0.5	100
19	2	0.625	0.375	160
20	3	0.7042254	0.2957746	284
21	4	0.7572203	0.2427797	528.247887
22	5	0.7946929	0.2053071	1006.67819
23	6	0.8224257	0.1775743	1945.46438
24	7	0.8437092	0.1562908	3792.77583
25	8	0.8605251	0.1394749	7437.31905
26	9	0.8741288	0.1258712	14643.1502
27	10	0.8853505	0.1146495	28915.1013

Figure 12

8. For comparison, in cell M19 compute N_{t+1} as the sum of the gametes in cells F16–G115 divided by 2.

9. Graph p , q , and N as a function of time.

We used the formula $=\text{SUM}(F16:G115)/2$ to compute the number of individuals (zygotes) in time step 2. Your results may not exactly match cell L19 because of the random number function used to determine which genotypes survived.

Use the XY scattergraph and label your axes fully. Your graph should resemble Figure 13. To create a secondary axis on the graph so that the frequencies are shown on the right axis and the number of individuals is on the left axis, double-click on the data in the graph that depicts p or q . A dialog box will appear. Click on the Axis tab, then select Secondary axis. Repeat for the other allele. To label the new axis, select the chart, then go to Chart | Chart Options | Titles and type in the labels for the primary y -axis (Number of individuals) and secondary y -axis (Frequency).

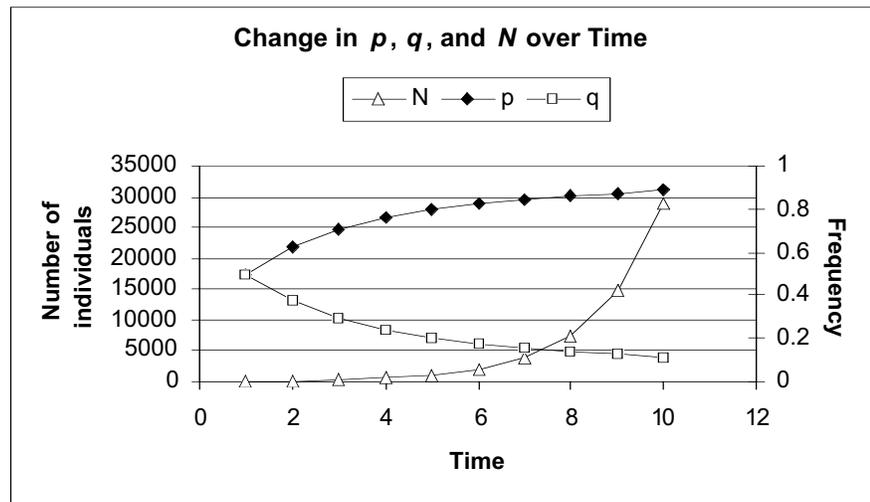


Figure 13

10. Save your work.

QUESTIONS

1. From your graphs in Section D of the exercise, describe the population in terms of natural selection *within* a generation (Figure 9). Describe the population in terms of W , w , and S (Figure 10).
2. In your model, you've selected against the A_2 homozygote. Yet the A_2 allele persists in the population, even after 10 years of constant selection. Extend your model to 100 years. At what frequency does the A_2 allele appear to stabilize? Why does the A_2 allele persist?
3. Modify your absolute fitness parameters by increasing the gamete contribution of the A_2A_2 genotype to 10 in cell E11. Examine your graph of relative fitness and selection coefficients. How did this change affect your population? What will happen to the frequency of the A_1 and A_2 allele over time?
4. Which affects the genetic rate of change in the population (change in A_1 allele from time step 1 to time step 2), relative fitness or absolute fitness? (Keep in mind that our calculations are based on absolute fitnesses.) Modify cells C10–E11 to answer this question. Enter the values shown below into your model:

Note that the absolute fitnesses have been changed, but the relative fitnesses (given in cells J12 and L12) remain the same. How does changing the absolute

	A	B	C	D	E
8			A1A1	A1A2	A2A2
9	Initial genotype frequencies =		0.25	0.5	0.25
10	Probability of genotype survival = l =		1	1	0.4
11	Half the # of gametes in next gen. = m =		4	4	2
12	Absolute fitness = $W = l * m$ =		4	4	0.8

fitness in the manner described affect p and q in the next generation? Modify your model so that the relative fitnesses are altered. How does changing the relative fitness affect p and q in the next generation?

5. Set up new entries as follows:

	A	B	C	D	E
8			A1A1	A1A2	A2A2
9	Initial genotype frequencies =		0.25	0.5	0.25
10	Probability of genotype survival = l =		0.6	0.2	0.4
11	Half the # of gametes in next gen. = m =		4	4	2
12	Absolute fitness = $W = l * m$ =		2.4	0.8	0.8

Compute the weighted average fitness (absolute fitness) in cell F12. The weighted average fitness can be computed by multiplying the absolute fitness of each genotype by its frequency in the population, and then summing these values together. Now compute λ for your population as N_{t+1}/N_t (given in cells L19 and L18). How do these numbers compare? Change some of your parameters in your model to see if your relationship holds no matter what parameters you change in your model. Why calculate a weighted average, rather than simply the average to predict population growth? Why is absolute fitness (weighted) used as indication of population growth rather than relative fitness?

6. Modify your absolute fitness parameters by selecting against the heterozygotes (absolute fitness = 0). Enter survival and reproductive values for the A_1A_1 and A_2A_2 homozygotes such that their absolute fitnesses are > 0 but equal in value. Change the genotype make-up of the initial population in the following manner:

How do p and q change over time? Next, change your values in cells B4–B6 as shown:

	A	B
3	Genotypes	# of individuals (zygotes)
4	A1A1	30
5	A1A2	50
6	A2A2	20

Update and graph your results. What happens to allele frequencies over time when $A_1 > 0.5$? When $A_1 < 0.5$? Explain your results.

	A	B
3	Genotypes	# of individuals (zygotes)
4	A1A1	20
5	A1A2	50
6	A2A2	30

7. Modify your absolute fitness parameters by selecting for the heterozygotes. Enter survival and reproductive values for the A_1A_1 and A_2A_2 homozygotes that result in an absolute fitness of 0, and values for the heterozygote > 0 as shown:

How does selection for the heterozygote affect p and q over time?

	A	B	C	D	E
8			A1A1	A1A2	A2A2
9	Initial genotype frequencies =		0.25	0.5	0.25
10	Probability of genotype survival = l =		0	1	0
11	Half the # of gametes in next gen. = m =		4	3	4
12	Absolute fitness = $W = l * m$ =		0	3	0

8. *(Advanced). Modify your model to include frequency dependent selection (the selection of a genotype depends on the frequency of the genotype in the population).

9. *(Advanced). Although you've entered survival and reproductive values for each genotype, these values remain fixed in your model. In reality, survival and reproductive rates are stochastic in nature. Modify your model to incorporate this element of stochasticity.

LITERATURE CITED

Futuyma, D. 1998. *Evolutionary Biology*, 3rd Edition. Sinauer Associates, Sunderland, MA.

Roughgarden, J. 1998. *Primer of Ecological Theory*. Prentice-Hall, Upper Saddle River, NJ.

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