

## A HARDY-WEINBERG EXCELL SPREADSHEET FOR GENE FREQUENCY CHANGES DUE TO SELECTION

John C. Bloom. Department of Computer Science, Miami University, Oxford, OH 45056  
Thomas G. Gregg, Department of Zoology, Miami University, Oxford, OH 45056

The famous Hardy-Weinberg equation shows the relationship between gene frequencies and genotype frequencies in random mating populations.

$$(pA + qa)^2 = p^2 AA + 2pq Aa + q^2 aa$$

This formula also serves as the starting point for understanding how different evolutionary forces, such as selection, drift, and migration bring about changes in gene and genotype frequencies. In this paper we are interested in the effects of selection on gene frequencies.

Every Genetics and Evolution textbook calculates selection-induced changes in gene frequencies in the same basic way as shown in Table 1 for a locus with complete dominance and selection against the homozygous recessive. Figure 1 extends this algebraic approach to other degrees of dominance and modes of selection.

This algebraic method can be very cumbersome, and perhaps intuitively unappealing to some in that it does not take into consideration the frequencies of the different kinds of mating or the fitness of their offspring. For instance if a heterozygote with a fitness of 1 mates with a homozygous recessive individual with a fitness of .5 half its offspring will have a fitness of 1 and half, a fitness of .5. But if the heterozygote were to have mated with a homozygous dominant all of its offspring would have a fitness of 1 so the two heterozygotes would have different probabilities of having their genes transmitted to future generations. However, it is possible to mimic any of the algebraic formulae in Table 1 and Figure 1 in an **excel spreadsheet** that immediately allows one to follow the changes due to selection from generation to generation when the initial conditions (fitnesses and initial gene frequencies) are set. Moreover, in the spreadsheet all the matings are accounted for, whereas in Table 1 and Figure 1 the matings are ignored in deriving the formulas for delta q. However, the different methods do not produce different results as one can verify by substituting values for p, q and s (1 – fitness) from the spreadsheet into the formulas in Table 1 and Figure 1.

SETTING UP THE SPREAD SHEET: Columns A - J.

LINE 1. Columns C, D and E are the frequencies of the three genotypes after 84 generations. We do this because most of the interesting changes will have taken place by the 84<sup>th</sup> generation and it saves scrolling down 424 lines. The formula for C1 is C424, D1 is D424, and E1 is E424.

LINE 2 Simply type in the genotypes A1A1, A1A2 and A2A2 into C2, D2 and E2.

LINE 3. Enter genotype fitnesses of your choice. Since by convention fitnesses are relative none should exceed 1. However, the program will work for absolute fitness values (any numbers) as well.

LINE 7. These are the zygote frequencies produced by random mating among the individuals in the previous generation. A1A1 zygotes in C7 will come from three kinds of mating: A1A1 X A1A1 (All the offspring from this mating will be A1A1), A1A1 X A1A2 ( There are 2 ways this mating can occur and half the offspring will be A1A1) , and A1A2 X A1A2 (one fourth of the offspring from this mating will be A1A1). So the formula for C7 is

$$C7 = (C4 * C4) + (C4 * D4 * .5 * 2) + (D4 * D4 * .0.25)$$

Four kinds of mating will produce A1A2 zygotes in D7 so the formula is

$$D7 = (C4 * D4 * .5 * 2) + (D4 * D4 * .5) + (D4 * E4 * .5 * 2) + (C4 * E4)$$

And three kinds of mating will produce A2A2 homozygotes so the formula For E7 is

$$E7 = (E4 * E4) + (E4 * D4 * .5 * 2) + (D4 * D4 * .25)$$

G7 is just the sum of the genotype frequencies so the formula is

$$G7 = (C7 + D7 + E7)$$

H7 and I7 Are the gene frequencies in the zygote population. The frequency of p is obtained by taking the frequency of A1A1 + half the frequency of A1A2. The frequency of q is half the frequency of A1A2 + the frequency of A2A2. This arithmetic calculation will give the gene frequencies whether or not the population is in equilibrium. The zygote population is in equilibrium as can be ascertained by substituting p and q from H7 and I7 into the Hardy-Weinberg equation and seeing that the gene frequencies correspond to the genotype frequencies in C7, D7 and E7.

$$H7 = (C7 + D7/2)$$

$$I7 = (D7/2 + E7)$$

LINE 8. To get the adult frequencies in line 8 each zygote frequency in multiplied by its fitness in Line 3. The formula for C8 is

$$C8 = (C7 * C\$3) \quad \text{The \$ results in the zygote frequency in any generation}$$

being multiplied by its fitness in C3 when the formula is copied into succeeding generations.

$$D8 = (D8 * D\$3)$$

$$E8 = (E8 * E\$3)$$

$$F8 = (C8 + D8 + E8) \quad \text{This will be the size of the population after selection also known as the mean fitness of the selected population designated by } \bar{W} \text{ as in Figure 1.}$$

LINE 9. This line shows the relative genotype frequencies of the adult population after selection and after the population has been normalized back to 1. To do this each genotype frequency in line 8 is divided by  $\bar{W}$ . Compare this to Table 1 where the relative genotype frequencies after selection are obtained by dividing each genotype frequency by  $1 - sq^2$ .

The formulae are  $C9 = (C8/F8)$

$$D9 = (D8/F8)$$

$$E9 = (E8/F8)$$

$$H9 = (C9 + D9/2)$$

$$I9 = (D9/2 + E9) \quad \text{Notice that the gene frequencies do not correspond to the genotype frequencies. This is because the adult population is not in equilibrium. This is a good place to point out that it takes only one generation of random mating to produce equilibrium because the zygotes produced by this population will be in equilibrium.}$$

$$J9 = (-I4 + I9)$$

At this point one simply “copies down” as many generations as desired and the gene and genotype frequencies will change automatically when the initial conditions are changed. Many combinations of initial conditions can be tried. An interesting one shows that the fittest does not always survive. For instance if we set the fitnesses at 1, .7 and .9 and the initial frequencies at .1, .5 and .4, A2A2 will fix. The fittest, A1A1 loses out because there are not enough of them to mate with each other and mating with the other genotypes breaks up the A1A1 combination and produces offspring of lower fitness. But if the initial frequencies are .2, .5 and .3 A1A1 goes to fixation.

SPREAD-SHEET Columns K, L, M and O

It is not necessary to go this extra step but the graph is a nice touch. To do this one laboriously identifies the cell containing each genotype in each generation. For example the formula for L3 is = C4, and L4 = C9 and so on. Once this is completed for the desired number of generations one uses the Excel graphing functions to graph the changes in genotype frequencies as seen in Table 3.

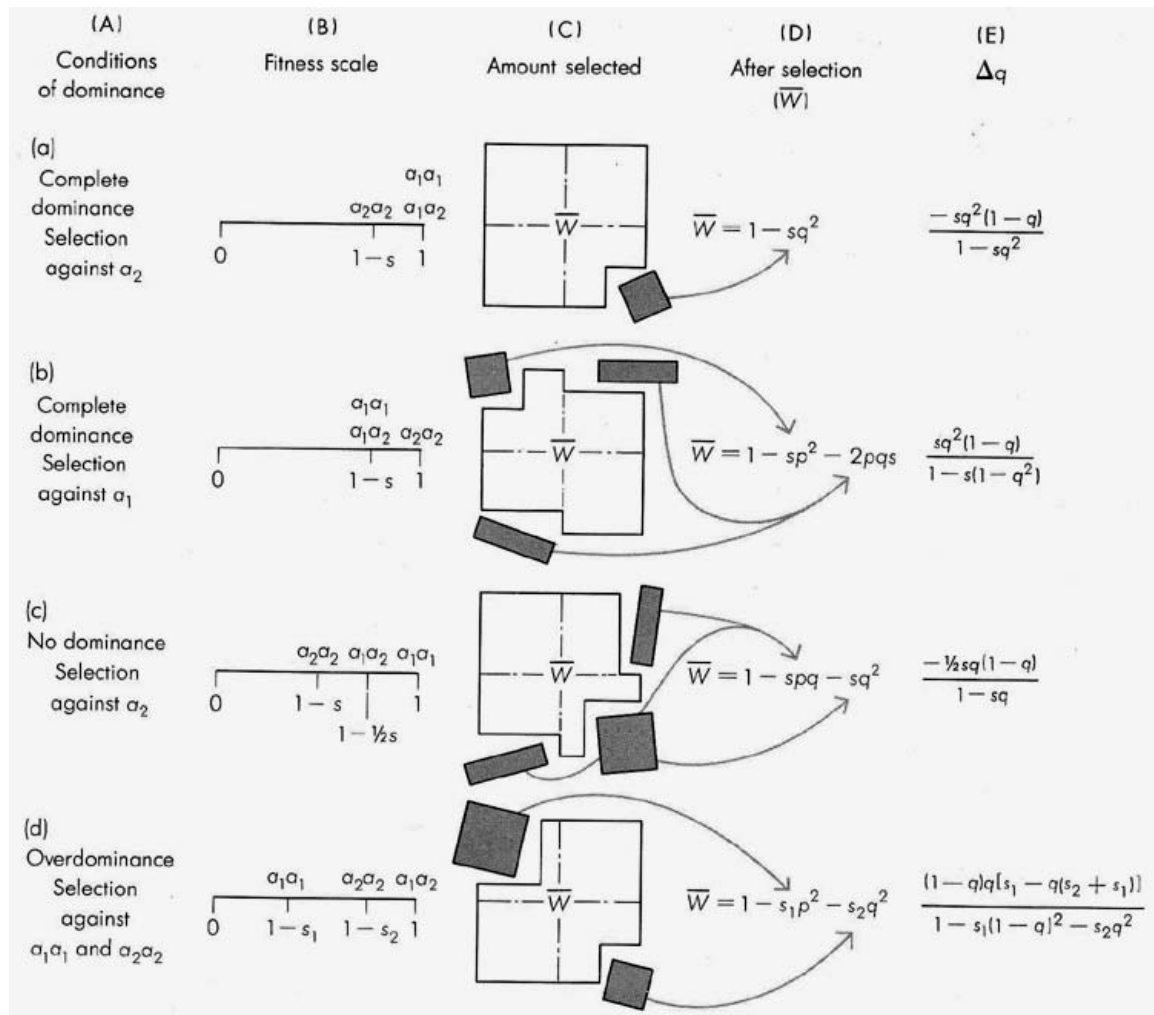
Table 1

Calculation of  $\Delta q$  for a deleterious recessive gene (*a*)

	AA	Aa	aa	Total	Frequency of <i>a</i>
Initial frequency	$p^2$	$2pq$	$q^2$	1	$q$
Adaptive value	1	1	$1 - s$		
Frequency after selection	$p^2$	$2pq$	$q^2(1 - s)$	$p^2 + 2pq + q^2 - sq^2$ $= 1 - sq^2$	
Relative frequency after selection	$\frac{p^2}{1 - sq^2}$	$\frac{2pq}{1 - sq^2}$	$\frac{q^2(1 - s)}{1 - sq^2}$		$\frac{pq + q^2(1 - s)}{1 - sq^2}$ $= \frac{pq + q^2 - sq^2}{1 - sq^2} = \frac{q(1 - sq)}{1 - sq^2}$
$\Delta q = \text{relative frequency of } a \text{ after selection} - \text{initial frequency of } a$					
$\Delta q = \frac{q(1 - sq)}{1 - sq^2} - q = \frac{q(1 - sq)}{1 - sq^2} - \frac{q(1 - sq^2)}{1 - sq^2} = \frac{q - sq^2 - q + sq^3}{1 - sq^2} = \frac{-sq^2 + sq^3}{1 - sq^2} = \frac{-sq^2(1 - q)}{1 - sq^2}$					

Classical calculation of delta q when “A” is completely dominant to “a” and selection is against the homozygous recessive.

Figure 1



A diagram showing the effects of different modes of selection with different degrees dominance.  $\bar{W}$  is the size of the population and is used to calculate the relative genotype frequencies as shown in table 1, which enables one to calculate the frequency of  $q$  after selection and the value of  $\Delta q$  as shown in Table 1. (Mettler, Gregg and Schaffer, 1987. Population Genetics and Evolution. Prentice-Hall, Englewood Cliffs, NJ.)

Table 2

	A	B	C	D	E	F	G	H	I	J
1		84th gen	0.0000	0.0000	0.0000	W	population	p	q	deltq q
2			A1A1	A1A2	A2A2					
3		fitness	1	1	0.8					
4		initial frequer	0.25	0.5	0.25			0.500	0.500	
5										
6										
7		zygotes	0.25	0.5	0.25		1	0.500	0.500	
8		adults	0.25	0.5	0.2	0.950				
9	1	generation	0.2632	0.5263	0.2105		1	0.526	0.474	-0.026
10										
11										
12		zygotes	0.2770	0.4986	0.2244		1	0.526	0.474	
13		adults	0.2770	0.4986	0.1795	0.955				
14	2	generation	0.2900	0.5220	0.1879		1	0.551	0.449	-0.025
15										
16										
17		zygotes	0.3036	0.4948	0.2016		1	0.551	0.449	
18		adults	0.3036	0.4948	0.1612	0.960				
19	3	generation	0.3164	0.5156	0.1680		1	0.574	0.426	-0.023
20										
21										
22		zygotes	0.3297	0.4890	0.1813		1	0.574	0.426	
23		adults	0.3297	0.4890	0.1451	0.964				
24	4	generation	0.3421	0.5074	0.1505		1	0.596	0.404	-0.022

Table 3

	K	L	M	N	O	P	Q	R	S	T	U	Y
1	Generation	A1A1	A1A2	A2A2	p	q	s					
2	0	0.25	0.5	0.25	0.500	0.500	0.2					
3	1	0.2632	0.5263	0.2105								
4	2	0.2900	0.5220	0.1879								
5	3	0.3164	0.5156	0.1680								
6	4	0.3421	0.5074	0.1505								
7	5	0.3670	0.4979	0.1351								
8	6	0.3909	0.4875	0.1216								
9	7	0.4138	0.4765	0.1097								
10	8	0.4357	0.4650	0.0993								
11	9	0.4566	0.4534	0.0900								
12	10	0.4764	0.4417	0.0819								
13	11	0.4953	0.4301	0.0747								
14	12	0.5131	0.4186	0.0683								
15	13	0.5300	0.4073	0.0626								
16	14	0.5461	0.3964	0.0575								
17	15	0.5613	0.3857	0.0530								
18	16	0.5757	0.3754	0.0490								
19	17	0.5893	0.3654	0.0453								
20	18	0.6023	0.3557	0.0420								
21	19	0.6145	0.3464	0.0391								
22	20	0.6262	0.3375	0.0364								
23	21	0.6372	0.3288	0.0339								
24	22	0.6477	0.3205	0.0317								
25	23	0.6577	0.3126	0.0297								

