

Practice problems I: **Solutions.**

A. Red and white are codominant in four o'clocks. A population consists of 35 red plants, 54 pink plants, and 67 white plants. Calculate the genotype and allele frequencies. Is this population in Hardy-Weinberg equilibrium? Show your work.

Let R = allele for red, r = allele for white.

$$f(RR) = 35 / 156 = 0.2244$$

$$f(Rr) = 54 / 156 = 0.3462$$

$$f(rr) = 67 / 156 = 0.4295$$

$$\text{Let } p = f(R). \quad p = f(RR) + 0.5 * f(Rr) = 0.3975$$

$$\text{Let } q = f(r). \quad q = f(rr) + 0.5 * f(Rr) = 0.6025$$

If in HWE, the number of RR =  $p^2 * 156$ ; Rr =  $2pq * 156$ ; rr =  $q^2 * 156$

$$\text{Expected number: } \text{Exp}(RR) = p^2 * 156 = 24.6$$

$$\text{Exp}(Rr) = 2pq * 156 = 74.2$$

$$\text{Exp}(rr) = q^2 * 156 = 56.6$$

Test for significant deviations using  $\chi^2$

$$\chi^2 = (35 - 24.6)^2 / 24.6$$

$$+ (54 - 74.2)^2 / 74.2$$

$$+ (67 - 56.6)^2 / 56.6$$

= 12. 12 > 3.84 (the critical value), so population is NOT at HWE.

B. A recent field experiment examined how the genotype at one locus affected the survival of sunflowers. 200 individuals with the AA, Aa, and aa genotypes were planted. Survival was surveyed at the time of flowering.

	AA	Aa	aa
number surviving	170	20	20

What are the relative fitnesses and selection coefficients for each genotype?

To calculate relative fitness, divide through by the highest fitness

	AA	Aa	aa
number surviving	170	20	20
divided by	170	170	170
relative fitness	1	0.12	0.12

selection coefficients show reduction in fitness compared to most fit genotype

relative fitness = 1 + s.

selection coefficient	-0.88	-0.88
-----------------------	-------	-------

C. Explain the difference between absolute and relative fitness, and why relative fitness is more useful for explaining evolution.

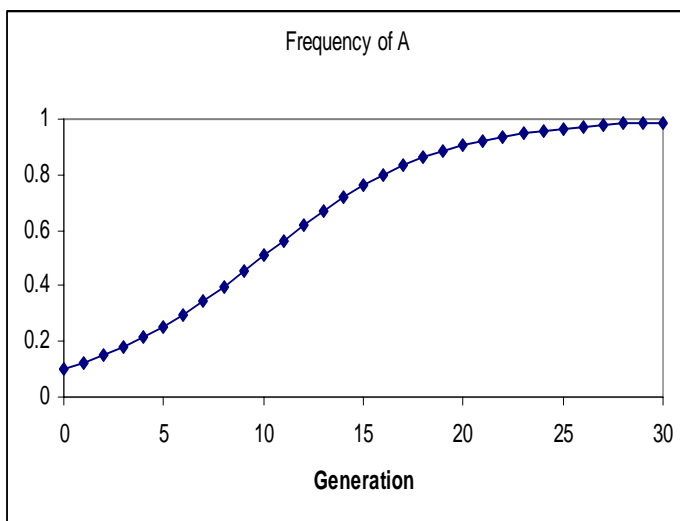
Absolute fitness is the number of offspring contributed to the next generation. Knowing this number for a genotype does not tell us whether that genotype is doing poorly or well, unless we know the fitness of other genotypes. Relative fitness shows how much fitness an allele or genotype has compared to the maximum fitness, and so whether it will increase or decrease.

D. One locus is expressed only in *Drosophila* sperm and not at any other point in the life cycle, with one allele having fitness higher than the other. The relative fitness for allele A is 1.0, and for allele a is 0.8.

i. Suggest one way to experimentally determine the fitness of fruit fly sperm.

There are several ways to do this. The most direct way is to mate a heterozygous male fruit fly to a female of known homozygous genotype. You will then be able to determine the number of each genotype of offspring, which will give you the relative fitness of each allele. For example, imagine that you mate an Aa male to a AA female. You count 320 Aa offspring and 400 AA offspring. The relative fitness of the a allele is 0.8 and that of A is 1.0. (Genotypes could be determined directly, via sequencing, or indirectly, by using a phenotype.)

ii. If the initial frequency of allele A is 0.1, sketch the change in allele frequencies over the next 50 generations. (No calculations needed, only general shape.)



iii . Why is the curve S-shaped? Explain why using the formula for  $\Delta p$ .

The shape of the curve depends on  $\Delta p$ . We know that

$$\Delta p = \frac{(W_A - W_a)p(t)q(t)}{W_A p(t) + W_a q(t)}.$$

First, if  $p$  is very small, the numerator will be very small. The numerator will increase as  $p$  approaches 0.5, as  $p(t)q(t)$  is at its maximum there. As  $p$  gets larger still,  $q$  gets smaller, so that the numerator declines as  $p$  approaches 1. This would suggest that the steepest part of the curve should be near the middle. This is where the response to selection,  $R$ , will be greatest.

The steepness of the curve also depends on the denominator:  $\Delta p$  increases as the denominator decreases.  $W_A > W_a$ , so the denominator will be minimized when  $p = 0$ , and will increase steadily in value after that.

Taking the two parts together, this suggests that the steepest part of the curve will be somewhere between  $p = 0$  (which would maximize things from the denominator side) and  $p = 0.5$  (which would maximize things from the numerator side), but it is hard to determine where exactly by this reasoning. In any case, the curve will begin relatively flat, will increase in steepness, and then will flatten out again.

This makes sense intuitively: if a favorable allele is rare, individuals with it may have higher fitness, but there are few of them, so the net change per generation will be low. When the unfavorable allele is rare, even if it is disappearing, the loss each generation will be comparatively small. When both alleles are common, the favored allele can gain many extra copies, and the rare allele will lose relatively many copies, each generation.

iv. How would this curve differ if the relative fitness for allele  $a$  were 0.9 rather than 0.8?

The steepness of the curve depends on the difference in fitnesses. The closer the two are to each other, the shallower the curve will be, and the longer it will take the favored allele to fix.

E. Explain the logic behind the equation for the frequency of allele A after selection:

$$p(t+1) = \frac{p(t)W_A}{p(t)W_A + q(t)W_a}$$

One way to think of this equation is to multiply both numerator and denominator by the total initial number of alleles: then, instead of frequencies, you would have numbers of alleles. The fitness would determine what proportion of the A allele survived, and likewise for the a allele. The new frequency is then the total number of A alleles after selection (= p \* W<sub>A</sub> \* initial total number) out of the total number of both alleles after selection (= p \* W<sub>A</sub> \* init. total + q \* W<sub>a</sub> \* init. total). You'll notice that the initial total is in both the numerator and denominator, so we can get rid of it - leaving our equation. The denominator is simply weighting by the new total.

III. Advanced level: most exam questions will not be this difficult.

F. Use excel to approximate the allele frequency where the curve is steepest in question D. (This is actually an easy question if you regularly use excel: if you are unsure how to approach this, come to office hours for some hints.)

$$\Delta p = \frac{(W_A - W_a)p(t)q(t)}{W_A p(t) + W_a q(t)}$$

The way to approach this is to set up a table in Excel. Here are the first two rows:

	A	B	C
1	p	$\Delta p$	fitness of a
2	0.05	$= (1 - \$c\$2) * a2 * (1-a2) / (a2 + ((1-a2)*\$c\$2)$	0.8

Excel uses locations in formulas: 'a2' refers to the value in cell A2. Locations can be referred to by relative position or absolute positions: \$c\$2 indicates that the program should always look in cell c2, no matter where it is.

You should be able to see how the formula in the cell matches the one given for  $\Delta p$ , if you substitute a2 for p, (1-p) for q, and 1 for W<sub>A</sub>. Fitness is \$c\$2.

Next, fill down column A by 0.05 steps until you reach 1.0. In column B, copy the formula into the cells below. You have now calculated the rate of change for each allele frequency. The maximum value should be at p = 0.45.

G. If allele A changes in frequency from 0.2 to 0.8 in 50 generations, what is the selection coefficient for allele a?

$$p(t) = \frac{p(0)W_A^t}{p(0)W_A^t + q(0)W_a^t}$$

$p(t) = 0.8$ ;  $t = 50$ ;  $p(0) = 0.2$ ;  $q(0) = 1 - p(0) = 0.8$ .  $W_A = 1$ . (This last step makes things nice:  $1^{50} = 1$ .)

$$0.8 = \frac{0.2}{0.2 + 0.8 * W_a^t}; 0.16 + 0.64 * W_a^t = 0.2.$$

$$0.04 = 0.64 * W_a^t; \frac{0.04}{0.64} = W_a^t; \ln(0.0625) = \ln(W_a^t); \frac{-2.776}{50} = \ln(W_a)$$

$$\exp(-0.055) = (W_a). \quad W_a = 0.946.$$

$$W_a = 1 + s; s = 0.054$$

The same thing could be done using the another route:

$$\frac{p(t)}{q(t)} = \frac{p(0)W_A^t}{q(0)W_a^t}$$

Using the values above, we obtain:

$$0.8 / 0.2 = 0.2 / 0.8 W_a^t;$$

$$1/16 = W_a^t; \ln(1/16) / 50 = \ln(W_a). \quad \exp(-0.055) = W_a .$$

$$W_a = 0.946 = 1 + s. \quad s = -0.054.$$

Same result, either way.