

Answer key to Problem set #2: selection

Biol 3410, 11 February 2009

Problem 1: Ethanol. Our quick-and-dirty equation of motion is $\Delta p \approx 0.25*(W_F - W_S)$, where W_F and W_S are the marginal fitnesses of the Fast and Slow alleles. Our job is to solve this for s , given the observed rate of allele-frequency change in the early generations of the experiment. Let the homozygous genotypic fitnesses be $W_{FF} = 1$ and $W_{SS} = 1-s$. If the F and S alleles interact in an “additive” or “codominant” way, then the fitness of the heterozygote will be half way between those of the two homozygotes: $W_{FS} = 1-0.5*s$, and the marginal fitnesses will be

$$W_F = [p^2(1) + \frac{1}{2}*2pq(1-0.5*s)]/p = 1 - \frac{1}{2}sq$$

$$W_S = [q^2(1-s) + \frac{1}{2}*2pq(1-0.5*s)]/q = 1 - \frac{1}{2}sq - \frac{1}{2}s$$

This implies that $(W_F - W_S) = \frac{1}{2}s$, so our equation becomes $\Delta p \approx 0.125*s$.

So what was Δp ? I estimate that p increased by $0.5-0.35 = 0.15$ over the first five generations. That would mean that the average one-generation change was $0.15/5 \approx 0.03$. Thus $0.03 \approx 0.125*s$. Dividing both sides by 0.125 gives $s \approx 0.24$. In other words, the Slow homozygotes appear to have suffered a fitness reduction of nearly one fourth relative to the Fast homozygotes.

Anything in this ballpark would be a perfectly good answer to the problem. However, it looks as though the initial rate at which Fast increased was greater than the rate of increase in subsequent generations. To find a value of s that fits the *whole experiment*, I wrote a program in the Python language to iterate the updating rule for 50 generations, beginning with $p = 0.35$. I then tried various values of s , looking for ones that seemed to give a good overall fit to the observations.

```
# adh.py
s = 0.13           # selection against SS homozygotes
p = 0.35          # initial frequency of fast allele (F)
for g in range(1,51):
    q = 1.0 - p
    wF = p + q*(1.0-0.5*s)
    w_bar = 1.0 - s*q
    p = p*wF/w_bar
    if g % 10 == 0:
        print "gen %2d    p = %5.3f" % (g,p)
```

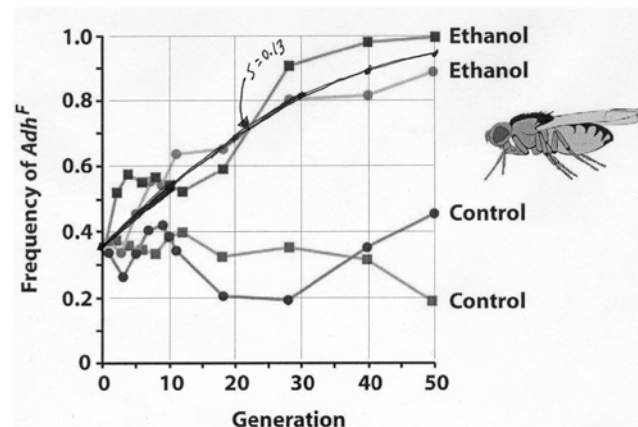
It produced the following output, which gives a fairly good overall fit to the whole experiment.

```
gen 10    p = 0.520
gen 20    p = 0.684
gen 30    p = 0.811
gen 40    p = 0.894
gen 50    p = 0.943
```

The predicted allele-frequency trajectory is plotted by hand on the graph. I like $s=0.13$ because it more or less splits the difference between the two replicate experiments over the last 10 generations.

Thus it uses the whole history, not just the first five or ten generations, which seem to have been somewhat atypical.

Estimates of s between 0.12 and 0.15 would all be just as defensible as the one I picked. But they're all a lot smaller than 0.24, my estimate by the quick and dirty (Q&D) method. The difficulty with Q&D is mainly that it uses only the first few generations, during which one replicate shot up in frequency, perhaps just by chance (*i.e.*, because of drift). If the trajectories weren't so noisy, Q&D would probably have given a lower estimate of s , closer to what I got from the full dynamics of the whole experiment. There are many possible reasons why the replicates differ from each other, but drift seems likely to be the main culprit. If you used the Q&D method, but with a larger number of generations (say, 10 or 20, as suggested in the problem, then you probably got a lower, more realistic estimate of s .



But please note that even if the “real” selective disadvantage of Slow (and advantage of Fast) were only around 0.1, that would still be *very strong selection!*

Problem 2: red and white flowers. In the first generation, the **r** allele has a frequency of $q = \sqrt{200/1000} = 0.447$, so **R** has a frequency of $p = 0.553$. In the second generation, $p = q = 0.5$. If the fitness of red-flowered plants is 1, then $w_{\text{bar}} = q^2(1+s) + (1-q^2)(1) = 1 + sq^2 = 1 + 0.2*s$. Thus $p' = 0.5 = 0.553*(1)/(1 + 0.2*s)$. This can be solved by elementary algebra, giving $s = 0.53$. Thus the relative fitnesses of red and white flowers are 1 and 1.53, in the first generation.

Note that this way of solving the problem takes s to be a selective advantage (not disadvantage). That’s perfectly fine, because fitnesses are *relative* rates or probabilities of reproduction.

Problem 3: Overdominance in the Mukai and Burdick experiment. If we let the frequency of the viable (V) allele be p , then the updating rule is $p' = p[W_V/W]$, where W_V is the marginal fitness of the V allele, and W is the population’s mean fitness (i.e., “W-bar”). The equilibrium occurs when $p' = p$, which is to say, when $W_V = W$. So we just need to solve for that condition.

$$W_V = [p^2 W_{VV} + \frac{1}{2} * 2pq(1)]/p = pW_{VV} + (1-p),$$

where W_{VV} is the fitness of VV homozygotes (what we want to solve for) and $W_{VL} = 1$ (by definition). We also know that $W_{LL} = 0$, again by definition. Thus

$$W_L = [\frac{1}{2} * 2pq(1) + q^2(0)]/q = p.$$

Now we can easily write down the population mean fitness and simplify, which gives

$$W = pW_V + (1-p)W_L = p^2 W_{VV} + 2p(1-p) \text{ (after a little bit of algebra).}$$

Or we could simply have written down the mean fitness directly from the *definition* of the mean fitness, in terms of the three (Hardy-Weinberg) genotype frequencies and their fitnesses:

$$W = p^2[W_{VV}] + 2p(1-p)[1] + q^2[0] = p^2 W_{VV} + 2p(1-p).$$

Now we can write the equilibrium condition, $W_V = W$, as a function of p and W_{VV} :

$$pW_{VV} + (1-p) = p^2 W_{VV} + 2p(1-p).$$

This is easily solved for W_{VV} , giving $W_{VV} = (2p-1)/p$. The graph shows $p \approx 0.8$ at equilibrium, which implies that $W_{VV} = (2*0.8 - 1)/0.8 = 0.75$.

Problem 4: Equilibrium frequencies of deleterious recessive alleles. This is very simple. We just plug the values of μ and s into the formula for the equilibrium frequency of the mutant: $q = (\mu/s)^{1/2}$.

μ	s	q	q^2	$2pq$	p^2
1.0e-005	1.0	0.003162	0.0000100	0.006305	0.99369
	0.1	0.010000	0.0001000	0.019800	0.98010
	0.01	0.031623	0.0010000	0.061246	0.93775
1.0e-006	1.0	0.001000	0.0000010	0.001998	0.99800
	0.1	0.003162	0.0000100	0.006305	0.99369
	0.01	0.010000	0.0001000	0.019800	0.98010
1.0e-007	1.0	0.000316	0.0000001	0.000632	0.99937
	0.1	0.001000	0.0000010	0.001998	0.99800
	0.01	0.003162	0.0000100	0.006305	0.99369

The equilibrium allele frequencies become much higher as selection against the mutant homozygotes becomes weaker, and lower as the mutation rate becomes lower. Most of the mutant alleles occur in heterozygotes, under all values of μ and s , but with weak selection even the homozygote frequencies can be quite high (for example, $q^2 = 0.001$ for $\mu = 1e^{-5}$ and $s = 0.01$). This means that even in an infinite population, typical individuals could be homozygous for quite a few mildly deleterious mutations, at different genetic loci. If many loci make only “marginal fitness contributions”, as implied by the yeast experiment of Thatcher *et al.* (1998), then many of these “mildly” deleterious mutations could be ones that completely abolish the gene’s function (for example, by *deleting* much or all of the gene)!