Homework 3: selection

Anthro/Biol 5221, October 24, 2008

There are two problems. Your written answers to both are due Monday, November 3rd

---

Problem 1. Strength of selection for *Adh*^{F} on ethanol-soaked fly food. Cavener and Clegg (*Evolution* **35**, 1-10, 1981) used wild-caught *Drosophila melanogaster* to establish experimental populations of flies that were grown either on standard fly food (cornmeal-molasses-agar) or on the same food supplemented with 10% ethanol. For 50 generations they tracked allele frequencies at 10 different enzyme loci that were polymorphic for alleles that could be assayed by starch-gel electrophoresis of the proteins. *Adh* (alcohol dehydrogenase) showed by far the strongest and most consistent pattern, which is summarized in the original figure from the paper (left) and in a recent textbook figure (right, from Freeman and Herron's *Evolutionary Analysis*, 4th edn, and reproduced in color in the lecture slides from October 10).

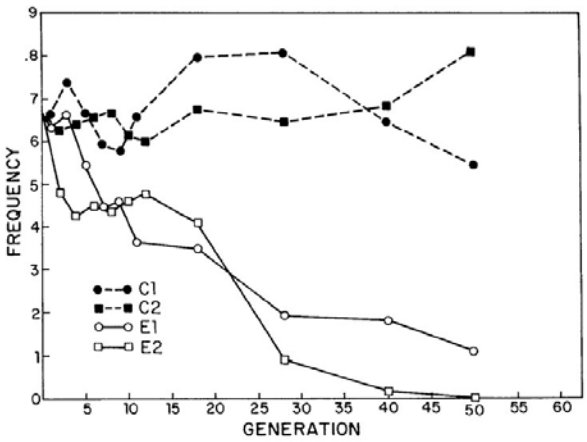
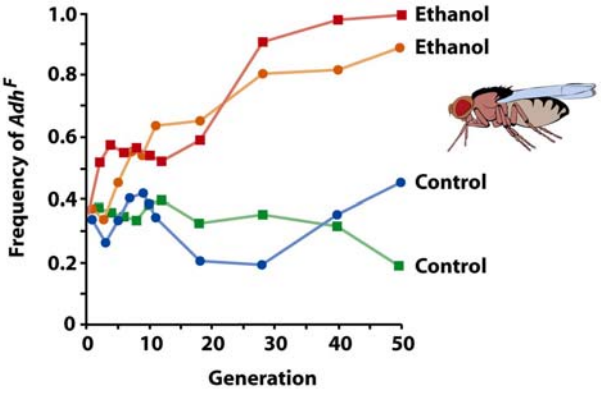


FIG. 1. *Adh*^{S}-frequencies for two control replicate populations (C_1 and C_2) and two ethanol selected replicate populations (E_1 and E_2).



The original figure plots the frequency of the *slow* allele on the vertical axis, but the textbook figure plots the frequency of *fast* (*i.e.*, it inverts the figure). Cavener and Clegg initiated all four experimental populations with flies taken from the same laboratory stock population, so they did not directly control the initial allele frequencies at any locus. Instead, the initial frequencies were those that occurred in the laboratory stock that was used to initiate the experiment. The initial frequency of *Adh*^{F} was $p \approx 0.35$, so the frequency of *Adh*^{S} was $q \approx 0.65$. (Note that the decimal points are missing from all but one of the numbers on the vertical axis of Cavener and Clegg's original figure.) The paper's main conclusion is that *fast* was strongly favored in the ethanol-soaked experimental populations: by generation 50 it had fixed in one and reached a frequency of almost 0.9 in the other. This was not particularly surprising, because *FF* homozygotes were already known to show twice as much Adh enzyme activity as *SS* homozygotes.

From the data as displayed in the graphs you can estimate the relative fitnesses of the two alleles in the high-ethanol environments by fitting a simple model of selection to the observed (average) change in allele frequencies over a given number of generations. To make things simple, let the fitness of the *FF* homozygotes be 1.0 and assume that the fitness of the *FS* heterozgotes is *intermediate* between the *FF* and *SS* homozygotes (*i.e.*, $h = ½$). Your job is to estimate the fitness of the SS homozygotes, which will be a number less than 1. (It will be 1-*s*, where *s* is the selection coefficient against SS homozygotes). You should do this in two different ways (as explained below) and then compare your estimates.

Method 1: quick-and-dirty calculation. The expected change in allele frequency (from one generation to the next) is a function of the marginal allelic fitnesses, the mean fitness, and the allele frequency itself. Since these quantities all change with the allele frequency, we can't easily calculate Δp for more than one generation at a time. But we can *approximate* the multi-generational allele-frequency change under certain circumstances. For example, with *additive* allelic interactions ($h = ½$), the *difference*

$$\Delta p = pq(\bar{W}_1 - \bar{W}_2)/\bar{W}$$

between the two marginal fitnesses depends on *s* but not *p*, so it does not change with the allele frequencies. And when the allele frequencies are near 0.5, the product *pq* is always close to 0.25. In addition, the mean fitness will always be fairly close to 1 unless *s* is very large, so the denominator can simply be ignored. Thus, if we consider just the first few generations of the experiment, when both alleles have frequencies near 0.5, we can approximate the one-generation change in allele frequency as 0.25*(difference between the marginal fitnesses).

So, here's how to do the quick-and-dirty estimate of *s*. First, derive an expression for the difference between the marginal fitnesses of the two alleles when $h = ½$. (It's a simple function of *s* only.) Second, from the graphs above, estimate the allele-frequency change in the first few generations of the high-ethanol treatments, while both alleles still have frequencies fairly near 0.5. (Obviously, you should average over the two replicates of this treatment. For example, after 5 generations I would say that the average frequency of the fast allele has risen to around 0.5. You may think the best estimate is a bit higher or lower than this, or you may prefer to use a different number of generations. That's fine.) Third, turn your estimate of the total allele-frequency *change* into a *per-generation* change (*i.e.*, divide it by the number of generations from the beginning of the experiment). Finally, set your expression for the one-generation change in allele frequency equal to your estimate of the average allele-frequency change in the early generations of the experiment, and solve for *s*. $W_{SS} = 1\text{-}s$. *Show your work!*

Method 2: model the full dynamics. To use more generations of the Cavener-Clegg experiment in our estimate of $W_{SS} = 1\text{-}s$, we need to do things a bit more exactly. Recall that

$$p' = p[\bar{W}_1]/\bar{W}$$

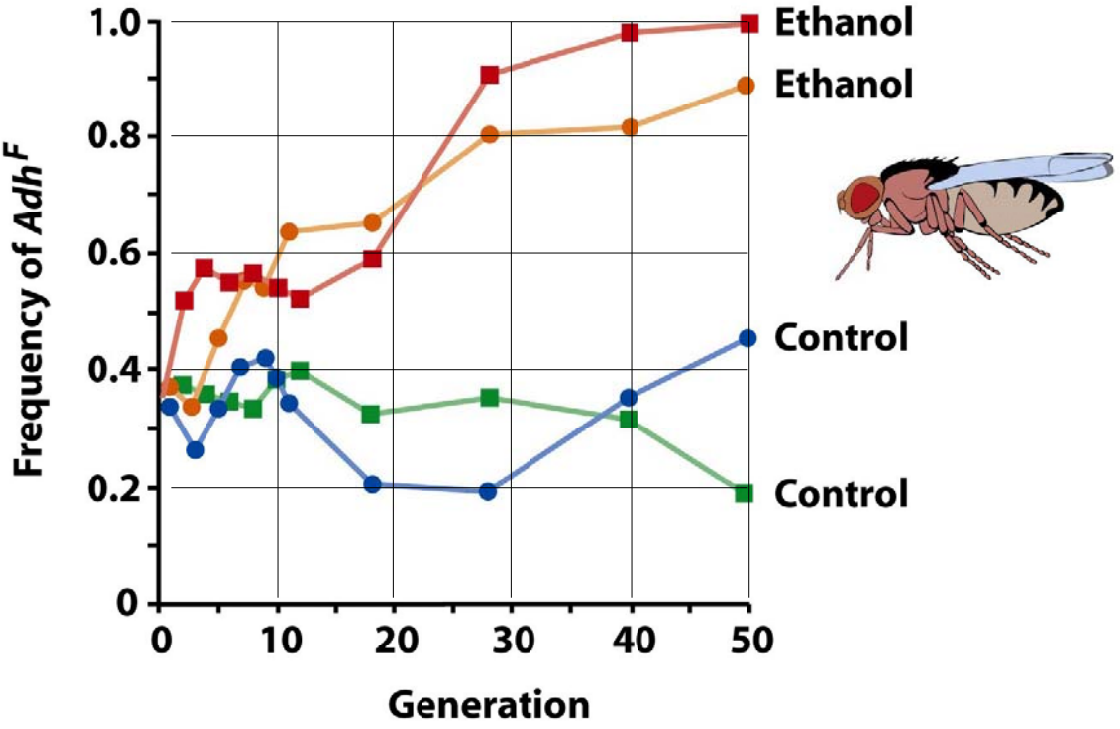
Again, this is just a one-generation updating rule. How can we extend it to 20 or 30 or 50 generations? Python to the rescue! Here's the outline of a program that *iterates* the selection model for 50 generations, starting with a *fast* allele frequency of $p = 0.35$. All you need to do is fill in the expressions indicated by the phrases in angle brackets, and then run the program.

```
# adh.py
s = <your guess here>    # selection against SS homozygotes
p = 0.35                 # initial frequency of the fast allele (F)
for g in range(1,51):
    q = 1.0 - p
    wF = <your expression for the marginal fitness of the fast allele>
    w_bar = <your expression for the population-wide mean fitness>
    p = p*wF/w_bar
    print "gen %2d   p = %5.3f" % (g,p)
```

When you've got this working, play with different values of ***s*** until you find one that seems to do a good job of modeling the Adh^F allele-frequency trajectories in the high-ethanol treatments, through all 50 generations of the experiment. *Show us your program and its output, and graph the expected trajectory on top of the observed one, using the gridded version of the figure (next page, and available in the PDF of this problem set on the course web site). Explain as clearly as you can why you think your favorite value of* **s** *is a plausible estimate of the "true" value. How does this estimate of* **s** *differ from the one you got with the quick-and-dirty method? Why do you think they differ in the way they do? And why do you think the observed trajectories differ from the predicted one, and from each other? Answer all these questions.*

Problem 2. A simpler case, but with dominance. White flowers (genotype *rr*) are recessive to red (*RR* and *Rr*) in an outbreeding plant species. In a large random sample, you count 200 white-flowered plants and 800 red-flowered plants. What's the frequency of the red allele (*R*)? One generation later, you count 250 white and 750 red plants. What was the change in allele frequency (of *r* and/or *R*)? If this change was caused by a preference of the local pollinators (night-flying moths) for white flowers, and not by drift, or selection at some linked locus, or magic or whatever, then what were the relative fitnesses of white-flowered and red-flowered plants in the first (parental) generation? *Please show your work.*

Hint: There are harder and easier ways to frame this problem. It will make your life easier if you set the fitness of red-flowered plants to 1.0 and focus on the frequency of the dominant *R* allele, which will then have a marginal fitness of 1.0 regardless of its frequency.


1.0
0.8
0.6
0.4
0.2
0
Frequency of Adh^F
0
10
20
30
40
50
Generation
Ethanol
Ethanol
Control
Control