BE/Bi 101: Order-of-Magnitude Biology
Homework 5
Due date: Friday, February 13, 2015

> "Problems worthy of attack prove their worth by hitting back."
>
> —Piet Hein

1. Snapshot—evolution by the numbers.
Cell Magazine has a short format (2 page) article known as "Snapshot" that gives a brief description of important areas of cell biology. One of us (RP) wrote such an article called "Key Numbers in Biology" that attempts to pull together 25 key numbers that any cell biologist needs to know. Read that article (which you can download here) to get a sense of how these features are structured. Now that you have seen an example, it is your turn to partially imagine how to construct such a Snapshot feature yourself. Specifically, your job is to help us write a Snapshot feature entitled "Evolution by the Numbers." First, make a list of five key headings for your numbers. To give you a sense of what we mean in the context of cell biology one could have "Length Scales," "Time Scales," "Concentrations," etc. Figure out what five categories you think would be most important for characterizing key numbers in evolutionary biology. Second, choose five key numbers that fall within any of these categories and try to give an impression of an order-of-magnitude estimate for your number and justification for why it is important. Please submit your answer to Rob, Justin, David and Pradeep in PDF form by email.

2. Population genetics and selection.
In class we talked about how allele frequencies change over time in the presence of selection. In this problem, you will flesh out the details of this interesting topic which was only considered in a cursory fashion in class.

a) Before considering the case of selection, we must first review the allele frequencies generation by generation for an effectively infinite population. Basically, this means we must remind ourselves of the key tenets of the Hardy-Weinberg model. As was done in class, imagine an urn that in generation M contains N_A copies of the allele A resulting in a frequency $p = N_A/2N$. Hence, there are $2N - N_A$ copies of allele a. What is the frequency of these alleles in the population? Now imagine an experiment in which to construct generation $M+1$ we draw N pairs (!!) of alleles with replacement in order to obtain the contents of our new urn for generation $M+1$. The concept of our experiment is intended to mimic the randomness of mating in actual biological populations. Given the way we have set up the urn drawing, what is the expected frequency of AA pairs? What is the frequency of Aa pairs? What is the frequency of aa pairs? Now, using the formula we derived in class that tells us how to obtain allele frequencies from genotype frequencies, find an expression for the allele frequencies in generation $M+1$ given the genotype frequencies. Specifically, demonstrate that

$$p' = p^2 + \frac{1}{2}\,2pq, \tag{1}$$

where p' is the frequency of the A allele in generation $M+1$ and p is its frequency in generation M and use this to demonstrate that allele frequencies are *invariant* under this simple model of mating. Do the same thing for q'. This jibes with what was claimed in class, namely, that Hardy-Weinberg is akin to Newton's first law of motion for the genetics of evolution—in the absence of any driving force, allele frequencies stay the same.

b) Now we are going derive the change in the frequencies Δp and Δq where we amend the procedure introduced in part (a) to include selection. Basically, we now imagine a two-step mating process in which after we draw a given allele pair (i.e. a genotype) we then decide if that genotype survives by keeping that allele pair with weight w_{ij}. Specifically, we keep AA pairs with weight w_{11}, we keep Aa pairs with weight w_{12} and we keep aa pairs with weight w_{22}. In class, we introduced the idea of the mean fitness as

$$\bar{w} = p^2 w_{11} + 2pq w_{12} + q^2 w_{22}. \tag{2}$$

This then allows us to ascribe genotype frequencies for generation $M+1$ as $x_{AA} = p^2 w_{11}/\bar{w}$, $x_{Aa} = 2pq w_{12}/\bar{w}$ and $x_{aa} = q^2 w_{22}/\bar{w}$, guaranteeing that the population size remains the same in each generation. Use all of these definitions and ideas to show that the change of allele frequency for A is given by

$$\Delta p = p' - p = \frac{p}{\bar{w}}\,a_1 \tag{3}$$

where

$$a_1 = p(w_{11} - \bar{w}) + q(w_{12} - \bar{w}) \tag{4}$$

is the average excess of fitness. Essentially, provide a clear derivation of the way in which allele frequencies change from one generation to the next in the case where selection is in effect. Find both Δp and Δq. Given those results for Δp and Δq, should we expect that $\Delta p + \Delta q = 0$ and if so, are your results for the changes in allele frequencies consistent with that constraint? To demonstrate this, you need to actually calculate $\Delta p + \Delta q$.

c) As a toy model for how allele frequencies associated with the sickle cell trait change assume that the fitnesses are $w_{AA} = 0.9$, $w_{AS} = 1.0$ and $w_{SS} = 0.2$, where we use A for the wild-type allele and S for the allele associated with sickle-cell. Imagine that the S allele begins with a very low frequency, q. First, give an intuitive argument for what you expect the *sign* of Δp and Δq to be in the next generation of reproduction. Then, by assuming that $q = 0.001$, find Δp and Δq by using the formulae you derived in the first part of the problem. Make sure you identify the mean fitness and the average excess of fitness. Do Δp and Δq have the signs you expect intuitively? Given this result, explain what will happen over time to the allele frequencies.

3. Mutation-selection balance.
We can continue to build on these same simple one-locus, two-allele models to examine combinations of different "forces of evolution". Specifically, here we think about the way in which mutation and selection compete to determine allele frequencies. In this problem we are going to work out the mathematics of this effect explicitly. Our strategy is similar in spirit to what we did in class where we introduced the idea of an urn from which we draw our alleles A_1 and A_2. We imagine that the allele A_1 is mutated into A_2 with probability μ each generation and that the rate of mutation of A_2 back to A_1 is negligibly small. This scenario could be realized if we think of A_2 really as the set of all the ways that the gene allele A_1 can be broken. There are thus many ways to break our allele but only one way that mutation can restore it and hence this rate is assumed small.

For the simple case in which there are only two possible alleles, we consider the process in two steps. To be concrete, consider the case in which the fitnesses are $w_{11} = 1$, $w_{12} = 1$ and $w_{22} = 1 - s$. First, show that we have

$$p_{\text{after selection}} = \frac{p}{1 - q^2 s}. \tag{5}$$

This essentially is a special case of what you did in the previous problem. We already worked this out in class, so your job is simply to recapitulate that discussion and to explain all of the steps leading up to that equation, both mathematically and conceptually. Now we need to impose the results of mutation. Impose the mutation by computing

$$p' = p_{\text{after mutation/selection}} = (1 - \mu)\,\frac{p}{1 - q^2 s}. \tag{6}$$

Explain what this equation means and how it captures the amendment of the allele frequency p due to mutation. For the case when mutation and selection exactly balance, we have $p' = p$. This implies that we have

$$p = (1 - \mu)\,\frac{p}{1 - q^2 s}. \tag{7}$$

Show that this steady state condition implies that

$$q = \sqrt{\frac{\mu}{s}}. \tag{8}$$

Examine your result qualitatively and provide an argument as to how increasing either mutation rate or selection coefficient will alter allele frequencies. Does this equation make sense intuitively? Why is this result referred to as "mutation-selection balance".