Bi/Ge105: Evolution
Homework 4
Due Date: Wednesday, February 11, 2026

"We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes."

–Pierre Simon Laplace, A Philosophical Essay on Probabilities

1. Population Genetics for Haploid Organisms (with thanks to Prof. Aleks Walczak)

In this problem, we build on what was done in class, asking you to do a detailed analysis of the mutation-selection balance in haploid organisms. Once again, we think about the one-locus, two-allele (A_1 and A_2) abstraction that we have already been using in class.

(A) In the presence of selection, with fitnesses w_1 and w_2 for our two genotypes, write down the expressions for the new values of p and q after a generation of selection. If you think of the urn idea we considered in class, then the number of A_1 alleles is N_1 and the number of A_2 alleles is N_2, and there are a total of $N = N_1 + N_2$ alleles in our urn. This means that we assign $p = N_1/N$ and $q = N_2/N$ to the probability of drawing an A_1 and an A_2, respectively. What we want is p' and q' which is the frequencies after a single generation of selection. Make sure you explain your notation and that you give an appropriate definition of the mean fitness in the context of this simplified haploid example.

(B) Now, consider the case in which the relative fitnesses are $w_1 = 1$ and $w_2 = 1 - s$, where s is the so-called selection coefficient. Work out an expression for Δp as a function of p, q and s. Now, assuming that the initial frequency of $p = 0.01$ and that $s = 0.1$, make a plot of Δp as a function of the generation number. Explain in what sense the system is evolving. These first two parts of the problem treat the finite population discretely and traffic in the language of discrete probability. Now we turn to an alternative view of these same kinds of problems using continuous variables and differential equations.

(C) Now imitate the concept we did in class in order to work out the mutation-selection balance. Here I am going to walk you through it step by step, but expect you to restate things in your own words and then to explain the significance of the results you obtain, including making plots to illustrate the outcome. We begin by noting that we have a population that harbors one of two alleles, with the number of cells with the A_1 allele labeled by n_1 and the number of cells with the A_2 allele labeled by n_2. In class we wrote equations for these two populations as

$$\frac{dn_1}{dt} = \mu_2 n_2 - \mu_1 n_1 + \gamma_1 (n_1, n_2) \tag{1}$$

and

$$\frac{dn_2}{dt} = \mu_1 n_1 - \mu_2 n_2 + \gamma_2 (n_1, n_2) \, . \tag{2}$$

Explain the first two terms in each equation in terms of mutation rates. We impose the constraint that the total number of cells is constant and given by

$$n = n_1 + n_2 \tag{3}$$

and also make the *ansatz* that the term describing growth and selection is given by

$$\gamma_i (n_1, n_2) = \underbrace{r_i, n_i}_{\substack{\text{usual}\\ \text{growth}}} - \frac{n_i}{n} \underbrace{f (n_1, n_2)}_{\substack{\text{choose}\\ \text{to impose}\\ \text{population size}\\ \text{constraint}}} \, . \tag{4}$$

Show that we can write the dynamics of the total number of cells as

$$\begin{aligned} \frac{dn}{dt} &= 0 = \gamma_1 (n_1, n_2) + \gamma_2 (n_1, n_2) \\ &= r_1 n_1 + r_2 n_2 - \frac{n_1}{n} f (n_1, n_2) - \frac{n_2}{n} f_1 (n_1, n_2) \end{aligned} \tag{5}$$

and then demonstrate that this implies

$$f (n_1, n_2) = r_1 n_1 + r_2 n_2. \tag{6}$$

In light of all of these assumptions, now show that the equation for $x = n_1/n$ is given by

$$\frac{dx}{dt} = \mu_2 - (\mu_1 + \mu_2)\,x + \underbrace{(r_1 - r_2)}_{\text{selection strength}}\,x(1-x). \qquad (7)$$

Now you are going to solve for the steady state frequency $x = n_1/n$ under three distinct scenarios. First, consider the case where $s = r_1 - r_2 = 0$ (i.e. no selection). Find x and explain what it means. Next, consider the case where there is no mutation, and only selection. Find x in steady state in this case too and explain what it means. Finally, solve for the steady state value of x in the regime of mutation-selection balance and once you have solved your quadratic equation, explain how the steady state frequency depends upon the s and $\mu = 1/2(\mu_1 + \mu_2)$.

The key point in this problem is that you "engage with the material." Though the problem description above sounds like a recipe, the key point is that you deeply understand the problem and its meaning. Specifically, we want to make sure you have intuition for selection-mutation balance. Thus, the goal is for you to create a pedagogically complete and clear description of what this problem is about, how we approach it and what it teaches us.

2. Population Genetics for Diploid Organisms

In the previous problem, we worked out the mathematical description of the change in allele frequencies for haploid organisms. Now we turn to the analogous approach for diploid organisms.

(A) Derive an expression for Δp, namely,

$$\Delta p = \frac{p}{\bar{w}}[p(w_{11} - \bar{w}) + q(w_{12} - \bar{w})], \qquad (8)$$

where $\bar{w} = p^2 w_{11} + 2pqw_{12} + q^2 w_{22}$. The way that I think of this is that the probability of the A_1A_1 genotype is $p^2 w_{11}/\bar{w}$, the A_1A_2 genotype has probability $2pqw_{12}/\bar{w}$ and the A_2A_2 genotype has probability $q_{22}^2/\bar{w}$. By using the definition of $\bar{w}$, demonstrate that this can also be written as

$$\Delta p = \frac{pq}{\bar{w}}[p(w_{11} - w_{12}) + q(w_{12} - w_{22})]. \qquad (9)$$

Make a plot of Δp vs p for the case of overdominance in which $w_{12} > w_{11} > w_{22}$. Make sure you explain what this graph demonstrates. Essentially, this is a phase portrait that shows how the allele frequency changes from one generation to the next.

(B) In class we claimed that in the case of overdominance, there is a fixed point p^*. Using the expression

$$\Delta p = \frac{pq}{\bar{w}}[p(w_{11} - w_{12}) + q(w_{12} - w_{22})], \tag{10}$$

find p^* by solving for the case in which $\Delta p = 0$. Your expression for p^* will be a simple function of w_{11}, w_{12} and w_{22}.

(C) Now repurpose your calculations to the case of the case of heterozygous advantage and choose our fitnesses symmetrically as $w_{11} = w_{22} = 0.1$ and $w_{12} = 1.0$. Once again, make a plot of Δp vs p and explain what the resulting graph shows us.