BIOB480/BIOE548 notes 9/19/2024

Introduction

- HW3 due Tuesday at 10:50 AM. Any questions?
- Revisions: in person or via email—if in person, shooting me an email doesn't hurt!
- Will post the reading for Tuesday right after class.
- Nancy Chen seminar (3:30 PM 346 Leon Johnson)

Selection recap:

Tuesday was a lot! Here are the most important models to remember.

The frequency of a recessive lethal allele in the next generation under complete dominance:

$$q_1 = \frac{q}{1+q}$$

The *change* in frequency of a recessive lethal allele and the alternate allele under complete dominance:

$$\Delta p = \frac{q^2}{1+q};\ \Delta q = \frac{-q^2}{1+q}$$

Alelle frequency changes under complete dominance and selection against q:

$$\Delta p = \frac{spq^2}{1-sq^2};\ \Delta q = \frac{-spq^2}{1-sq^2}$$

Alelle frequency changes under partial dominance and selection against q:

$$\Delta p = \frac{spq((p-q)h+q)}{1-2hspq-sq^2};\ \Delta q = \frac{-spq((p-q)h+q)}{1-2hspq-sq^2}$$

These models let us predict allele frequency changes under scenarios where selection is the only evolutionary force acting on a given population. For example, consider the mice studied by Nachman et al. 2003. Coat color appears to show complete dominance: Individuals with the genotypes DD and Dd at Mc1r are dark, while those with the genotype dd are light. If $f(D) = 0.3$ and $f(d) = 0.7$, and the relative fitness of white mice on lava rocks is 0.3 ($W(dd) = 0.3$, which means a selection coefficient of $s = 1 - 0.3 = 0.7$), how much will allele frequencies in that habitat change in one generation? We can solve for this with simple substitution:

$$\Delta p = \frac{spq^2}{1-sq^2};\ \Delta q = \frac{-spq^2}{1-sq^2}$$

$$\Delta p = \frac{0.7*0.3*0.7^2}{1-0.7*0.7^2};\ \Delta q = \frac{-0.7*0.3*0.7^2}{1-0.7*0.7^2}$$

$$\Delta p = 0.1566;\ \Delta q = -0.1566$$

Time required for a given allele frequency change

As discussed above, the frequency of a lethal deleterious allele in the next generation is $q_1 = \frac{q}{1+q}$. It follows that its frequency in the generation *after that* is $q_2 = \frac{q_1}{1+q_1}$. Since we know what q_1 is in terms of q (the original allele frequency), we can substitute it and solve for a general relationship:

$$q_2 = \frac{\frac{q_0}{1+q_0}}{1+\frac{q_0}{1+q_0}} = \frac{q_0}{1+q_0} \div 1 + \frac{q_0}{1+q_0}$$

$$q_2 = \frac{q_0}{1+q_0} \div \frac{1+q_0}{1+q_0} + \frac{q_0}{1+q_0} = \frac{q_0}{1+q_0} \div \frac{1+q_0+q_0}{1+q_0}$$

$$q_2 = \frac{q_0}{1+q_0} \div \frac{1+2q_0}{1+q_0} = \frac{q_0}{1+q_0} * \frac{1+q_0}{1+2q_0} = \frac{q_0(1+q_0)}{(1+q_0)(1+2q_0)} = \frac{q_0}{1+2q_0}$$

By extension, $q_t = \frac{q_0}{1+tq_0}$.

We can rearrange this to get the number of generations (t) required to change the allele frequency from q_0 to q_t:

$$q_t = \frac{q_0}{1+tq_0};\ q_t(1+tq_0) = q_0$$

$$q_t + tq_0q_t = q_0;\ tq_0q_t = q_0 - q_t$$

$$t = \frac{(q_0 - q_t)}{q_0q_t} = \frac{1}{q_t} - \frac{1}{q_0}$$

We can apply this to determine how long it would take for a recessive lethal allele to drop from a frequency of $f(q_0) = 0.7$ to $f(q_t) = 0.2$:

$$t = \frac{1}{0.2} - \frac{1}{0.7} = 3.57 \text{ generations}$$

Mutation as an evolutionary force

The next violation of HWP assumptions we will examine is mutation. Mutations are a sudden genetic change in an allele or chromosome. They can be caused by the impacts of oxygen, radiation, chemicals, or cellular processes. DNA repair fixes many mutations but is imperfect. Importantly, mutations are the origin of all genetic diversity.

Common types of mutations include:

- Base substitutions
- Additions or deletions in DNA sequences
- Gene duplications
- Insertion of mobile DNA elements (**transposons**)

While we often think of mutations as inherently negative, they can have a variety of fitness effects:

- **silent substitutions**: a mutation in a coding region that does not result in an amino acid substitution in a protein
- **neutral mutations**: mutations with little to no impact on fitness
- **deleterious mutations**: mutations that negatively impact fitness
- **beneficial mutations**: mutations that positively impact fitness

Importantly, some mutations may be beneficial in some circumstances but not others!

A key parameter for a given mutation type is its rate, typically denoted by u or μ ("myu"). Mutation rates describe the probability of mutation per unit per generation. Different genetic markers have different rates. As a set of benchmarks:

- 1 x 10^{-8} to 1 x 10^{-9} per base per year: DNA nucleotides
- mammalian mtDNA typically has 5 - 10x the nuclear DNA mutation rate
- 1 x 10^{-4} per locus: mammalian microsatellites
- 1 x 10^{-5} per locus: morphological mutations in fruit flies

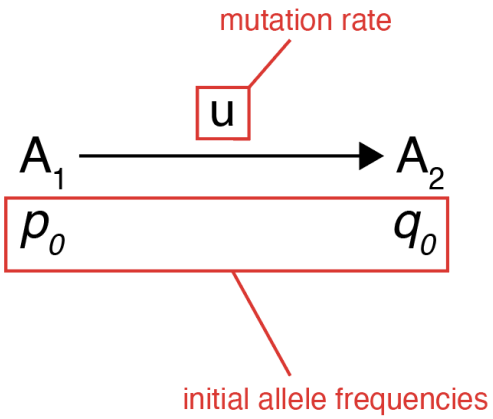
For example, a per nucleotide rate of 10^{-5} (equivalent to $\frac{1}{100000}$) leads us to expect a single mutation in a stretch of 100,000 bases each generation (because the independent probabilities of mutation at each base are summed). If instead our mutation rate is per *gene* and we know there are roughly 20,000 genes in the human genome, we would expect $10^{-5} * 20,000 = 0.2$ mutations per individual (so, roughly every 5 individuals we'll find one new mutation in a coding region).

The probability of an individual NOT having a mutation in one of their 20,000 genes is a little more complicated. To determine it, we need to account for both the probability of no mutation in their maternally-inherited chromosomes and the probability of no mutation in their paternally inherited chromosomes. Each haploid genome requires multiplying the independent probability of NO mutation ($= 1 - \mu$) by itself 20,000 times, as these independently probabilities must occur concurrently. We then must multiply the probabilities of no mutation in each haploid genome by each other, as these two events must also occur together:

$$p(\text{no mutation}) = (1-\mu)^{20000} + (1-\mu)^{20000} = (1-\mu)^{40000} = (1 - \frac{1}{100000})^{40000} \sim 0.67$$

Conversely, to find the probability of at least one new mutation in a coding region in a human's diploid genome, we can either take the complement of no mutation (i.e., $1 - 0.67 = 0.23$).

But what does this mean for allele frequencies? We start with a simple model where allele A_1 (occuring at frequency p_0) mutates to become A_2 (occuring at frequency q_0) at rate u / μ. (For now, we assume mutation is only one-dimensional.)



The frequency of A_1 in the next generation (p_1) will be its initial frequency multiplied by the complement of the mutation rate, i.e. the expected proportion of alleles that do NOT mutate:

$$p_1 = p_0(1-\mu)$$

We can use this formula in the following example:

Q: What is the frequency of allele A_1 in the next generation if its mutation rate u is 2.5 x 10^{-5} and it occurs at an initial frequency of 0.7?

A:

$$p_1 = p_0(1-u);$$
$$p_1 = 0.7(1-0.000025);$$

$$p_1 = 0.6999825$$

This shows us that among other things, mutation is a *slow* way to change allele frequencies and restore genetic diversity!

This same model helps us determine the change in A_1 from generation to generation: it is the difference between p_1 and p_0, which simplifies to the negative product of the mutation rate and initial frequency of A_1:

$$\Delta p = p_1 - p_0 = p_0(1-\mu) - p_0 = p_0 - \mu p_0 - p_0 = -\mu p_0$$

We can also use this relationship to build a model of how mutation changes allele frequencies t generations in the future. If $p1 = p_0(1-\mu)$, then $p2 = p_1(1-\mu) = p_0(1-\mu) * p_0(1-\mu) = p_0(1-\mu)^2$. Therefore, $p_t = p_0(1-\mu)^t$, which means that $p_t \sim p_0 e^{-ut}$. (This equation is an approximation of exponential decay.)

An immediate application of our model is to address the question of how long it takes for mutation to restore lost genetic diversity. We can do this by isolating t in the equation above:

$$p_t \sim p_0 e^{-ut}$$
$$\frac{p_t}{p_0} = e^{-ut}$$
$$ln(\frac{p_t}{p_0}) = -ut$$
$$\frac{ln(\frac{p_t}{p_0})}{-u} = \frac{ln(p_t) - ln(p_0)}{-u} = \frac{-ln(p_0) - ln(p_t)}{u} = t$$

We can apply this equation if (for example) we have a mutation rate of 10^{-6} and want to know how long it takes for p to go from a frequency of 1 to 0.9:

$$\frac{-ln(1) - ln(0.9)}{10^{-6}} = t = 105,360 \text{ generations}$$

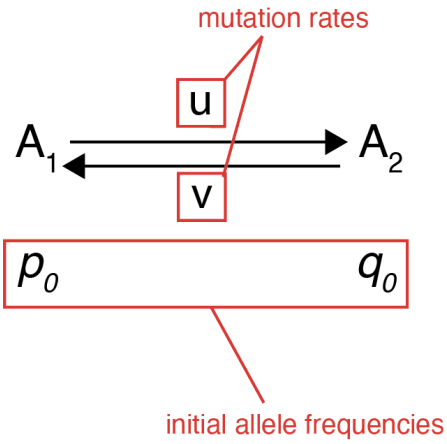
But what happens if we have mutation in two directions? In this model, allele A_1 (occuring at frequency p_0) still mutates to become A_2 (occuring at frequency q_0) at rate u, but we also permit A_2 to mutate back to A_1 at rate v:

In this scenario, we would expect the equilibrium frequency of each allele to be determined by the relative strength of mutation in each direction:

$$\hat{p} = \frac{v}{u+v};\ \hat{q} = \frac{u}{u+v}$$

For example: what is the equilibrium frequency of A_2 if $u = 10^{-5}$ and $v = 3 * 10^{-6}$?

$$\hat{q} = \frac{10^{-5}}{10^{-5} + (3 * 10^{-6})} = 0.77$$


mutation rates
u
A1
A2
v
p0
q0
initial allele frequencies