Problem Set #6

Solutions

1. (1 mark) Migration among four subpopulations occurs according to the island model at a rate $m = 0.05$. If the initial allele frequencies in the four populations are 0.2, 0.4, 0.6, and 0.8, what are the expected frequencies after 10 generations assuming no other evolutionary processes are operating? What will the equilibrium allele frequencies be in each population?

p_bar = (0.2 + 0.4 + 0.6 + 0.8)/4 = 0.5, which is also the equilibrium allele frequency in each population.

After 10 generation, use $p_t = \overline{p} + (1 - m)^t (p_o - \overline{p})$

to solve for p_t in each population to get 0.320, 0.440, 0.560, and 0.680 in the four populations respectively.

For example, in the first population in which p_0 = 0.2:

$$p_t = \overline{p} + (1 - m)^t (p_o - \overline{p})$$

$$p_t = 0.5 + (1 - 0.05)^{10} (0.2 - 0.5)$$

$$p_t = 0.320$$

2. (2 marks) Imagine a population of size N_e = 2000 that receives and average of two migrants per generation. By what proportion is heterozygosity reduced in this population on average at equilibrium between drift and migration? How does this compare to the proportionate reduction in heterozygosity at mutation-drift equilibrium if $\mu = 1 \times 10^{-9}$? In one sentence, explain why these results are so different.

migration-drift

$$\hat{F} \approx \frac{1}{1+4N_e m}$$

$$\hat{F} \approx \frac{1}{1+4(2)} = \frac{1}{9} = 0.111$$

mutation-drift

$$\hat{F} \approx \frac{1}{1+4N_e \mu}$$

$$\hat{F} \approx \frac{1}{1+4(2000)(1\times 10^{-9})} = \frac{1}{1+8\times 10^{-6}} = 0.9999$$

These differ so dramatically because migration is introducing new alleles that are not IBD to any alleles in the population at a much higher rate than is mutation, thereby maintaining a much higher heterozygosity.

3. (2 marks) The human immunodeficiency virus HIV causes AIDS. In this virus, the rate of nucleotide evolution has been estimated at approximately 0.01 substitutions per synonymous site per year. Two viruses isolated in Zaire and San Francisco in 1983 differ in 1/3 of their synonymous sites. Estimate the YEAR in which these viruses last shared a common ancestor.

$$\hat{K} = -\frac{3}{4}\ln(1 - \frac{4}{3}\hat{p}) = -\frac{3}{4}\ln\left(1 - \frac{4}{3}\left(\frac{1}{3}\right)\right) = 0.4408$$

$$t = \frac{\hat{K}}{2\hat{k}} = \frac{0.4408}{2(0.01)} = 22 \text{ years}$$

Therefore, these viruses last shared a common ancestor in 1983 – 22 years = 1961.

4. (1 mark) Assume that the amino acid sequence of a gene in two species differs at 60 of 200 sites. Calculate the expected number of substitutions per site over this time period, and the rate of amino acid replacement per site per unit time if they last shared a common ancestor 3×10^8 years ago.

expected # of substitutions:

$$p_t = 60/200 = 0.30$$

$$\begin{aligned}\hat{K} &= -\ln(1-p_t) \\ &= -\ln(1-0.30) \\ &= 0.357\end{aligned}$$

rate of amino acid replacement per site per unit time:

$$K = 2kt$$

$$k = \frac{K}{2t} = \frac{0.357}{2(3\times 10^8)}$$

$$\approx 5.95\times 10^{-10}$$

5. (2 marks) Synonymous sites in the Histone H3 gene have been found to evolve on average at a rate of 5 $\times 10^{-9}$ per year. If two species were found to differ at ¾ of the synonymous sites for this gene, what would you conclude about the date since they last shared a common ancestor and why? What strategy might you suggest to help date their divergence?

$$\hat{K} = -\frac{3}{4}\ln(1-\frac{4}{3}\hat{p}) = -\frac{3}{4}\ln\left(1-\frac{4}{3}\left(\frac{3}{4}\right)\right) = -\frac{3}{4}\ln\left(0\right) = undefined$$

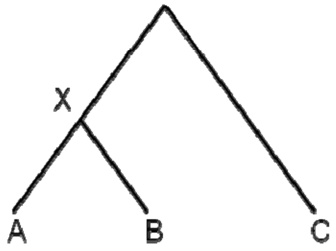
This answer is not biologically informative (indeed, we can’t get an answer!). It is caused by the fact that these two nucleotide sequences are saturated with mutations and are no longer any more similar than would be expected by chance alone. In other words, so many substitutions have occurred that any evidence of their shared evolutionary history has been erased.

To better date their divergence, sites with a lower substitution rates could be analyzed (e.g., non-synonymous sites at this gene evolve at approximately 1/100 the rate – see figure from lecture). Such sites will accumulate substitutions more slowly and therefore may not yet be saturated.

6. (2 marks) Given the phylogeny indicated below:

a) Explain, in four sentences or less, the basic concept behind a relative rate test for consistency in nucleotide substitution rates in the lineages leading from X to A and from X to B.

b) Perform this test using the sequence data below. With reference to the chi-square table on the course website (in the ‘Problem sets and midterm’ folder), do you accept or reject the null hypothesis that of constant substitution rates in these two lineages?



A	ATG	CTA	GCA	TGC	ATG	CTA	GCA	ATA	ATA	GCT	CTA	ATC
B	TTG	CTA	GCA	TGG	ATG	GTA	GTA	TAC	ATT	GCA	CTT	ATG
C	ATG	CTA	TCA	TGC	TTG	GTA	GCA	ATC	ATA	GCA	CTA	ATG

A	ATG	CTA	GCA	TGC	ATG	CTA	GCA	ATA	ATA	GC**T**	CTA	ATC
B	**T**TG	CTA	GCA	TG**G**	ATG	GTA	GTA	**TA**C	AT**T**	GCA	CT**T**	ATG
C	ATG	CTA	TCA	TGC	TTG	GTA	GCA	ATC	ATA	GCA	CTA	ATG

a) If A and B have had the same substitution rate since their split from X, the expected # of substitutions along the branch from X to A should be the same as along the branch from X to B. The relative rate test infers these substitutions by comparison with an outgroup (C) and uses a chi-square test to determine whether the # along each branch are consistent with the null hypothesis of rate constancy.

b)

B=C≠A is 4 A=C≠B is 7

Expected # of sites differing = (4 + 7)/2 = 5.5

Chi-square (1 d.f.) = $(4 - 5.5)^2/5.5 + (7 - 5.5)^2/5.5 = 0.818$

This value is less than the critical value, so the conclusion is that you accept the null hypothesis of rate constancy.