

FINAL EXAMINATION

Write your answers and show your work only in the space provided below each question. When there is a line labelled "Answer", write the answer on the line and show calculations above the line. I will not grade anything written anywhere else, and I will not grade anything that is written illegibly. When calculations are necessary to get an answer, you must show the calculations to get credit. All quantitative answers must have units (e.g. 27 substitutions/site X year). Partial credit will be given when appropriate. Note that sometimes we can't answer a scientific question with a number, but can say something is "greater than" or "equal to" or "less than" something else.

NOTE: I usually took one point off for errors in arithmetic or algebra, or incorrect units.

1. (15) The *lgt* gene in dragons codes for an enzyme required for the synthesis of lighter fluid. Dragons are diploid. Suppose that you cloned both copies of this gene from all of the 50 dragons still alive (they are an endangered species) and sequenced a 20-bp segment of each gene. You found three different sequences, as follows:

sequence type	sequence	number of genes
1	C A T T G C A A T G A T G A C G G G C A	20
2	G A T T G C A A T G T T C A C G G G C A	40
3	C A T T G C A A C G A T C A C G G C C A	40

(a) The base pair polymorphism at the eleventh site is nonsynonymous and can be detected by electrophoresis, with T giving a slower-migrating protein than A. Calculate the expected heterozygosity if this difference is selectively neutral.

$$f(1,3) = (20+40)/100 = 0.6 = p \quad f(2) = 0.4 = q \quad H = 2pq \text{ or } H = 1 - (p^2 + q^2) = 0.64$$

(b) Assuming that $N_e = N$ and that the differences between the *lgt* alleles are selectively neutral, use your answer to estimate the mutation rate for electrophoretic variants; be sure to give the units of your answer.

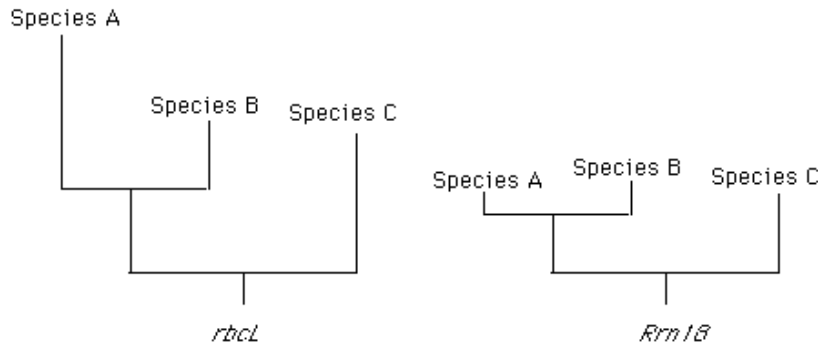
$$H = 4N_e u \quad u = 0.48/(4 \times 50) = 0.0024 \text{ mutations per gene per generation or mutations/gene X generation}$$

$$\text{If you used } H = 4N_e u / (1 + 4N_e u), \text{ then } u = 0.0067 \text{ mutations/gene X generation}$$

(c) What would you predict about the expected heterozygosity of this region if a new mutation that is very strongly advantageous should occur in position 10? Momentarily increased by the mutation, then decreased due to hitchhiking; eventually coming back up.

2. (22) Below are scaled phylogenetic trees of the *rbcL* and *Rrn18* genes of three species of plants. The *rbcL* gene resides in the plastid and codes for a major protein required

for photosynthesis. The *Rrn18* gene is the nuclear gene that codes for the 18S rRNA. Species B and C are photosynthetic; species A is not. The *rbcL* gene of species A has a frameshift mutation in the fourth codon; the *rbcL* genes of species B and C have no frameshift mutations. There are two striking features of these trees: (i) the branch leading to species A is much longer than the other branches in the *rbcL* tree; and (ii) all the branches are shorter in the *Rrn18* tree.



(a) Circle the coalescent of species B and C in the *Rrn18* tree. Circle should be around the root node.

(b) Could the accelerated evolution in the *rbcL* gene of A be due, partly or entirely, to a change in mutation rate? random drift? selection? Explain each answer *briefly*.

mutation Yes, mutation rate could have increased in the lineage leading to A.

random drift No, if N_e decreased in the lineage leading to A so that more detrimental mutations were fixed, then the branch length to A should also be increased in the *Rrn18* tree.

selection Yes, *rbcL* lost its function in the nonphotosynthetic species and is a pseudogene so its evolutionary rate is expected to accelerate.

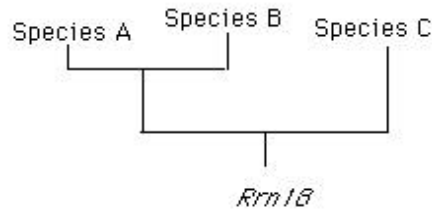
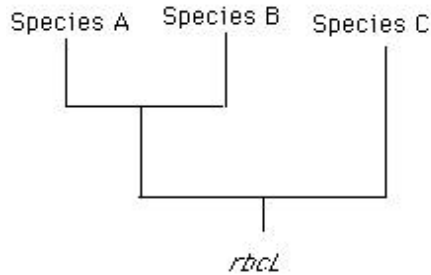
(c) Could the lower rate of evolution of the *Rrn18* gene, relative to the *rbcL* gene, be due, partly or entirely, to a change in mutation rate? random drift? selection? Explain each answer *briefly*.

mutation Yes, the mutation rate could be lower in the nuclear genome.

random drift Yes, the effective population size is larger for the nuclear genome because the plastid genome is inherited uniparentally and is effectively haploid; consequently selection would be stronger on nuclear genes.

selection Yes, *rbcL* and *Rrn18* may be subject to different selective pressures.

(d) Suppose that you made a new tree based only on the third codon position of *rbcL* and found the tree below, in which the species A branch is much shorter and all the rest of the branches are slightly shorter. Would this change your answer to (b), and if so, how and why?



I deleted this question because it should have said that the species A branch is the same length and all the rest of the branches are longer. Then the answer would have been that the accelerated evolution of A as seen in the total sites must be mainly or entirely due to mutation, because the synonymous substitution rate is approximately equal to the mutation rate and also to the pseudogene rate.

3. (15) The following two sequences were obtained from two species of dragons found on Procyon IV.

green dragon AATGCTTGCTGGG-GCGCAATTA
blue dragon ATTGCTG-CTGAGTGCAGTT-

(a) Calculate the sequence difference (to four decimal places).

$d = 4 \text{ differences} / 20 \text{ sites present in both species} =$
Answer: 0.2000

(b) Calculate the sequence divergence (to four decimal places).

$K = -(3/4)\ln[1 - (4/3)d] =$
Answer: 0.2326

(c) The rate of sequence evolution in these dragons is 10^{-8} substitutions per bp per year. Approximately how long ago was the most recent common ancestor of these two species?

$E = K/2T \quad T = K/2E = 0.2326/2 \times 10^{-8} =$
Answer: 1.16×10^7 years

4. (5) Using one or more runs of the driftworm simulations, how could you simulate the difference between evolution in a gene and in a pseudogene? (Answer in one sentence.)

Do one run with mutation and selection to simulate the gene and one with mutation but without selection to simulate the pseudogene.

5. (10) *Salmonella* is a bacterial human pathogen of great medical importance. *Salmonella* cells may be sensitive to the antibiotic chloramphenicol (genotype *cap-s*) or resistant (*cap-r*). If one isolates cells from many different sources that have not been exposed to chloramphenicol, most of them will be *cap-s*., even though such cells mutate to *cap-r*

with easily detectable frequency. But since the widespread use of chloramphenicol began, patients with *Salmonella* infections are increasingly found to be infected with *cap-r* cells. Based on these observations, if the fitness of *cap-s* is 1, will the selection coefficient s of the *cap-r* mutants be > 0 , 0 , or < 0 :

(a) in the presence of chloramphenicol? > 0

(b) in the absence of chloramphenicol? < 0

6. (10) Species become endangered when their populations drop to very low values. Suppose the effective population size of a diploid organism decreases from 10^6 to 10^3 and stays at the new low value long enough to reach equilibrium between mutation and drift.

(a) What happens to neutral genetic diversity, as measured by expected heterozygosity or nucleotide diversity, in this species? Use an appropriate equation to give a quantitative (although approximate) answer.

Using subscripts L and S for large and small populations:

$$H_L/H_S = 4N_Lu/4N_Su = 10^6/10^3 = 10^3 \quad \text{Diversity decreases 1,000-fold.}$$

(b) What happens to the effectiveness of natural selection in this species, i.e. to the extent to which selection can eliminate detrimental mutations and fix advantageous mutations? Explain your answer.

The effectiveness of selection decreases because the smaller population has more drift.

7. (15) If K_N is the rate of nonsynonymous substitution in a gene, and K_S is the rate of synonymous substitution, the ratio K_N/K_S be > 1 , 1 , or < 1 :

(a) in a pseudogene? 1 (no selection)

(b) in a functional gene? < 1 (most selected mutations are detrimental)

(c) at a site where an advantageous mutation has recently been fixed? > 1