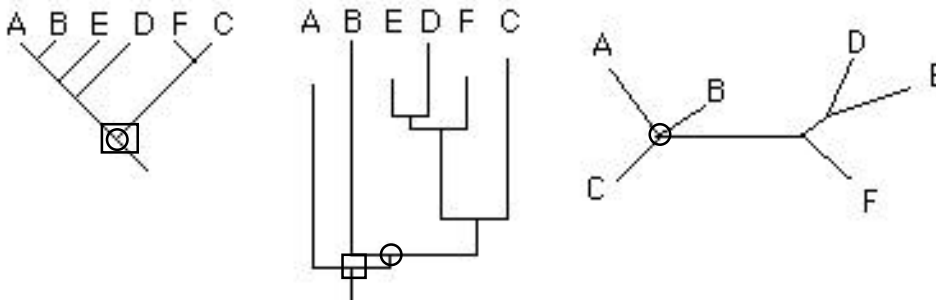


MCB/GENE/ECOL/BIOC/INSC 545
FALL 2002
FINAL EXAMINATION

- PRINT YOUR NAME LEGIBLY AT THE TOP OF EACH PAGE AND SIGN IT ON THE FIRST PAGE.
- 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).
- Approximate point values for each question are given in parentheses; these are subject to change. Partial credit will be given when appropriate.

1. (10) In each of the phylogenetic trees below in which it is possible to do so, (a) circle the coalescent (the most recent common ancestor) of genes B and C, and (b) draw a square around the root node (coalescent of all taxa).

2 points each circle or square. -1 if a circle or square is on a branch instead of a node as specified in the question.



2. (10) Evolutionary models. 5 points each part

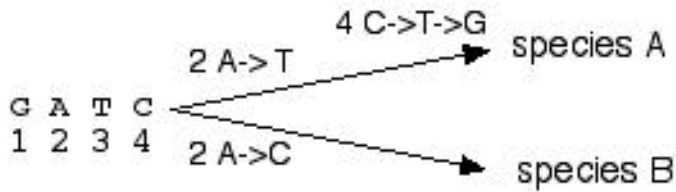
(a) The Jukes-Cantor model has only one parameter: what is it? (A couple of words or a phrase, please, not just a symbol.)

Answer substitution rate

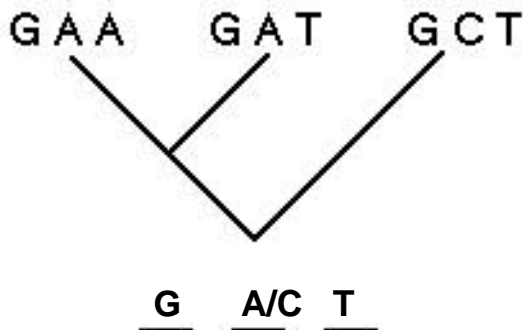
(b) Name one additional parameter or kind of parameter that is found in more complex models.

Answer transition/transversion ratio, or 6 different mutation rates, or base frequencies, or proportion of invariant sites, etc., etc.

3. (10) On the diagram below, indicate two different kinds of multiple hits. Indicate the number of the site at which each occurs.
5 points each for any of several different kinds of multiple hits; two examples are:



4. (9) Below is the sequence of the second codon in a gene in three organisms, and the cladogram describing their relationships. Write the most parsimonious reconstruction of the ancestral sequence at the base of the tree. If a site is ambiguous, write both possibilities separated by /. 3 points each site

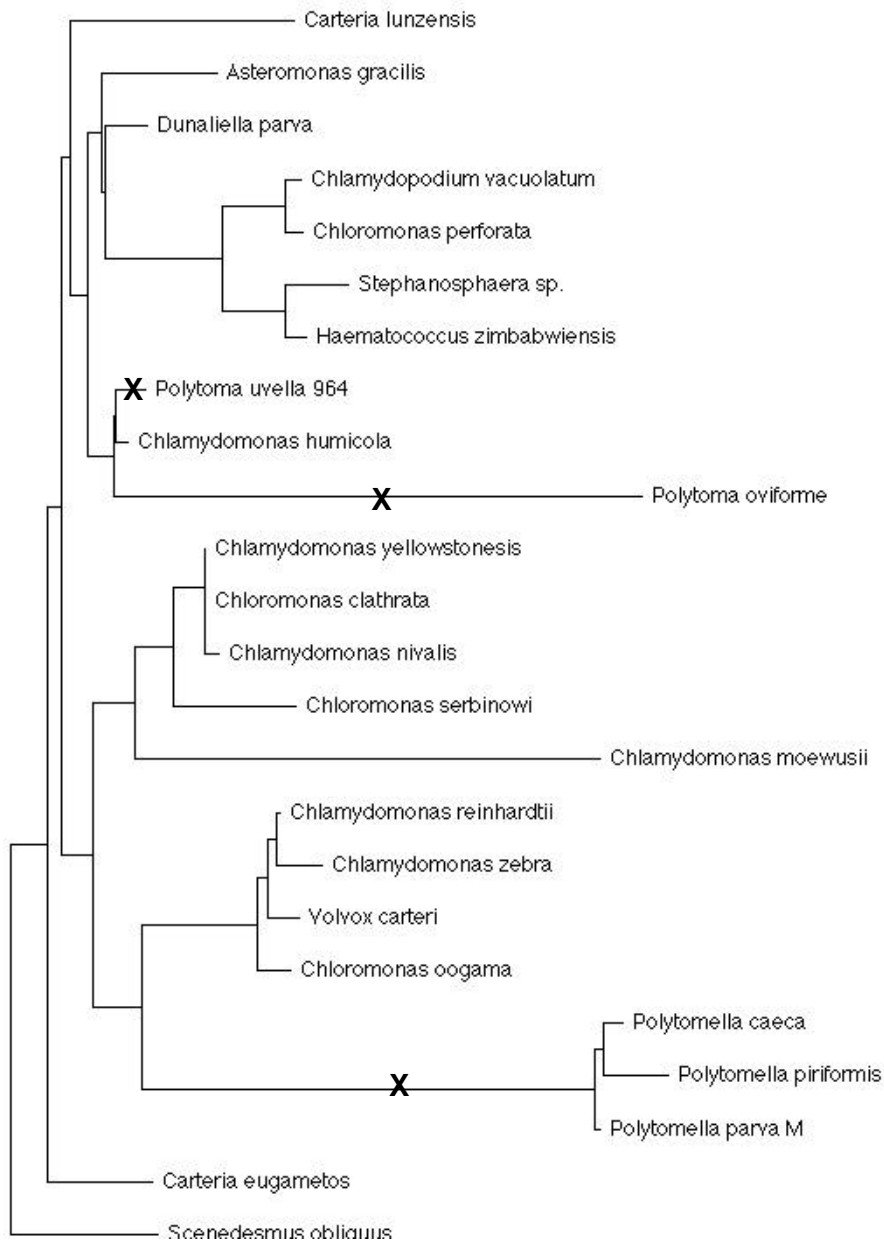


5. (16) The tree on the next page was made with *Rrn18* sequences by neighbor-joining with a maximum likelihood evolutionary model. *Scenedesmus* is the outgroup.

- (a) 6 pts Is the genus *Chlamydomonas* a clade? Circle your answer: yes no
- (b) 5 pts *Polytoma* and *Polytomella* are non-photosynthetic algae. How many times was photosynthesis lost in this group of algae? three Put an X on each branch where photosynthesis was lost.
- (c) 5 pts Is the rate of evolution of this gene constant in this group of algae? yes no

Explain your answer in one sentence.

The terminal nodes are not all the same distance from the basal node, even though they are all extant taxa.



6. (5) Phillippe calculated the nucleotide diversity of the protist *Paramecium aurelia* as = 1.14. How can you tell that his calculation is wrong? (One sentence, please.)

Nucleotide diversity is a probability, and probabilities cannot be greater than 1.

7. (20) George sequenced the alpha amylase gene from a human and a robin. He found that there are 31 synonymous differences and 4 nonsynonymous differences between the genes of the two species. For each of the following, briefly explain why it can, or cannot, explain the larger number of synonymous differences: **4 points each. Many people somehow thought the question was about differences in synonymous or nonsynonymous sites between humans and robins, even though no such differences were given in the question. The question is about the difference between the numbers of synonymous and nonsynonymous differences.**

(a) differences in mutation rate between synonymous and nonsynonymous sites

No, on average mutation rates are the same for all sites in a gene.

(b) differences in population size

No, the population size is the same for synonymous and nonsynonymous sites in the same organism.

(c) differences in effective population size

No, the effective population size is the same for synonymous and nonsynonymous sites in the same organism.

(d) differences in directional selection

Yes; synonymous mutations are subject to weak or no selection, while many nonsynonymous substitutions are subject to selection, and most of those are detrimental.

(e) differences in balancing selection (I didn't discuss this, but let's see what you can do with your intuition)

I gave full or partial credit for almost any sensible answer. A good answer would be that balancing selection acts to maintain nonsynonymous differences within a population and would result in more nonsynonymous differences than synonymous differences, the opposite of what is specified in the question. Another good answer would be that balancing selection generally acts on only one site in a gene, if any, and only affects a few neighboring linked sites.

8. (20) Consider a population of plants in which there are two alleles of *mfg* (*my favorite gene*); *mfg1* and *mfg2* have frequencies 0.9 and 0.1 respectively. **4 points each part; -1 point for arithmetic errors**

(a) What is the expected heterozygosity of this gene?

$$H = 2 \times 0.9 \times 0.1 = 0.18$$

(b) What would be the observed heterozygosity if the population is random mating?

Observed frequency of heterozygotes = $2 \times 0.9 \times 0.1 = 0.18$ (the Hardy-Weinberg equation)

(c) What can you say about the observed heterozygosity if the population is inbred?

< 0.18

(d) What would you have predicted the expected heterozygosity would be if all mutations are selectively neutral, the effective population size is 10^5 , and the mutation rate is 10^{-7} ?

$$H \approx \frac{4N_e u}{1 + 4N_e u} = 4 \times 10^5 \times 10^{-7} / (1 + 4 \times 10^5 \times 10^{-7}) = 0.038$$

$$\text{or } H \approx 4N_e u = 4 \times 10^5 \times 10^{-7} = 0.04$$