

1-5: Multiple choice. Choose the single best answer. (2 pts. each)

- 1. In a 1978 paper, "Cases in which parsimony or compatibility methods will be positively misleading", Felsenstein described analytical conditions under which maximum parsimony would find the incorrect tree with increasingly strong support as the number of characters increased, a phenomenon called "inconsistency". Which of the following types of characters would you expect to be *least* prone to inconsistency in parsimony analysis?**
 - a. Morphological characters with two character states
 - b. DNA sequence data coded with four character states
 - c. DNA ambiguity codes with a total of eight character states.
 - d. Amino acid data coded with 20 character states

- 2. Which of the following adequately defines *homologous* characters?**
 - a. Characters are homologous if they are more similar to each other than to any other identifiable character.
 - b. Characters are homologous if they are modified versions of the same character in a common ancestor.
 - c. Characters are homologous if it can be shown that there are no intermediate stages or forms found in nature.
 - d. Characters are homologous if they were independently acquired via organic evolution under similar conditions of natural selection.

- 3. Zuckerkandl and Pauling distinguish between semantic, episemantic, and asemantic molecules. What is the critical distinction between these three classes of molecules?**
 - a. The degree to which they are able to convey information about the evolutionary history of the organism.
 - b. The degree to which they are involved in the reproduction and fitness of the organism.
 - c. The degree to which they are subject to natural selection.
 - d. None of the above.

- 4. Which of the following is *not* an assumption shared by maximum parsimony, distance, and maximum likelihood methods?**
 - a. Descendants inherit their character-states from their ancestors.
 - b. Characters cannot assume character-states that they hold elsewhere on the tree.
 - c. Character-state changes occur independently of each other.
 - d. The evolutionary process can be represented by a dichotomously branching tree.

5. What is the distinction between an algorithmic and an optimality method?

- a. Optimality methods do not use algorithms.
- b. Algorithmic methods do not optimize an objective function.
- c. Optimality methods are faster than algorithmic methods.
- d. Algorithmic methods are more consistent than optimality methods.

6-10: Fill in the Blank (2 pts. each)

- 6. On a phylogenetic tree, a _____ defines a taxon bipartition.
- 7. In cladistic (parsimony) terminology, a character that shows a unique character state in one taxon, and has another character state in all other taxa is considered to be _____, and is considered to be one type of uninformative character.
- 8. An _____ is a defined series of steps that can be used to solve an analytical problem.
- 9. _____ algorithms are intended to find good solutions to a problem, but are not guaranteed to find the best possible solution.
- 10. The _____ uses synthetic oligonucleotide primers, a thermostable DNA polymerase, and temperature cycling to generate usable quantities of a selected region of the genome from small quantities of genomic template.

11. Discuss why the use of models of sequence evolution can improve phylogenetic analyses of DNA data. (5 points)

12. Describe the following models of sequence evolution: Jukes-Cantor; Kimura Two Parameter; Felsenstein 1984 (or Hasegawa, Kishino, and Yano 1985); General Time Reversible. Your discussion should both identify the fundamental shared features of these models and describe the differences among them. Use the fact that these are nested models to save words! (10 points)

- 13. What is branch swapping? Diagram Nearest Neighbor Interchange (NNI), Subtree Pruning and Regrafting (SPR), and Tree Bisection and Reconnection (TBR). (15 points).**

14. According to the article "An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis", by D.M. Hillis and J.J. Bull [Syst. Biol. 42:182-192 (1993)], which refers to nonparametric bootstrapping, what is the relationship between bootstrap proportions and the probability that the corresponding clade is correct? (5 pts)

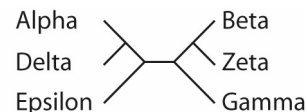
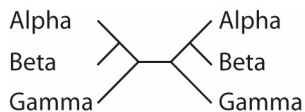
- If you reach this point you have written too much -

- 15. Imagine that you are sitting on an NSF panel considering dissertation improvement proposals. Summaries of several proposals are given below. For each of these, indicate whether or not you feel that the student has an adequate understanding of phylogenetic methods to justify funding their proposal, and indicate why you are of this opinion. (5 pts each)**
- a. Gary has determined 300 base pairs of the mitochondrial gene *coxII* from 30 male and 30 female parasitic flies with flightless females. He would like to use these data to study whether or not these flies interbreed with a second population of flies that lives nearby, and would like to have funding to collect *coxII* data from this second population.
- b. Jimin has 300 DNA sequences of the genes encoding EF-Tu and SSU rRNA from representative bacteria, and would like to use these data to test whether the protein-coding sequences give the same phylogeny for bacteria as do the rRNA sequences. She proposes to analyze these data using GTR + I + Gamma distances and minimum evolution. She argues that maximum likelihood would be too computationally intensive for this project, and that parsimony would yield unreliable results.

- c. Scott has actin sequences from 30 species of endomycorrhizal fungus, and is requesting funding to sequence tubulin from the orchid hosts for these fungi. He would like to test whether or not the fungi have swapped hosts during the course of evolution. He intends to concatenate the sequences into a single dataset and use p-distances and UPGMA to analyze these data.
- d. Marsha has 8000 sequences of the gene *atpB* from tracheophytes, and proposes to use these sequences to perform a maximum likelihood analysis with the Jukes-Cantor model using an exhaustive search because it is guaranteed to find the shortest possible tree.

15a. Using unweighted maximum parsimony, calculate the length of the trees shown, given the alignment below. Use the algorithm we discussed in class, and make your calculations clear (draw your own trees – don't try to fit all your calculations on these small diagrams!). Treat all character-state transformations as equally likely and reversible. (15 pts)

Alpha	ATGGC	GGGAA	AAAGT
Beta	ATGTC	AAGAA	ACTCA
Gamma	ATGTC	AAGAA	ACTCA
Delta	ATGGC	GGGGC	GAGAT
Epsilon	ATGGC	GGGGC	GAGGT
Zeta	ATGGC	TGGGA	ACGGA



15b. Which of the trees is favored (better) according to parsimony? (5 pts)

15c. How many of the characters in the matrix above are considered *informative* according to parsimony? (5 pts)

-end-