

Comprehensive Exam  
Biological Sequence Analysis  
Sample Questions

1. (20 points) Given the distance matrix below, define the neighbor relations one would use for quartet puzzling. To get full credit show all the quartets and the neighbor relation that comes from each.

	A	B	C	D	E
A	0	2	10	10	10
B	2	0	10	10	10
C	10	10	0	2	6
D	10	10	2	0	6
E	10	10	6	6	0

2. (20 points) Consider the following neighbor relations:  
{ $N(ac;db), N(ae;db), N(ce;db), N(ac;eb), N(ac;ed)$  }  
Use quartet puzzling to determine the correct tree topology. (Show all steps to get full credit.)

3. When performing dynamic programming to determine global sequence alignment, the computational time increases.
  - a. (10 points) If we divided both sequences in half and compared first halves and second halves, would this affect computation time? If so, how?
  - b. (10 points) Would there be any potential problems with the final alignment results if we did this? If so, how might we fix these problems?

4. The following two questions concern hidden Markov models
  - a. (10 points) Describe what the Viterbi algorithm and the posterior probabilities are and what they tell us.
  - b. (10 points) If we have an idea for the topology of a hidden Markov model, and we have a sequence of emissions, what procedure would we use to determine the emission and transition probabilities?

5. Suppose that we do find life on Mars and it has DNA similar to our DNA. Also assume that this Martian DNA uses 3-base codons to code for amino acids that differ from the ones used on Earth. Also suppose we find the DNA of many organisms and want to create a phylogeny.
  - a. (10 points) How would you go about constructing a substitution matrix to use for sequence comparison?
  - b. (10 points) How would you prove or verify (the method not the proof) that your method was valid?