

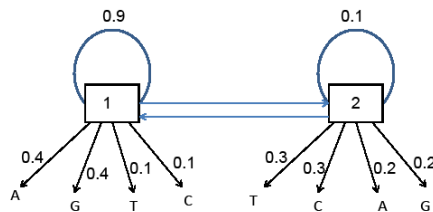
CAP5510 Introduction to Bioinformatics

Fall 2009

Homework #4

(Assigned Nov.4 Due: Nov. 16)

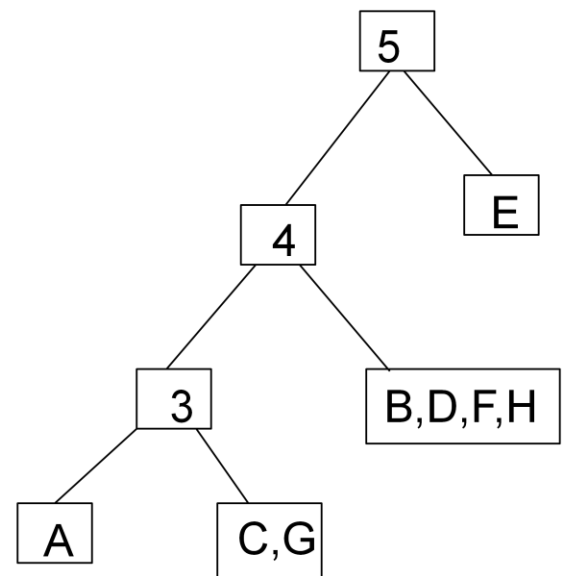
1. Show that every binary tree with n leaves has $2n-3$ edges.[10 points]
2. Consider a two-state (1, 2) Hidden Markov Model as shown below. When in state 1, it is more likely to emit purines A and G. When in state 2 it is more likely to emit pyrimidines C and T. Decode the most likely sequence of states (1 or 2) for sequence GGCT. Use log-scores rather than product of probability scores. [25 points]



3. While we showed in the class how to solve the Sankoff's Weighted Small Parsimony problem using dynamic programming, we did not show how to construct the equivalent of "edit graph" for this problem. Cast the Weighted Small Parsimony problem in terms of finding a path in an appropriate directed acyclic graph and write the recurrence relations with respect to this graph. Design a backtracking procedure to reconstruct the optimal assignments of characters in the Sankoff algorithm for the Weighted Small Parsimony problem. Illustrate the procedure with the example on p.373 for the binary tree and δ given in Fig.10.18. [30 points]

4. Evolutionary trees have been constructed based on clustering approach, distance based approach, and character based approach and phylogenetic tree approach. Discuss their advantages and disadvantages and discuss what kind of data are used in these methods and how such data can be obtained. [10 points]
5. (a) Fill in the table below so that the matrix is ultrametric and obtain the ultrametric tree. [15 points]

	A	B	C	D	E	F	G	H
A	0	4	3	4	5	4	3	4
B								
C								
D								
E								
F								
G								
H								



5 (b) Define phylogenetic tree. State a necessary and sufficient condition for a $n \times m$ binary matrix to have a phylogenetic tree.

Can you construct a phylogenetic tree for the following matrix? (10 points)

	1	2	3	4	5
A	1	1	0	0	0
B	0	0	1	0	1
C	1	1	0	0	1
D	0	0	1	1	0
E	0	1	0	0	1

