

Consider the following distance matrix defined for 7 species A, B, C, D, E, F, G

- 1. Given the distance matrix provided, construct a tree, T, as follows:
- a) Begin with an empty edge list, E
- b) Select the minimum-distance edge, e, not in E, such that the graph, T, formed by the edges E U {e} has no cycles. If no such edge exists, return T
- c) Set E = E U {e}
- d) Goto step b

Draw the resulting tree, T.

2. Which of the seven algorithm-design techniques mentioned in Chapter 2 of the textbook does this tree-construction algorithm best illustrate? Explain your answer.

3. Given a graph with N vertices, how many times will step c in question 1 be executed?

4. A minimum spanning tree of a graph connects all of the graph's vertices with a minimal sum of the edge distances. Does the algorithm from question 1 achieve this objective? Justify your answer.

5. Does the algorithm given question 1 solve the Hamiltonian path problem (a minimum distance path through all vertices)? Explain your answer.

- 6. An alternative algorithm for computing a minimum spanning tree from a fully connected graph is: Remove the most expensive edge on each cycle, until no cycle remains. In each iteration, this algorithm removes one edge. How many iterations will be executed in a graph with N vertices?
- 7. What is the suffix array for the following sequence? TAGAGAGACAT.

8. Draw the suffix tree.

9. Create the profile matrix for the following 5 strings.

1234567 5, GCACGCT 52GGAGGGT 53TCAGAGT 54TCGGACT 55TCAGGCA

10. Create the consensus strand for the 5 strings above.

11. What is the p most probable 7-mer using the profile GCATCAGACT?

mine five out	eotide sequences
OW-9	
1234507	
GCACGET	
GGAGGGT	
TCAGAGT	
TCCCACT	
Tractica	
TORDER	
	owing five nock 1234567 GGAGGGT GGAGGGT TCAGAGT TCAGGCA TCAGGCA

12 Use 0 to encode the majority allele and 1 to encode the minority allele at every position. Write down the SNP matrix for the given sequences.

-	1.	12	1			6	T.
1			-				
1		-	-		-		
31	_	_	-	-	-		
31		_			_		
5	_			1			
5							

 Does this matrix admit a perfect phylogery tree? If your answer is YES, draw the perfect phylogeny tree, otherwise, show the matrix is not compatible.

14. Decode the most likely sequence states (R,S) for the sequence GCATCAGACT using the following HMM.

