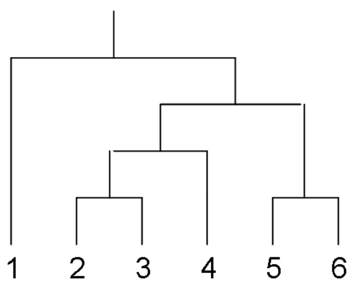
Introduction to bioinformatics, Autumn 2007, Exercise 5

10.10.2006

Update 8 October: assignment 4 refers to the matrix D in assignment 3, **not** in assignment 2.

1. (Chapter 12, Exercise 5) Find out the parsimony score with the algorithm given in lectures for the tree

   given the following set of sequences

   ```
   1 TCAA
   2 ACTT
   3 TTTT
   4 TATA
   5 ACTA
   6 TGAG
   ```

   Indicate the F set at each node of the tree.

2. Give a phylogenetic tree for the sequences in the assignment 1 which achieves better parsimony score than the tree in the above figure. Calculate the parsimony score using the algorithm given in lectures and indicate the F sets.

   You may use any method available to build the tree, including guessing. Explain your method.

3. (Chapter 12, Exercise 4)

(a) Find an additive phylogenetic tree corresponding to the distances in the following matrix D

	a	b	c	d	e
a	0	3	10	10	5
b		0	11	11	6
c			0	4	7
d				0	7
e					0

Explain how did you find the solution.

(b) Is the data in D ultrametric? Why or why not?

4. Perform UPGMA clustering on data in the matrix D given in assignment 3. Calculate the distance matrices and intermediate trees corresponding to the iterations of the algorithm.

   Draw the resulting tree and compare against the tree you found in assignment 3. Explain the differences in the trees, if any.

5. (Chapter 12, Exercise 11) This exercise illustrates an important principle in calculating likelihoods on trees. Probability of observing particular bases at the leaves of a tree with $n = 3$ species is

$$p(i_1, i_2, i_3) = \sum_a \sum_b \pi_a q_{ai_3}(t_2) q_{ab}(t_2 - t_1) q_{bi_2}(t_1) q_{bi_1}(t_1).$$

   This can also be written in the form

$$p(i_1, i_2, i_3) = \sum_a \pi_a q_{ai_3}(t_2) \sum_b q_{ab}(t_2 - t_1) q_{bi_2}(t_1) q_{bi_1}(t_1).$$

   Evaluate carefully how many addition and multiplication operations are performed in these two formulae, and deduce that the first form is less efficient than the second.

6. Answer the course questionnaire at `http://ilmo.cs.helsinki.fi/kurssit/servlet/Valinta?kieli=en`.

   You do not have to answer the questionnaire before the exercise session, as long as you do it within a week after the course exam.