## Algorithms for Bioinformatics (Autumn 2012)

## Exercise 5 (Thu 11.10, 10-12, B119, Niko Välimäki)

1. Ultrametric and additive distances.
a) Which of the matrices below are ultrametric?
b) Which of the matrices below are additive?

|  | A | B | C | D |
| :---: | :---: | :---: | :---: | :---: |
| A | 0 | 4 | 10 | 10 |
| B |  | 0 | 10 | 10 |
| C |  |  | 0 | 8 |
| D |  |  |  | 0 |


|  | A | B | C | D |
| :---: | :---: | :---: | :---: | :---: |
| A | 0 | 8 | 7 | 4 |
| B |  | 0 | 5 | 6 |
| C |  |  | 0 | 5 |
| D |  |  |  | 0 |


|  | A | B | C | D |
| :---: | :---: | :---: | :---: | :---: |
| A | 0 | 4 | 6 | 3 |
| B |  | 0 | 4 | 6 |
| C |  |  | 0 | 7 |
| D |  |  |  | 0 |

## 2. UPGMA.

Simulate the UPGMA algorithm with the distance matrix given below. Check that the distances given by the resulting tree correspond to the distance matrix.

|  | A | B | C | D | E |
| :---: | :---: | :---: | :---: | :---: | :---: |
| A | 0 | 6 | 10 | 10 | 6 |
| B |  | 0 | 10 | 10 | 2 |
| C |  |  | 0 | 4 | 10 |
| D |  |  |  | 0 | 10 |
| E |  |  |  |  | 0 |

## 3. Neighbor joining.

Simulate the neighbor joining method with the distance matrix given below. (Check the updated lecture slides for correct formulas!) Check that the distances given by the resulting tree correspond to the distance matrix.

|  | A | B | C | D |
| :---: | :---: | :---: | :---: | :---: |
| A | 0 | 4 | 4 | 5 |
| B |  | 0 | 6 | 3 |
| C |  |  | 0 | 7 |
| D |  |  |  | 0 |

## 4. Simulating small parsimony.

a) Solve the small parsimony problem using Sankoff's algorithm, sequences ACAC, ATAT, CTCT, GTGT being the leaves (from left to right) of a balanced binary tree.
b) In the large parsimony problem the leaves can be in any order and the tree shape is not fixed. Does any other tree give better parsimony score for our example?

## 5. Ultrametric condition.

Consider the three-point condition: A symmetric distance matrix $D=\left\{d_{i j} \mid\right.$ $1 \leq i \leq n, 1 \leq j \leq n\}$ corresponds to an ultrametric tree if and only if $d_{i j} \leq$ $\max \left(d_{i k}, d_{k j}\right) \mathrm{f}$ or all $i, j, k$. An ultrametric tree for $D$ is an edge-weighted tree (positive weight associated to each edge) such that the sum of weights in the path from leaf $i$ to node $v$ and from leaf $j$ to $v$ are both $\frac{1}{2} d_{i j}$, where $v$ is the lowest common ancestor of $i$ and $j$. (Notice this is an alternative but semantically identical definition to what was used in the lectures).
a) Prove that the three-point condition can identically be stated as follows: two of the three values $d_{i j}, d_{i k}$, and $d_{k j}$ are equal and one is smaller than the others.
b) Prove the three-point condition.

