Computational Biology HW 3

Probst Jennifer, 16703423

7. November 2020

Question 1

The probability at node 6 is the probability of a nucleotide N at node 6 given the nucleotides / probabilities of nucleotides of the children nodes. We do not observe a G in any of the child nodes of node 6, whereas 3 is an A, which leads to a higher probability of node 6 being an A. The probabilities are furthermore influenced by different branch lengths for t_3 and t_7, which lead to different P matrices and therefore affect the calculation of the internal node 6 likelihoods.

Question 2

As the nearest-neighbour interchange move switches subtrees and the intermediary calculations on internal nodes depend on subtrees of internal nodes, you can only reuse the calculated probabilites for the individual subtrees.

Question 3

The UPGMA algorithm only computes pairwise distances, but ML tree search takes higher order correlations into account. Moreover branch lengths in the tree have different meanings: time in the case of UPGMA and number of mutations in the ML tree search. As a third difference, the UPGMA algorithm runs has polynomial time complexity to find the optimal tree, whereas the ML tree inference is NP-hard.

Question 4

We define x as the number of nucleotides we have and n as the number of internal sites. We need a pruning step for each internal node, which each takes two times of x additions and x multiplications, so 2*(n*2*x) operations. To calculate the log likelihood we need a sum of n sums with each x multiplications and x additions, so n*2*x operations. In total this are 6xn operations, which leads to 24n operations in the case of 4 nucleotides and 30n operations for 5 nucleotides.

Question 5

No, we cannot. In the script 6.3.3.3 shows how the condition of time-reversibility implies that the likelihood remains the same no matter where one places the root of the tree. If the substitution model is not time reversible, the detailed balance equation $\pi_i * p_{i,j}(t) = \pi_i * p_{j,i}(t)$ is not fulfilled and we can therefore not reorder the summation the likelihoods don't remain the same.