23. October 2013

Exercise 1:   In the lecture we saw how the Fitch algorithm can be used to find the minimum number of mutations needed to generate given sequences along a given tree. How can you extend this algorithm to also get the corresponding optimal assignment of nucleotides to the inner nodes of the tree?

Exercise 2:   Given a tree with tips are labled with homologous sequences and a cost function $S(.,.)$ that assings a penalty $S(a, b)$ to each possible type of mutations $a \to b$, how can you find the minimal costs needed to generate the sequences by evolution along the tree.

Exercise 3:   (Mainly for bioinformaticians and biostatisitcians) Assume that the two tips i and k of a binary tree form a monophyletic group. Furthermore, assume that the pairwise distances of all tips of the tree are precisely known. Proof that NeighborJoining will not join i with any tip j other than k in its first step.

Exercise 4:   (Mainly for biologists) Assemble a dataset of published DNA sequences of several primates (including humans, chimpanzees and gorillas) and other mammals to reconstruct the species' phylogenetic tree using NeighborJoining, UPGMA and Maximum Parsimony. Use at least 5, better 10 different loci. Explore how the results depend on the selection of loci. (Several alignment programs are installed on turtle, e.g. mafft, muscle, probcons, and prank).