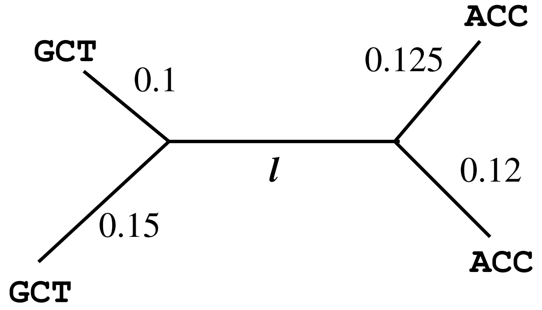
Exercise 1: Simulate sequence datasets with different trees with 5 and with 10 taxa and with different mutation models (e.g. JC and HKY). Explore how the accuracy of the ML tree found by DNAML depends on the substitution model used in DNAML and the sequence length. Apply also RAxML to the data.

Exercise 2: (Mainly for bioinformaticians) Find a way to simulate data for a gene that has two exons and one intron. In the coding region, the third codon position evolves faster than the first two, and the intron evolves even faster. You can use seq-gen and postprocess the output. For simplicity, neglect that the mutation probabilities of the third codon position depend on the current states of the first two. Simulate several phylogenetic datasets and analyze them with RAxML with and without partitioning. How is the effect of partitioning on the accuracy of the result and how does this depend on the sequence lengths, the number of taxa, and the branch lengths of the tree?

Exercise 3: (Mainly for biologists) Repeat exercise 3 of exercise sheet 2 with RAxML. Try different options of this program, including partitionings of the data. Compare your results to published trees of these species.

Exercise 4: For the tree given in Exercise 4 in sheet 2 compute the conditional expectation values for the numbers of mutations of all possible types on the central branch (the one of length 0.2) given the sequences at the tipps of the tree.

Exercise 5: Find the central branch lenght ℓ that maximizes the likelihood of the tree shown below (for fixed lengths of the other branches). For the substitution process assume a Jukes-Cantor model with $\lambda = 1$, such that the rate of a change from nucleotide x to any *other* nucleotide y is $\lambda/4$.


ACC
GCT
0.1
0.125
l
0.12
0.15
ACC
GCT