Exercises – Phylogenetics

Universität Bielefeld, SS 2018,
Dr. Roland Wittler

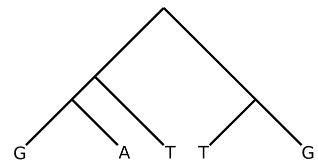
https://gi.cebitec.uni-bielefeld.de/Teaching/2018summer/Phylogenetik

Exercise Sheet 4 — 3.05.2018

Due: 17.05.2018

Task 1 Small Parsimony – Fitch-Algorithm. **(3 points)**

(a) Apply the Fitch-Algorithm (as presented in the lecture / lecture notes) on the tree on the right. Write down all solutions and their parsimony cost that can be found with the algorithm. Specify the set S for each internal node (see figure on page 28).



(b) Now, we have a look at the original work of Walter M. Fitch: "Towards Defining the Course of Evolution: Minimum Change for a Specific Tree Topology", published in the journal "Systematic Zoology". You can find this article online: http://www.jstor.org/stable/2412116. In this article, Fitch introduces an extra step after the bottom-up phase such that the top-down phase will find *all* optimal labelings.
Use this algorithm to enrich the set S for the tree above. Are there new solutions that were not found in task (a)? Indicate all such solutions and their parsimony costs.

Task 2 Small Parsimony – Sankoff-Algorithm. **(3 points)**

Apply the Sankoff-Algorithm (**with unit costs**) on the tree from Task 1 in order to determine a most parsimonious labeling for the internal nodes. Specify the values for $C(u, a)$ for each internal node (like in the figure on page 30).

Write down all solutions.

1.5 Bonus Points: Repeat the exercise with the following cost function:

cost	A	C	G	T
A	0	2	4	4
C	2	0	4	4
G	4	4	0	2
T	4	4	2	0

Task 3 Number of binary trees. **(3 points)**

Conni Count, the most unsuccessful bioinformatician in his time, wants to find a most parsimonious phylogenetic tree by specfying all possible unrooted trees and calculating the parsimony cost for each tree.

(a) He is able to calculate 1.000.000 trees in one second with his implementation of the Fitch-Algorithm. Conni is 30 years old. How old does Conni have to become, if he wants to get the result for a dataset with 17 species?

(b) Conni received a compute cluster for Christmas. He is now able to calculate $10^{12} = 1.000.000.000.000$ trees per second with one terahertz and one calculation per clock. Our universe is roughly 15.000.000.000 years old. If Conni would have started his program on this computing cluster at the big bang, how many leaves could have been processed at most until today?

Hint: Do not try to solve the formula $U_n = \prod_{i=3}^{n} \Pi(2i-5)$ to n. Calculate U_n for increasing n instead. (A spreadsheet works wonders.)