

# Exercises – Phylogenetics

Universität Bielefeld, SS 2018

Dr. Roland Wittler

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

## Exercise Sheet 10 — 05.07.2018

Due: 12.07.2018 (This is the second last exercise sheet.)

### Task 1 Modeling Amino Acid Replacements. (4 points)

Consider the following alignment of sequences  $x$  and  $y$  over the simplified amino acid alphabet  $\mathcal{A} = \{C, E, S\}$ .

$$\begin{aligned} x &= \text{S C C S E E E C S S E C C C E C S C E C} \\ y &= \text{S S C S E E S C S S E C C C C S S C E C} \end{aligned}$$

- Write down the values for all  $m_{i,j}$ , all  $f_i$  and  $N$ . Use them to calculate the transition matrix  $P$ .
- Obviously (without computing it),  $P$  is not calibrated to 1 PAM. Why?
- Calculate the score matrix  $S$  using  $P$  and  $\pi_i = f_i$ . If you do not have any results from (a) use

$$P = \begin{pmatrix} \frac{14}{17} & \frac{1}{17} & \frac{2}{17} \\ \frac{1}{10} & \frac{4}{5} & \frac{1}{10} \\ \frac{2}{13} & \frac{1}{13} & \frac{10}{13} \end{pmatrix} \text{ and } \pi = \left( \frac{17}{40}, \frac{1}{4}, \frac{13}{40} \right).$$

### Task 2 Maximum Likelihood Estimation of Evolutionary Distances. (4 points)

Assume the evolution of the sequences  $x$  and  $y$  from Task 1 follows an EMP with the following transition matrix:

$$P = \begin{pmatrix} 1 - \frac{1}{2}t & \frac{1}{4}t & \frac{1}{4}t \\ \frac{1}{4}t & 1 - \frac{1}{2}t & \frac{1}{4}t \\ \frac{1}{4}t & \frac{1}{4}t & 1 - \frac{1}{2}t \end{pmatrix}$$

Determine the time  $t$  such as to maximize the likelihood  $Pr(A, B | t)$ .

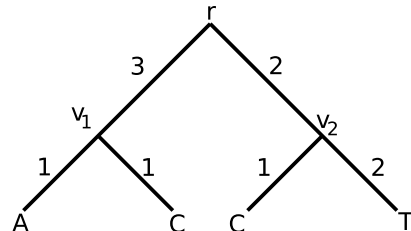
### Task 3 Computing the Likelihood of a Given Tree. (4 points)

Consider the given transition probability matrix  $P(t)$  and the given tree  $T$ . Compute the *likelihood* of  $T$ .

$$P(t) = \begin{pmatrix} 1 - 3a_t & a_t & a_t & a_t \\ a_t & 1 - 3a_t & a_t & a_t \\ a_t & a_t & 1 - 3a_t & a_t \\ a_t & a_t & a_t & 1 - 3a_t \end{pmatrix}$$

where

$$a_t = \frac{1 - \exp(-4t/30)}{4}$$



Turn over! Bitte wenden!

The following exercises are due in **two** weeks. Prepare for the exam (and earn missing points).

**Task 4 Perfect Phylogeny.**

Use the  $\mathcal{O}(mn)$  Algorithm (lecture notes, pages 22–23) to create a PP from the given matrix. Write down the intermediate results after the steps 1 and 2.

**(3 bonus points)**

	1	2	3	4	5
<i>A</i>	1	1	0	1	0
<i>B</i>	0	0	1	1	0
<i>C</i>	1	0	0	1	0
<i>D</i>	0	0	1	1	1
<i>E</i>	0	0	0	0	0

**Task 5 Small Parsimony.**

**(3 bonus points)**

Compare the algorithms from Fitch (version in the lecture notes) and Sankoff. Point out their similarities and differences.

Explain the run-time of the algorithms.

**Task 6 Maximum Parsimony.**

**(3 bonus points)**

Write a pseudo-code for the *row-wise Branch-and-Bound*.

**Task 7 Distances.**

**(3 bonus points)**

If  $d^T$  is an ultrametric on  $n$  taxa, there are at most  $n - 1$  different entries  $d_{i,j}^T$  for all  $i \neq j$ .

- Justify this statement. (Hint: Each internal node is an LCA of some leaf nodes. What is the pairwise distance of those?)
- Find a counter example with four taxa to show that the reverse is not true. (“Each path metric with at most  $n - 1$  different entries  $d_{i,j}^T$  for all  $i \neq j$  forms an ultrametric.”)