

Exercises – Phylogenetics

Universität Bielefeld, WS 2015/2016,
Dipl.-Inform. Damianos Melidis · B. Sc. Kevin Lamkiewicz
<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2015winter/Phylogenetik>

Exercise List 12 — 26.01.2016

Due to: 02.02.2016

Exercise 1 Jukes-Cantor correction.

(5 Points)

Consider an alignment with length 900 and 75 substitutions (no InDel included). We assume an evolution of the sequences following the Jukes-Cantor model.

- What is the Jukes-Cantor corrected distance d (in PEM) of the sequences?
- How many mutation events occurred approximately?
- How many mutation events occur and how many substitutions in the alignment would appear, if the distance d between the sequences is 20 PEM?

Exercise 2 Modeling amino acid replacements.

(6 Points)

Assume the alphabet for amino acids $\mathcal{A} = \{A, B, C\}$. Consider the following alignment of the sequences X and Y:

```
X = B C C B A A A B C B A C C B A B A C A C  
Y = B C A C A A A B C B A C C A A B A C A A
```

- Write down the values for all $m_{i,j}$, all f_i and N . Use them to calculate the transition matrix P .
- Is P calibrated to 1 PAM? Elaborate.
- Calculate the score matrix S using P and $\pi_i = f_i$.

Exercise 3 Maximum likelihood estimation of evolutionary distances.

(4 Points)

Assume the evolution of the sequences X and Y from exercise 2 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 that the likelihood $Pr(A, B; t)$ is maximal.