

# Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2012/2013, Dr. Roland Wittler  
<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2012winter/Phylogenetik>

## Blatt 12 vom 17.01.2013

Abgabe in einer Woche zu Beginn der Vorlesung oder vorab im Briefkasten bei U10-151.

### Aufgabe 1 Jukes-Cantor Korrektur.

(5 Punkte)

In einem Alignment der Länge 600 finden sich 40 Substitutionen (und keine Indels). Wir nehmen eine Entwicklung der Sequenzen nach dem Jukes-Cantor Modell an.

- Wie groß ist die Jukes-Cantor korrigierte Distanz  $d$  (in PEM) der beiden Sequenzen?
- Wie viele Mutationsereignisse haben demnach in etwa stattgefunden?
- Wie viele Mutationsereignisse und wie viele Substitutionen im Alignment würden sich ergeben, wenn die Sequenzen einen Abstand  $d$  von 20 PEM hätten?

### Aufgabe 2 Modellierung von Aminosäure-Austausch.

(6 Punkte)

Zur Vereinfachung nehmen wir im Folgenden das Aminosäure-Alphabet  $\mathcal{A} = \{X, Y, Z\}$  an. Betrachte das folgende Alignment der Sequenzen A und B:

A = Z Z Y X X Y Z Y X Y Z X X Y Z Y Z X X Z  
B = Z Z Y X Z X Z Y X Y Z X X Z Z Y Z X Z Z

- Gib die Werte für alle  $m_{ij}$ , alle  $f_i$  und  $N$  an und berechne daraus die Übergangsmatrix  $P$ .
- Ist  $P$  auf 1 PAM kalibriert? Begründe.
- Berechne aus  $P$  die Score-Matrix  $S$  (mit  $\pi_i = f_i$ ). Wenn du keine Werte aus (a) hast, verwende

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

- Die folgenden Fragen sollen im Tutorium diskutiert werden. Du musst keine Antworten aufschreiben, sollst dir aber Lösungswege überlegen.

Wie kann  $P$  auf 30 PAM kalibriert werden? Wie kann daraus eine PAM-30 Similarity-Score Matrix bestimmt werden?

### Aufgabe 3 Maximum-Likelihood Abschätzung von Distanzen.

(4 Punkte)

Nimm an, die Entwicklung der Sequenzen A und B aus Aufgabe 1 unterlag einem EMP mit der folgenden Übergangsmatrix:

$$P = \begin{pmatrix} 1 - \frac{2}{3}t & \frac{1}{3}t & \frac{1}{3}t \\ \frac{1}{3}t & 1 - \frac{2}{3}t & \frac{1}{3}t \\ \frac{1}{3}t & \frac{1}{3}t & 1 - \frac{2}{3}t \end{pmatrix}$$

Bestimme die Zeit  $t$  für die die Likelihood  $Pr(A, B; t)$  maximal ist.