

Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2011/2012, Dr. Roland Wittler
<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011winter/Phylogenetik>

Blatt 13 vom 19.01.2012

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

Aufgabe 1 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 = X Y X Z X Y X X X Y Z Z X Y Y Y Z X X Z
B = X Y X Z Z X X X X Y Z Z X Z Y 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

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

Aufgabe 2 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.

Aufgabe 3 Pulley Principle.

(6 Punkte)

Das *Pulley Principle* von Joseph Felsenstein besagt, dass die Likelihood für einen phylogenetischen Baum (unter gewissen Voraussetzungen) unabhängig von der Wahl der Wurzel ist.

Wir betrachten das Alphabet $\{A, G\}$ und einen Baum, der lediglich aus zwei Blättern besteht, die durch einen Wurzelknoten verbunden sind. Blatt M ist beschriftet mit A und Blatt N mit G . Die Kante von der Wurzel zu M habe die Länge t_M und die andere Kante t_N .

Zeige, dass die Likelihood für diesen Baum unabhängig von der genauen Position der Wurzel ist, also nur abhängt von der Summe der Kantenlängen: $t_M + t_N$.

- Gib dazu zunächst die Likelihood in Abhängigkeit von t_M und t_N an. (Siehe Beispiel im Skript auf Seite 80, oben.)
- Verwende nun die Voraussetzung, dass der zugrundegelegte Evolutionsprozess *reversibel* ist, um die Likelihood so umzuformen, dass sie nur noch ein π_i (z.B. π_G) enthält.
- Da P eine stochastische Matrix ist, können wir die *Chapman-Kolmogorov Gleichung* verwenden: $P(t_M + t_N) = P(t_M)P(t_N)$. Mit einem der vier Einträge in $P(t_M + t_N)$ kann die Likelihood so vereinfacht werden, dass sie schließlich nur noch von der Summe $t_M + t_N$ abhängt und nicht mehr von den einzelnen Kantenlängen.