Übungen zur Vorlesung Bioinformatik 1
Wintersemester 2003/2004

Übungsblatt 6

Abgabe bis 28.11.03 , 16.30 INF 580

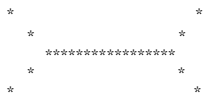
Aufgabe 2

Versetzen wir uns in das Jahr 1977 und in Dayhoffs Lage. Wir haben vor, eine Scorematrix a la Dayhoff zu entwickeln (PAM), probieren aber erst einmal mit einem etwas verkleinerten Sequenzdatensatz, i.e., mit 4 Sequenzen anstatt mehreren 100:

1) F R E I Z E I T
2) Z E I T G E I S T
3) F R E I H E I T
3) Z E I T H A S T

a) Aligniere sie von Hand global, ohne Gaps, schneide zu lange Sequenzen ab, so dass ein Block entstehtindem nur AS und keine Gaps stehen. (0P)

b) Erstelle den bestmöglichen Parsemony-Baum (Baum mit niedrigsten Kosten, wenn jede Substitution eine Einheit kostet). Die Wurzel kann vernachlässigt werden. O.B.d.A. betrachten wir nur Bäume mit paarweisen Verzweigungen, i.e. , unsere Bäume sehen topologisch so aus: (1P)



c) Erstelle eine Matrix der Übergangswahrscheinlichkeiten aus diesem Baum (ohne Zeiteichung) (1P)

d) Addiere zu jedem Eintrag in der Matrix der Übergangswahrscheinlichkeiten den Wert 0.001 dazu (wegen des Problems des Nulllogarithmierens...). Erstelle damit eine Scorematrix wie die PAM1, also mit den logarithmierten relativen Wahrscheinlichkeiten $s(x_i, y_i)$. Am Ende wird in der Scorematrix auf eine Stelle gerundet. Wir verwenden den natürlichen Logarithmus. (2P)

Tipp: verwende R !