

Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WS 2011/2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011winter/SequenzAnalyse>

Übungsblatt 7 vom 16.12.2011

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Substitutions-Scorematrix)

(4 Punkte)

Gegeben seien folgende Proteinsequenzen mit dem Alphabet $\Sigma = \{A, I, Q, R\}$:

RIIAA

ARIIQ

RRIRQ

RRIRA

RIRRA

Bestimme für diese Sequenzen die

1. Observationsmatrix M^t (Verteile Pseudocounts wenn nötig)
2. Hintergrundverteilung π ,
3. Likelihood-Ratio $\left(\frac{M^t(x,y)}{\pi_x \pi_y}\right)$,

um schließlich daraus die log-Likelihood Scorematrix $\left(s^t(x,y)\right) := \left(\log_2\left(\frac{M^t(x,y)}{\pi_x \pi_y}\right)\right)$ zu berechnen.

Aufgabe 2 Alignmentklassen

(4 Punkte)

Gegeben sind die Sequenzen $x = AB$ und $y = CD$.

1. Bestimme alle 13 möglichen Alignments von x und y .
2. Welche davon sind *echt* verschieden?

Aufgabe 3 (Die Anzahl globaler Alignments)

(4 Punkte)

Erstelle ein Programm in Java, welches

1. $N(m,n)$ rekursiv berechnet,
2. $N(n,n)$ approximativ berechnet,
3. $N'(m,n)$ exakt berechnet und
4. $N'(n,n)$ approximativ berechnet.

Gib die Ergebnisse für $n = m \in \{1, 5, 10, 100\}$ grafisch an und vergleiche diese. Diskutiere den Unterschied zwischen $N(m,n)$ und $N'(m,n)$.