

Übungen zur Vorlesung Grundlagen der Sequenzanalyse

Universität Bielefeld, WS 2006/07

Dr. Sven Rahmann · Dipl.-Bioinf. Katharina Jahn

<http://gi.cebitec.uni-bielefeld.de/teaching/2006winter/sequenzanalyse/>

Blatt 6 vom 23.11.2006

Abgabe am 30.11.2006 vor der Vorlesung um 8:30 in H3

Aufgabe 1 Gegeben ist folgende Häufigkeitstabelle von beobachteten Substitutionen über einen gewissen (nicht spezifizierten) evolutionären Zeitraum t .

	A	B	C
A	60	1	1
B	1	20	3
C	1	3	10

Berechne die Austauschmatrix M^t , die Hintergrundverteilung π , sowie die log-odds Scorematrix S^t .

Aufgabe 2 Gegeben sind $x = \text{TCCATACA}$ und $y = \text{TACCAACACG}$. Berechne den lokalen Alignment-Score $s(x, y)$ und ein optimales lokales Alignment (mit Stift und Papier) mit Hilfe des Smith-Waterman Algorithmus, wobei die Score-Funktion wie folgt gegeben ist: 2 Punkte für ein Match, -3 für ein Mismatch, -2 für ein Indel (lineare Gapkosten).

Aufgabe 3 Wie muss der Alignment-Graph am Beispiel der vorigen Aufgabe modifiziert werden, wenn man auch Flips mit Score 0 zulassen will? Gib genau an, welche zusätzlichen Kanten benötigt werden. Schreibe explizit den zugehörigen DP-Algorithmus auf.

Aufgabe 4 (ein Nachtrag zum Problem, welche Wörter dasselbe q -gram Profil wie ein gegebenes $x \in \Sigma^*$ haben:)

Der *De Bruijn Graph* $B(x, q)$ zum Paar (x, q) ist wie folgt definiert¹:

- Knoten sind alle $(q - 1)$ -gramme in x .
- Die Kanten stellen die q -gramme von x dar, indem für jedes q -gram azb in x mit $a \in \Sigma$, $z \in \Sigma^{q-2}$, $b \in \Sigma$ die Kante von dem Knoten mit dem $(q - 1)$ -gram az zum Knoten mit dem q -gram zb gezogen wird. Es gibt also ggf. Mehrfachkanten; die Vielfachheit einer Kante entspricht der Vielfachheit es betreffenden q -grams in x .

Ein *Euler-Pfad* ist ein Pfad in $B(x, q)$, der jede Kante (gemäß ihrer Vielfachheit) genau einmal benutzt. Überlege: Es gibt eine 1-zu-1-Beziehung zwischen Euler-Pfaden in $B(x, q)$ und Wörtern mit demselben q -gram Profil wie x .

Zeichne $B(x, q)$ für Aufgabe 1 von Blatt 4: $x = \text{cbabbcabcbabaa}$ und $q = 3$. Wie viele und welche Wörter gibt es mit demselben q -gram Profil wie x ?

¹de Bruijn, N. G. (1946). "A Combinatorial Problem". Koninklijke Nederlandse Akademie v. Wetenschappen 49: 758–764.