

Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SS 2010
Prof. Dr. Jens Stoye · Dr. Inke Herms

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2010summer/SequenzAnalyse>

Blatt 11 vom 02.07.2010

Abgabe bis zum 09.07.10 in U10-151 (Briefkasten)
Bitte gib auch den Namen deines Tutors an.

Aufgabe 1 (Steinerbaum)

2 Punkte

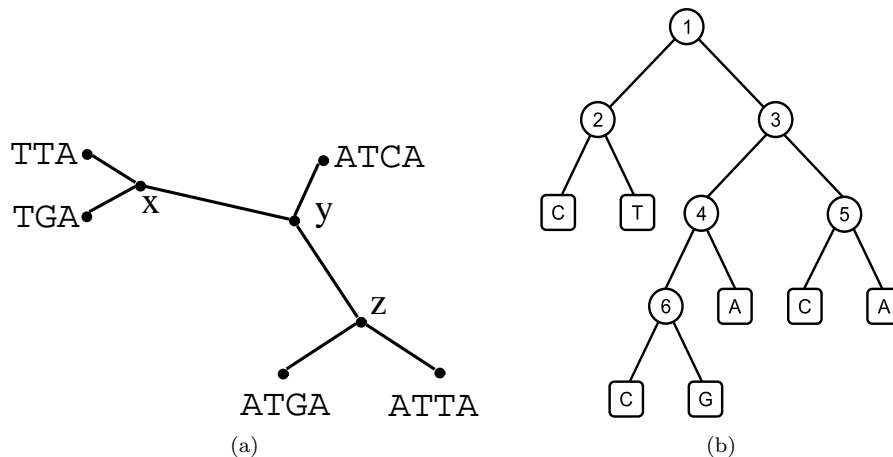
Gib für die gegebenen Sequenzen über dem Alphabet $\Sigma = \{A, C, G, T\}$ jeweils einen Steinerbaum und die entsprechenden Gesamtkosten an.

- a) $s_1 = \text{CCTG}$, $s_2 = \text{CGAA}$, $s_3 = \text{CGAG}$, $s_4 = \text{GCAG}$.
b) $s_1 = \text{CGAT}$, $s_2 = \text{GGCC}$, $s_3 = \text{GGA}$, $s_4 = \text{CGT}$.

Aufgabe 2 (Baumalignment)

4 Punkte

Gegeben seien die folgenden Bäume:



- a) Um die Kosten für ein Baumalignment exakt zu berechnen, kann der Sankoff-Algorithmus verwendet werden. Erkläre in 3–5 Sätzen, wie dieser Algorithmus funktioniert. Erläutere auch, welche Rolle der Fitch-Algorithmus dabei spielt.
b) Beschrifte die inneren Knoten x , y und z des Baumes in Abbildung (a) so, dass die Kosten des Baumalignments möglichst gering sind. Berechne die Kosten des entsprechenden Alignments.
c) Berechne für den phylogenetischen Baum in Abbildung (b) die sparsamste Beschriftung der inneren Knoten mit Hilfe des Fitch-Algorithmus. Gib dabei für die inneren Knoten jeweils die mit ihnen assoziierten Informationen der Bottom-Up- und der Top-Down-Phase an.

Aufgabe 3 (Aligning Alignments)

3 Punkte

Für die auf Alignmentsspalten verallgemeinerten Einheitskosten soll ein optimales multiples Alignment A_3 aus den folgenden beiden Alignments A_1 und A_2 berechnet werden:

$$A_1 = \begin{pmatrix} \text{ACTC} \\ \text{AG-C} \end{pmatrix}, \quad A_2 = \begin{pmatrix} \text{GACT-} \\ \text{TAGTC} \\ \text{G-CTC} \end{pmatrix}$$

- a) Berechne die minimalen Alignmentkosten.
b) Gib A_3 an.