

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2014/2015

Dr. Roland Wittler · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014winter/SequenzAnalyse>

Übungsblatt 11 vom 06.01.2015

Abgabe 13.01.2015

## Aufgabe 1 (Steinerbaum)

(4 Punkte)

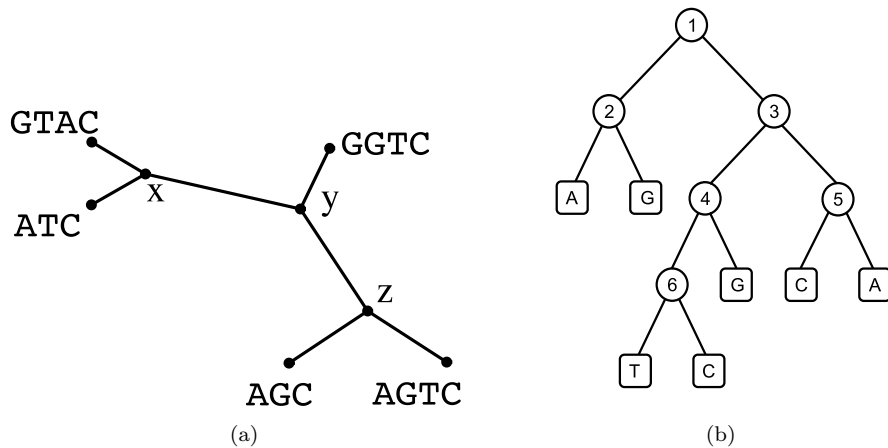
Bestimme (durch Knobeln) jeweils einen minimalen Steinerbaum für die folgenden Mengen von Sequenzen. Benutze dabei die Einheitskostendistanz.

1.  $s_1 = TCA$ ,  $s_2 = CCTA$ ,  $s_3 = GTAC$ ,  $s_4 = CAG$ .
2.  $s_1 = GTCG$ ,  $s_2 = GCGT$ ,  $s_3 = AGG$ ,  $s_4 = AGTC$ .

## Aufgabe 2 (Baumalignment)

(6 Punkte)

Gegeben seien die folgenden Bäume:

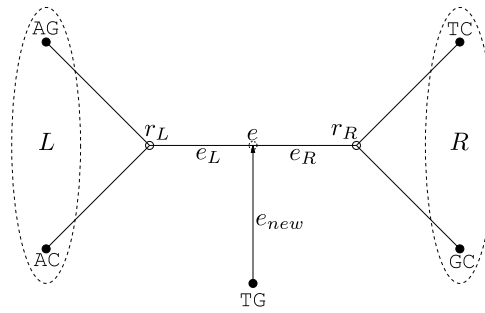


1. Um die Kosten für ein Baumalignment exakt zu berechnen, kann der Sankoff-Algorithmus verwendet werden. Erkläre in 3–5 Sätzen, wie dieser Algorithmus funktioniert. Erläutere auch, welche Rolle der Fitch-Algorithmus dabei spielt.
2. Beschrifte die inneren Knoten  $x$ ,  $y$  und  $z$  des Baumes in Abbildung 1(a) so, dass die Kosten des Baumalignments möglichst gering sind. (Hier soll kein Algorithmus angewendet werden, sondern “gut geraten” werden.) Berechne die Kosten des entsprechenden Alignments.
3. Berechne für den phylogenetischen Baum in Abbildung 1(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.

Bitte wenden!

**Aufgabe 3 (Greedy Three-Way Tree Alignment)**

(4 Punkte)

Betrachte folgenden Zwischenschritt in einer *Greedy Three-Way* Baumalignment-Konstruktion.

$$A = \begin{pmatrix} A & G & - \\ A & - & C \\ T & - & C \\ - & G & C \\ T & G & - \\ \vdots & & \end{pmatrix}$$

1. Bestimme die Kantenlänge  $e_{new}$  unter Einheitskosten.
2. Ordne diesen Zwischenschritt kurz (in 3–5 Sätzen) in den Gesamtalgorithmus ein.