

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2022

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

<https://gi.cebitec.uni-bielefeld.de/teaching/2022summer/sa>

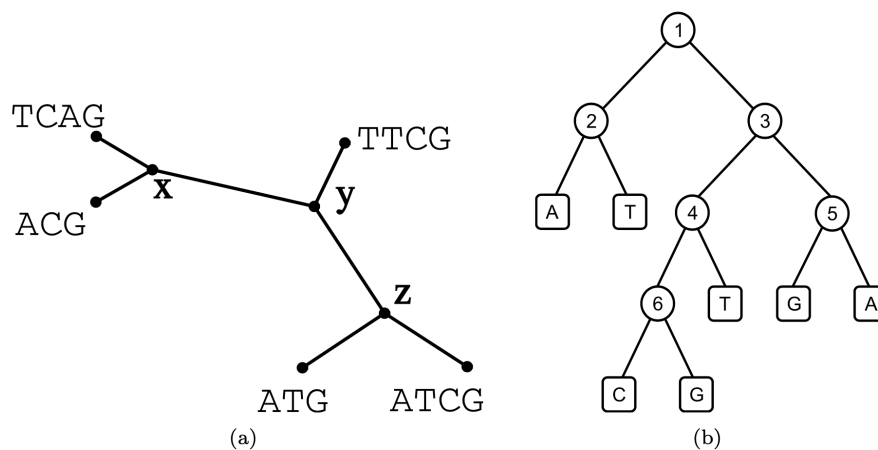
**Übungsblatt 12 vom 30.6.2022**

**Abgabe am 7.7.2022 bis 12:00 Uhr**

## Aufgabe 1 (Baumalignment)

(10 Punkte)

Gegeben seien die folgenden Bäume:



- 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.
- Beschrifte die inneren Knoten  $x$ ,  $y$  und  $z$  des Baumes in Abbildung (a) so, dass die Kosten des Baumalignments möglichst gering sind. (Hier soll kein Algorithmus angewendet, sondern “gut geraten” werden.) Berechne die Kosten des entsprechenden Alignments.
- 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 2 (Genomalignment mit MUMs)

(10 Punkte)

Gegeben seien die Genome:  $\begin{cases} \mathcal{G} = \text{GTGCCCATATTCAGAA} \\ \mathcal{H} = \text{TGCTTATTCCACGAA} \end{cases}$

- Verwende den verallgemeinerten Suffixbaum  $\mathcal{G}\#\mathcal{H}\$$ , um alle MUMs der Mindestlänge 3 von  $\mathcal{G}$  und  $\mathcal{H}$  zu finden.
- Berechne ein globales Alignment von  $\mathcal{G}$  und  $\mathcal{H}$ , indem du die MUMs verkettest (Gewicht=Länge) und die Lücken durch paarweise optimale Alignments füllst.