

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2019
Dr. Daniel Dörr · Michel T. Henrichs

<https://gi.cebitec.uni-bielefeld.de/teaching/2019summer/sa>

Übungsblatt 13 vom 1.7.2019
Abgabe am 8.7.2019 bis 12:00 Uhr (mittags)

Aufgabe 1 (Wiederholung)

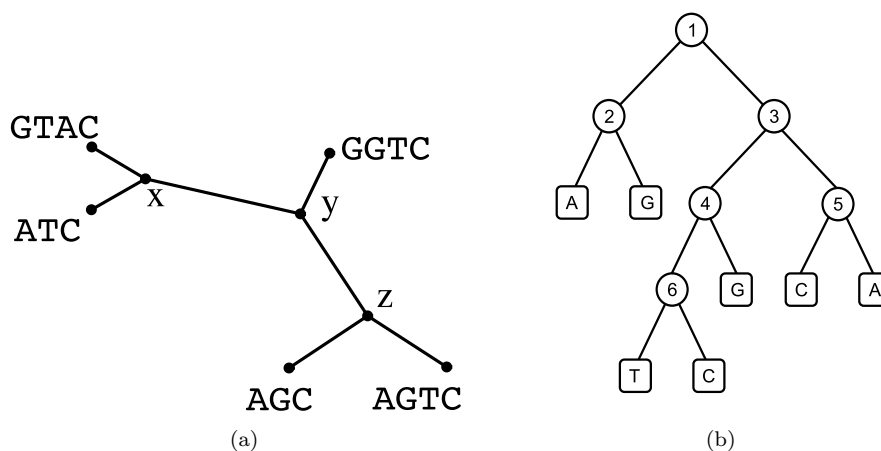
(2* Punkte)

Welche Themen würdest du gerne für die Klausur wiederholen? Schreibe zwei Themen oder konkrete Fragen auf, die dir noch nicht ganz klar sind.

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.

Aufgabe 3 (Algorithmen-Tabelle)

(8* Punkte)

Mache dir klar, welche Laufzeiten die folgenden Algorithmen haben:

Berechnung der q -gram-Distanz; WOTD-Algorithmus; Berechnung der Maximal-Matches-Distanz; Needleman-Wunsch-Algorithmus; Smith-Waterman-Algorithmus; Gotoh-Algorithmus; Waterman-Eggert-Algorithmus; Sellers' Algorithmus; Hirschberg-Algorithmus; Algorithmus, um einen String exakt in einem Suffixbaum zu finden; Algorithmus, um mit Hilfe eines Suffixbaums den kürzesten eindeutigen String zu finden; Algorithmus, um maximale Repeats mit einem Suffixbaum zu finden; exakte Berechnung eines multiplen Alignments; Center-Star-Approximation; Divide-and-Conquer-Alignment-Algorithmus; Sankoffs Algorithmus; Fitch-Algorithmus.

Fülle dazu eine Tabelle in der unten stehenden Form aus. Nenne den Namen des Algorithmus und beschreibe kurz, was er macht, wenn man es nicht schon am Namen erkennen kann. Schreibe die asymptotische Laufzeit auf und erkläre die Bedeutung der Parameter. Verstehe, wie die Laufzeit zu Stande kommt, aber schreibe es nicht auf.

Algorithmus	Zweck	Laufzeit	Parametererklärung
Berechnung der q -gram-Distanz	–	$\mathcal{O}(\sigma^q + m + n)$	σ : Alphabetgröße, q : Länge des q -grams, m und n : Länge der Sequenzen
WOTD-Algorithmus