

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2019

Dr. Daniel Dörr · Michel T. Henrichs

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

Übungsblatt 7 vom 20.5.2019

Abgabe am 27.5.2019 bis 12:00 Uhr (mittags)

Aufgabe 1 (Paarweises Alignment mit linearem Speicherbedarf)

(4 Punkte)

Gegeben sind die Strings $x = \text{AGCT}$ und $y = \text{AGT}$. Verwende den Hirschberg-Algorithmus und Einheitskosten, um ein Alignment der beiden Strings x und y in linearem Platz zu berechnen. Du kannst Alignments direkt berechnen, wenn einer der beiden Strings eine Länge von ≤ 1 hat. Gib das endgültige Alignment explizit an.

Aufgabe 2 (Dot-Plot)

(3+1 Punkte)

Gegeben seien die beiden Sequenzen $x = \text{GCAATGACTTAGTATTGTA}$ und $y = \text{GCAATTATTCAGGTATTGTA}$.

1. Erstelle einen ungefilterten und einen gefilterten ($q=5$) Dot-Plot für x und y .
2. Was ist im gefilterten Dot-Plot zu sehen? Wie könnte es entstanden sein?

Anstatt die Plots im ersten Aufgabenteil von Hand zu zeichnen, kann auch ein Programm geschrieben werden, welches zwei Strings sowie q entgegennimmt und dann einen gefilterten Dot-Plot anzeigt oder ausgibt. Dafür gibt es einen Extrapunkt.

Aufgabe 3 (BLAST)

(5 Punkte)

Gegeben sei die Query $x = \text{PAPALAPAP}$ und $q = 4$. Der Score für einen Match sei $+3$ und für einen Mismatch -1 . Indels sollen nicht möglich sein, haben also Kosten ∞ . Betrachte nun die Nachbarschaft von x :

1. Gib für $i = 3$ und $i = 5$ jeweils zwei Tupel (z, i) an, die sich in der Nachbarschaft $N_8(x)$ befinden, aber nicht in $N_9(x)$. Die Indizierung der Sequenz x beginnt hier bei 0.
2. Bestimme die folgenden Listen $P_k(z)$ (Beachte die verschiedenen Werte für $k!$):
 - (a) $P_4(\text{AAAA}), P_8(\text{AAAA})$
 - (b) $P_8(\text{PAPA}), P_9(\text{PAPA})$
 - (c) $P_8(\text{LALA}), P_9(\text{LALA})$