

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2016/2017

Prof. Dr. Jens Stoye · M.Sc. Tizian Schulz

<https://gi.cebitec.uni-bielefeld.de/teaching/2016winter/sa>

Übungsblatt 4 vom 22.11.2016

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Dot-Plot) (4 Punkte)

Gegeben seien die beiden Sequenzen $x = \text{CGTTACTGAATACATAACAT}$ und $y = \text{CGTTAATAAGTCCATAACAT}$.

1. Erstelle einen *unfiltered* und einen *filtered* ($q=5$) *dot plot* für x und y .
2. Was ist im *filtered* dot plot zu sehen? Wie könnte es entstanden sein?

Aufgabe 2 (BLAST) (3 Punkte)

Informiere Dich genauer über BLAST (Skript S. 58/59).

1. Welche Rolle spielt der *X-Drop-Algorithmus* und wie funktioniert er?
2. Was sind die Unterschiede zwischen BLAST Version 1.4 und den Versionen ab 2.0?
3. Was bedeutet der Begriff *Sensitivität* im Kontext eines Matches?

Aufgabe 3 (Approximatives Stringmatching) (5 Punkte)

1. Finde die Endpositionen aller Vorkommen des Patterns $x = \text{AGAG}$ im Text $y = \text{CTGAGGTAGAG}$ mit maximal $k = 1$ Fehlern. Verwende die Cutoff-Variante von *Sellers' Algorithmus* mit **Einheitskosten** und markiere die *last essential indices*.
2. Gib zu jeder gefundenen Endposition alle zugehörigen Alignments an.
3. Warum macht es Sinn, in der Praxis nicht *alle* gefundenen Endpositionen auszugeben? (Stichwort: *Runs*.) Welche Endpositionen würde man im obigen Beispiel nicht ausgeben?

Aufgabe 4 (Forward-Backward Technik) (4 Punkte)

Gegeben sind die Sequenzen $s = \text{GTGCAATC}$ und $t = \text{CTCAGAT}$. Berechne, jeweils mit **Einheitskosten**,

- die Edit-Matrix D ,
- die Edit-Matrix D^{rev} der reversen Strings,
- die Gesamtkostenmatrix T und
- die Zusatzkostenmatrix C .

Gib ein optimales globales Alignment von s und t an und markiere es in D und C .

Aufgabe 5 (Paarweises Alignment mit linearem Speicherbedarf) (4 Punkte)

Zeige anhand der Sequenzen aus Aufgabe 4, wie das optimale Alignment berechnet werden kann, während nur linearer Speicherplatz benötigt wird. Simuliere dazu die Schritte des in der Vorlesung besprochenen Divide-and-Conquer Verfahrens auf Kopien der Edit-Matrix. (Es kann teilweise auf die in Aufgabe 4 berechneten Werte zurückgegriffen werden.)

1. Markiere in einer neuen, leeren Matrix der entsprechenden Größe, für welche Bereiche D und D^{rev} berechnet werden würden bzw. berechne die Werte wenn nötig neu und trage sie ein.
2. Trage die Werte von T für die mittlere Zeile $m' = \lceil m/2 \rceil$ ein.
3. Markiere ein Minimum in T und die entsprechende Spalte, an der die Matrix geteilt wird.
4. Führe rekursiv die oberen Schritte für die beiden Teilmatrizen aus bis jeweils nur noch eine Zeile übrig ist.