

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2024

Prof. Dr. Jens Stoye · Leonard Bohnenkämper

<https://gi.cebitec.uni-bielefeld.de/teaching/2024winter/sa>

Übungsblatt 6 vom 21.11.2024

Abgabe am 28.11.2024 bis 10:00 Uhr (morgens)

Aufgabe 1 (Dot-Plot)

(3 Punkte)

Gegeben seien die beiden Sequenzen $x = \text{GATGCCTGGACTTGCC}$ und $y = \text{AGGACTTGGATGCCTT}$. Erstelle die folgenden *dot plots* für x und y :

- ungefiltert;
- gefiltert mit $q = 3$;
- gefiltert mit $q = 6$.

Aufgabe 2 (BLAST)

(7 Punkte)

Gegeben seien die Query $x = \text{MYPMMYMP}$ und die Datenbanksequenz $y = \text{PMYYMYPMMPPYMYPP}$. Das Alphabet ist $\Sigma = \{\text{M, P, Y}\}$. Der Score für einen Match sei $+3$ und für einen Mismatch -1 . Indels sollen nicht möglich sein, haben also Score $-\infty$. Benutze $q = 3$ und $k = 6$.

- Berechne die 6-Nachbarschaft $N_6(x)$.
- Gib die BLAST-Hits von x in y an.

Aufgabe 3 (Paarweises Alignment mit linearem Speicherbedarf)

(6 Punkte)

Zeige anhand der Sequenzen aus Aufgabe 3 des vorigen Zettels ($x = \text{ATGCAATC}$ und $y = \text{CTCAGAT}$), wie das optimale Alignment berechnet werden kann, während nur linearer Speicherplatz benötigt wird. Simuliere dazu die Schritte des in der Vorlesung besprochenen Hirschberg Verfahrens auf Kopien der Edit-Matrix. (Es kann teilweise auf die auf dem vorigen Zettel berechneten Werte zurückgegriffen werden.)

Führe rekursiv die oberen Schritte für die beiden Präfixe und Suffixe des vorigen Zettels, bis jeweils nur noch eine Zeile übrig ist.

Aufgabe 4 (Approximative Textsuche)

(4 Punkte)

Gegeben seien das Muster $s = \text{ACCT}$ und der Text $t = \text{CTAACCTCCT}$.

- Verwende die *Cutoff*-Variante von Sellers' Algorithmus mit Einheitskosten und berechne die Edit-Matrix D von s und t mit bis zu $k = 1$ Fehlern. Markiere in jeder Spalte den *last essential index*.
- Markiere in D alle Positionen, an denen ein Vorkommen von s mit bis zu $k = 1$ Fehlern in t endet. Gib für jede dieser Endpositionen ein korrespondierendes Alignment an.

