

Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, WiSe 2009/2010

Prof. Dr. Jens Stoye · Dipl.-Inform. Nils Hoffmann

<http://wiki.techfak.uni-bielefeld.de/gi/GILectures/2009winter/SequenzAnalyse>

Blatt 11 vom 07.01.2010

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 Dotplots

(3 Punkte)

Gegeben sind die Worte $x = \text{AGACTTC}$ und $y = \text{AGATTTG}$.

1. Zeichne den Dotplot für x und y .
2. Zeichne den Dotplot für die Wortlänge $q = 2$.

Aufgabe 2 Datenbanksuche

(2 Punkte)

1. Welche unterschiedlichen Annahmen werden bei

- a) der Suche nach einem optimalen Alignment von zwei Sequenzen und bei
- b) der Suche nach einer ähnlichen Sequenz in einer Datenbank

getroffen? Welche Anforderungen werden dadurch an die Software gestellt, die diese Probleme lösen soll?

2. Erkläre die Unterschiede zwischen *on-line* und *index-basierter* Datenbanksuche. Wie sind die Laufzeiten zur Vorverarbeitung und bei der Suche? Welche Laufzeit ergibt sich bei k Suchen? In welchen Fällen lohnt sich eine index-basierte Suche, wann ist die on-line Variante vorzuziehen?

Aufgabe 3 FASTA

(2 Punkte)

1. Erstelle per Hand einen q -Gram Index für $x = \text{aaabcc}$ mit $q = 2$.
2. Schiebe danach ein Fenster der Länge 2 über $y = \text{abaabcab}$ und verwende den q -Gram Index, um $c(x, y; d)$ für alle $d \in \{-|x| \dots |y|\}$ zu berechnen. (Hinweis: `c[j-i]++`). Wie ist der FASTA-Score $C(x, y)$?

Aufgabe 4 BLAST

(3 Punkte)

Informiere Dich genauer über BLAST.

1. Welche Rolle spielt der **X-Drop Algorithmus** und wie funktioniert er genau?
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?