

Präsenzübungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SoSe 2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2012summer/SequenzAnalyse>

Präsenzübungsblatt 1, Woche 16/2012

Aufgabe 1 (BLAST)

Gegeben sei die Query $x = \text{PAPALAPAP}$ und $q = 4$. Der Score für einen Match sei $+3$ und für einen Mismatch -1 . 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 $N_9(x)$.
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) $P_8(\text{LALA}), P_9(\text{LALA})$

Aufgabe 2 (Statistik auf Sequenzen)

Eine DNA Sequenz T hat die relativen Buchstabenhäufigkeiten $f_a = f_t = 0,3$ und $f_g = f_c = 0,2$.

1. Wie wahrscheinlich ist dann die Sequenz $x = \text{tgacaatg}$ mit den oben angegebenen Häufigkeiten?
2. Wie hoch ist der Erwartungswert für die Anzahl der Treffer von x in T , wenn T die Länge 1000 hat?

Aufgabe 3 (Index-basierte Datenbanksuche)

Beschreibe jeweils, wie die folgenden Algorithmen im Vergleich zu BLAST arbeiten.

1. BLAT
2. SWIFT

In welchem Kontext werden die Algorithmen verwendet?