

Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SS 2010
Prof. Dr. Jens Stoye · Dr. Inke Herms

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2010summer/SequenzAnalyse>

Blatt 8 vom 04.06.2010

Abgabe am 18.06.10 vor Beginn der Vorlesung.
Bitte gib auch den Namen deines Tutors an.

Aufgabe 1 (Sum-of-Pairs Score)

3 Punkte

Gegeben sind die vier Sequenzen $s_1 = \text{GAACT}$, $s_2 = \text{GCCT}$, $s_3 = \text{CAGT}$, $s_4 = \text{GACT}$, sowie lineare Gapkosten von 2 und die folgende Substitutions-Scorematrix:

	A	C	G	T
A	3	-2	-3	-1
C	-2	5	-1	-2
G	-3	-1	3	-3
T	-1	-2	-3	5

Berechne den Sum-of-Pairs Score des folgenden multiplen Alignments:

G	A	A	C	T
G	C	-	C	T
-	C	A	G	T
G	-	A	C	T

Aufgabe 2 (Sum-of-Pairs Alignment)

3 Punkte

Finde drei Sequenzen, für die mind. eine der Projektionen $\pi_{(1,2)}$, $\pi_{(1,3)}$ und $\pi_{(2,3)}$ des optimalen multiplen Alignments auf die paarweisen Alignments *nicht* dem optimalen paarweisen Alignment entspricht.

Aufgabe 3 (Multiples Alignment von drei Sequenzen)

4 Punkte

Beschreibe, wie der Needleman-Wunsch-Algorithmus für paarweises Alignment verallgemeinert wird, um ein direktes Alignment von drei Sequenzen zu berechnen. Verwende die Standard-Editdistanz und

- a) gib die Rekursion explizit an,
- b) implementiere den Algorithmus in einer Programmiersprache deiner Wahl,
- c) gib die Kosten eines optimalen Alignments von $a = \text{CTGT}$, $b = \text{AGT}$ und $c = \text{CAG}$ an.

Aufgabe 4 (Laufzeit-Abschätzung)

3 Punkte

Gegeben sind k Sequenzen der Länge $n = 1000$. Die Berechnung eines exakten paarweisen Alignments mittels *dynamic programming* dauert ca. zwei Sekunden. Ein Tag hat 86400 Sekunden und ein Jahr im Durchschnitt 365,2425 Tage.

- a) Wie viele paarweise Alignments dieser Form kann man berechnen, bevor in 5 Milliarden Jahren die Sonne zu einem roten Riesen wird?
- b) Wie viele Sequenzen kann man mit einem direkten multiplen Alignment ($\mathcal{O}(2^k \cdot n^k)$) in dieser Zeit alignieren?