

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2019

Dr. Daniel Dörr · Michel T. Henrichs

<https://gi.cebitec.uni-bielefeld.de/teaching/2019summer/sa>

Übungsblatt 11 vom 17.06.2019

Abgabe am 24.06.2019 bis 12:00 Uhr (mittags)

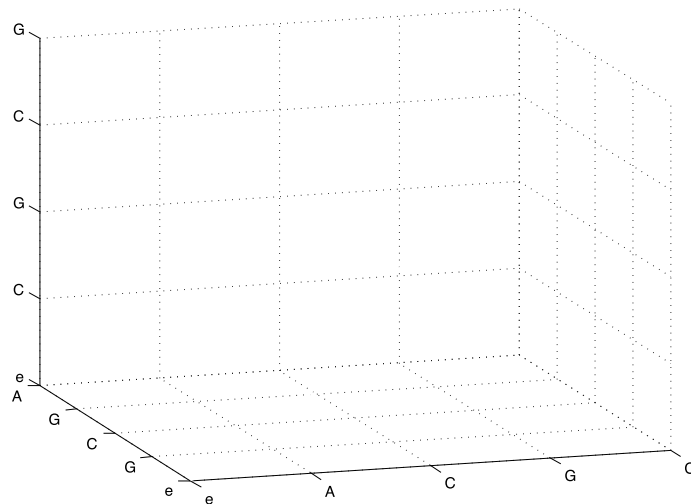
Aufgabe 1 (Geometrische Projektion des Alignments)

(2 Punkte)

Gegeben sei das multiple Alignment der Sequenzen $s_1 = \text{CGCG}$, $s_2 = \text{ACGC}$, $s_3 = \text{GCGA}$:

$$\begin{pmatrix} - & C & G & C & G \\ A & C & G & C & - \\ G & C & G & A & - \end{pmatrix}$$

Zeichne das Alignment im 3-Dimensionalen Raum und die jeweiligen Projektionen auf den 2-Dimensionalen Unterräumen ein:



Aufgabe 2 (Sum-of-Pairs Score)

(4 Punkte)

Schreibe eine Funktion in einer Programmiersprache deiner Wahl, die für gegebene Match-, Mismatch- und lineare Gapscores und ein gegebenes multiples Alignment dessen Sum-of-Pairs Score berechnet. Gaps im Alignment kannst du einfach als „-“ in den Sequenzen kodieren. Deine Funktion soll also eine Liste oder ein Array von Sequenzen, die das Alignment darstellen, sowie die drei Score-Werte als Eingabe erhalten und den Sum-of-Pairs Score zurückgeben.

Berechne mit deinem Programm die Kosten des folgenden Alignments bei einem Matchscore von 5, Mismatchkosten von 1 sowie linearen Gapkosten von 2.

$$\begin{pmatrix} T & C & T & A & G \\ T & - & T & - & C \\ - & C & C & - & - \\ - & C & C & A & G \end{pmatrix}$$

Aufgabe 3 (Sum-of-Pairs Alignment)

(2 Punkte)

Finde drei Sequenzen, für die mindestens 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 4 (Laufzeit-Abschätzung)

(4 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.

1. Wie viele paarweise Alignments dieser Form kann man berechnen, bevor in 5 Milliarden Jahren die Sonne zu einem roten Riesen wird?
2. Wie viele solcher Sequenzen kann man mit einem direkten multiplen Alignment ($\mathcal{O}(2^k \cdot n^k \cdot k^2)$) in dieser Zeit alignieren, wenn man dabei annimmt, dass ein multiples Alignment der Länge $n = 1000$ für $k = 2$ zwei Sekunden benötigt?
3. Erkläre die Laufzeit des direkten multiplen Alignments.