

# 9. Übung zur Vorlesung "Sequenzanalyse"

Sebastian Böcker, Marcus Ludwig, Kai Dührkop, Fleming Kretschmer

Ausgabe: 12.1.2022

Abgabe: 16.1.2022

## Aufgabe 1 (5 Punkte)

Gegeben ist das folgende multiple Sequenzalignment:

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

- Erstellen Sie das Profil für Alignment  $A$ .
- Was ist das optimale Alignment (Einheitskosten) dieses Profils mit der Sequenz  $CGAAC$ ? Geben Sie die DP-Matrix mit an.

## Aufgabe 2 (10 Punkte)

Gegeben sind drei Sequenzen:  $AACGTCATATGCAGTTCG$ ,  $TCATAGGAACGGTGCACG$  und  $GTTGGCATAGCCAACGCCACG$ . Die Wahrscheinlichkeit, dass zwischen zwei Sequenzen ein Segment der Länge  $l$  mit  $m$  oder mehr Treffern auftritt, ist  $P(l, m) = \sum_{i=m}^l \binom{l}{i} p^i (1-p)^{l-i}$  mit  $p = 1/4$  (Nukleotide). Das Gewicht eines Segments sei  $w_D = -\log_{10}(P(l, m))$ .

- Welche Segmentlängen  $l$ ,  $4 \leq l \leq 5$ , mit wie vielen Treffern  $m$  ermöglichen ein Gewicht  $> 1,5$ ?
- Finden Sie alle paarweisen Segmente der Länge 4 bis 5 und Gewicht  $> 1,5$ . Zum Visualisieren und Finden der Segmente empfehlen sich Dotplots.