

# Klausur Algorithmische Bioinformatik

FSU Jena, SS 2019

Name:

Matrikelnummer:

## Paarweises Sequenzalignment

1. Gegeben zwei Sequenzen  $A = (A_1, \dots, A_m)$  und  $B = (B_1, \dots, B_m)$  und eine Scorematrix  $S_{ij}$ . Geben Sie Initialisierungs- und Rekursionsformel und den Start des Tracebacks für den Smith-Waterman Algorithmus mit linearen Gapkosten an.
2. Was muss am Needleman-Wunsch-Algorithmus geändert werden, wenn man im Alignment Gaps am Ende nicht bestrafen will?
3. Wie berechnet man einen DotPlot, so dass die Laufzeit unabhängig von der Fenstergröße ist?
4. Wie berechnet man das optimale Needleman-Wunsch Alignment durch jeden festen Punkt  $(k, l)$  ohne die asymptotische Laufzeit zu verändern?
5. Wieviele Alignments zweier Sequenzen der Länge vier gibt es? Dabei zählen wir die Alignments  $\begin{array}{cc} A & - \\ - & B \end{array}$  und  $\begin{array}{cc} - & A \\ B & - \end{array}$  getrennt.

## Alignmentstatistik

1. Geben Sie die Varianz einer Poisson-verteilten Zufallsvariable mit dem Erwartungswert  $\lambda$  an.
2. Geben Sie ein Beispiel für eine binomialverteilte Zufallsvariable an. Welche Werte kann die Zufallsvariable annehmen?
3. Was ist und wie funktioniert die Momentenmethode (Method of Moments)?
4. Für eine feste Wahl der Match- bzw. Mismatchkosten hat die Karlin-Altschul-Statistik einen bestimmten Gültigkeitsbereich für die Gapkosten. Wie wächst der Score in Abhängigkeit von der Länge der Sequenz in diesem Bereich?

## Multiple Alignments, Markov-Modelle

1. Die Zustandsfolge einer homogenen Markovkette (erster Ordnung) mit drei Zuständen ist gegeben mit *ACBAABAACCACB*. Schätzen Sie die Übergangswahrscheinlichkeiten für *CB* und *BA*.
2. Generiert eine Markovkette (erster Ordnung) über dem Basenalphabet im allgemeinen eine iid Sequenz und warum? 'iid' steht dabei für 'identically and independently distributed'.
3. Wie werden im Clustal-Algorithmus die multiplen Alignments berechnet (Prinzip)?