

# Klausur zum Modul

## Algorithmen in der Bioinformatik

Sommersemester 2018  
25.07.2018

Name: .....

Matrikelnummer: .....

Studiengang: .....

**Geben Sie den Lösungsweg immer mit an!**

Nur mit blauem oder schwarzem Kugelschreiber schreiben.

Schreiben Sie auf jeden Zettel Ihre Matrikelnummer.

Geben Sie für jede (Teil-)Aufgabe nur eine einzige Lösung ab. Bei mehreren, alternativen Lösungen zu einer Aufgabe wird die Schlechteste bewertet.

Teilnahme an der Klausur erfolgt unter Vorbehalt einer vorhandenen Zulassung.

Aufgabe Nr.:	Punktzahl:	Davon erreicht:
1	17	
2	10	
3	15	
4	8	
$\Sigma$	50	

**Es sind keinerlei Hilfsmittel erlaubt. Bitte schreiben Sie deutlich mit einem schwarzen oder blauen Stift.**

## 1. Alignments

- a) Geben Sie die Dynamische Programmierungsmatrix mit Backtrackingpointern an, um optimale globale Alignments zwischen den Sequenzen GCTA und CCA zu bestimmen. Ein Match wird mit +3 bewertet, ein Mismatch und ein Gap werden mit jeweils  $-1$  bestraft.

9

Geben Sie die optimalen Alignments und deren Scores an.

- b) Was muss am Needleman-Wunsch-Algorithmus geändert werden, wenn man im Alignment Gaps am Ende der Sequenzen nicht bestrafen will?
- c) Im Allgemeinen: Wie verhält sich der Score eines optimalen lokalen Alignments gegenüber dem Score eines optimalen globalen Alignments bei gleichem Scoring-schema? Begründen Sie Ihre Antwort.

4

4

## 2. Assembly

Gegeben ist der String XYZWVYZSTYZU.

- a) Zeichnen Sie den De-Bruijn-Graphen  $G$ , dessen Kanten alle 3-mere des Strings sind.
- b) Finden Sie einen Eulerweg in  $G$ . Wieviele verschiedene Eulerwege gibt es in dem Graphen? Zu welchen Strings korrespondieren diese Wege?

6

4

## 3. Suffix-Bäume

Gegeben seien ein Suffixbaum  $\mathcal{T}$  für einen Text  $T$  der Länge  $n$  über dem Alphabet  $\{C, G, T, A\}$  sowie ein Pattern  $p$  der Länge  $m$ . Gesucht ist die Menge aller Matchpositionen von  $p$  in  $T$ , d. h. die Indizes  $M_p = \{i \mid T[i, \dots, i + m - 1] = p\}$ .

- a) Beschreiben Sie einen Algorithmus, der diese Aufgabe in Zeit besser als  $O(nm)$  löst, und schätzen Sie dessen Laufzeit ab.
- b) Nun genügt es, *eine* Matchposition  $j$  aus  $M_p$  zu bestimmen. Beschreiben Sie einen effizienten Preprocessing-Algorithmus auf  $\mathcal{T}$ , so dass  $j$  in Zeit  $O(m)$  gefunden werden kann.

8

7

## 4. Phylogenetische Bäume

8

Gegeben ist folgende Distanzmatrix  $D$  für drei Blätter  $A$ ,  $B$  und  $C$ :

$$D = \begin{pmatrix} 0 & 7 & 10 \\ 7 & 0 & 11 \\ 10 & 11 & 0 \end{pmatrix}$$

Zur Erinnerung: Eine Distanzmatrix ist genau dann additiv, wenn ein kantengewichteter Baum existiert, so dass für jedes Paar  $i, j$  von Blättern der Baumabstand gleich dem Abstand in der Matrix ist.

Ist obige Matrix  $D$  additiv? Beweisen Sie Ihre Antwort.