

Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SS 2011
Prof. Dr. Jens Stoye · Daniel Dörr

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011summer/SequenzAnalyse>

Blatt 9 vom 08.07.2011

Abgabe entfällt

Aufgabe 1 (MUMs und MEMs in Suffixbäumen)

4 Punkte

Berechne den generalisierten Suffixbaum der folgenden Sequenzen:

$$s_1 = \text{GCATAAATAAA}, s_2 = \text{GAATAAATAGA} \text{ und } s_3 = \text{GAATAAGTAAA}$$

- Finde mit Hilfe des generalisierten Suffixbaums alle *maximal unique matches* (MUMs) und gib sie an.
- Ermittle alle *maximal multiple exact matches* (*multi-MEMs*) und gib sie an.

Aufgabe 2 (Multiples Genomalignment)

4 Punkte

- a) Diskutiere die Vor- und Nachteile von filterbasierten multiplen Genomalignment-Algorithmen.
- b) Was ist die grundsätzliche Problematik bei Programmen wie MGA oder MUMmer bzw. welchen Vorteil bietet MAUVE?
- c) Verallgemeinere den Graphen basierend auf der partiellen Ordnung von MUMs (siehe Abschnitt 16.2 im Skript Seite 175) für multiMUMs in globalem *multiplen* Genomalignment.
- d) Gegeben sei ein Graph über der partiellen Ordnung von multiMUMs in n Sequenzen. Schlage einen Algorithmus vor, welcher diesen Graphen aktualisiert wenn ein zusätzliches $n + 1$ -tes Genom zum Alignment hinzugefügt wird.