Prof. Dr. Knut Reinert
Dr. Roland Krause
Matthias Winkelmann
Patrick Pett

Institut für Informatik
AG Algorithmische Bioinformatik

Algorithmische Bioinformatik
Erstes Übungsblatt WS 11/12

Abgabe immer am Tag vor der Übung

Name: Matrikelnummer:

Übungsgruppe: 1☐ 2☐ 3☐

Ich kann Aufgabe ☐ nicht vorrechnen.

Aufgabe 1: Anzahl ungezeichnter Bäume

Betrachten wir nur Topologien der Bäume, ignorieren also die Auszeichnung der Blätter.Wie viele Blätter benötigt ein es mindestens für *ungewurzelte* Bäume mit mehr als einer Topologie? Wie viele Blätter benötigt ein es mindestens für *gewurzelte* Bäume mit mehr als einer Topologie? Finden Sie die Rekursionsgleichung für die Anzahl der Topologien von gewurzelten Bäumen.

Aufgabe 2: Parsimony

Gegeben ist das folgende Alignment und die Bäume $T_1 = ((A, B), C, (D, E))$ und $T_2 = ((A, E), B, (C, D)$.

b)

```
A     AGTCATTTGATTAC
B     AGAGAGTTCGATTC
C     AGCCGTTTCGATTC
D     AGAGGGTTCGACTC
E     ATACGGTTCGACTT
```

a) Welche Spalten wären sicher nicht informativ, wenn Sie den Parsimony-Algorithmus nach *Fitch* anwenden würden?

b) Welche Spalten sind möglicherweise informativ, wenn Sie den Parsimony-Algorithmus nach *Sankoff* anwenden würden und die verwendete Kostenfunktion c nicht kennen?

c) Finden Sie den Baum mit dem kleinsten Parsimony-Score. Benutzen Sie den Sankoff-Algorithmus mit einer Kostenfunktion von $c(a_v, a_w) = 1 \forall a \in A$

Begründen Sie Ihre Antworten.

Aufgabe 3: Hierarchisches Clustering

Implementieren Sie hierarchisches Clustering einer Distanzmatrix.

a) für single linkage clustering.

b) für UPGMA.

c) Überlegen Sie sich die Komplexität Ihrer Lösung.

Sie können eine Programmiersprache Ihrer Wahl benutzen.