

# Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2014/15

Martin Vingron · Juliane Perner · Annkatrin Bressin

**Blatt 1 · Ausgabe am 20.10.2014**

**Abgabe am 27.10.2014 vor Beginn der Vorlesung**

Name:

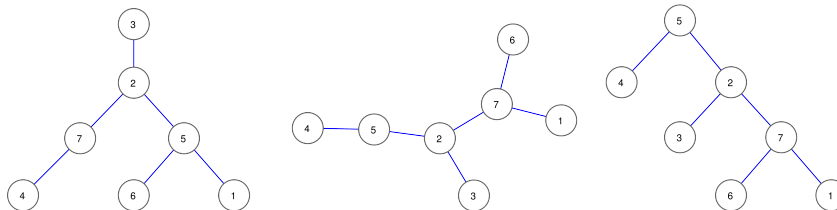
Matrikelnummer:

Übungsgruppe:

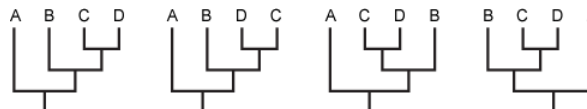
**Aufgabe 1** (30 Punkte; Theorie). In folgender Aufgabe sollen Sie sich mit der graphischen Darstellung, der Implementierung und der Abzählbarkeit von Bäumtopologien beschäftigen.

1. Entscheiden und begründen Sie welche der Bäume in den folgenden zwei Gruppen äquivalent sind.

- Gruppe A:



- Gruppe B:



2. Beweisen Sie die in der Vorlesung vorgestellte Formel zur Berechnung der Anzahl an unterschiedlichen Topologien von nicht gewurzelten binären Bäumen mit  $n$  Blättern:

$$U_n = \prod_{i=3}^n (2i - 5)$$

3. Machen Sie Vorschläge für Datenstrukturen um phylogenetische Bäume zu implementieren. Welche Operationen und Suchanfragen wird man beim Arbeiten mit phylogenetischen Bäumen häufig brauchen?

**Aufgabe 2** (20 Punkte; Theorie). Gegeben seien fünf Taxa a, b, c, d, e und das folgende multiple Alignment von Nukleotidsequenzen mit drei Spalten bzw. Merkmalen.

	1	2	3
a	G	G	A
b	T	A	A
c	A	A	A
d	G	G	A
e	C	T	T

sowie eine Baumtopologie (im Newick Format):  $((a, b), (e, (c, d)))$ .

- Bestimmen Sie mit Hilfe des in der Vorlesung vorgestellten Parsimonie-Algorithmus für obige Baumtopologie eine Belegung der inneren Knoten mit Nukleotiden, so dass der Baum minimale Länge hat.
- Welche Spalten des Alignments sind in Bezug auf Maximum Parsimony phylogenetisch informativ? Informativ bedeutet hier, dass die Spalte eine Entscheidung zwischen verschiedenen Baumtopologien ermöglicht.

**Aufgabe 3** (10 Punkte; Programmieren). Berechnen Sie die Jukes-Cantor Korrektur der berechneten Distanzen aus der 1. Aufgabe von Übungsblatt 1 und vergleichen Sie diese graphisch mit der Hamming-Distanz. Was beobachten Sie?

**Aufgabe 4** (40 Punkte; Theorie). In der Mensa gibt es jeden Tag entweder ein Fleischgericht ( $M$ ), ein Fischgericht ( $F$ ) oder ein Gemüsegericht ( $V$ ). Aus langjähriger Beobachtung weiß der Student, dass mit einer gewissen Wahrscheinlichkeit zum Beispiel auf einen Fleischtag ein Fischtage ( $P(F|M)$ ) oder ein Gemüsetag ( $P(V|M)$ ) folgt. Er hat die folgenden Übergangswahrscheinlichkeiten für alle möglichen Kombinationen  $P_{i,j}, i, j \in \{M, F, V\}$  an Gerichten beobachtet:

$$P(V|M) = 0.1, P(F|M) = 0.1, P(M|V) = 0.03,$$

$$P(F|V) = 0.02, P(M|F) = 0.2, P(V|F) = 0.75$$

- Modellieren Sie die Essensausgabe als Markovkette erster Ordnung und geben Sie die Transitionsmatrix  $P$  an. Skizzieren Sie die Markovkette.
- Berechnen Sie den Wahrscheinlichkeitsvektor  $\pi$  des stationären Zustandes.
- Sie gehen an einem beliebigen Tag in die Mensa und wissen nicht, was für ein Essen es am vorherigen Tag gab. Mit welcher Wahrscheinlichkeit wird Gemüse angeboten?
- Sie gehen am Montag in die Mensa und wissen nicht, was es am Freitag zu Essen gab. Mit welcher Wahrscheinlichkeit gibt es diese Woche  $VVVMMF$ ?
- Wenn es an einem Tag Fleisch gab, mit welcher Wahrscheinlichkeit gibt es dann welches Essen vier Tage später?