



Einführung in die Bioinformatik

Prof. Dr. Enno Ohlebusch, Prof. Dr. Hans Kestler, Dr. Karlheinz Holzmann,
Tobias Badura

WS 16/17

Übungsblatt 4

Abgabe bis 25.01.2017, 12:00. Lösungen bitte elektronisch (an tobias.badura@uni-ulm.de) mit Namen in Text und Quelltextdateien abgeben.

1. Aufgabe (5): UPGMA

Gegeben ist eine Matrix mit den paarweisen Distanzen von 4 Organismen. Bestimmen Sie den Baum, der sich mit dem UPGMA-Verfahren ergibt.

D	a	b	c	d
a	0			
b	3	0		
c	6	7	0	
d	5	6	3	0

- Welche paarweisen Distanzen ergeben sich durch die Kantenlängen?
- Zeichnen Sie einen Baum der die paarweisen Distanzen korrekt widerspiegelt.

2. Aufgabe (10)

Gegeben ist eine DNA-Sequenz S des Salmonella-Genoms:

```
1 atgctgattc tgactcgtcg agttggtgag accctcatga ttggcgatga ggtcaccgtg
61 acagtttttag ggggtgaaggg caaccagggtg cgtattggcg tgaacgcccc gaaagaagtt
121 tctgtccatc gtgaagagat ctaccagcgt atccaggctg aaaaatccca gcagtcagtt
181 tactaa
```

1. Bestimmen Sie für alle $a \in \{A, C, G, T\}$ die relative Häufigkeit p_a von a in S

$$p_a = \frac{\text{Anzahl der } a \text{ in } S}{|S|},$$

wobei $|S|$ die Länge von S (Anzahl der Basen) ist.

2. Bestimmen Sie für alle $a, b \in \{A, C, G, T\}$ die Erwartungswerte $E[X_{a,b}] = p_a \cdot p_b \cdot |S|$ für die Anzahl der Vorkommen der Dinucleotide.
3. Vergleichen Sie die Erwartungswerte $E[X_{a,b}]$ mit der Anzahl der Vorkommen der Dinucleotide in S .
4. Bestimmen Sie für alle $a, b, c \in \{A, C, G, T\}$ die relative Häufigkeit $p_{a,b,c}$ der Codons abc .
5. Vergleichen Sie die Erwartungswerte $E[X_{a,b,c}] = p_a \cdot p_b \cdot p_c \cdot |S|$ mit der tatsächlichen Anzahl der Codons.

3. Aufgabe (10): Biologische Grundlagen Teil 3

- (a) Wie und wozu werden die Nucleotide bei einem Microarray-Experimentes markiert?
- (b) Wieso sollten aus einem Microarray-Experiment gewonnene Daten normalisiert werden?
- (c) Welche Normalisierungsverfahren für Microarrays kennen Sie?
- (d) Beschreiben Sie ein Normalisierungsverfahren für Microarrays.
- (e) Beschreiben Sie stichwortartig den Ablauf eines cDNA Microarray-Experimentes.