

Übungen zur Vorlesung Algorithmische Bioinformatik

Freie Universität Berlin, WS 2016/17

Martin Vingron · Annalisa Marsico · Alena van Bömmel · Edgar Steiger · Thimo Wellner

Blatt 3 · Ausgabe am 31.10.2016

Abgabe am 7.11.2016 vor Beginn der Vorlesung

Name:

Matrikelnummer:

Übungsgruppe:

Aufgabe 1 (40 Punkte; Theorie/Rechnen). Gegeben sind die folgenden Distanzmatrizen:

(a)		a	b	c	d	(b)		a	b	c	d	e
	a	0	8	9	5		a	0	8	8	6	3
	b		0	9	8		b		0	4	7	8
	c			0	9		c			0	7	8
	d				0		d				0	6
							e					0

1. Überprüfen Sie, ob diese Matrizen jeweils eine additive Metrik und/oder Ultrametrik repräsentieren.
2. Wenden Sie auf beide Matrizen jeweils Single Linkage Clustering und Neighbour Joining an. Ist das Clustering in allen Fällen eindeutig? Haben Sie eine Erklärung dafür?

Aufgabe 2 (15 (+10) Punkte; Praxis). Installieren Sie die Software PHYLIP¹ zum Rekonstruieren von phylogenetischen Bäumen. Nutzen Sie PHYLIP einmal mit dem Maximum-Parsimony- und einmal mit dem Maximum-Likelihood-Algorithmus um phylogenetische Bäume für die mitochondrialen DNA-Sequenzen² zu konstruieren. Geben Sie die resultierenden Bäume im Newick-Format an und kommentieren Sie kurz die Ergebnisse.

Bonus-Aufgabe: Installieren Sie außerdem die Software RAxML³ und berechnen Sie mit dieser einen phylogenetischen Baum für die mitochondrialen DNA-Sequenzen mit dem Maximum-Likelihood-Algorithmus. Vergleichen Sie mit den Resultaten von PHYLIP (zum Beispiel, indem Sie die Bäume wie in Aufgabe 3 von PHYLIP zeichnen lassen).

Aufgabe 3 (30 Punkte; Programmieren). Erstellen Sie ein Programm, welches...

1. ein Alignment im FASTA-Format einliest.
2. aus diesem Alignment und mittels der Jukes-Cantor-Korrektur mit normierter Hamming-Distanz eine Distanzmatrix erstellt. Betrachten Sie dabei die Gaps „-“ als gleichwertige Mutation.
3. aus dieser Distanzmatrix mit dem UPGMA-Algorithmus einen phylogenetischen Baum im Newick-Format erzeugt.
4. mit folgendem Befehl aufrufbar ist: `program_name inputfile outputfile`

¹<http://evolution.genetics.washington.edu/phylip.html>

²Material 1: <https://www.molgen.mpg.de/3698377/mitodna.txt>

³<http://sco.h-its.org/exelixis/web/software/raxml/>

Testen Sie Ihr Programm anhand des auf der Vorlesungsseite gegebenen Alignmentfiles⁴ und zeichnen Sie den resultierenden Baum mit PHYLIP. Dazu benötigen Sie ein fontfile, welches Sie im src-Ordner von PHYLIP finden.

Aufgabe 4 (15 Punkte; Programmieren). In der jeweils 3. Aufgabe vom Übungsblatt 1 und 2 haben Sie eine zufällige DNA-Sequenz und Mutationen in 10.000 Generationen generiert sowie die Entwicklung der Hamming-Distanz D berechnet und grafisch dargestellt.

- A) Berechnen Sie in jeder Generation die Jukes-Cantor-Korrektur d der berechneten Distanzen. Wie muss der Parameter α der Ratenmatrix gewählt werden, damit er zu unserem Modell passt?
- B) Vergleichen Sie die Jukes-Cantor-Korrektur grafisch mit der Hamming-Distanz D . Was beobachten Sie?

⁴Material 2: <https://www.molgen.mpg.de/3698386/test88.txt>