

Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2014/2015, Dr. Roland Wittler, Kevin Lamkiewicz

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014winter/Phylogenetik>

Blatt 8 vom 26.11.2014

Abgabe in einer Woche zu Beginn der Vorlesung oder vorab bei deinem Tutor oder beim Veranstalter.

Aufgabe 1 Rekonstruktion additiver Bäume.

(4 Punkte)

Die nebenstehende Distanzmatrix ist *additiv*.

Rekonstruiere den entsprechenden Baum mit Hilfe des Algorithmus von Waterman (Skript, Abschnitt 7.3.1).

Gehe lixikographisch vor, d.h. beginne mit der Kante $\{A, B\}$, füge dann die Taxa C , D , und E ein. Wähle als Blattpaar zum Einfügen des jeweils nächsten Taxons zunächst immer $\{A, B\}$, bevor du gegebenenfalls auf ein anderes Paar ausweichst.

Gib alle Zwischenschritte an.

| | <i>A</i> | <i>B</i> | <i>C</i> | <i>D</i> | <i>E</i> |
|------------|----------|----------|----------|----------|----------|
| <i>A</i> : | 0 | 11 | 7 | 9 | 10 |
| <i>B</i> : | | 0 | 8 | 8 | 9 |
| <i>C</i> : | | | 0 | 6 | 7 |
| <i>D</i> : | | | | 0 | 3 |
| <i>E</i> : | | | | | 0 |

Aufgabe 2 Neighbor Joining.

(6 Punkte)

Verwende *Neighbor Joining* (siehe Skript, Abschnitt 7.3.4), um einen phylogenetischen Baum aus der folgenden Matrix zu rekonstruieren. Gib alle Zwischenschritte an.

| | <i>A</i> | <i>B</i> | <i>C</i> | <i>D</i> | <i>E</i> |
|------------|----------|----------|----------|----------|----------|
| <i>A</i> : | 0 | 8 | 8 | 6 | 6 |
| <i>B</i> : | | 0 | 4 | 8 | 4 |
| <i>C</i> : | | | 0 | 8 | 4 |
| <i>D</i> : | | | | 0 | 6 |
| <i>E</i> : | | | | | 0 |

Bitte wenden!

Aufgabe 3 Fitch-Margoliash.

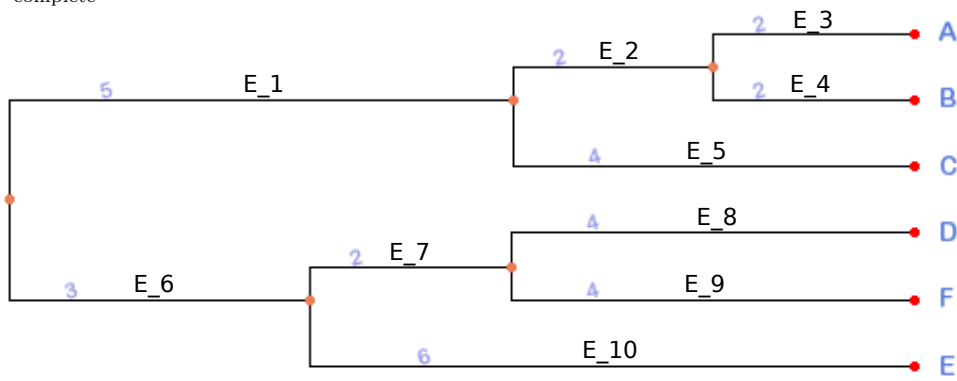
(4 Punkte)

Auf Übungsblatt 7 wurden *Complete linkage* und *UPGMA* als Clustering-Verfahren zur Baumrekonstruktion benutzt. Unten ist je ein möglicher Baum abgebildet, wenn die gegebene Matrix d^M zur Rekonstruktion verwendet wird.

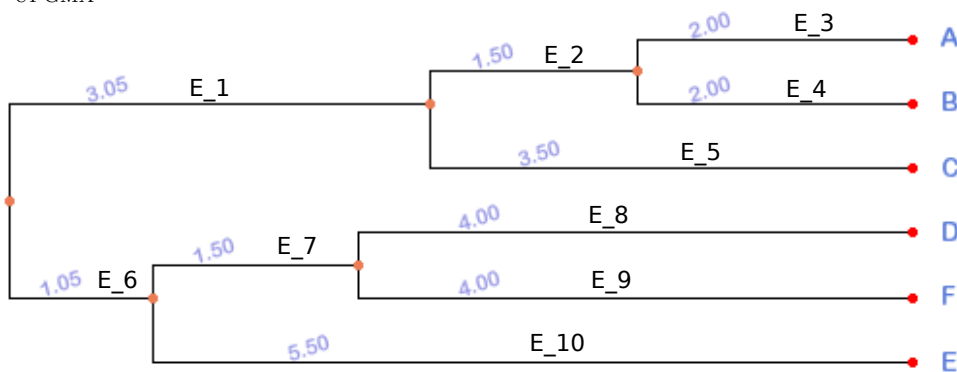
$$d^M :=$$

| | A | B | C | D | E | F |
|----|---|---|---|----|----|----|
| A: | 0 | 4 | 8 | 18 | 18 | 6 |
| B: | | 0 | 6 | 12 | 8 | 8 |
| C: | | | 0 | 18 | 18 | 12 |
| D: | | | | 0 | 10 | 8 |
| E: | | | | | 0 | 12 |
| F: | | | | | | 0 |

$\mathcal{T}_{\text{complete}} =$



$\mathcal{T}_{\text{UPGMA}} =$



Berechne den *Least Squares* Fehler $E := \|\vec{d}^{\mathcal{T}} - \vec{d}^M\|^2$ (nach Fitch und Margoliash) für beide Bäume. Dazu gib zuerst d^M in Vektorschreibweise an. Danach führe die folgenden Schritte jeweils für beide Bäume durch:

- Schreibe $M^{\mathcal{T}}$ und \vec{w} bezüglich \mathcal{T} auf. (Diesen Schritt darfst du für $\mathcal{T}_{\text{UPGMA}}$ auslassen.)
Tipp: Fasse die zwei Kanten, die inzident mit der Wurzel sind, zu einer Kante zusammen.
- Berechne daraus $\vec{d}^{\mathcal{T}}$. (Bzw. lies die Werte für $\mathcal{T}_{\text{UPGMA}}$ aus dem Baum ab.)
- Berechne schließlich $E := \|\vec{d}^{\mathcal{T}} - \vec{d}^M\|^2$.

Welcher der beiden Bäume ist „besser“?