

# Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2013/2014, Dr. Roland Wittler, M.Sc. Kai Stadermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2013winter/Phylogenetik>

## Blatt 8 vom 2.12.2013

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

### Aufgabe 1 Eigenschaften von Distanzen.

(4 Punkte)

Bei jeder Distanzfunktion  $d$  gilt der folgende Zusammenhang:

$$„d \text{ ist ultrametrisch}“ \Rightarrow „d \text{ ist additiv}“ \Rightarrow „d \text{ erfüllt die Dreiecksungleichung}“$$

Beweise diesen Zusammenhang, indem du die folgenden Punkte zeigst:

- (a) „ $d$  genügt der *Drei-Punkt-Bedingung*“  $\Rightarrow$  „ $d$  genügt der *Vier-Punkt-Bedingung*“

Beweisskizze: Für beliebige vier Elemente, gehen wir wie folgt vor. Wir benennen zwei der Elemente, die minimalen Abstand haben mit  $a$  und  $b$ . Wir benennen die beiden anderen Elemente  $c$  und  $d$ , so dass gilt  $d_{ac} \leq d_{ad}$ . Dann können  $a, b, c$  und  $d$  in genau zwei verschiedenen binären, ultrametrischen Bäumen angeordnet werden. **Welche?** In jedem der zwei Fälle gelten nun mehrere Drei-Punkt-Bedingungen (**Welche?**), die zur gesuchten Vier-Punkt-Bedingung kombiniert/umgeformt werden können. **Wie?** (Der Fall, dass die Distanzen einer nicht-binären Topologie entsprechen, muss **nicht** gesondert betrachtet werden. Diesen erfassen wir, indem wir Kanten der Länge Null erlauben. Zum Beispiel stellen wir  $((a:2,b:2,c:2):1,d:3)$ ; binär als  $((a:2,b:2):0,c:2):1,d:3)$ ; dar.)

- (b) „ $d$  genügt der *Vier-Punkt-Bedingung*“  $\Rightarrow$  „ $d$  erfüllt die *Dreiecksungleichung*“

Tipp: Die Vier-Punkt-Bedingung kann auch auf drei Punkte angewendet werden.

### Aufgabe 2 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.

	A	B	C	D	E
A :	0	11	7	9	10
B :		0	8	8	9
C :			0	6	7
D :				0	3
E :					0

Bitte wenden!

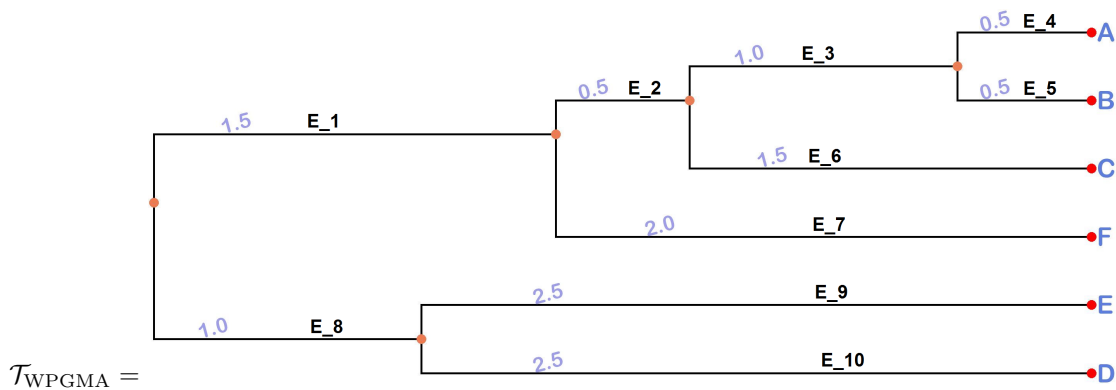
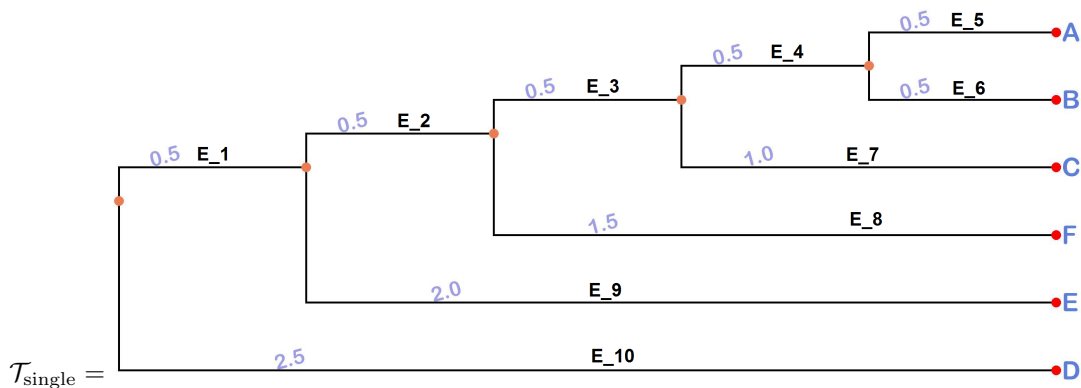
### Aufgabe 3 Fitch-Margoliash.

(4 Punkte)

Auf Übungsblatt 7 wurden *Complete linkage* und *UPGMA* als Clustering-Verfahren zur Baumrekonstruktion benutzt. Werden anstatt dessen *Single linkage* und *WPGMA* Clustering verwendet, kommen zu der gegebenen Matrix  $d^M$  die folgenden Bäume  $\mathcal{T}_{\text{single}}$  und  $\mathcal{T}_{\text{WPGMA}}$  heraus:

$$d^M :=$$

	A	B	C	D	E	F
A:	0	1	4	9	9	6
B:		0	2	9	9	4
C:			0	9	9	3
D:				0	5	6
E:					0	4
F:						0



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

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

Welcher der beiden Bäume ist „besser“?