Universität Bielefeld
Technische Fakultät

AG Genominformatik
Dr. Katharina Jahn
M.Sc. Christina Ander

Vorlesung: Phylogenetik
Wintersemester 2010/2011

Übungen

Übung 8, Abgabe: 23.12.2010

1. **Ultrametriken** **(4 Punkte)**
   Ist eine Pfadmetrik d^T gegeben, die die Entfernung von n Taxa zueinander angibt, dann gilt:
   Beschreibt d^T eine Ultrametrik, dann gibt es maximal $n-1$ unterschiedliche Einträge $d^T_{i,j}$ für alle $i \neq j$.

   (a) Begründe, dass diese Aussage gilt. (Hinweis: Überlege für jeden inneren Knoten eines ultrametrischen Baumes welche Distanzen die Taxa, die diesen Knoten als LCA haben, haben können.)

   (b) Finde ein Gegenbeispiel mit vier Taxa, das zeigt, dass der Umkehrschluss nicht gilt. (Umkehrschluss: "Jede Pfadmetrik mit maximal $n-1$ unterschiedlichen Einträgen $d^T_{i,j}$ für alle $i \neq j$ beschreibt eine Ultrametrik")

2. **Rekonstruktion additiver Bäume.** **(4 Punkte)**
   Die folgende Distanzmatrix ist *additiv*:

	A	B	C	D	E	F
A	0	9	7	9	7	9
B		0	5	5	9	5
C			0	5	7	5
D				0	9	3
E					0	9
F						0

   Rekonstruiere den entsprechenden additiven Baum mit Hilfe des Algorithmus von Waterman (Skript, Abschnitt 6.3.1). Gib alle Zwischenschritte an.

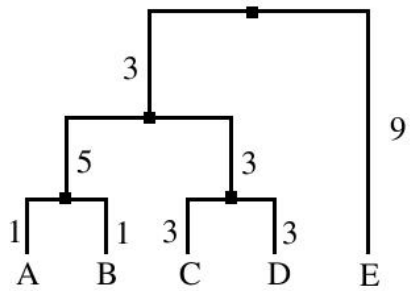
(Bitte umblättern.)

3. **Fitch-Margoliash.** **(6 Punkte)**
Auf dem Übungsblatt 7 wurde *UPGMA* als Clustering-Verfahren zur Baumrekonstruktion benutzt. Werden anstatt dessen *Complete linkage* und *WPGMA* Clustering verwendet, kommen zu der gegebenen Matrix d^M die folgenden Bäume $\mathcal{T}_{\text{complete}}$ und $\mathcal{T}_{\text{WPGMA}}$ heraus:

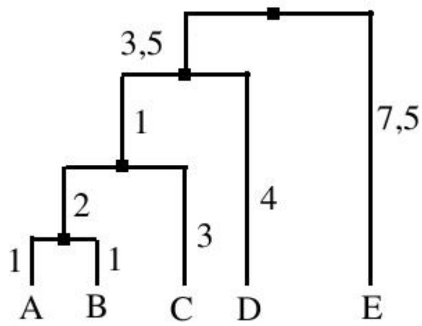
$d^M :=$

	A	B	C	D	E
A	0	2	8	12	18
B		0	4	8	18
C			0	6	18
D				0	12
E					0

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



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



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:
$\mathcal{T} \in \{\mathcal{T}_{\text{complete}}, \mathcal{T}_{\text{WPGMA}}\}$

1. Schreibe $M^{\mathcal{T}}$ und $\vec{w}$ bezüglich $\mathcal{T}$ auf.
2. Berechne daraus zunächst $\vec{d^{\mathcal{T}}}$.
3. Berechne schließlich $E := ||\vec{d^{\mathcal{T}}} - \vec{d^M}||^2$.

Welcher der beiden Bäume ist „besser"?