

Vorlesung: Phylogenetik  
Wintersemester 2010/2011

Übungen

Übung 5, Abgabe: 02.12.2010, bis 10 Uhr

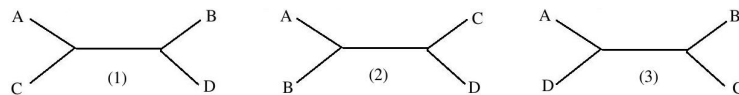
1. Maximum Parsimony Branch-and-Bound

(3 Punkte)

Gegeben sei die nebenstehende Matrix.

Gesucht ist ein **binärer** Baum mit der minimalen Benennung der inneren Knoten unter Berücksichtigung der **Einheitskosten**. Für diese vier Taxa gibt es die möglichen Topologien:

	1	2	3	4
A	a	e	$\beta$	$\beta$
B	b	f	$\alpha$	$\beta$
C	a	e	$\beta$	$\alpha$
D	b	f	$\alpha$	$\gamma$



- (a) Nutze den **Column-wise Branch-and-Bound Algorithmus** von Seite 32 um aus den möglichen Topologien die mit den minimalen Kosten zu berechnen. Gib dabei alle Teilschritte an und mache deutlich, wann eine bestimmte Topologie nicht mehr weiter untersucht werden muss. (Hinweis: Die Reihenfolge der vorgegebenen Topologien soll euch diese Aufgabe erleichtern.)
- (b) Bildet der gefundene Baum eine perfekte Phylogenie? Begründe deine Entscheidung.

2. Greedy Sequential Addition.

(4 Punkte)

Betrachte die *Greedy Sequential Addition*-Heuristik (GSA) zur Annäherung einer Lösung des *Maximum Parsimony Problems*. Der Algorithmus verfährt für gegebene Taxa  $t_1, t_2, t_3$  bis  $t_n$  wie folgt:

$GSA([t_1, t_2, t_3, \dots, t_n])$

1. construct tree  $T_3 := (t_1, t_2, t_3)$ ;
2. for  $i = 4 \dots n$ :
  - a. for each edge  $e$  in  $T_{i-1}$ :
    - i. connect taxon  $t_i$  to edge  $e$
    - ii. compute the weight of this tree
  - b. select  $T_i$  with minimal weight among all above considered trees of size  $i$ .
3. return  $T_n$

Abbildung 1 zeigt den gesamten Suchraum für ein Beispiel mit  $n = 5$  Taxa:  $A$ ,  $B$ ,  $C$ ,  $D$  und  $E$ . Im ersten Durchlauf der Schleife (Zeile 2 des Pseudocodes) werden beim Hinzufügen des 4. Taxons drei Bäume erzeugt und deren Kosten berechnet. Ein Baum mit minimalen Kosten wird ausgewählt und im nächsten Schleifendurchlauf werden ausgehend von diesem Baum fünf weitere analysiert.

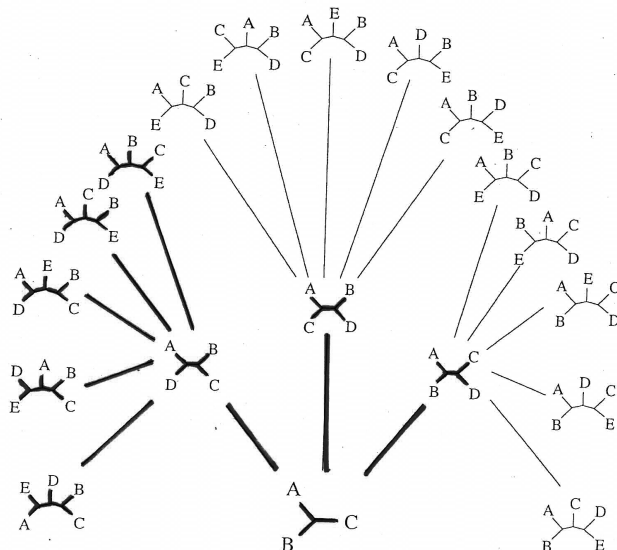


Abbildung 1: Die fettgedruckten Bäume werden betrachtet, die anderen Bäume werden unter Verwendung des *GSA*-Algorithmus nicht berechnet<sup>1</sup>.

- (a) Wie viele Bäume werden allgemein beim Hinzufügen des  $i$ -ten Taxons zunächst berücksichtigt, bevor einer ausgewählt wird? Erkläre diese Anzahl. (Tipp: Der Beweis auf Seite 13 im Skript kann euch einen entscheidenden Hinweis liefern.)
- (b) Wie viele Bäume werden in allen Schritten  $i = 4 \dots n$  insgesamt betrachtet? Zeige, dass diese Anzahl quadratisch mit  $n$  zunimmt. (Tipp: Verwende die Formel des kleinen Gauß.)

### 3. Minimum Spanning Tree

(3 Punkte)

Ein minimaler Spannbaum  $T$ , in einem Graph  $G = (V, E)$  hat die Eigenschaft: Sei  $e$  eine Kante in  $T$ , die  $T$  in zwei Teilbäume  $T_1$  und  $T_2$  aufteilt. Dann hat  $e$  das geringste Gewicht aller Kanten aus  $E$ , die in  $G$  einen Knoten aus  $T_1$  und einen Knoten aus  $T_2$  miteinander verbinden. (Skript Seite 35)

Beweise diese Eigenschaft.

<sup>1</sup>Aus: J. Felsenstein, *Inferring Phylogenies*, 2004