

Übungen zur Phylogenetik Vorlesung

Universität Bielefeld, WS 2012/2013, Dr. Roland Wittler
<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2012winter/Phylogenetik>

Blatt 5 vom 15.11.2012

Abgabe in einer Woche zu Beginn der Vorlesung oder vorab im Briefkasten bei U10-151.

Aufgabe 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. ((A,C),(B,D));
2. ((A,B),(C,D));
3. ((A,D),(B,C));

	1	2	3	4
A	α	γ	ϵ	η
B	β	δ	ζ	η
C	α	γ	ϵ	θ
D	β	δ	ζ	θ

Nutze den *Column-wise Branch-and-Bound* Algorithmus 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 Reihenfolgen der vorgegebenen Topologien und Merkmale sollen dir diese Aufgabe erleichtern.)

Aufgabe 2 Greedy Sequential Addition.

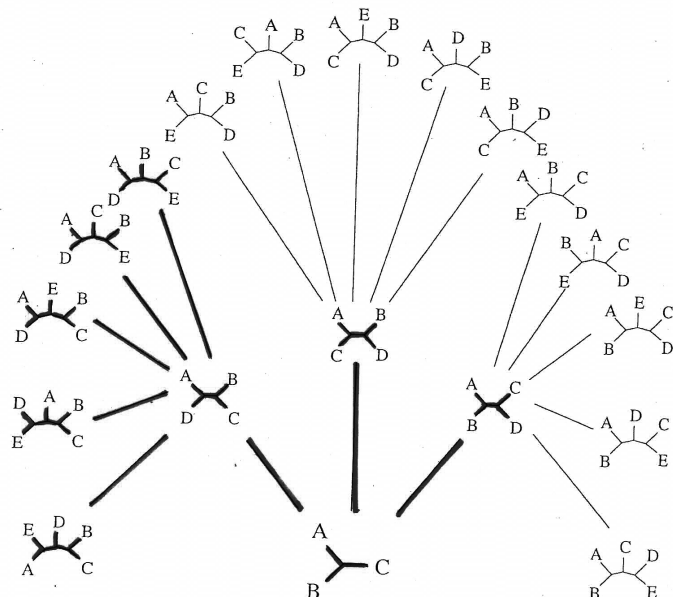
(3 Punkte)

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

Die untenstehende Abbildung 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. Die betrachteten Bäume sind in der Abbildung hervorgehoben.



- (a) Wie viele Bäume werden allgemein beim Hinzufügen des i -ten Taxons zunächst berücksichtigt bevor einer ausgewählt wird?
- (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.)

Aufgabe 3 Nearest Neighbor Interchange.

(3 Punkte)

Verwende die *Nearest Neighbor Interchange*-Heuristik (NNI) zur Annäherung einer Lösung des *Maximum Parsimony Problems*. Der Algorithmus verfährt für einen gegebenen Baum T wie folgt:

NNI(T)

1. compute $W(T)$, the parsimony score of tree T
2. compute neighbors $T_1, \dots, T_n$ of T
3. compute score $W(T_i)$ for all neighbors $T_i, 1 \leq i \leq n$
4. if ($W(T_i) \geq W(T)$ for all $1 \leq i \leq n$)
5. then return T
6. else
 - a. select any $T_{next} \in \{T_1, \dots, T_n\}$ with $W(T_{next}) = \min_i(W(T_i))$
 - b. return NNI(T_{next})

In der untenstehenden Abbildung sind alle 15 ungewurzelten Topologien für die Spezies A bis E abgebildet. Darin ist jeder Baum durch eine Nummer gekennzeichnet und durch Linien mit seinen Nachbarn verbunden. Verwende diese Nachbarschaften und die in der untenstehenden Tabelle gegebenen *Parsimony Scores* um mit Hilfe des NNI-Algorithmus einen optimalen Baum zu finden. Beginne mit Baum 8 und gib die Zwischenschritte und die Lösung an. Gibt es in Schritt 6.a des Algorithmus mehrere Alternativen, verfolge alle Möglichkeiten. Wird der optimale Baum in jedem Fall gefunden?

Baum Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Kosten	4	2	4	4	4	4	4	4	2	2	2	3	1	3	2

