

Einführung in die Bioinformatik WS2008/2009

Übung 9: Phylogenie

06.12.2008 (Abgabe: 11.12.2008)

1. Newick tree file format:

2.5 Pkt.

- (a) Lesen nach und verstehen Sie das Newick-Baum-Format zum kodieren evolutionärer Baumstrukturen (ebenfalls bekannt unter den Namen New Hampshire- oder Phylyp Baum-Format).

Mögliche Quellen:

- Der *Background information* Teil aus *UNIT 6.2 Visualizing Phylogenetic Trees Using TreeView* aus dem Buch *Current Protocols in Bioinformatics* (page2002c.cpbj.pdf bei den Übungsunterlagen).
- Joe Felsensteins Beschreibung des Newick-Formats auf der Homepage des PHYLIP-Software-Packets: <http://evolution.genetics.washington.edu/phylyp/newicktree.html>

- (b) Zeichnen Sie die Phylogenie, die dem Newick-Format ((Seq1, Seq2), (Seq3, Seq4), Seq5); entspricht.

2. Maximum Parsimony: Anhand des folgenden Alignments:

5 Pkt.

```
Seq1:  A G G C T T A A A G C G G
Seq2:  A C G C A T A A G A G C G G
Seq3:  A G G C T T C A T A G G G G
Seq4:  A T G C A T C A A A G G G A
Seq5:  A C G C T T C A C A G C G G
```

- (a) markieren Sie alle Parsimony-informativen Spalten,
(b) berechnen Sie die minimal notwendige Anzahl an Mutationen für den Baum ((Seq1,Seq2), (Seq3,Seq4),Seq5); (derselbe wie in 1) und
(c) geben Sie eine Nukleotid-Konfiguration für die internen Knoten des Baumes an, für die die Baum-Länge (Anzahl an notwendigen Mutationen) immernoch minimal ist.

3. **Distanz-Methoden:** Ist die folgende Distanz-Matrix

5 Pkt.

	Seq1	Seq2	Seq3	Seq4
Seq1	0	11	11	8
Seq2	11	0	2	15
Seq3	11	2	0	9
Seq4	8	15	9	0

(a) ultrametrisch, d.h. erfüllt sie die Ultrametrik-Ungleichung?

(b) additiv, d.h. erfüllt sie die 4-Punkt-Bedingung?

4. **Maximum Likelihood Bäume:** Für ein hypothetisches Alignment aus vier Sequenzen erhalten Sie für die drei möglichen Bäume die folgenden log-Likelihood-Werte:

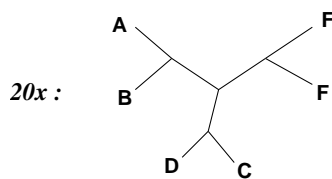
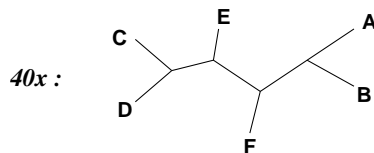
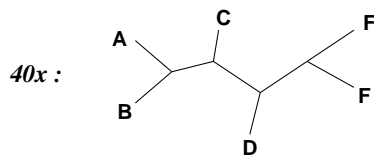
2.5 Pkt.

$$T_1: -5721.39, T_2: -7381.44, T_3: -4209.51$$

Welcher der drei Bäume ist der Maximum-Likelihood-Baum für unsere Daten?

5. **Konsensus-Bäume:** Aus eine Bootstrap-Analyse mit 100 Pseudo-Samples erhalten Sie folgenden 100 Bäume:

5 Pkt.



(a) Extrahieren daraus alle (nicht-trivialen) Splits und zählen Sie wie oft diese in den 100 Bäumen vorkommen.

(b) Konstruieren Sie daraus den 50%-Majority-Consensus-Baum und den Strict-Consensus-Baum. Schreiben Sie in den Consensus-Bäumen an jede interne Ast (= Split) die seiner Häufigkeit entsprechende Prozentzahl.

(c) Begründen Sie kurz, warum die Splits, die im 50%-Majority-Consensus-Baum inkludiert werden, in $> 50\%$ und nicht in $\geq 50\%$ aller Eingabebäume vorkommen müssen.