

## Algorithmische Bioinformatik Übungsblatt 10

Ausgabe: 21. Januar 2020 · Besprechung: 28. Januar 2020

**Aufgabe 10.1** Löse das kleine Maximum Parsimony Problem auf dem folgenden Baum und gib eine optimale Lösung (Kosten, Belegung) an.

$((((TC)(AG))((G((TT)C))(CG)))$

**Aufgabe 10.2** (Fitch-Hartigan für Maximum Parsimony) Beweise, dass für Einheitskosten der Fitch-Hartigan-Algorithmus zur Lösung des Kleinen Maximum-Parsimony-Problems (in einem genau zu formulierenden Sinn) äquivalent ist zu Sankoff's Dynamic-Programming-Algorithmus.

**Aufgabe 10.3** (Inkonsistenz von Maximum Parsimony) Wir betrachten einen ungewurzelten Baum (Quartett) mit sehr ungleichen Kantenlängen (wobei die Zahlen keine Distanzen, sondern proportional zu Mutationswahrscheinlichkeiten für DNA-Nukleotide sein sollen):  $(A:3, B:30) - 4 - (C:30, D:3)$ . Überlege, wie bei dieser Baumtopologie typische DNA-Sequenzen in  $B, C, D$  aussehen, wenn die Sequenz in  $A$  aus lauter As besteht. Wie würde der von Maximum Parsimony (großes Parsimony-Problem) rekonstruierte Baum aussehen?

**Aufgabe 10.4** (Entwicklung der Hamming-Distanz im Zeitverlauf) Die Sequenz eines Markergens verändert sich unabhängig in zwei Arten mit gemeinsamem Vorfahren. Es gelte die Hypothese einer *molekularen Uhr*: Wir nehmen an, dass im Mittel pro einer Million Jahre fünf von hundert Nukleotiden mutieren (dabei können Null-Mutationen, z.B.  $A \rightarrow A$  vorkommen). Die Sequenzen haben die Länge  $n$ ; dabei ist  $n$  sehr groß. Wie sieht die Hamming-Distanz (Anzahl der Unterschiede) der Sequenzen nach  $t$  Millionen Jahren (relativ zu  $n$ ) für kleine  $t$  und für  $t \rightarrow \infty$  aus? Ist daher die Hamming-Distanz ein guter Zeitschätzer in der Phylogenetik? (Hinweis: Betrachte und analysiere eine passende Markovkette.)