

Übungen zur Phylogenetik Vorlesung

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

Blatt 2 vom 18.10.2012

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

Aufgabe 1

(1 Punkt)

Beschrifte deine Abgabe gut leserlich mit deinem Namen und deinem Tutorium (TutorIn, Uhrzeit). Besteht deine Abgabe aus mehreren Blättern, hefte sie mit einer Heftklammer (keine Büroklammer) zusammen.

Aufgabe 2 Spezies- und Genbäume.

(2 Punkte)

- Gib alle Paare von orthologen und paralogen Genen in Bild 2.3 (rechts) im Skript an.
- Recherchiere und erkläre die Begriffe *In-paralog* und *Out-paralog*.

Aufgabe 3 Gene Tree/Species Tree Reconciliation.

(3 Punkte)

Gegeben sei folgender Genbaum G und Speziesbaum S :

$$\begin{aligned} G &= (((g_4, g_5)g_3, g_6)g_2, g_7)g_1; \\ S &= ((A, B)s_2, C)s_1; \end{aligned}$$

Hinweis zur Newicknotation: g_1, g_2, g_3, s_1 und s_2 sind die Bezeichnungen der inneren Knoten. Knoten g_3 ist der Elternknoten von g_4 und g_5 , Knoten g_2 der Elternknoten von g_3 und g_6 , und Knoten g_1 ist die Wurzel von G . Analog ist s_2 der Elternknoten von A und B , und s_1 die Wurzel von S .

Gen g_4 ist in Spezies A enthalten, g_5 und g_6 in B , und g_7 in C .

- Zeichne G und S nebeneinander, beschrifte jeweils die inneren Knoten und beschrifte die Blätter in G zusätzlich mit den entsprechenden Speziesnamen.
- Gib $\gamma(g_1)$, $\gamma(g_2)$ und $\gamma(g_3)$ an. Gib $M(\cdot)$ für alle Gene (auch die Blätter) in G an und füge entsprechende Pfeile von Knoten in G zu Knoten in S in dein Bild ein. Gib für die inneren Knoten g_1, g_2 und g_3 jeweils an, ob es sich um eine Duplikation oder Speziation handelt.
- Zeichne das erhaltene Szenario aus Speziesbaum und enthaltenem Genbaum, ähnlich zu dem in der Vorlesung bzw. Bild 2.3 (rechts) im Skript. Tipp: Das Szenario enthält eine Genlöschung.

Aufgabe 4 Lowest Common Ancestors.

(2 Punkte)

Nicht verwirren lassen: *Lowest* und *last common ancestor* sind im Prinzip dasselbe. In der Phylogenetik bezeichnet man den letzten gemeinsame Vorfahren als *last common ancestor* – abgekürzt *LCA*. Abstrahiert auf Knoten in Bäumen entspricht dies genau dem graphentheoretischen Begriff des *lowest common ancestor* (niedrigster gemeinsamer Vorfahrknoten), ebenfalls mit *LCA* abgekürzt.

Gegeben sei ein gewurzelter Baum $T = (V, E)$. Überlege, wie man für zwei gegebene Blattknoten u und v den *lowest common ancestor*, $LCA(u, v)$, bestimmen kann. Gib deinen Algorithmus stichwortartig an (Pseudocode in beliebiger Notation; siehe z.B. Skript, Seiten 11, 21, 22, 27, 29, etc.) und erkläre die Laufzeitkomplexität im *worst case*.