

## Aufgabenblatt 14

### Multiples Alignment und Phylogenie

---

Abgabe: **15.02.2005** bis **17 Uhr** über Goya  
Maximal: 20 Punkte  
Namen nicht vergessen!

---

1. Wir wollen mit der Proteinsequenz, die eine 'Death domain' enthält, (Q05652 ist auf der Übungs-webseite gegeben) nach weiteren Mitgliedern der Familie mit PSI-BLAST suchen. Verwendet für die PSI-BLAST suche dabei die Datenbank 'swissprot', sonst die Standard-Parameter. Iteriert dabei bis keine neue Sequenz mit einem E-value besser als der threshold hinzukommt.
  - (a) Wie viele Sequenzen sind nach jedem Iterationsschritt in dem Ergebnisset? **3 P**
  - (b) Wie hoch ist dabei die Spezifität und Seinsitivität, wenn man als Vorgabe die Sequenzen nimmt, die PROSITE unter <http://us.expasy.org/cgi-bin/nicesite.pl?PS50017> als Mitglieder dieser Familie ausweist (49 True positive hits + 28 False negative hits)? **2 P**
  - (c) Woran kann es liegen, dass auch nach mehreren Iterationen nicht alle / mehr Vertreter der Familie gefunden werden? **2 P**
  
2. Gebt für das folgende multiple Sequenzalignment von 7 DNA-Sequenzen das Profil an.

Identifizier		Sequenz
embl K00287 SCK2		TCAGTTGGTAGAGC
embl K00298 SCM3		TCAGTAGGAAGAGC
embl K00321 SCMI		TGGCGCAGT-GGAA
embl K00323 CUMI		TGGCGCAGT-GGAA
embl K00316 NCMI		TGGCGCAGC-GGAA
embl K00344 SPF		TGGTGTAGTTGGGA
embl K00357 CUP		TGGTCTAGT-GGTA

**2 P**

3. Fügt zu dem multiplen Sequenzalignment aus der vorherigen Aufgabe die Sequenz, S=TCATCACGTGAA hinzu. Gebt dazu das / ein optimales Alignment zwischen k' und S' an und dessen Score  $s(A)$ . (Hinweis: kann auch durch probieren verschiedener Kombinationen gefunden werden). Ähnlichkeitsmatrix, m:

	A	C	G	T	-
A	1	-3	-2	-3	-1
C	-	1	-3	-2	-1
G	-	-	1	-3	-1
T	-	-	-	1	-1

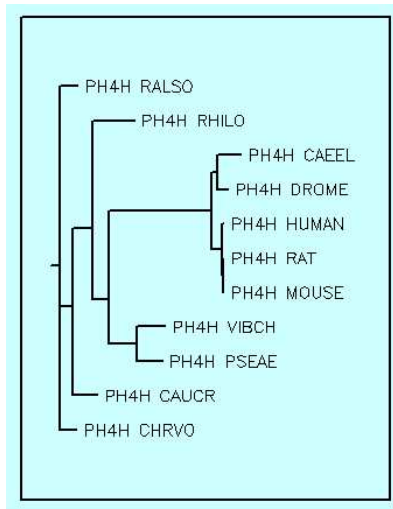
**3 P**

4. Gegeben ist die folgende Matrix M

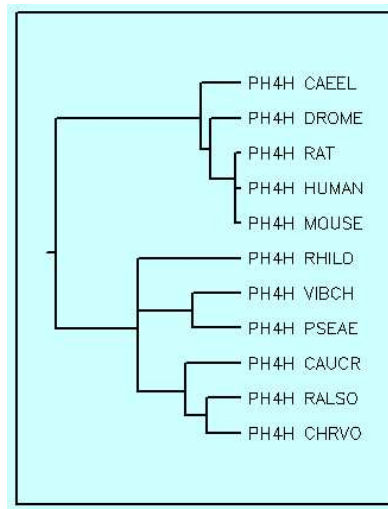
	A	B	C	D	E
A	0	7	13	9	7
B	7	0	12	8	6
C	13	12	0	6	10
D	9	8	6	0	6
E	7	6	10	6	0

- (a) Ist die Matrix M eine additive Matrix? 1 P
- (b) Ist die Matrix M auch eine ultrametrische Matrix? Wenn ja, dann konstruiert den phylogenetischen Baum nach der UPGMA-Methode, wenn nein, begründet eure Entscheidung! 3 P

5. Gegeben sind die beiden phylogenetischen Bäume von Sequenzen des Proteins 'Phenylalaninehydroxylase' aus verschiedenen Organismen. Die Bäume sind mit unterschiedlichen Methoden von Phylip, einem Programm zur Berechnung von phylogenetischen Bäumen, erstellt worden.



(a) Maximum likelihood



(b) Maximum likelihood + clock

Welcher der beiden Bäume entspricht eher der Baumstruktur der NCBI Taxonomy <http://www.ncbi.nlm.nih.gov/Taxonomy/>?

Gebt ausserdem für den besten Baum die möglichen Beschriftungen der inneren Knoten an. 4 P

CAEEL: Caenorhabditis elegans; CAUCR: Caulobacter crescentus; CHRVO: Chromobacterium violaceum; DROME: Drosophila melanogaster; HUMAN: Homo sapiens; MOUSE: Mus musculus; PSEAE: Pseudomonas aeruginosa; RALSO: Ralstonia solanacearum; RAT: Rattus norvegicus; RHILD: Rhizobium loti; VIBCH: Vibrio cholerae