

Prof. Dr. Knut Reinert  
Dr. Roland Krause  
Matthias Winkelmann  
Patrick Pett

Institut für Informatik  
AG Algorithmische Bioinformatik

## Algorithmische Bioinformatik WS 11/12

### Übungsblatt 3

Abgabe am 7.11.2011, 12 Uhr

Name:

Matrikelnummer:

Übungsgruppe:

1  2  3

---

#### Aufgabe 1: Orthologe

Klassifizieren Sie die internen Knoten des Genbaumes  $G = (((A, B), C), ((A, B), (D, D)))$  anhand des Speziesbaumes  $S = (((A, B), C), D)$  nach Speziations- und Duplikationsereignissen. Erstellen Sie den vereinigten Gen- und Speziesbaum und fügen Sie Genverlust-Ereignisse ein. Nennen Sie Beispiele für Orthologe, Ko-Orthologe, In-Paralogue und Out-Paralogue.

#### Aufgabe 2: PhoP

In den pathogenen Bakterien *Salmonella typhimurium* und *Mycobacterium tuberculosis* existiert jeweils ein Gen namens phoP, das für den *Response regulator* eines *two-component systems* kodiert. Beide Proteine sind mit der Virulenz der Organismen in Verbindung gebracht worden und werden in einigen Publikationen als Orthologe geführt.

Stellen Sie fest, ob die beiden Proteine wirklich ortholog zu einander sind.

- a) Verwenden Sie dazu die gegebenen Sequenzen, die relevanten Proteine in *Salmonella typhimurium* und *Mycobacterium tuberculosis* sowie *Bacillus subtilis* enthalten<sup>1</sup>. Es lohnt die Identifier der Sequenzen zu vereinheitlichen (siehe 1. Übungszettel). Überlegen Sie sich die Phylogenie der Organismen und erstellen Sie einen Stammbaum.

---

<sup>1</sup>[http://lectures.molgen.mpg.de/Algorithmische\\_Bioinformatik\\_WS1011/dokumente/phop\\_homologs.fasta](http://lectures.molgen.mpg.de/Algorithmische_Bioinformatik_WS1011/dokumente/phop_homologs.fasta)

- b) **COG** Überprüfen Sie, zu welcher Orthologen Gruppe in der COG-Datenbank die jeweiligen Proteine gehören. Da die Annotation auf der Originalseite der Datenbank unter <http://www.ncbi.nlm.nih.gov/COG> schlecht zu durchsuchen ist, können Sie die STRING-Datenbanken im COG-Mode verwenden <http://string.embl.de>.
- c) **BLAST** Verwenden Sie die “reciprocal best hit”-Methode. Klassifizieren Sie die relevanten (!) Proteine.
- d) **Stammbaum** Erstellen Sie einen phylogenetischen Baum der Proteine mit einer Methode Ihrer Wahl und fügen Sie die relevanten(!) Duplikations- und Speziationsereignisse in den Baum ein.
- e) **Auswertung** Diskutieren und bewerten Sie die Ergebnisse. Warum sind die Ergebnisse unterschiedlich und was bedeutet das für die Interpretation der Funktion? Wie können Sie die Verlässlichkeit angeben?

### Aufgabe 3: HMMs

Gegeben sei das HMM in Fig. 1.

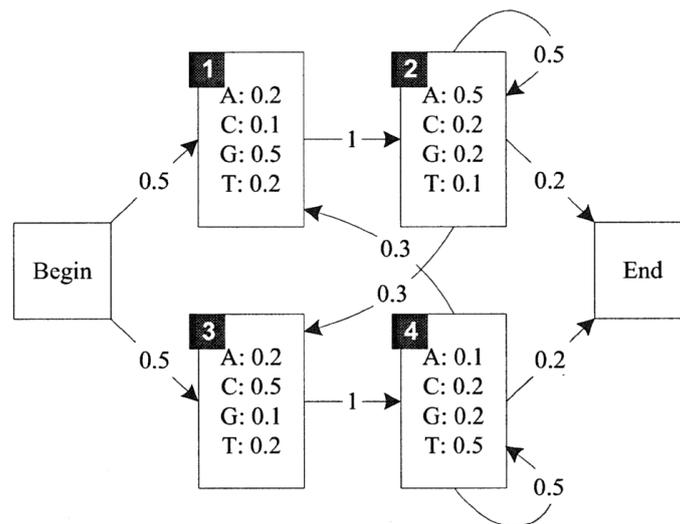


Abbildung 1: HMM

- a) Geben Sie zu diesem HMM das Alphabet  $\Sigma$ , die Zustandsmenge  $Q$  und die Matrix  $A$  der Übergangswahrscheinlichkeit an.
- b) Bestimmen Sie mit dem Viterbi-Algorithmus den Viterbi-Pfad für die Beobachtung CTG.
- c) Berechnen Sie die Forward und Backward Matrix für CTG.