

Freie Universität Berlin  
Fachbereich Mathematik & Informatik  
AG Medical Bioinformatics  
Tim Conrad

## 2. Praktikumsteil

# ALGORITHMISCHE BIOINFORMATIK

WS 13/14

18. November 2013

**Abgabe: Sonntag, den 15.12.2013 bis 23:59 Uhr via Subversion.**

## Einführung

Ziel dieses Praktikums ist es mit Hilfe von Werkzeugen der Bioinformatik aus gegebenen Sequenzen einen *Phylogenetischen Baum* zu erzeugen und die Qualität des Baumes via Bootstrapping zu evaluieren. Als Konstruktionsverfahren für den Baum soll UPGMA benutzt werden. Für dieses Praktikum steht ihnen ein Datensatz zur Verfügung, den Sie von der Vorlesungs-Webseite herunterladen können. In der Datei befinden sich mehrere genomische Sequenzen.

Bemerkungen:

- Wir empfehlen die Verwendung von BioJava und den EBI Web Services<sup>1</sup>.
- Der Bericht soll mit L<sup>A</sup>T<sub>E</sub>X erstellt werden.

---

<sup>1</sup><http://www.ebi.ac.uk/Tools/webservices/>

## Aufgabenstellung

Entwickeln Sie ein Programm, das folgende Aufgaben löst. (Die Verwendung von BioJava und NCBI ist nur eine Empfehlung - Sie können auch andere Programmiersprachen oder Tools benutzen.)

### 1. Aufgabe

1. Lesen Sie die gegebenen Sequenzen mit Hilfe der BioJava API ein.
2. Führen Sie mit Hilfe des NCBI Web Services ein multiples Alignment der Sequenzen durch (benutzen Sie dazu z.B. den MUSCLE Web Service oder den Clustal Web Service)
3. Leiten Sie aus dem Sequenzalignment eine Distanz-Matrix ab.
4. Implementieren Sie den UPGMA Algorithmus, um aus der Matrix einen *Phylogenetischen Baum* zu erzeugen. (*Eine Erklärung des UPGMA Algorithmus finden Sie im Buch zur Vorlesung*)
5. Validieren Sie den Baum mit Hilfe des Bootstrapping Verfahrens.

### 2. Aufgabe

Diese Aufgabe stellt eine Erweiterung der 1. Aufgabe dar und soll mit mindestens fünf unterschiedlich stark ausgeprägten Mutationsstufen durchgeführt werden. Das benutzte Mutationsmodell muss dabei im Bericht beschrieben werden.

1. Verändern Sie die gegebenen Sequenzen indem Sie einzelne Basen mutieren lassen.
2. Wiederholen Sie die Schritte 2 bis 5 aus der 1. Aufgabe, um den Baum erneut zu validieren.

Wie verändert sich die Qualität des Baums? Wiederholen Sie die Prozedur mit steigender Mutationsrate und beobachten Sie dabei die Qualität des Phylogenetischen Baums.

### Auswertung

Der Bericht soll aus zwei Teilen bestehen. Im ersten Teil sollen Sie auf verständliche Art und Weise die Implementierung Ihres Programms beschreiben. Beachten Sie, dass der Programmcode ebenfalls bewertet wird. Dieser sollte also ebenfalls gut kommentiert sein. Im zweiten Teil des Berichts sollen die Ergebnisse aus den Aufgaben 1 und 2 dokumentiert werden.