

Übungen zur Vorlesung  
**Einführung in die angewandte  
Bioinformatik**  
Sommersemester 2009

**Übungsblatt 7**  
Bearbeitungszeit:  
04.06.2009

**Aufgabe 7.1 – Multiples Alignment mit ClustalW**

Führen Sie ein multiples Sequenzalignment mit ClustalW durch. Laden Sie sich dazu die von der Übungs-Webseite die Datei `six_proteins.fasta` herunter. Lassen Sie alle Einstellungen der ClustalW-Seite unverändert, kopieren Sie nur die Sequenzen in das entsprechende Feld (oder geben Sie den Dateinamen an) und starten Sie das Alignment.

Betrachten Sie zunächst die paarweisen Alignments (unter *Scores table*). **Hinweis:** Falls hier Sequenzpaare fehlen, klicken Sie auf *View Output File*. Für welches Sequenzpaar ist der Score am höchsten und wie hoch ist dieser? \_\_\_\_\_

Betrachten Sie nun das multiple Alignment. Welcher Eindruck drängt sich auf den ersten Blick auf?

\_\_\_\_\_

Welche Sequenz enthält die längste Insertion gegenüber den anderen Sequenzen?

\_\_\_\_\_

Spiegelt sich dies auch in den paarweisen Alignments wider?

\_\_\_\_\_

Betrachten Sie auch das Phylogramm. Wer ist der nächste Verwandte von *Rattus norvegicus*?

\_\_\_\_\_

Klicken Sie zuletzt auf *View Alignment File* und lassen Sie die angezeigte Seite geöffnet oder speichern Sie sie in eine Datei. Sie werden Sie in der nächsten Aufgabe brauchen

**Aufgabe 7.2 – Phylogenie mit Phylip**

In der letzten Aufgabe wurde Ihnen zum multiplen Alignment, das ClustalW berechnet hat, ein Phylogramm präsentiert. Nun sehen wir uns die Schritte zur Erstellung eines solchen phylogenetischen Baums genauer an. Gehen Sie dazu auf die Phylip-Seite und wählen Sie zunächst das Programm `protdist`, um im ersten Schritt die Distanzen zwischen den Proteinen zu berechnen. Fügen Sie das Alignment von ClustalW ein und starten Sie das Programm. Klicken Sie auf der Ergebnisseite auf `outfile`, um sich die Distanztabelle anzusehen. Zwischen welchen Sequenzen ist die Distanz am größten?

\_\_\_\_\_

Reichen Sie nun zum Erstellen des Baums die Distanzen an den kitsch-Algorithmus weiter (*Run the selected program...*) und starten Sie diesen (*Run kitsch*). Sehen Sie sich auch hier in `outfile` und `outtree` zwei verschiedene Darstellungsformen desselben Ergebnisses an.

Wählen Sie dann das Programm `drawgram`, um den Baum darzustellen. Wählen Sie als tree style zunächst Cladogram. Speichern Sie die Datei `plotfile.ps`, die Sie auf der Ergebnisseite finden (Sie brauchen diese Datei für die nächste Aufgabe). Öffnen Sie die Datei, um sich den Baum anzusehen. Wer ist nach den beiden Homo-sapiens-Sequenzen am engsten verwandt?

Gehen Sie nun zurück auf die drawgram-Seite und wählen Sie statt des Cladograms andere Baumarten (*tree styles*), wie Curvogram, Eurogram oder Circular tree. Sehen Sie sich natürlich auch die *plotfiles* an.

Wählen Sie nun drawtree statt drawgram und sehen Sie sich den resultierenden Baum an. Was ist der wesentliche Unterschied? \_\_\_\_\_

### **Aufgabe 7.3 – Tree of Life**

Gehen Sie nun auf die Tree-of-Life-Webseite, suchen Sie nach Homo sapiens und wählen Sie den ersten Treffer. Sie befinden sich nun an einem Blatt des Lebensbaums. Navigieren Sie nun mittels des Links hinter *Containing group* im Baum immer weiter in Richtung der Wurzel. In welcher Gruppe sind wir zum ersten Mal mit Ratten gemeinsam? \_\_\_\_\_

Navigieren Sie nun in die umgekehrte Richtung (d.h. in Untergruppen hinein), sodass Sie in der Gruppe landen, in der sich Schweine und Rinder trennen. Wie heißt die Gruppe? \_\_\_\_\_

Öffnen Sie die in Aufgabe 7.2 gespeicherte Datei. Erkennen Sie den Weg, den Sie in diesem Baum zurückgelegt haben? \_\_\_\_\_

### **Aufgabe 7.4 – Pfam**

Gehen Sie auf die Pfam-Webseite. Suchen Sie nach der Proteinfamilie mit der Accession-Nummer PF09507. Wie viele Sequenzen sind in Pfam für diese Familie gespeichert? \_\_\_\_\_

Machen Sie sich mit den Unterpunkten, die auf der linken Seite verlinkt sind, vertraut.

Sehen Sie sich den phylogenetischen Baum an (Unterpunkt "Trees"). Mit welcher Methode wurde der angezeigte Baum erzeugt?

\_\_\_\_\_

Sie möchten nun statt dieser Methode lieber den UPGMA-Algorithmus verwenden. Leider bietet Pfam diese Funktion nicht an, verwenden Sie daher phylip mit den von Pfam zur Verfügung gestellten Rohdaten. Im Unterpunkt "Alignments" bekommen Sie das dazu nötige multiple Alignment. Stellen Sie die "Formatting options" auf "Seed", "FASTA", "Alphabetical", "Gaps as dashes" und wählen Sie "Generate" aus. Speichern Sie die Datei und verfahren Sie nun ähnlich wie in der vorigen Aufgabe: Lassen Sie also zunächst protdist laufen, wählen Sie dann "neighbor" und stellen Sie "UPGMA" als "Distance method" ein. Lassen Sie dann drawgram auf den Daten laufen (belassen Sie hier die Parameter auf ihren voreingestellten Werten).

Vergleichen Sie den Baum, den Sie erhalten, mit dem, den Pfam anzeigt.

Betrachten Sie, wie die beiden Blätter Q6CE56\_YARLI/16-452 und Q5AA19\_CANAL/22-404 einsortiert werden. Was ist hier der Unterschied zwischen den Bäumen?

\_\_\_\_\_

Sind die Bäume insgesamt ähnlich?

\_\_\_\_\_

Erstellen Sie nun mit phylip einen weiteren Baum auf den gleichen Daten, aber verwenden Sie den Neighbor-Joining-Algorithmus. Ist der Baum, den Sie so erhalten, der gleiche wie der von Pfam angezeigte?

### **Aufgabe 7.5 – Theorieaufgabe: Anzahl möglicher Bäume**

Diese Aufgabe können Sie mit Hilfe der Vorlesungsfolien beantworten. Wie viele ungewurzelte Bäume mit 7 Blättern gibt es?

\_\_\_\_\_

Wie viele gewurzelte Bäume mit 7 Blättern gibt es?