

Übungen zum Sequenzanalyse-Praktikum

Universität Bielefeld, SoSe 2014
Prof. Dr. Jens Stoye · M.Sc. Nina Luhmann · M.Sc. Linda Sundermann
<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014summer/SequaPrak>
praktikum-seqan@CeBiTec.Uni-Bielefeld.DE

Übungsblatt 11 vom 30.06.2014
Abgabe am Donnerstag, den 03.07.2014

Aufgabe 1 (MSA mit T-Coffee und Clustal Omega)

1. Gehe auf die Seite www.tcoffee.org und klicke auf *T-Coffee*, um zum Eingabe-Interface zu kommen. Füge die Hämoglobinsequenzen von letzter Woche ein (du findest sie nun auch auf der Sequenzanalysepraktikums-Seite) und starte das Programm mit Standardeinstellungen. Füge die gleichen Sequenzen noch einmal auf der *Clustal-Omega*-Seite in das Eingabe-Interface ein und lasse das Programm mit Standardeinstellungen laufen. Vergleiche nun die beiden multiplen Alignments per Augenmaß. Erkennst du Unterschiede?
2. Vergleiche auch das *T-Coffee*-Layout mit dem von *Clustal Omega*. Was ist gleich? Was ist anders? Erkläre bei Unterschieden, was sie bedeuten.
3. *Clustal Omega* berechnet aus einer paarweisen Distanzmatrix einen *Guide Tree* und baut ein multiples Alignment auf, indem es die Sequenzen progressiv dem *Guide Tree* folgend aligniert. Dabei wird ein zuvor festgelegtes Scoring-Schema benutzt. *T-Coffee* benutzt auch einen *Guide Tree*, besitzt für jedes zu alignierende Sequenzpaar jedoch einen spezifischen Score, der aus Konsistenzberechnungen in einer früheren Phase resultiert. So werden bei der Alignmentberechnung homologe Bereiche schneller gefunden und nicht so schnell Fehler gemacht. Wie erklärst du dir aber nun, dass *Clustal Omega* das gleiche Alignment wie *T-Coffee* gefunden hat?

Aufgabe 2 (T-Coffee Flavors)

Der Algorithmus von *T-Coffee* wurde dir im Seminar vorgestellt. Es gibt auch noch einige Erweiterungen, zum Beispiel *M-Coffee*. Das Programm hierzu benutzt gleich zu Beginn mehrere MSA-Algorithmen, um ein besseres multiples Alignment zu erhalten.

1. Benutze das Programm *T-Coffee* mit den Beispielsequenzen auf der Eingabeseite und Standardeinstellungen. Wie hoch ist der Score des Alignments?
2. Benutze nun *M-Coffee* mit den gleichen Sequenzen und ebenfalls Standardeinstellungen. Wie hoch ist der Score hier? Welches der beiden Alignments ist besser?
3. Welche multiplen Alignmentmethoden benutzt *M-Coffee* standardmäßig?
4. Schau nach, ob es diese oder ähnliche Methoden auch bei *T-Coffee* gibt. Wenn ja, wähle sie alle aus. Wie sieht nun das Ergebnis aus, das du erhältst?

Aufgabe 3 (T-Coffee Core)

1. Nimm nun die Sequenzen aus Aufgabe 2 und berechne mithilfe von *Clustal Omega* ein multiples Alignment. Was fällt dir im Vergleich zu den beiden multiplen Alignments von *T-* und *M-Coffee* auf? Kann man eventuell an gewissen Eigenschaften des Alignments erkennen, wie die Algorithmen arbeiten? Wenn ja, woran?
2. Die [tcoffee.org](http://www.tcoffee.org)-Plattform erlaubt es, für bereits fertige multiple Alignments einen Score zu berechnen. Wähle dafür das Programm *Core* aus. Schau dir anhand der Beispiel-Datei an, in welchem Format das multiple Alignment eingegeben werden muss. Beschreibe kurz den Aufbau des Formats.
3. Füge nun deine Ergebnisse von *Clustal Omega* so ein, dass der Score berechnet werden kann. Wie hoch ist er? Wie ist die Qualität des Alignments verglichen mit der von *T-* und *M-Coffee*?

Aufgabe 4 (T-Coffee auf der Kommandozeile)

1. Lade dir über www.tcoffee.org *T-Coffee* für dein System herunter. Lies dir den ersten Teil der README-Datei durch. Probiere aus, ob *T-Coffee* bei dir läuft. Benutze dazu den einfachsten Aufruf und die Datei mit den Hämoglobinsequenzen.
2. Beschreibe kurz und zusammenfassend, welche Informationen du auf der Kommandozeile ausgegeben bekommst.
3. Vergleiche das Alignment mit dem aus Aufgabe 1. Unterscheidet sich die Darstellung? Ist das Alignment identisch?
4. Ein Teil der Ausgabe beschreibt den Baum, der für das progressive Alignment benutzt wurde. Stelle diesen beschriebenen Baum in deinem Protokoll dar. (Zeichne ihn entweder per Hand oder direkt am Computer.)
5. Schau dir nun die Dokumentation des Kommandozeilentools an. Du findest sie auch auf der *T-Coffee*-Seite. Welche Dateien erzeugt der einfache Aufruf von *T-Coffee*? Was muss gegeben sein, wenn du *M-Coffee* benutzen möchtest?
6. Betrachte nun nochmal die Ausgabe des Programms. Welchen Parameter bekommt *-matrix*? Schlage in der Dokumentation nach, was das bedeutet.
7. Nenne einen Vorteil, den die Benutzung des Kommandozeilentools gegenüber der Onlineversion bietet. Was ist ein Nachteil? Fasse dich kurz (maximal vier Sätze).