

Übungen zum Sequenzanalyse-Praktikum

Universität Bielefeld, WiSe 2017/18
Prof. Dr. Jens Stoye · Dr. Daniel Dörr · M. Sc. Tizian Schulz
[http://gi.cebitec.uni-bielefeld.de/teaching/2017winter/sequapra-
praktikum-seqan@CeBiTec.Uni-Bielefeld.DE](http://gi.cebitec.uni-bielefeld.de/teaching/2017winter/sequapra-
praktikum-seqan@CeBiTec.Uni-Bielefeld.DE)

Übungsblatt 11 vom 16./17.01.2018
Abgabe bis Sonntag bzw. Montag, 24:00 Uhr.

Benutze für diesen Zettel *Clustal Omega* auf dem EBI Webserver: <http://www.ebi.ac.uk/Tools/msa/clustalo/>.

Aufgabe 1 (MSA verwandter Sequenzen mit Clustal Omega)

1. Suche dir die Proteinsequenzen des Haemoglobin-Alpha-Gens des Menschen, der Hausmaus, der Schopffente, der Spitzschwanzente und des Atlantischen Lachses heraus. Verwende dazu die *Accession*-Nummern: *CAA00181.1*, *NP_032244.2*, *ACT80500.1*, *ACM67760.1* und *NP_001117134.1*.
2. Berechne mit den Sequenzen ein multiples Aligment mit *Clustal Omega* mit Standardeinstellungen. Beschreibe die Ausgabe kurz.
3. Als letzte Zeile im multiplen Alignment findest du verschiedene Zeichen. Welche sind es und was bedeuten sie?
4. Lasse dir jetzt die Farben im multiplen Alignment anzeigen. Was zeigen sie an? (Eine kurze Erklärung genügt.)
5. Auf der Ausgabenseite kann man einen phylogenetischen Baum der Sequenzen darstellen lassen. Dabei kann man zwischen den Optionen „Cladogram“ und „Real“ wählen. Wie unterscheiden sich die beiden Darstellungen voneinander?
6. Bilde die beiden phylogenetischen Bäume in deinem Protokoll ab. Ist die Haemoglobin-Sequenz des Menschen ähnlicher der von Vögeln oder der von Fischen? Woran kannst du das erkennen?

Aufgabe 2 (MSA weit verwandter Sequenzen mit Clustal Omega)

1. Suche dir nun die Proteinsequenz mit der *Accession*-Nummer *NP_001186363.1* heraus. Um was für eine Sequenz handelt es sich? Aligniere nun die fünf Haemoglobinsequenzen aus Aufgabe 1 zusammen mit dieser Sequenz.
2. Betrachte zunächst den Guide Tree. Welche Struktur hättest du erwartet?
3. Wie können die Parameter `--guidetree-in` bzw. `--profile1` und `--profile2` eingesetzt werden, um eine bestimmte Aliginierungs-Abfolge zu forcieren? Informiere dich dazu in der Dokumentation von Clustal Omega. (Die Webversion des EBI bietet diese Parameter nicht an.)

Aufgabe 3 (Untersuchung unbekannter Sequenzen / Vergleich mit DIALIGN 2)

Benutze für diese Aufgabe die Datei `unknown_query.fas`, die unter `/vol/seqan/Praktium/Thema11` im CeBiTec-System abgelegt ist.

1. Die Datei `unknown_query.fas` enthält vier bekannte und eine unbekannt Sequenz. Aligniere alle fünf Sequenzen mit *Clustal Omega*.
2. Zu welcher Sequenz hat die Query die höchste Ähnlichkeit? Um welche Sequenz handelt es sich bei dieser? Aus welchem Organismus stammt sie?
3. Was stellen die anderen drei Sequenzen dar und aus welchen Organismen kommen sie?
4. Um was für eine Sequenz wird es sich bei der Query wahrscheinlich handeln? Kannst du deine Vermutung mit BLAST bestätigen? Welche Sequenz findest du?
5. Aligniere die fünf Sequenzen auch noch mit DIALIGN 2¹. Vergleiche die beiden Alignments. Diskutiere Ähnlichkeiten und Unterschiede.

¹<https://bibiserv.cebitec.uni-bielefeld.de/dialign/>