

Übungen zum Sequenzanalyse-Praktikum

Universität Bielefeld, SoSe 2016

Dr. Roland Wittler · M.Sc. Linda Sundermann

<http://gi.cebitec.uni-bielefeld.de/teaching/2016summer/sequaparak>

praktikum-seqan@CeBiTec.Uni-Bielefeld.DE

Übungsblatt 10 vom 21./22.06.2016

Abgabe bis Sonntag bzw. Montag, 24:00 Uhr.

Benutze für diesen Zettel jeweils die *DIALIGN*-Version vom „Bielefelder Bioinformatik-Server“ (<http://bibiserv.techfak.uni-bielefeld.de/dialign>) und die *DIALIGN-Pfam*-Version vom „Göttingen Bioinformatics Compute Server“ (<http://dialign.gobics.de>).

Aufgabe 1 (*DIALIGN 2* – Segmente)

Auf der Praktikums homepage findest du die Datei *dialign_sequences.fas*, mit der du in dieser Aufgabe arbeiten sollst.

1. Benutze die ersten beiden Sequenzen der Datei um mit *DIALIGN* ein Alignment mit Standardinstellungen zu berechnen. Auf eine Validierung des Dateiformats kannst du verzichten. Betrachte nach der Berechnung den Output im *Custom Format*. Was bedeuten die kleinen und die großen Buchstaben im Alignment?
2. Wie viele Segmente wurden aligniert? Fertige eine grobe Skizze an.
3. Berechne nun ein Alignment mit allen drei Sequenzen aus der Datei. Wie sieht das Alignment nun aus? Fertige ebenfalls eine grobe Skizze an, die auf die Elemente aus der vorherigen Skizze eingeht. Betrachte dazu die nicht alignierten Regionen der dritten Sequenz genau. Warum kann sie nicht problemlos an die anderen beiden aligniert werden?

Aufgabe 2 (*DIALIGN 2*)

1. Auf der Veranstaltungshomepage findest du die multiple FASTA-Datei *four_seq.fas*. Um was für Sequenzen handelt es sich und aus welchen Organismen stammen sie?
2. Berechne mit *DIALIGN* ein multiples Alignment für die gegebenen Sequenzen. Beschreibe den Aufbau des multiplen Alignments im *Custom Format* in drei bis fünf Sätzen.
3. Betrachte nun die Zahlen, die du unter jeder Alignmentsspalte findest. (Achtung: Es kann sein, dass die Alignmentsspalten nicht richtig übereinander stehen, wenn das Format verschoben ist.) Wofür sind die Zahlen da? Was bedeutet ein Wert von 9, was einer von 0? Löse diese Aufgabe alleine mit diesem Alignment, du musst die Information nicht im Internet heraus suchen.
4. Betrachte nun den Bereich des Alignments, der von Position 848 bis einschließlich 898 (in Bezug zu der Sequenz *AB612241.1*) reicht. Ist er gut konserviert? Vergleiche den Bereich mit *megablast* gegen die *Nucleotide collection*. Verwende dazu die Sequenz aus *AB612241.1*. Beschreibe das Ergebnis kurz. Unterstützt das Ergebnis deine Feststellung zur Konserviertheit des Bereiches?
5. Woran erkennst du Bereiche, die nicht gut konserviert sind? Suche dir nun einen nicht gut konservierten Bereich im Alignment aus, schreibe die Position im Bezug zu einer Sequenz und die Basen in diesem Bereich raus. Vergleiche auch diese Sequenz mit *megablast* gegen die *Nucleotide collection*. Vergleiche dein Ergebnis mit dem aus Aufgabe 2.4. War dieses Ergebnis zu erwartet?

Aufgabe 3 (DIALIGN-Pfam at GOBICS)

1. Welche Versionen von *DIALIGN* findest du auf dem „Göttingen Bioinformatics Compute Server“?
2. Beschreibe kurz, wie *DIALIGN-Pfam* funktioniert.
3. Lasse *DIALIGN-Pfam* mit den auf der Seite gespeicherten Beispielsequenzen laufen. Betrachte die Proteindomänen und das Alignment zu ihnen genauer. Welche Domänen werden gefunden. An welchen Positionen in den Sequenzen startet der Match zur Domäne in *Pfam*? Welche Sequenz macht beim Alignment zum *Pfam*-Match Probleme und warum?
4. Beschreibe das fertige Alignment kurz.