

Übungen zum Sequenzanalyse-Praktikum

Universität Bielefeld, WS 2017/18

Dr. Daniel Dörr · M. Sc. Tizian Schulz · Prof. Dr. Jens Stoye

<http://gi.cebitec.uni-bielefeld.de/teaching/2017winter/sequaprak>

praktikum-seqan@CeBiTec.Uni-Bielefeld.DE

Übungsblatt 5 vom 14./15.11.2017

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

Auf diesem Zettel sollen die Datenbanksuchprogramme BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) und FASTA (http://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi?rm=select&pgm=fa) kennen gelernt werden. Benutze, wenn es nicht anders angegeben ist, immer die Standardwerte der Suchmaschinen.

Aufgabe 1 (BLAST)

Benutze für die folgende Aufgabe die Sequenz *unknown_gene.fas* von der Veranstaltungshomepage.

1. Nutze den Algorithmus *blastn*, um ähnliche Sequenzen zu finden. Wie ist die Ausgabe von BLAST generell aufgebaut?
2. Welches Gen wurde hier sequenziert und woran erkennst du das?
3. Ist die komplette mRNA des eingegebenen Gens vorhanden? Wie kannst du das herausfinden?
4. Suche jetzt mit *blastx* in der *SwissProt*-Datenbank. Vergleiche die Anzahl an gefundenen Sequenzen mit deinen Treffern in der Nukleotid-Datenbank. Diskutiere dein Ergebnis.
5. Warum kann es Sinn ergeben, eine Nukleotidsequenz erst in eine Aminosäuresequenz zu übersetzen, bevor man sie mit anderen Sequenzen vergleicht?

Aufgabe 2 („Word Size“ in BLAST)

1. Suche dir die FASTA-Sequenz des Gens mit der Accession Nummer *A01590.1* heraus. Um was für ein Gen handelt es sich?
2. Nutze *megablast* mit Standardeinstellung, um deine Sequenz gegen die Nukleotid-Datenbank zu vergleichen. Beschreibe die Ergebnisse kurz.
3. Wie lauten die Standardeinstellung für *megablast* im Abschnitt „General Parameters“ und was bedeuten sie?
4. Vergleiche nun noch einmal mit *megablast* deine Sequenz mit der Nukleotid-Datenbank, stelle dabei aber den Parameter „Word Size“ auf 256. Vergleiche das Ergebnis mit der vorherigen Suche und erkläre, wie der Unterschied zu Stande kommt.

Aufgabe 3 (FASTA und BLAST)

Auf der Veranstaltungshomepage findest du eine unbekannte Proteinsequenz mit dem Namen *unknown_protein.fas*.

1. Nutze das Programm *FASTA* für eine Suche in der *SwissProt*-Datenbank.
2. Suche auch mit *blastp* in der *SwissProt*-Datenbank.
3. Vergleiche und diskutiere die Ergebnisse kurz.