

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/SequaPraktikum-seqan@CeBiTec.Uni-Bielefeld.DE>

Übungsblatt 6 vom 19.05.2014
Abgabe am Donnerstag, den 22.05.2014

Aufgabe 1

1. Lade die Sequenz von <http://bibiserv.cebitec.uni-bielefeld.de/sadr2/databasesearch/blast/exercise1/sequence.fas> herunter (Sie wurde in Bielefeld sequenziert!)
2. Nutze *blastn* auf dem *NCBI*-Server, um ähnliche Sequenzen zu finden, die bereits bekannt sind.
3. Welches Gen wurde hier sequenziert?
4. Ist das die komplette mRNA? Wie kannst du das herausfinden?
5. Suche jetzt mit *blastx* in der *SwissProt*-Datenbank. Vergleiche die Anzahl an gefundenen Sequenzen mit deinen Treffern in der Nukleotid-Datenbank.
6. Schau dir die Alignments von *blastx* genauer an. Warum sind manche Teile der Sequenz in kleinen, grauen Buchstaben geschrieben? (Schau dir die *BLAST* FAQs an.)

Aufgabe 2

1. Benutze die Sequenz aus Aufgabe 1 erneut. Suche mit *blastx* nach ähnlichen Sequenzen des *Homo sapiens* in den Datenbanken *nr* und *SwissProt*.
2. Suche einen Treffer für beide Datenbanken mit moderaten E-values und vergleiche sie. Sind sie unterschiedlich? Warum?

Aufgabe 3

1. Benutze die unbekannt Proteinsequenz von <http://bibiserv.cebitec.uni-bielefeld.de/sadr2/databasesearch/fasta/exercise1/sequence2.fas> und nutze *FASTA* für eine Suche in *SwissProt*.
2. Suche auch mit *blastp* in der Datenbank *SwissProt*.
3. Vergleiche die Ergebnisse.

Aufgabe 4

1. <http://bibiserv.cebitec.uni-bielefeld.de/sadr2/databasesearch/fasta/exercise2/sequence3.fas> ist eine Sequenz mit einem Sequenzierfehler. Identifiziere die Position dieses Fehlers in der Sequenz.
2. Suche mit *blastx* in *SwissProt*.
3. Suche mit *fasty* in *SwissProt*.
4. Vergleiche die Ergebnisse.

Aufgabe 5

1. Führe eine "Sequenzsuche" gegen die *Pfam* Datenbank (*HMMER* Suche) durch. Nutze das unbekannt Protein aus der vorherigen Aufgabe.
2. Wie viele unterschiedliche Treffer bekommst du?
3. Warum gibt es im Vergleich zu den *FASTA* Ergebnissen nur so wenige Treffer?
4. Was kannst du über die Proteinzusammensetzung und gegebenenfalls die Funktion herausfinden?