

# Übungen zum Sequenzanalyse-Praktikum

Universität Bielefeld, SoSe 2015

Dr. Roland Wittler · M.Sc. Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2015summer/SequaPrak>

[praktikum-seqan@CeBiTec.Uni-Bielefeld.DE](mailto:praktikum-seqan@CeBiTec.Uni-Bielefeld.DE)

**Übungsblatt 5 vom 11.05.2015**

**Abgabe bis Donnerstag, 24:00 Uhr.**

## Aufgabe 1 (BLAST)

Lade die Sequenz von <http://bibiserv.cebitec.uni-bielefeld.de/sadr2/databasesearch/blast/exercise1/sequence.fas> herunter (Sie wurde in Bielefeld sequenziert!)

1. Nutze *blastn* auf dem *NCBI*-Server, um ähnliche Sequenzen zu finden, die bereits bekannt sind. Welches Gen wurde hier sequenziert?
2. Ist das die komplette mRNA? Wie kannst du das herausfinden?
3. Suche jetzt mit *blastx* in der *SwissProt*-Datenbank. Vergleiche die Anzahl an gefundenen Sequenzen mit deinen Treffern in der Nukleotid-Datenbank. Diskutiere dein Ergebnis.
4. 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.)
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 GI Nummer 21694053 heraus. Um was für ein Gen handelt es sich?
2. Nutze *megablast* mit Standardeinstellung, um deine Sequenz gegen die Nukleotid-Datenbank zu vergleichen. Wie lauten die Standardeinstellung im Abschnitt „General Parameters“ und was bedeuten sie?
3. 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 den Unterschied.

## Aufgabe 3 (FASTA und BLAST)

Auf <http://bibiserv.cebitec.uni-bielefeld.de/sadr2/databasesearch/fasta/exercise1/sequence2.fas> findest du eine unbekannte Proteinsequenz.

1. Nutze *FASTA* (auf der EMBL-Seite) für eine Suche in der *SwissProt*-Datenbank.
2. Suche auch mit *blastp* in der *SwissProt*-Datenbank.
3. Vergleiche und diskutiere die Ergebnisse.

## Aufgabe 4 (Sequenzierfehler finden)

Auf <http://bibiserv.cebitec.uni-bielefeld.de/sadr2/databasesearch/fasta/exercise2/sequence3.fas> findest du eine Sequenz mit einem Sequenzierfehler. Identifiziere die Position dieses Fehlers in der Sequenz anhand der folgenden Datenbanksuche:

Suche mit *blastx* in *SwissProt*. Beschreibe, was du bei den Alignments beobachtest und diskutiere, wie du den Fehler feststellen kannst. Erkläre, wie du die genaue Position ermitteln kannst.

### Aufgabe 5 (HMMER)

Nutze für diese Aufgabe wieder die Proteinsequenz aus Aufgabe 3.

1. Was für Sequenzen findest du in der *Pfam*-Datenbank und wie sind sie repräsentiert?
2. Nutze *phmmer*, um die Proteinsequenz gegen die *Pfam*-Datenbank zu vergleichen.
3. Betrachte die „sequence features“ und die „hit details“. Welche und wie viele unterschiedliche Treffer bekommst du?
4. Warum gibt es im Vergleich zu den *FASTA*-Ergebnissen nur so wenige Treffer?
5. Was kannst du über die Proteinzusammensetzung und gegebenenfalls die Funktion herausfinden?