

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/SequaPrak>
praktikum-seqan@CeBiTec.Uni-Bielefeld.DE

Übungsblatt 5 vom 12.05.2014
Abgabe des Protokolls bis Donnerstag, 15.05.2014

Aufgabe 1 (NCBI/GenBank)

1. Lade dir von der NCBI-Datenbank die Nukleotidsequenz des *Leptin*-Proteins der Maus mit der *Accession Number* U18812 herunter. Stelle dabei sicher, dass du die Sequenz der Original-Veröffentlichung verwendest.
2. Beschreibe den grundlegenden Aufbau des GenBank-Eintrags.
3. Gib an, welche anderen Formate noch verfügbar sind.
4. Der *Leptin*-Eintrag wurde am 30. März 1995 ersetzt. Wieso?
5. Kannst du eine Sequenz finden, die gemeinsam mit dem *Nature*-Eintrag vom 21. Dezember 1994 veröffentlicht wurde?

Aufgabe 2 (EMBL SRS)

1. Verwende die Accession Number U18812.1, um von der EMBL-Datenbank denselben Eintrag wie in der vorherigen Aufgabe aufzurufen.
2. EMBL verwendet SRS (*Sequence Retrieval System*), um seine Datenbanken abzufragen. SRS wurde ursprünglich von Thure Etzold am EBI entwickelt. Klicke auf den **Text**-Knopf oben auf der Seite, um den ursprünglichen EMBL-Eintrag im Text-Format zu erhalten. Beschreibe die grundlegende Struktur dieses EMBL-Eintrags.
3. Ist die Information anders als die von GenBank am NCBI?

Aufgabe 3 (NCBI Taxonomy)

1. Gehe mit deinem Web-Browser zur Datenbank *NCBI Taxonomy*, wo wir etwas über unsere Vorfahren herausfinden wollen. Suche nach *neanderthalensis*. Ist der Neandertaler nicht ausgestorben? Wie kann es dann Proteinsequenzen von ihm geben?
2. Welche Proteine wurden sequenziert? (Nenne mindestens fünf.) Was ist über *Osteocalcin* bekannt, im Vergleich zum modernen Menschen?
3. Gibt es auch von anderen ausgestorbenen Organismen Einträge in der Datenbank?
4. Die *Taxonomy*-Datenbank ist sehr nützlich, wenn man alle zu einer Spezies verfügbaren Sequenzen herunterladen möchte. Wie viele Sequenzen gibt es für den Menschen?

Aufgabe 4 (PDB)

1. Gehe zur Website der PDB. Ist die Proteinstruktur von *Leptin* experimentell verifiziert? Wenn ja, von wem und wo wurde sie publiziert?
2. Welche Art von Experiment wurde durchgeführt, um die Struktur zu erhalten?
3. Lade den PDB-Eintrag herunter. Verwende *rasmol* oder *jmol* innerhalb der PDB-Seite, um die Struktur zu visualisieren. Füge einen Screenshot in dein Protokoll ein.

Aufgabe 5 (UNIX-Kommandos)

Verwende UNIX-Kommandos oder selbstgeschriebene Skripte in einer Sprache deiner Wahl, um die folgenden Aufgaben zu bearbeiten.

1. Gehe noch einmal zur NCBI-Taxonomy-Datenbank und lade nun im FASTA-Format die Nukleotidsequenzen aller Mitglieder der Familie der Rattenkängurus (*Potoroidae*) herunter.
2. Wie viele Sequenzen erhältst du? War der Download erfolgreich? (Manchmal brechen große Downloads unvermittelt ab, daher ist es immer sinnvoll, zu überprüfen, ob auch tatsächlich alle Sequenzen in der Datei enthalten sind.)
3. Schreibe die Protein-IDs in eine Extra-Datei und verwende NCBI's *Batch Entrez*, um die entsprechenden Proteinsequenzen herunterzuladen.
4. Extrahiere die GI-Nummern für alle Sequenzen.
5. Die Sequenzen von wie vielen verschiedenen Genera und Spezies sind in der Datei vorhanden?

Aufgabe 6 (readseq)

Ein sehr nützliches Werkzeug, um Sequenzdateien zwischen verschiedenen Formaten zu konvertieren, ist *readseq*. Verwende die Web-Version <http://www.ebi.ac.uk/cgi-bin/readseq.cgi>, um in der folgenden Aufgabe Formate zu konvertieren und Features zu extrahieren. (Alternativ kannst du auch UNIX-Kommandos und einen Texteditor verwenden.)

1. Verwende den *ENA*-Service auf der EBI-Webseite, um alle Nukleotidsequenzen der Mitglieder der Familie der Rattenkängurus (*potoroidae*) im EMBL-Format herunterzuladen. Hierbei ist es wichtig, *ENA* (<http://www.ebi.ac.uk/ena/>) zu verwenden und nicht die allgemeine Suche.
2. Wie viele Sequenzen sind verfügbar? War der Download erfolgreich?
3. Wie viele der Einträge enthalten das Feature *5'UTR*?
4. Wie viele und welche tRNA-Gene sind annotiert?