

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

Übungsblatt 4 vom 04.05.2015

Abgabe bis Donnerstag, 24:00 Uhr.

Aufgabe 1 (NCBI/GenBank)

Suche die vollständige *coding sequence* des menschlichen P53 Gens in GenBank.

1. Wie lauten die *Accession* und die *GI Number*?
2. Wie viele Basenpaare hat die Sequenz?
3. Wer ist für die initiale Eintragung der Sequenz in GenBank verantwortlich?
4. Was ist P53?
5. Finde mit Informationen des GenBanks Eintrags heraus, wie häufig Mutationen im P53 Gen in Krebs auftreten. Wo findest du diese Angabe?

Aufgabe 2 (EMBL SRS)

1. Welchen Eintrag findest du in der EMBL-Datenbank unter der *Accession Number* AB082923?
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?
4. Was bedeutet der Eintrag */db_xref="PDB:1A1U"*?

Aufgabe 3 (PDB)

1. Suche in PDB den Eintrag mit der ID *1A1U*. Was findest du?
2. Auf welchen Artikel bezieht sich dieser Eintrag?
3. Welche Art von Experiment wurde durchgeführt, um die Struktur zu erhalten?
4. Suche nun nach „p53“. Wie viele Einträge gibt es zum *tumor suppressor*? Wie viele stammen aus dem Menschen? In welchen Jahren wurden die meisten Einträge hinzugefügt?

Aufgabe 4 (NCBI Taxonomy)

1. Gehe zur Datenbank *NCBI Taxonomy*, wo wir etwas über unsere Vorfahren herausfinden wollen. Suche nach *Homo neanderthalensis*. Ist der Neandertaler nicht ausgestorben? Wie kann es dann Proteinsequenzen von ihm geben?
2. Wie viele Proteine wurden vom Neandertaler sequenziert? Was ist über *Osteocalcin* bekannt, im Vergleich zum modernen Menschen?
3. Von welchen anderen ausgestorbenen Organismen gibt es auch noch Einträge in der Datenbank?
4. Die *Taxonomy*-Datenbank ist sehr nützlich, wenn man alle zu einer Spezies verfügbaren Sequenzen herunterladen möchte. Wie viele Nukleotid-EST-Sequenzen findest du beim Menschen?

Aufgabe 5 (UNIX-Kommandos)

Verwende UNIX-Kommandos, um die folgenden Aufgaben zu bearbeiten. Gib bei der Beantwortung deiner Fragen auch die Kommandos an, die du benutzt hast.

Gehe noch einmal zur *NCBI Taxonomy*-Datenbank und lade nun im GenBank-Format die Nukleotidsequenzen aller Mitglieder der Familie der Rattenkängurus (*Potoroidae*) herunter.

1. 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.)
2. Schreibe die Protein-IDs in eine Extra-Datei und verwende NCBIs *Batch Entrez*, um die entsprechenden Proteinsequenzen angezeigt zu bekommen und lade sie im FASTA-Format herunter. Wie viele Einträge bekommst du?
3. Die Sequenzen von wie vielen verschiedenen Spezies sind in der Datei vorhanden?