

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/sequapra>

praktikum-seqan@CeBiTec.Uni-Bielefeld.DE

Übungsblatt 4 vom 07./08.11.2017

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

Aufgabe 1 (NCBI/GenBank)

Suche in GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) nach dem Eintrag mit der *Accession.Version*-Nummer: *KR230391.1*.

1. Von welchem Organismus stammt der Eintrag? Um was für eine Organelle handelt es sich? Ist das Genom vollständig?
2. Wie lautet die GI-Nummer und ist es sinnvoll, sie als einzige identifizierende Angabe für den Eintrag zu benutzen?
3. Wer hat den Eintrag in die Datenbank eingebracht?
4. In welcher Zeitschrift wurde das Genom veröffentlicht?
5. Mit welcher Sequenzieretechnologie wurde das Genom sequenziert?
6. Wie viele Basenpaare hat die Sequenz?

Aufgabe 2 (Sequenzlängenhistogramm)

Schreibe ein Programm, das eine Datei im multiple FASTA-Format einliest und entweder ein Histogramm der Sequenzlängen erstellt oder eine Ausgabe hat, aus der mit einem anderen Programm, z. B. mit Excel, ein Histogramm erstellt werden kann. Es sollen nur Sequenzlängen beachtet werden, die kürzer oder gleich 1000 bp sind. Wähle eine Balkenbreite von 100 bp in deinem Histogramm und eine sinnvolle Achsenbeschriftung.

Lade dir nun die kodierenden Sequenzen des Eintrags mit der *Accession.Version*-Nummer *KR230391.1* im FASTA-Format herunter und wende dein Programm darauf an. Bilde das erhaltene Histogramm bitte in deinem Protokoll ab.

Aufgabe 3 (UNIX-Kommandos)

Lade dir den Genbank-Eintrag mit der *Accession.Version*-Nummer: *KR230391.1* herunter. Verwende UNIX-Kommandos, um die folgenden Aufgaben zu bearbeiten. Gib bei der Beantwortung deiner Fragen auch die Kommandos an, die du benutzt hast.

1. Wie viele Gene werden in dem Eintrag gefunden?
2. Wie viele tRNAs kommen in komplementärer Richtung vor?
3. In dem Features-Bereich kommen verschiedene Attribute, wie z. B. *organism* oder *gene* vor. Zähle alle dieser unterschiedlichen Attribute auf. (Tipp: Die Zeilen mit den Attributen sind die einzigen, in denen ein Gleichheitszeichen vorkommt.)

Aufgabe 4 (EMBL/ENA)

1. Welchen Eintrag findest du in der EMBL-Datenbank (<http://www.ebi.ac.uk/ena>) unter der *Accession.Version*-Nummer *KR230391.1*?
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 in zwei bis drei Sätzen.
3. Ist die Information anders als die von GenBank am NCBI?