

Softwarewerkzeuge der Bioinformatik

Prof. Dr. Volkhard Helms
PD Dr. Michael Hutter, Markus Hollander,
Andreas Denger, Marie Detzler, Velik Velikov
Winter Semester 2021/2022

Universität des Saarlandes
Zentrum für Bioinformatik

Übungsblatt 1

Sequenzanalyse: Biologische Datenbanken

Lernziel: Sie sollen lernen, wie Sie mit Datenbanken Informationen wie z.B. Sequenzen, IDs, Mutationen, Proteinfamilien, Genompositionen, Transportmechanismen, Drug Targets etc. über bestimmte Gene, Proteine und Organismen in Erfahrung bringen können.

Aufgabe 1.1: NCBI und UniProtKB

In dieser Aufgabe verwenden Sie NCBI (<http://www.ncbi.nlm.nih.gov>) und UniProtKB (<http://www.uniprot.org/>), um Fragen über ein Protein zu beantworten. Die Identifikationsnummer (*accession number*) des Proteins in der UniProtKB/Swiss-Prot Datenbank lautet **Q38856**.

- a) Um welches Protein handelt es sich?
- b) Aus welchem Organismus stammt es?
- c) Welche Funktion hat es?
- d) In welchem Zellkompartiment ist es zu finden?
- e) Wurden für dieses Protein Mutationen gefunden, die einen Einfluss auf die Funktion haben? Wenn ja, welche (Position und Aminosäure-Austausch) und was ist deren funktionelle Konsequenz?
- f) Wie heißt das zugehörige Gen und auf welchem Chromosom befindet es sich?
- g) Speichern Sie die Proteinsequenz im FASTA-Format.
- h) Zu welcher Proteinfamilie gehört das Protein?
- i) Finden Sie ein weiteres Protein derselben Familie aus demselben Organismus. Geben Sie die *accession number* des Proteins an.
- j) Suchen Sie ein Protein, das aus dem Organismus *Oryza sativa* stammt und dieselbe Funktion, Zellkompartiment und Familie wie **Q38856** hat. Geben Sie die *accession number* an.

Aufgabe 1.2: TCDB – Transporter Classification Database

In dieser Aufgaben benutzen Sie die TCDB (<http://www.tcdb.org>) zur Klassifizierung des Proteins **Q38856** aus Aufgabe 1.1.

Hinweis: Die Transport Classification ID (TC-ID) besteht aus fünf Komponenten X.A.Y.Z.V, wobei X, Y, Z und V Zahlen sind und A ein Buchstabe.

X: Transporter-Klasse
A: Transporter-Subklasse
Y: Transporter-(Super)Familie
Z: Transporter-Subfamilie

- a) Welche TC-ID hat das Protein?

- b) Wie heißt die Transporter-Klasse, zu der das Protein gehört?
- c) Zu welcher Transporter-Familie gehört das Protein?
- d) Beschreiben Sie die strukturelle Gemeinsamkeiten von Proteinen dieser Familie.
- e) Welche Substrate werden durch Proteine dieser Familie transportiert?
- f) Was ist über den Transportmechanismus bekannt? Was ist der Unterschied zu ATP gestütztem Transport?

Aufgabe 1.3: PATRIC - Pathosystems Resource Integration Center

In dieser Aufgabe nutzen Sie PATRIC (<http://patricbrc.org/>), um Informationen über die Gattung *Staphylococcus* herauszufinden.

- a) Was können Sie über die Abstammung (*lineage*) der Staphylokokken sagen?
- b) Im Folgenden geht es um Antibiotikaresistenzen und Drug Targets im Bakterium *Staphylococcus aureus*, genauer *Staphylococcus aureus subsp. aureus N315*.
 - i. Welche Datenbanken werden bzgl. Antibiotikaresistenz von PATRIC abgefragt?
 - ii. Wie viele Resistenzgene werden angezeigt? Nennen Sie ein paar Beispiele.
 - iii. Was ist die Funktion von *mecA* bzw. des entsprechenden Proteins?
 - iv. Welche Drug Targets finden Sie und aus welchen Datenbanken stammen diese?

Viel Spaß!