Softwarewerkzeuge der Bioinformatik

Prof. Dr. Volkhard Helms   Saarland University
PD Dr. Michael Hutter, Markus Hollander,   Center for Bioinformatics
Marcial Josef Paszkiel
Winter Semester 2022/2023

Exercise Sheet 1

Sequence Analysis: Biological Databases

Learning objective: *The goal is to learn how to use databases to find information about specicfic genes, proteins and organisms, such as sequences, IDs, mutations, protein families, genome positions, transport mechanisms, drug targets etc.*

Exercise 1.1: NCBI and UniProtKB

Use UniProtKB (https://www.uniprot.org/) to answer questions about a protein. The *accession number* of the protein in the UniProtKB/Swiss–Prot database is **Q38856**.

a) What is the protein?

b) What is the name of the organism?

c) What is its function?

d) In which cellular compartment can it be found?

e) Have mutations been found for this protein that affect its function? If so, which ones (position and amino acid exchange) and what are the functional consequences?

f) What is the name of the corresponding gene and on which chromosome is it located?

g) Save the protein sequence in FASTA format.

h) To which protein family does the protein belong?

i) Find one more protein of the same family in the same organism. Name the *accession number.*

j) Find another protein from the organism *Oryza sativa* with the same function, cellular compartment and family as **Q38856**. Write down the *accession number.*

Exercise 1.2: TCDB – Transporter Classification Database

In this exercise you use TCDB (https://www.tcdb.org) to classify protein **Q38856** of the previous exercise.

Hint: The transporter classification ID (TC–ID) is comprised of five components X.A.Y.Z.V, where X, Y, Z and V are number and A a letter.

> X: transporter class
> A: transporter subclass
> Y: transporter (super)family
> Z: transporter subfamily

a) What is the TCDB–ID of the subfamily to which the protein belongs?

b) What is the name of the transporter class to which the protein belongs?

c) What is the name of the transporter family?

d) Describe the structural similarities of proteins in this family.

e) Which substrates are transported by proteins of this family?

f) What is known about the transport mechanism? What is the different to ATP based transport?

Exercise 1.3: BV–BRC – Bacterial and Viral Bioinformatics Resource Center

In this exercise you use BV–BRC (https://www.bv-brc.org/) to find information about the species *Staphylococcus*.

a) What can you find out about the lineage of staphylococci?

b) The following is about antibiotic resistances and drug targets in the bacterium *Staphylococcus aureus*, specifically *Staphylococcus aureus subsp. aureus N315*.

  i. Which databases are accessed by BV–BRC regarding antibiotic resistance?

  ii. How many genes associated with antibiotic resistance are listed? Name some examples.

  iii. What is the function of mecA or of the corresponding protein?

  iv. Which drug targets can you find and from which databases?

Have fun!