Use of the UniProtKB (protein sequence) database and homology search with FASTA and BLAST/PSI-BLAST

niProtKB

- Go to https://legacy.uniprot.org
- Type AQP11 – see how many entries you find
- Now type AQP11 AND horse - see how many entries you find; how many of them are reviewed, how many are unreviewed
- Access the Advanced search, select "gene name" while typing "aqp11" AND select "Organism" while typing "horse" – see how many reviewed/unreviewed entries you find
- Explore the Entry with Uniprot ID: **F6S3G9**, by clicking on it

niProtKB

- Go to https://legacy.uniprot.org

- After clicking on the entry **F6S3G9**
- See in what chromosome is located the corresponding gene
- See where the protein is located in the cell. Cytoplasm? Membrane? Extracellular?
- In what tissue is it expressed?
- Visualize the GeneTree in Ensembl (through the provided link)
- Visualize the protein sequence – how long it is?
- Select the FASTA format for the protein sequence (we'll use it for the homology search)

ASTA
put

- Go to https://www.ebi.ac.uk/Tools/sss/fasta/

& select **Protein**

- Select the UniProtKB/Swiss-Prot(swissprot) database
- In the input window copy the protein FASTA sequence (ID: **F6S3G9)**
- Have a look at the algorithm parameters by clicking on "More Options" under "Enter your parameters ("; take note of the range for the E(Expectation)-value
- Run the Job, by clicking on **Submit**

ASTA
tput

- Go to https://www.ebi.ac.uk/Tools/sss/fasta/

& select **Protein**

- Take note of the top 5 hits; what values are reported for each hit in the output summary table?
- Have a look at the "Visual Output": how many sequences are significantly similar to the query?
- Have a look at the "Functional Predictions"
- Go back to the "Summary Table" and click on "show" below "Alignments", on the left
- Have a look at the alignment with the first human hit and to all the values reported above the alignment itself (scores, E-value etc.). What do they mean?

SI-)BLAST
put

- Go to https://blast.ncbi.nlm.nih.gov/Blast.cgi

& select **Protein BLAST**

- In the input window copy the protein FASTA sequence or the Uniprot ID (**F6S3G9)**
- Select the UniProtKB/Swiss-Prot(swissprot) database
- Select the PSI-BLAST algorithm
- Run the Job, by clicking on **BLAST**

SI-)BLAST
tput

- Go to https://blast.ncbi.nlm.nih.gov/Blast.cgi & select **Protein BLAST**

- See the species, sequence identity, Score, E-value and Query cover of the top 5 hits
- Visualize the **graphic summary**: how many hits are found? How many very good hits are found?
- Compare results with those obtained by FASTA
- Visualize the alignment with the human sequence; take note of the sequence similarity and gaps
- Run PSI-BLAST Iteration 2

SI-BLAST
tput

- Go to https://blast.ncbi.nlm.nih.gov/Blast.cgi

& select **Protein BLAST**

- How many hits are found at PSI_BLAST Iteration 2
- How many new hits have a low enough E-value (let's say below $1x10^{-10}$)?
- Select only the hits with a low enough E-value for building the PSSM for PSI-BLAST iteration 3
- Were new hits found at Iteration 3? If so, few or many?
- Have a look at the top-1 new hit (sequence identity, etc) if any
- Visualize the alignment
- Visualize the Graphic Summary now; how different is it from that of Iteration 1?

HPRED
put

- Go to https://toolkit.tuebingen.mpg.de/tools/hhpred

- In the input window copy the protein FASTA sequence or the Uniprot ID (**F6S3G9)**
- Select the PfamA database (after removing default-selected databases)
- Run the Job, by clicking on **Submit**

HPRED
tput

- Go to https://toolkit.tuebingen.mpg.de/tools/hhpred

- Was the job faster or slower than the previous ones?
- Take note of the job steps visualized on the screen while in progress
- How many hits are found? How significant?
- Explore the first hit in Pfam

(PF00230.23)