Introduction to Bioinformatics
BIOL-3951 / COMP-3550 / BIOL-7941
Lab 3: Pairwise Sequence Alignment

Read the document "HowTo_BLASTGuide" available in D2L and then complete all the tasks described below.

1. **Question:** You have to find out fungi proteins similar to the human beta globin (NP_000509).

    Resources: Go to NCBI BLAST page https://blast.ncbi.nlm.nih.gov/Blast.cgi and click on Protein BLAST.

    Analysis:

    First, perform a blastp search against the Refseq database restricted to fungi (taxid: 4751) by entering the accession number and restricting the database. Note how many significant hits you get and their E-values.

    Now repeat the search using either PSI-BLAST or DELTA-BLAST (select Refseq as the database to use and restrict the search to fungi). Note how many significant hits you get now. If you are using PSI-BLAST, perform at least three iterations. Which search returns the most significant hits and more globin-like sequences?

    Download the PSSM [click download on the top of the results page and select PSSM to restart search]. Go to the PSSM viewer to display the PSSM. Can you identify highly conserved positions? Write down, the first five highly conserved positions you identified and make sure you understand the PSSM contents.

Highly conserved positions:	

2. **Question:** Find out whether a wolf (*Canis lupus*) protein is significantly similar to a mouse E3 ubiquitin-protein ligase Itchy (Q8C863). Any similar protein should contain the protein domain WW/rsp5/WWP.

    Resources: PHI-BLAST, PROSITE.

    Analysis:

    Get the pattern for the protein domain WW/rsp5/WWP. To do this, navigate to PROSITE web page and search for the domain using the identifier PDOC50020. Browse the result page and copy/paste the consensus pattern to be entered as PHI pattern.

    Using the mouse E3 ubiquitin-protein ligase Itchy (Q8C863) as query and the PROSITE pattern for the domain WW/rsp5/WWP (PDOC50020) perform a PHI-BLAST search against the reference proteins database.

    Fill the following table with the information for the wolf protein you identified.

Accession	Description	E-value	% identity	% Coverage	Alignment Coordinates

3. **Question:** You need to quickly characterize the sequence below:

   MSMLPSFGFTQEQVACVCEVLQQGGNLERLGRFLWSLPACDHLHKNESVLKAKAVVAFHR
   GNFRELYKILESHQFSPHNHPKLQQLWLKAHYVEAEKLRGRPLGAVGKYRVRRKFPLPRT
   IWDGEETSYCFKEKSRGVLREWYAHNPYPSPREKRELAEATGLTTTQVSNWFKNRRQRDR
   AAEAKERENTENNNSSSNKQNQLSPLEGGKPLMSSSEEEFSPPQSPDQNSVLLLQGNMGH
   ARSSNYSLPGLTASQPSHGLQAHQHQLQDSLLGPLTSSLVDLGS

   Resources: SMART-BLAST.

   Analysis:
   Enter the sequence in the [SMART-BLAST](#) and explore the results page.

   - What protein is this sequence?
   - From which species are the closest orthologs?
   - From which type of organism would you think the sequence came from? Plant?
     Animal? Bacteria? Be a specific as possible.

4. **Question:** Complete Cases 4 "Annotate a Metagenomic Contig" and 5 "Examine Conserved
   Domains and Solved Structures to Support a Protein Annotation" described in the
   document "Five examples for NCBI BLAST" available in D2L.

When you are finished, answer the lab quiz through D2L.