

Next week we'll learn about xml parsing and parse an xml file to help us quantify genetic diversity in our sequence. To prepare, this week you'll go to NCBI's BLAST website and run a nucleotide BLAST using BLASTn.

Adjusting parameters as necessary, and making note of all settings below:

- 1) Limit the results returned to 50.
- 2) Limit your search to an organism relevant to the known biology of your gene.

For example:

- * if your gene is involved in a human disease, limit your results to humans.
- * if your gene is involved in mad cow disease, limit your results to cows (*Bos taurus*).

Begin by limiting your search to the RefSeq mRNA database (refseq_rna). This query may not return any results - that's ok.

If this limited BLAST does not return any results:

- * Add a raw text cell below and write a few sentences on why this might be. Things that might be relevant to address include: is the organism highly studied (humans or a model organism), is the condition/disease associated with the gene very rare, might the nucleotide sequence be very conserved.

- * Re-run your BLAST using a broader taxonomic limitation. For example, instead of limiting results to *Homo sapiens*, expand the limit to primates (taxid:9443). Instead of limiting results to *Bos taurus*, expand the limit to the genus *Bos*, or even further to even-toed ungulates (taxid:91561).

- 3) Download the results of your BLAST as an xml file. This option can be found on the results page in the "Descriptions" section, under "Sequences producing significant alignments:", click on "Select All", then click the "v" to the right of "Download", select "XML" and click "Continue". Save this file to your thumbdrive. Rename the file "BLASTn_" + your gene abbreviation + .xml. For example a BLASTn using the gene PRNP as a query would be named "BLASTn_PRNP.xml".