**BIFX-550  
S. Ravichandran, Ph.D.  
Biology, Hood College, Frederick, MD 21702**

Ongoing Assignment-bifx550b

Your Name: James Jedediah Smith

There will be three ongoing assignments, bifx550a-c. Assignments will be released as we progress through the class. These assignments are setup to help the students complete BIFX550’s find-a-gene project. You will use your find-a-gene query (gene/protein) to answer these questions. If you are working on bifx550a, please download the file, rename it as your firstinitiallastname\_bifx550a.docx (example, SRavichandran\_bifx550a.doc). Complete the assignment and upload the document in Hood College BIFX550 class BlackBoard site before the deadline

*Please note when I say, “this gene” or “this protein”, it means your query gene/protein respectively.*

**Assignment bifx550b (for due date lookup Blackboard announcement pages)**

1. Can you identify any homologs (report up to ten) for your gene? From the results can you say whether they are orthologs or paralogs?

* [MID1, *M. mulatta*](https://www.ncbi.nlm.nih.gov/gene/709683) (Ortholog)
* [MID1, *C. lupus*](https://www.ncbi.nlm.nih.gov/gene/491737) (Ortholog)
* [MID1, *B. taurus*](https://www.ncbi.nlm.nih.gov/gene/534604) (Ortholog)
* [Mid1, *M. musculus*](https://www.ncbi.nlm.nih.gov/gene/17318) (Ortholog)
* [Mid1, *R. norvegicus*](https://www.ncbi.nlm.nih.gov/gene/54252) (Ortholog)
* [MID1, *G. gallus*](https://www.ncbi.nlm.nih.gov/gene/373920) (Ortholog)
* [mid1, *X. tropicalis*](https://www.ncbi.nlm.nih.gov/gene/448651) (Ortholog)
* [mid1, *D. rerio*](https://www.ncbi.nlm.nih.gov/gene/100330952) (Ortholog)

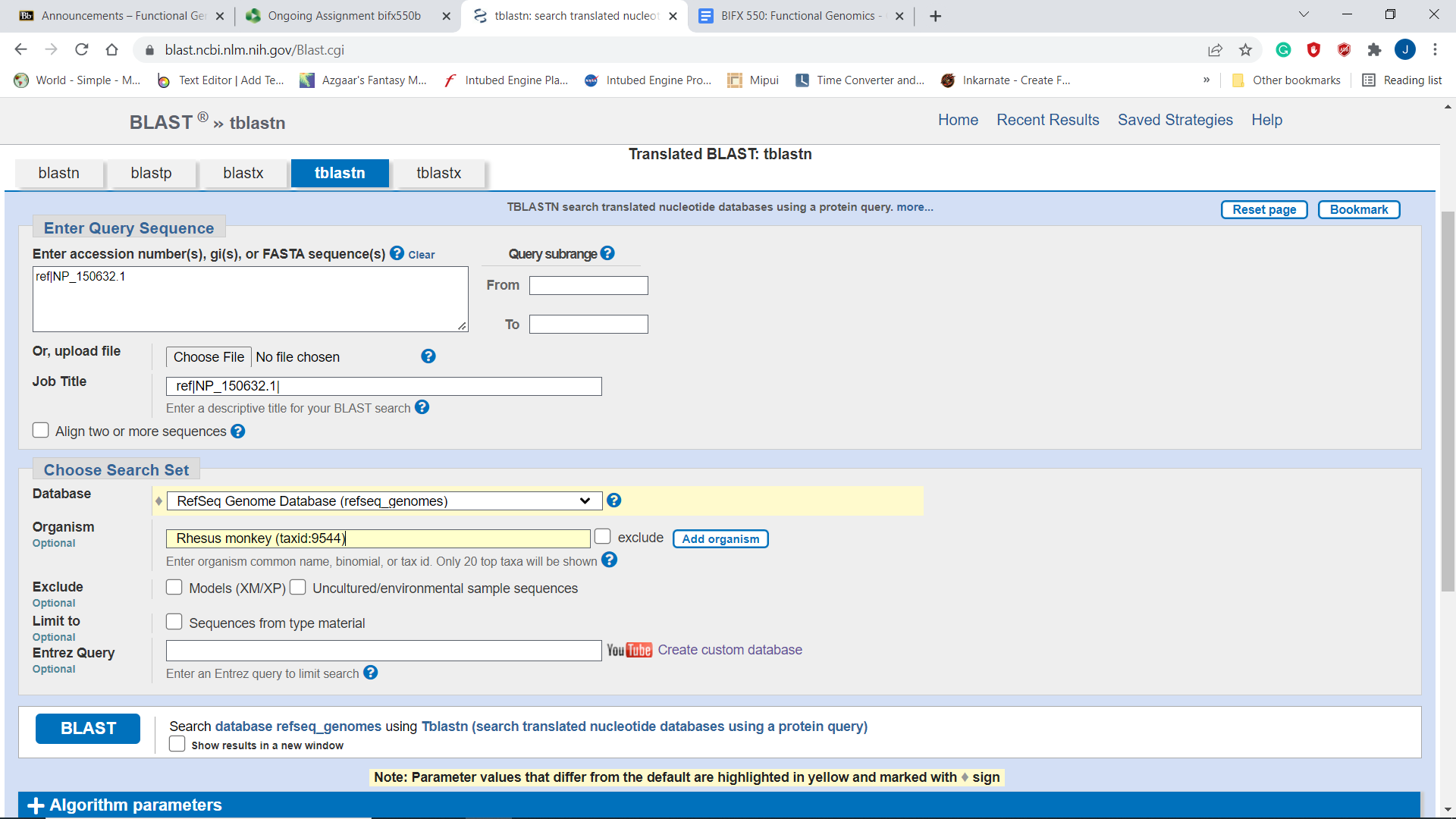
1. We are on a journey to find a novel gene. Here is the definition of a novel gene, “*For our purposes a “new” gene refers to the discovery of some DNA sequence in a database that is not annotated (described).”* You will be doing lots and lots of BLAST (and many of its flavors, BLASTN etc.) searches in this class. Do you understand, how TBLASTN, BLASTX, and BLASTP works? Please explain the different flavors (i.e., the ones shown above) of BLAST.

* TBLASTN: Identifies which nucleotides are most related to protein query.
* BLASTX: Identifies which proteins are most related to nucleotide query.
* BLASTP: Identifies which proteins are related to a protein query.

1. For this problem, you are going to select a homolog (or specifically a ortholog) that is known for your query gene. For example, your protein could be a human protein (from Gene-X) and you might be choosing a mouse protein (from Gene-X) as a homolog for your query protein. You are going to use BLAST search to show that you know how to identify this homologous protein. Here are the steps:
   1. Use your starting query protein (same as Section I. Question 1) and carry out a TBLASTN using an organism (that you have already identified) and DNA database (ex. HTGS, dbEST or genomic) to re-identify the homolog that you had chosen. This exercise will mimic the process that you will be using many times until you find a novel gene. Show the BLAST input page (including algorithm parameters) used for this search, and also show the results (see below for additional information)
      1. For the BLAST results, change the font to Courier size 10 so that the results are displayed neatly. You can also screen capture a BLAST output. It is not necessary to print out all the BLAST results if there are many pages.

Graphical user interface, application, table

Description automatically generated



* 1. Check and report the E-value for the homolog

The E-Value was 9e-142.

* 1. Did you see any false positives or false negatives?

No. The only one with anywhere near 100% identity was the one previously identified as a homolog.

* 1. Take the homolog (DNA) and do a BLASTx search against nr protein database, do you get hit (100%) identity from the same organism? Answer (Yes or no) and explain the result.

Yes, it got 100% identity. This makes sense because you are using DNA from one organism to search for the associated protein within that same organism. It seems like a less than 100% identity might be cause for concern here.

Graphical user interface, text, application, email

Description automatically generated

* 1. Take the homolog (protein; identified in step a) and do a BLASTP search, did you get any hit(s)? Explain the result.

Yes, there was a 100% hit on the original Human protein used to identify the ortholog in the first place. This makes sense because it wouldn’t be an ortholog if it wasn’t very similar to the original.

Graphical user interface, text, application, email

Description automatically generated