

Biol5705  
Module: Gene Sequence Analysis  
Assignment 1

Email results to [morgan.g.i.langille@gmail.com](mailto:morgan.g.i.langille@gmail.com) before start of next class (Jan. 23).

- Download gene sequence "assignment\_1.fasta" from:  
<http://morganlangille.com/teaching/biol5705/assignment1.fasta>
- Open the file in a text editor.

Q1: How many sequences are in the file?

Q2: Are the sequences DNA or protein?

- You are told that this gene is from humans, but you don't know what the gene does or where in the genome it is located.

Q3: What type of BLAST program should you use to search with this sequence?

Q4: What NCBI BLAST database should you choose?

Q5: Are there any other obvious non-default options you should specify?

- Use the NCBI BLAST web server (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) to run your chosen BLAST program with the options you specified.

Q6: What is the e-value, total bit score, identity, alignment length and query coverage of your top hit?

Q7: What is the gene name and gene product from your top hit?

Q8: Describe the differences in your sequence compared to the best one found in the database in terms of gaps and kinds of substitutions.

Q9: Identify the best scoring sequence that has more than a single hit. What is the name of this gene? Describe the locations of the similarity to your sequence.

Q10: Identify the closest ortholog to your sequence within the refseq database and name that organism as well as answer Q6 for this ortholog.