

Biol5705
Module: Gene Sequence Analysis
Assignment 1

Email results to morgan.g.i.langille@gmail.com before start of next class.

- 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? (1pt)

Q2: Are the sequence(s) DNA or protein? (1pt)

- 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? (1pt)

Q4: What NCBI BLAST database should you choose? (1pt)

Q5: Are there any other obvious non-default options you should specify? (1pt)

- 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 best hit? (1pt)

Q7: What is the gene description of your top hit? (1pt)

Q8: Describe the differences in your sequence compared to the best one found in the database in terms of gaps and kinds of substitutions. (1pt)

Q9: Identify the best scoring sequence that has more than a single hit (i.e. two or more aligned regions to the same sequence). What is the description of this gene? Describe the locations of the similarity to your sequence. (1pt)

Q10: Identify the closest ortholog to your sequence within the RefSeq database and name that organism. (1pt)