

## Problem Set 2

**DUE: September 11, 11:59 PM**

Answer **all** of the following questions related to performing the requested tasks and then include your answers in an electronic file, save as a pdf, and merge together as a single pdf with bookmarks added to identify different sections. Upload your pdf file to the Blackboard Assignment page for Problem Set 2 by 5 pm on the due date. Show your work or explain how you found your answer where possible.

1. Download a nucleotide sequence from GenBank for a gene of interest. What gene did you download? Was the sequence published? Where, when, and by whom? What format is your nucleotide sequence in? Why is this gene of interest?
2. Perform a BLAST search with your DNA sequence. What does it match to (show the top 10 hits)? Are they from the same study or different studies? Is your sequence protein coding or not? What is the E-value of hit number 10 compared to hit number 1? What is an E-value?
3. Produce a pdf file with a short fragment (no more than 500 bps) of your sequence and the top 5 hits from your BLAST search showing them in FASTA format plus hits from at least 3 different species. What species are included in your data? List the taxonomic hierarchy of these species?
4. How many characters are in your FASTA file? How much space would you need to store a FASTA file of a human genome? A bacterial genome? A viral genome? How did you calculate this?
5. Translate the following sequence to amino acids (assuming the reading frame starts at the first letter):

```
ccgtcgtagc accgagcctc agcaccacga aagagattga  
agtagttcct cggaaagtgc ttcgactctt ccttgaaaca  
tgtcttcctg gagcaaccaa cctgccatgg atgattatgg
```

What does this sequence encode? From what organism does this sequence come? What is the function of this gene? What are the proportions of polar, non-polar, and charged amino acids (attach a link to or copy of the resource you used to help you determine this)?