Problem Set 1

Answer 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.

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 separate PDF file with a short fragment (no more than 50 bps) of your sequence and the top 5 hits from your BLAST search showing them in FASTA format.

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. Design a pair of PCR primers to amplify the first 50 bps of the first sequence (list the sequence first and then the PCR primers). What complications did you worry about in your primer design? Will your primers amplify other genes? The same gene from other species? How did you test this?

6. Using the command-line, create a directory called "BISC2584" and then a subdirectory called "Homework". Move your FASTA file with your data into this subdirectory. List the directory and take a screen shot and save it to your pdf file to include in your homework write-up. Your screenshot must show all of these steps.

7. Translate the following sequence to amino acids (assuming the reading frame starts at the first letter):

```
CCGTCGTAGCACCGAGCCTCAGCACCACGAAAGAGATTGA
AGTAGTTCCTCGGAAAGTTCTTCGACTCTTCCTTGAAACA
TGTCTTCCTGGAGCAACCAACCTGCCATGGATGATTATGG
```