BIOL 5153, Practical Programming for Biologists
Assignment #4
Due Monday, 02 March 2020, by 11:59 PM.

Note: READ AND FOLLOW THESE INSTRUCTIONS CAREFULLY. DO EXACTLY WHAT IS
ASKED OF YOU, AND TURN IN EXACTLY WHAT IS ASKED OF YOU.

Note: Record your answers in the file, 'last_name_assn04.xlsx',
available from the 'assignments/assn_04' folder on Blackboard.
Substitute your last name (all lower case) for 'last_name'.

1. Using the default parameters, use BLASTN to search the watermelon
   nad4L **nucleotide** sequence to a database consisting of the watermelon
   nad4L **nucleotide** sequence only. Record results for the **TOP** hit only
   in last_name_assn04.xlsx. Do this on Razor or Trestles, and use a
   job submission script called last_name_assn04.1.pbs. Turn in the
   merged STDOUT+STDERR file called last_name_assn04.1.out.

2. Using the default parameters, use BLASTN to search the watermelon
   nad4L **nucleotide** sequence to a local database consisting of the
   nucleotide sequences of all the genes (protein-coding, rRNA, and
   tRNA) in the watermelon mitochondrial genome. Record results for the
   TOP hit only in last_name_assn04.xlsx. Do this on Razor or Trestles,
   and use a job submission script called last_name_assn04.2.pbs. Turn
   in the merged STDOUT+STDERR file called last_name_assn04.2.out.

3. Using the default parameters, use BLASTN to search the watermelon
   nad4L **nucleotide** sequence to a database consisting of the complete
   watermelon mitochondrial genome ('watermelon.fsa'). Record results
   for the **TOP** hit only in last_name_assn04.xlsx. Do this on Razor or
   Trestles, and use a job submission script called
   last_name_assn04.3.pbs. Turn in the merged STDOUT+STDERR file called
   last_name_assn04.3.out.

4. Using the default parameters, use BLASTN to search the watermelon
   nad4L **nucleotide** sequence to a nucleotide database of sequenced
   plant mitochondrial genomes (located in
   watermelon_files/mt_genomes). Record results for the **TOP** hit only in
   last_name_assn04.xlsx. Do this on Razor or Trestles, and use a job
   submission script called last_name_assn04.4.pbs. Turn in the merged
   STDOUT+STDERR file called last_name_assn04.4.out.

5. Using the default parameters, use BLASTN to search the watermelon
   nad4L **nucleotide** sequence to the NCBI 'nr' nucleotide database.
   Record results for the **TOP** hit only in last_name_assn04.xlsx. Do
   this on Razor or Trestles, and use a job submission script called
   last_name_assn04.5.pbs. Turn in the merged STDOUT+STDERR file called
   last_name_assn04.5.out.

6. Using the default parameters, use BLAST to search the watermelon
   nad4L **nucleotide** sequence against the watermelon mitochondrial
   genome ('watermelon.fsa'). Choose an algorithm that will search the
   translated nad4L nucleotide sequence against the genome. Record
   results for the **TOP** hit only in last_name_assn04.xlsx. Do this on

Razor or Trestles, and use a job submission script called last_name_assn04.6.pbs. Turn in the merged STDOUT+STDERR file called last_name_assn04.6.out.

7. Look at your data from questions 2–6, what is the relationship between the raw score and the size of the database? The bit score and the size of the database? The e-value and the size of the database? Why? Explain this pattern in **no more than 4 clear, complete sentences**. Put your answer in a file called last_name_assn04.7.txt.

8. One way to find repetitive sequences in a genome is to BLAST a genome against itself. Find repeats in the watermelon mitochondrial genome using (1) default parameters, (2) "somewhat sensitive" parameters (available from the BLAST website), and (3) sensitive parameters. Sensitive parameters are as follows: match reward = 5, mismatch penalty = -4, gap open = 8, gap extension = 6, and word size = 7. Excluding the whole-genome match to itself, how many repeats do you find under the three settings? What is the size of the smallest repeat under each setting? What parameter does this size directly relate to? Run these analyses on Razor or Trestles. Put your answer in a text file called 'last_name_assn04.8.txt' (all lower case). Be clear and concise.

9. Put all of the required files in a folder called 'assn04' and push it to your class GitHub repository.