Assignment #1

Due Wednesday, August 26 before midnight (California time)

1. Write a Python function that takes as input a FASTA file and returns a sequence string [2pt].

2. Write a Python function that takes as input a sequence string and returns a list with 4 entries that are the number of A, C, G, and T in the sequence [2pt].

3. Write a Python function that takes two inputs: a sequence string and a string of two letters (e.g., “CG” or “CT”). This function returns the number of times the two letters occur consecutively in the sequence [2pt].

4. Explore the NCBI website, go to the following two pages, and download the FASTA files for the human gene *PTPN11* and it’s *Drosophila* orthologue *csw*.

   https://www.ncbi.nlm.nih.gov/nuccore/NM_002834
   https://www.ncbi.nlm.nih.gov/nuccore/NM_057783.3

   For each of the two FASTA files, print the output of function #2 and function #3 with input “CG”. Compare the results and describe your finding [2pt].

5. [Bonus] Write another Python function that takes as input a sequence string and returns a list with 16 entries that are the outputs of function #3 for all 16 possible two letter strings [Bonus 1 pt].

Turn in the code for the three (or four) Python functions and the answer for question #4 into one file in Jupyter Notebook format (.ipynb). Use the “Turnitin” link on Blackboard/Assignments/Assignment 1 to submit this file.