CSCI 5481 Fall 2021

Sched088

# Readme for Homework 2:

The main file requires two input arguments to work, but only utilizes the second (system 3rd) argument as the whole program runs until one .py file. To run the program input:

*[program\_name.py fasta\_filename]*

E.g. ‘cbg\_islands.py HMC\_NT\_011515.fasta’

The main file takes the input arguments and then runs two import functions, one to get the sequence import (from the fasta\_filename) and the other to import the gene sequence (for this homework, Chr21.txt).

Next, the Viterbi **HMM** function is called to identify the high CpG regions of the sequence import. The output from this file is then fed into the **GeneIdentifier** function which compares the position of the start and end of the high CpG regions against the start and end positions, as well as the + or - direction of the provided gene.

Global Variables declared outside of main function

The transition and emission tables are hardcoded as global variables. Due to the sequence offset of the HMC21\_NT\_011515.fasta file, an offset variable ‘seq\_start’ was also declared as a global variable. The value of this variable is 43507093, rather than the 43507094 provided in the assignment to account for the python 0-indexing.