

Homework 7

CIS 160 FS2015

Due: Thursday December 3rd at the *beginning of lecture*

Points available: 50pts

For this assignment you will turn in:

In class(10pts):

1. A statement of the problem (typed)
2. An explanation of your solution (typed)
3. A flowchart (hand-drawn or computer generated)
4. Pseudocode (typed)

Via BlackBoard(40pts):

1. C program named `<username>_gc.c`

Assignment:

Follow the steps that we have outlined in class for algorithm development to generate a program that reads in DNA sequences from a file and determines the content of A, T, C, and G in the sequence. Specifically, I am interested in the GC content (the percentage of the sequence that is G or C). The first line of the file will be an integer that tells you how many sequences there are in the file. Each line following will contain a single sequence. You will need to store the percent of A, T, C, G in a 2D array, this is because you need to know the average GC content of the genome to determine whether a bacterial gene is, or is not, pathogenic. If a bacterial gene has a higher GC content than the genome as a whole, then it is likely that that gene is pathogenic.

The Wikipedia page on GC content gives additional explanation: <https://en.wikipedia.org/wiki/GC-content>

Specifications:

Inputs:

- File called `sequences.txt` (contains a plasmid of *Yersinia pestis*)

Outputs:

- File called `content.txt` containing A, T, C, G, and GC content of each sequence along with a pathogenicity prediction:

EX:

%A	%T	%C	%G	%GC	pathogenic?
10	20	40	30	70	Y
20	50	10	20	30	N

Functions:

1. `void printToFile(int seq, float content[seq][4], float avgGC)`
 - a. prints the results out to a file
 - b. You should open and close your file in this function
2. `float averageGC(int seq, float content[seq][4])`
 - a. calculates the average GC content for the whole genome
3. `char isPathogenic(float avgGC, float seqGC)`
 - a. returns Y if pathogenic, N if not

* This is the minimum functions that you must use. You may use others if you like.

Other:

1. This is individual work. You may NOT work in groups.
2. Please staple all work together.
3. You are expected to error check.
4. For code: No compile = No points, no exceptions!
5. You must use the functions exactly as described
6. You must have your input file called "sequences.txt" and output file called "content.txt"