Homework#4 (40/100 points)
Due: 5/10/2019 11.55pm

Instructions: Submit your solution+code for this assignment to the Blackboard system (under Assignment->Homework#4).

1. Longest common subsequence algorithm can be used to find the longest common subsequence of two sequences of DNA strands. A strand of DNA consists of a string of molecules called bases. The Adenine(A), Guanine (G), Thymine (T) and Cytosine (C) are the four types of bases found in DNA molecule. For example,

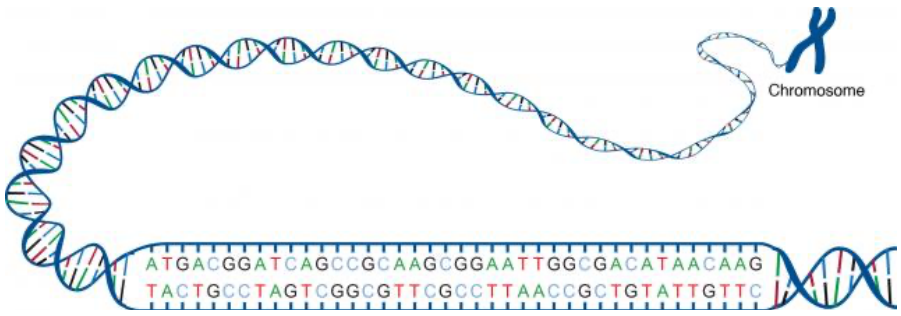

Image source: https://www.genome.gov/

Your task is to write a C++ program to implement an $\theta(mn)$ algorithm (where m is the length of the first sequence and n is the length of the second sequence) to solve the longest common subsequence problem of two DNA strings. Complete the code provided on the Bb. You should implement the same algorithm provided in the code (the CLRS book). Your algorithm should determine the length of the longest common subsequence.

The complete genome sequence of the Coronavirus is provided in coronavirus_genome.txt (length = 29,903). There are 4 more sequences of unknown viruses. Complete the code provided on Blackboard to find out which of these strands might also be Coronavirus. That is, you need to determine which of the sequences give us the longest common subsequences with the sequences provided in coronavirus_genome.txt. Find the two sequences that are similar to the Coronavirus or in another word find the sequences that have the longest common sequence with the coronavirus_genome.txt.

*** Do not post any homework questions or the solutions on any online forum/discussion board. ***

Deliverables – what you need to submit:

1. Complete the table below and submit it on Bb.

Sequence (file)	**Size**	**LCS with coronavirus_genome.txt**	**Similar strands?**
coronavirus_genome.txt	29,903	29,903	Yes (exact same strand)
sequence1.txt	29,748	?	?
sequence2.txt	?	?	?
sequence3.txt	?	?	?
sequence4.txt	?	?	?

2. Submit the complete source code on Bb **in .cpp format only**.

Sample output - Here's a sample output for strings "GFSGTGUGFSG" and "ZGFFGUGFFGUEH" (from exercise#11)

```
Length of the first string is: 11
Length of the second string is: 13
LCS of the first and second string is: 7
```

Grading rubric

Accuracy: implement the algorithm provided and able to solve (for any given 2 strings) correctly

Performance: θ (mn)

*** Do not post any homework questions or the solutions on any online forum/discussion board. ***