Computational Biology and Advanced Topics

Homework – 4

(Smith-Waterman Algorithm - Local Alignment of Sequences)

➢ Upload your project (to Perculus System) in a *.rar* file with the following name:

"<the student number><the name and surname>_hw4.rar"

For example: **"20202020yilmazatay_hw4.rar".**

➢ The program should be in programming languages "C, C++, C# or JAVA".

➢ Don't use the ready string match functions.

➢ Copy homeworks will be punished with -50p.

➢ *You will explain your homeworks on the online platform. In order for the assignment to be accepted, full answers to the questions to be asked are required.*

a) Given sequence *A* of length *n* and sequence *B* of length *m*, how many possible alignments are there? Give an expression in *n* and *m*. Here, *A is the first sequence; B is the second sequence.* Report the sample possibilities for the entered A and B sequences in a pdf file.

b) Implement ***Smith-Waterman Algorithm*** and show your execution time for calculating the optimal local alignment score of two DNA/RNA/Protein sequences —*A* and *B*.

   ✓ The values of the Match, Mismatch, and Gap parameters will be dynamically entered by the user.

   ✓ Your algorithm should take two text files *seqS.txt* and *seqT.txt* as input. The data will be kept in files <u>as in homework 3</u>.

   ✓ The output of your algorithm should be the optimal local alignment score of the given two sequences *A* and *B*.

   ✓ You will present the dynamic table visually in your program. **Other rules and required features <u>in Homework 3</u> also apply to <u>this homework</u>.**

   ✓ *The deadline for the assignment is at 23:00 on May 7, 2020.*