

Computational Biology and Advanced Topics

Homework - 3

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

“<the student number><the name and surname>_hw3.rar”

For example: “20202020yilmazatay_hw3.rar”.

- The program should be in programming languages “C, C++, C# or JAVA”.
- Don’t use the ready string match functions.
- Refer the PDF file “*info_Needleman_Wunsch_Alg.pdf*”.
- 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 .

- b) Implement ***Needleman-Wunsch Algorithm*** and show your execution time for calculating the optimal global alignment score of 2 DNA/RNA/Protein sequences A and B . A is the first sequence; B is the second sequence.

- ✓ Use the following values for Match = 5, Mismatch = -3, and Gap = -5:
- ✓ Your algorithm should take two text files *seqS.txt* and *seqT.txt* as input. Where the two text files should contain the DNA/RNA/Protein sequences A , and B , respectively. Each *.txt* file should consist of two lines. The first line has an integer l denoting the length of the sequence, and the second line lists the DNA (or RNA/Protein) sequence of length l , using symbols from the set {A, C, G, T}.
- ✓ The output of your algorithm should be the optimal global alignment score of the given two sequences A and B .
- ✓ Give the execution time of your algorithm.
- ✓ You will present the dynamic table visually in your program (as mentioned in the lesson).
- ✓ ***The deadline for the assignment is at 23:00 on May 7, 2020.***

Your program should present visually a dynamic table as follows (for DNA nucleotides).
The code should also provide the steps given. The score of each step is shown.

		T	A	T	A	G	C
	O						
G							
T							
T							
A							
T							
C							

Step	Scoring Rubric/Assessment Criteria	Points
1	Setup labeled alignment matrix	10
2	Include initial column and row for GAPS	10
3	All alignment matrix elements filled in	10
4	Evidence for correct use of scoring scheme	10
5	Direction arrows drawn between all cells	10
6	Evidence of multiple arrows to a given cell if appropriate	10
7	Correct optimal score position in matrix used	10
8	Correct optimal score obtained for given scoring scheme	10
9	Traceback path(s) clearly highlighted	10
10	Correct alignment(s) yielding optimal score listed	10

(100 Total points)