# BIO310 Introduction to Bioinformatics
# Homework 2 Spring 2021

March 25, 2021

**Instructions:**

- We expect you to submit the final version till the due date and this will be your homework 2 grade, which is out of 100.
- For the homework submission, submit a PDF document for the answers of the write-up questions, the plots should be appropriately labeled, figures should have captions and should be appropriately cited within the main text. Name your submission as `BIO310-HW2-YourName.pdf` where you substitute in your first and last names into the filename in place of `'YourName'` and submit online through `SUCourse` as a single file. Upload your final report on `SuCourse` by the due date.
- Upload the code online on SuCourse by the due date. The code you submit should be in a format that is ready to run. In submitting the code on SuCourse, compress it as a ZIP file with the name BIO310-HWXcode-YourName.zip where you substitute in your first and last names into the file name in place of 'YourName' and X with the current homework number.
- An ipynb with code and report together is also acceptable.
- If you are considering to submit the homework late, please see the late submission policy in the syllabus
- Please follow the submission instructions, not adhering the submission standards will lead to point deduction.

# 1 Crime Investigation [20 pts.]

You have been called to assist in a crime scene investigation: the body of a tourist was found at the airport. He seems to have suffered from convulsions and internal bleeding. Detectives at the crime scene found a drink carton with some sort of beverage: it still contained some fluid which looks like milk. This may be key evidence. The fluid was sent to the lab and you receive a list of the components of the beverage. Some small molecules such as sugar were found, but also four unidentified proteins were detected. It is your job to analyze these proteins to see if you can help figuring out how the tourist died. A list containing the amino acid sequences of the 4 proteins (called suspect1 through 4) is given in a separate document. You now have enough information to start your investigation. For each of the unidentified proteins, answer these four questions :

1. Which protein is it?
2. From which organism does it originate?
3. What is the function of this protein?
4. Is this protein guilty? Could it be responsible for the death of the tourist? Why (not)?

How did the victim die?

# 2 Global Alignment [40 pts.]

1. Complete the given global sequence alignment code.
2. You should write the output into an output file. The file should include the alignment of the two sequences, the scoring matrix, and the values for mismatch, gap opening penalty, gap extension penalty and match scores used to generate the alignment and the score achieved by the alignment.
3. Test your program with several test cases. We have provided additional global alignment test cases.
   (a) Run your algorithm with the linear gap penalty.
   (b) Run your algorithm with the affine gap penalty. Include gap opening and gap extension penalty.
   (c) Compare your results with linear vs affine gap penalty.

# 3 Local Alignment [40 pts.]

1. Implement the local sequence alignment algorithm with linear gap penalty.You can modify the global alignment algorithm.
2. You should write the output into an output file. The file should include the alignment of the two sequences,the scoring matrix, and the values for mismatch, gap penalty and match scores used to generate the alignment and the score achieved by the alignment.
3. Test your program with several test cases. Especially test edge cases carefully. For example, how would your algorithm run if two very short strings are input, for example 'A' vs 'T' alignment. We have provided additional local alignment test cases separately. Submit the output of these test cases.