CSE 5370: Bioinformatics Homework-2

Divya Boggavarapu(102086719)

October 2022

Collaboration Statement

I have collaborated with Vijitha Kotapati(1001860730) and Tulasi Sridevi Navuluru(1002010740) to do the programming questions in 2,3 and 4 by understanding the Needleman-Wunsch, Smith-Waterman algorithms.

1 Substitution Matrices

Suppose that transition mutations (A ←→G and T ←→C) are less common than tranversions (A ←→ T , A ←→ C, G ←→ T , and G ←→ C). Following is the substitution matrix that reflects this.

	A	*G*	*T*	*C*
A	1	−5	−1	−1
G	−5	1	−1	−1
T	−1	−1	1	−5
C	−1	−1	−5	1

2 Global Alignment

we have conducted global alignment with the Needleman-Wunsch algorithm and implemented it in a single python file called **1002086719_NW.py**

The global alignment function in the file will take in sequence A and sequence B as strings to be aligned (assume that these strings only contain the chars "acgt"), a gap penalty, and a substitution matrix and returns a list of tuples representing possible alignments.

2.1 An example

In **1002086719_NW.py**, we executed the global alignment function with input strings "GATA" and "CTAC", substitution matrix d and gap penalty as parameters which returned the tuples (”GATA", ”-CTAC"),("GATA-","C-TAC").

3 Local Alignment

we have conducted local alignment with the Smith-Waterman algorithm and implemented it in a single python file called **1002086719_SW.py**

The local_alignment function in the file will take in sequence A and sequence B as the strings to be aligned (assume that these strings only contain the chars “acgt”), a gap penalty, and a substitution matrix and returns a list of tuples representing possible alignments.

4 A Custom Alignment

- I have taken my first name **divya**, last name **boggavarapu** in lowercase and concatenated them to be **’divyaboggavarapu’**.
- I have written code to create a custom substitution dict to substitute all the 26 English alphabets as characters and included code for this problem in the file **”1002086719_CUSTOM.py”**

  The file **”1002086719_S.txt”** will provide the output of my pretty matrix.
- I ran “local_alignment” function from Problem 3 with the custom S defined by my name, a gap penalty of -2, my concatenated name (i.e. “divyaboggavarapu”) as the first string, and the pangram “thequickbrownfoxjumpsoverthelazydog” as the second string. Following are the output tuples:

  (’_______ggavarapu’, ’thequicwnfoxjump’), (’_____________arapu’, ’thequickobvroerthe’)
- The file **”1002086719_D.txt”** will provide the output of my pretty matrix.

5 Difficulty Adjustment

Please find my feedback below:

- How long did this assignment take you to complete?
  - **I completed my assignment in 14 hours.**
- If the assignment took you longer than the 10 hours, which parts were overly difficult?
  - **Faced difficulty in understanding the First and Custom alignment questions.**