CSE 5370: Bioinformatics Homework-3

Harshini Kandimalla(1001960046)

1 Substitution Matrices

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

	A	G	T	C
A	1	-1	-2	-2
G	-1	1	-2	-2
T	-2	-2	1	-1
C	-2	-2	-1	1

Global Alignment

we have conducted global alignment with the Needleman-Wunsch algorithm and implemented it in a single python file called **1001960046_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.

1.1 An example

In **1001960046_NW.py**, the global alignment is performed on function with input strings "GATA" and "CTAC", substitution matrix h and gap penalty as parameters which returned the tuples (" GATA","-CTAC"),("GATA-","C-TAC").

2 Local Alignment

we have conducted local alignment with the Smith-Waterman algorithm and implemented it in a single python file called **1001960046_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.

3 A Custom Alignment

- I have taken my first name **harshini**, last name **kandimalla** in lowercase and concatenated them to be **'harshinikandimalla'**.
- 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 **"1001960046_CUSTOM.py"**

  The file" **1001960046_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., "harshinikandimalla") as the first string, and the pangram "thequickbrownfoxjumpsoverthelazydog" as the second string. Following are the output tuples: **[('rshini', 'rthela'), ('la', 'la')]**
- The file **"1001960046_ D.txt"** will provide the output matrix calculated prior to traversal.

4 Difficulty Adjustment

Please find my feedback below:

- How long did this assignment take you to complete?
  - **I completed my assignment in 12 hours.**
- If the assignment took you longer than the 10 hours, which parts were overly difficult?
  - **Faced difficulty in understanding the Custom alignment questions and doing first one.**