

## CSE 5370: BIO-INFORMATICS

### HOMEWORK – 3

#### Substution Matrices

The transition mutations ( $A \leftrightarrow G$  and  $T \leftrightarrow C$ ) are less common than transversions ( $A \leftrightarrow T$ ,  $A \leftrightarrow C$ ,  $G \leftrightarrow T$ , and  $G \leftrightarrow C$ ). To create a substitution matrix that reflects the lower frequency of transition mutations, we can use the following values:

	A	G	T	C
A	3	-1	2	2
G	-1	3	2	2
T	2	2	3	-1
C	2	2	-1	3

Here, the transition mutations have a penalty of -1, while the transversion mutations have a penalty of 2 and match score is 3. This results in a higher penalty for transversions, reflecting their lower frequency.

#### Global Alignment

Implemented the Needleman-Wunsch algorithm and returned the possible alignments, and also took multiple examples that we discussed in class to crosscheck the correctness.

#### Local Alignment

Implemented the Smith-Waterman algorithm and returned the possible alignments, and also took multiple examples that we discussed in class to crosscheck the correctness.

And as discussed in class, it is printing the output as expected and used pretty print.

#### Custom Alignment

Written code to create substitution matrix that will store in 1002059166\_S.txt, and used that substitution matrix, local alignment to print the possible alignments and matrix in 1002059166\_D.txt.

Here is the below possible Alignments for my name sravanchandaka and given string.

[( 'ch', 'ck' )]

[( 'vanch', 'verth' )]

[( 'chan', 'ckbr' )]

[( 'ak', 'ck' )]

## 5) Difficulty Adjustment

In total assignment took more than 23 hours to complete, initially to understand the concept took some time.

The implementation of Custom Alignment and understanding the concept is difficult, as multiple arrows need to be printed in the matrix it is challenging.