Bioinformatics
Programming Assn 3
—

Kushagra Agarwal

2018113012

CODE USAGE

The code for the assignment is present in **pairwise_aligner.py**

Run: **python3 pairwise_aligner.py**

```
kushagra@kushagra:~/Documents/A2k20/Bioinformatics/Assign4/MyAligner$ python3 pairwise_aligner.py

This Pairwise Alignment code can perfome both:
1: Needleman-Wunch Algorithm
2: Smith-Waterman Algorithm
for both Protein and DNA sequences

Default Parameters used:
i) -10 for gap penalty
ii) Scoring Matrices: dnafull for DNA sequences and blossom62 for Proteins

Do you wish to use the default fasta files for alignment?
1) Yes
2) No
Enter your choice: █
```

Press 1 to proceed with default sequences:

```
Enter your choice: 1

Using Default files:
DNA: DNA1.fasta (Accession ID: AF358259.1), DNA2.fasta (Accession ID: L03306.1)
Protein: Protein1.fasta (Accession ID: Q9BYF1), Protein2.fasta (AccessionID: Q5EGZ1)

Processing inputs and calculating alignments

Computing Global Alignment for DNA seqeunces .....
Done: Check output in GlobalAlignmentDNA.txt

Computing Local Alignment for DNA seqeunces .....
Done: Check output in LocalAlignmentDNA.txt

Computing Global Alignment for Protein seqeunces .....
Done: Check output in GlobalAlignmentProtein.txt

Computing Local Alignment for Protein seqeunces .....
Done: Check output in LocalAlignmentProtein.txt

Bye!
```

Defaults used

Gap penalty:

1) **-10 for Proteins**
2) **-10 for DNAs**

Scoring matrices:

1) **Blossom62 for Proteins**
2) **DNAfull for DNAs**

DNA Sequences:

1) **DNA1.fasta:** AF358259.1 Homo sapiens glutaredoxin pseudogene 2
2) **DNA2.fasta:** L03306.1 Mus musculus core-binding factor mRNA

Protein Sequences:

1) **Protein1.fasta:** Q9BYF1 ACE2_HUMAN Angiotensin-converting enzyme 2
2) **Protein2.fasta:** Q5EGZ1 ACE2_RAT Angiotensin-converting enzyme 2

Comparing with

EMBOSS NEEDLE

Parameters used in Options:

For protein

MATRIX	GAP OPEN	GAP EXTEND	END GAP PENALTY	END GAP OPEN	END GAP EXTEND
BLOSUM62	10	10.0	true	10	10.0

For DNA

MATRIX	GAP OPEN	GAP EXTEND	END GAP PENALTY	END GAP OPEN	END GAP EXTEND
DNAfull	10	10.0	true	10	10.0

Protein1.fasta vs Protein2.fasta

Emboss:

```
#=======================================
#
# Aligned_sequences: 2
# 1: ACE2_HUMAN
# 2: ACE2_RAT
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 10.0
#
# Length: 805
# Identity:     664/805 (82.5%)
# Similarity:   725/805 (90.1%)
# Gaps:           0/805 ( 0.0%)
# Score: 3594.0
#
#
#=======================================
```

My output: (GlobalAlignmentProtein.txt)

```
ALIGNED SEQUENCES:

MSSSSWLLLSLVAVTAAQSTIEEQAKTFLDKFNHEAED
MSSSCWLLLSLVAVATAQSLIEEKAESFLNKFNQEAED

Number of Matches: 664
Number of Mismatches: 141
Number of Gaps: 0

FINAL SCORE: 3594
```

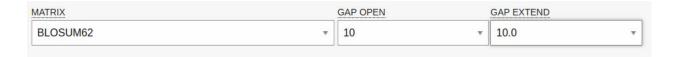## MATCHED

DNA1.fasta vs DNA2.fasta

Emboss:

```
#
# Aligned_sequences: 2
# 1: AF358259.1
# 2: L03306.1
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 10.0
#
# Length: 1608
# Identity:      795/1608 (49.4%)
# Similarity:    795/1608 (49.4%)
# Gaps:          193/1608 (12.0%)
# Score: -435.0
#
```

My output: (GlobalAlignmentDNA.txt)

```
ALIGNED SEQUENCES:

_TTCAG_TAAACATTAC_GA_TTGTGCA_AAA
ATGCCGCGCGTCGTCCCGGACCAGAGGAGCAA

Number of Matches: 795
Number of Mismatches: 620
Number of Gaps: 193

FINAL SCORE: -435
```

MATCHED

Comparing with

EMBOSS WATER

Parameters used in Options:

For protein

MATRIX	GAP OPEN	GAP EXTEND
BLOSUM62	10	10.0

For DNA

MATRIX	GAP OPEN	GAP EXTEND
DNAfull	10	10.0

Protein1.fasta vs Protein2.fasta

Emboss:

```
#
# Aligned_sequences: 2
# 1: ACE2_HUMAN
# 2: ACE2_RAT
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 10.0
#
# Length: 805
# Identity:      664/805 (82.5%)
# Similarity:    725/805 (90.1%)
# Gaps:            0/805 ( 0.0%)
# Score: 3594.0
#
```

My output: (LocalAlignmentProtein.txt)

```
ALIGNED SEQUENCES:

1
MSSSSWLLLSLVAVTAAQSTIEEQAKTFLDKFNHEAED
805
1
MSSSCWLLLSLVAVATAQSLIEEKAESFLNKFNQEAED
805

FINAL SCORE: 3594
```

MATCHED

DNA1.fasta vs DNA2.fasta

Emboss: https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=emboss_water-I20200916-100902-0844-75136608-p2m

```
# Aligned_sequences: 2
# 1: AF358259.1
# 2: L03306.1
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 10.0
#
# Length: 265
# Identity:     149/265 (56.2%)
# Similarity:   149/265 (56.2%)
# Gaps:          31/265 (11.7%)
# Score: 95.0
```

My output: (LocalAlignmentDNA.txt)

```
ALIGNED SEQUENCES:

294
ACAAG_ATTCAAGAT_TATTTGGAACA
553
270
AGAAGCAGGCAAGGTATACTT_GAA_G
508

FINAL SCORE: 95
```

MATCHED