Algorithms in Bioinformatics

BLAST (Basic Local Alignment Search Tool) for COVID 19 Genome Strains

Group Members

Anmol Kumar (2018382)
Prutyay Gautam (2018403)
Sanskar Sachdeva (2018411)
Sarthak Pal (2018412)

Program execution instructions

To run the following program, execute it with the following command:

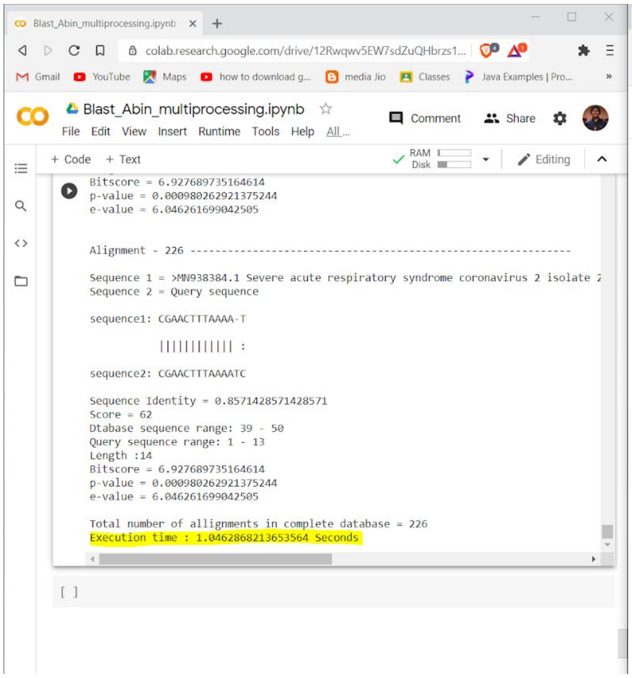
```
python3 file_name.py
```

You are prompted to input the following options :

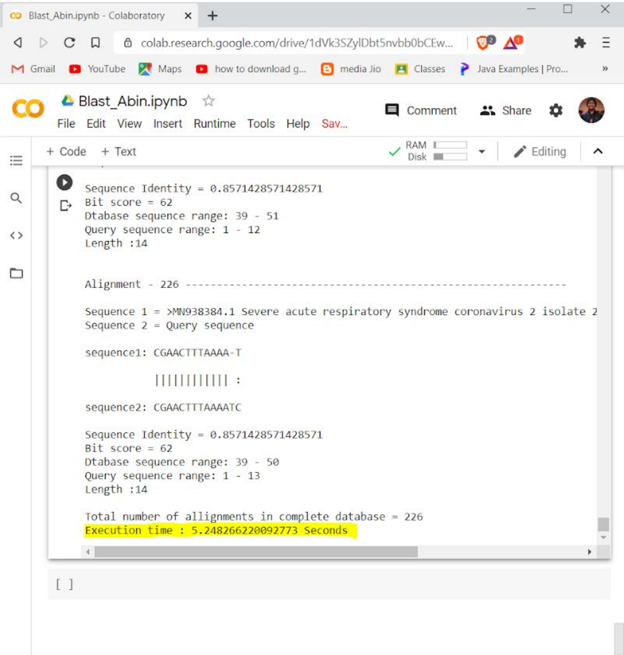
- **Query Sequence:** String input
- **Word Length:** Integer Input
- **HSSP Score Threshold** : Integer Input
- **Extension Threshold**: Integer Input

Optimizations

- Optimization was done on improving the computation time, by applying multiprocessing. 5 times less time was used by the program using multiprocessing.


Blast_Abin_multiprocessing.ipynb
colab.research.google.com/drive/12Rwqwv5EW7sdZuQHbrzs1...
Gmail
YouTube
Maps
how to download g...
media Jio
Classes
Java Examples | Pro...
Blast_Abin_multiprocessing.ipynb
File Edit View Insert Runtime Tools Help All ...
Comment
Share
+ Code + Text
RAM
Disk
Editing

```
Bitscore = 6.927689735164614
p-value = 0.000980262921375244
e-value = 6.046261699042505


Alignment - 226 -------------------------------------------------------------

Sequence 1 = >MN938384.1 Severe acute respiratory syndrome coronavirus 2 isolate 2
Sequence 2 = Query sequence

sequence1: CGAACTTTAAAA-T

           |||||||||||| :

sequence2: CGAACTTTAAAATC

Sequence Identity = 0.8571428571428571
Score = 62
Dtabase sequence range: 39 - 50
Query sequence range: 1 - 13
Length :14
Bitscore = 6.927689735164614
p-value = 0.000980262921375244
e-value = 6.046261699042505

Total number of allignments in complete database = 226
Execution time : 1.0462868213653564 Seconds
```


Blast_Abin.ipynb - Colaboratory
colab.research.google.com/drive/1dVk3SZylDbt5nvbb0bCEw...
Gmail
YouTube
Maps
how to download g...
media Jio
Classes
Java Examples | Pro...
Blast_Abin.ipynb
File Edit View Insert Runtime Tools Help Sav...
Comment
Share
+ Code + Text
RAM
Disk
Editing

```
Sequence Identity = 0.8571428571428571
Bit score = 62
Dtabase sequence range: 39 - 51
Query sequence range: 1 - 12
Length :14


Alignment - 226 -------------------------------------------------------------

Sequence 1 = >MN938384.1 Severe acute respiratory syndrome coronavirus 2 isolate 2
Sequence 2 = Query sequence

sequence1: CGAACTTTAAAA-T

           |||||||||||| :

sequence2: CGAACTTTAAAATC

Sequence Identity = 0.8571428571428571
Bit score = 62
Dtabase sequence range: 39 - 50
Query sequence range: 1 - 13
Length :14

Total number of allignments in complete database = 226
Execution time : 5.248266220092773 Seconds
```

- The Smith-Waterman algorithm was applied on a strict sub range of database and query which was developed by identifying the adjacent words. The algorithm was not extended outside of these ranges as they would have led to less score or highly gapped local alignment. Therefore computation time was saved by not calculating such low scoring local alignments. If in case user wanted to get these alignments they can get it by applying a smaller HSSP Thershold.

References

- Blast: The Algorithm: https://www.youtube.com/watch?v=wfi_KimrNQM
- Heuristics of Blast: https://www.youtube.com/watch?v=jzSIC2UzxZ4
- Basic Local Alignment Search tool: https://www.sciencedirect.com/science/article/pii/S0022283605803602