How to use DNA_Sequence_Detector

2020 August 31st

Shintaro Miyazaki
Satoshi Okada

1) The four files below have to be placed in the same folder.
   SD_internal.ipynb
   SD_external.ipynb
   Sequence file 1 (query sequence, in FASTA format)
   Sequence file 2 (subject sequence(s), in FASTA format, multi-FASTA format is acceptable)

2) The libraries below have to be installed.
   numpy
   tqdm
   matplotlib
   scipy
   numba
   datetime

3) Open the “SD_external.ipynb” file in Jupyter Notebook or Jupyter Lab.

4) Input the sequence file names. In the example below, the name of the query sequence file is “repeatUnit.txt” and the name of the subject sequence file is “testSequence.txt”.

```
repeatUnitFileName='CUP1_repeatUnit.txt'
f = open(repeatUnitFileName, 'r')
repeatUnitList = f.readlines()
f.close()
```

```
sequenceFileName='testSequence.txt'
f = open(sequenceFileName, 'r')
sequenceList = f.readlines()
f.close()
```

5) Run all cells.

6) DNA_Sequence_Detector returns 2 sets of output files.
   SD.txt contains the result of detection of the query sequence. (in TSV format)
   A number of PNG image files contain graphs showing the position of the query sequence in the subject sequence.