Nathaniel Evans

Mike Mooney

BMI665

10/23/18

HW#4

This program can be run from bash command line using the command:

$ python evans\_hw4.py <DNA-path-name>

eg.

$ python evans\_hw4.py "dna.txt"

Where the dna.txt file is stored in: ./data/

Alternatively, running without an argument will result in a default DNA seq test.

The purpose of this script is to:

A. Read in the IUPAC codes from the url:

https://en.wikipedia.org/wiki/Nucleic\_acid\_notation

B. Initialize a DNA object representing a given DNA sequence, read from file.

C. Analyze a user provided DNA sequence and print to console any restriction enzyme sites that are found.

If the provided DNA sequence contains characters other than [ATCG] a SeqTypeError will be thrown.

The result of this script is to print subsequences of our given DNA string to the console, along with their index within the greater sequence and the IUPAC code that resulted in the matched. These subsequences are matches from several IUPAC codes, which are specific to enzyme restriction sites. The restriction sites are as follows:

'EcoRI': 'GAATTC'

'EalI': 'YGGCCR'

'ErhI': 'CCWWGG'

'EcaI': 'GGTNACC'

'FblI': 'GTMKAC'

As an example. Given no argument, the python script will run a default test case using the DNA string input of:

"GAATTCAAAAAACGGCCAAAAAAAAACCTTGGAAAAAAAAGGTAACCAAAAAAAAAGTATAC"

The resulting console output is:

running test case

------------------------------------------------------

DNA pattern match(es) to code: EcoRI

index span: (0, 6)

DNA sequence match: GAATTC

-------------------------------------------------------

------------------------------------------------------

DNA pattern match(es) to code: EalI

index span: (12, 18)

DNA sequence match: CGGCCA

-------------------------------------------------------

------------------------------------------------------

DNA pattern match(es) to code: ErhI

index span: (26, 32)

DNA sequence match: CCTTGG

-------------------------------------------------------

------------------------------------------------------

DNA pattern match(es) to code: EcaI

index span: (40, 47)

DNA sequence match: GGTAACC

-------------------------------------------------------

------------------------------------------------------

DNA pattern match(es) to code: FblI

index span: (56, 62)

DNA sequence match: GTATAC

-------------------------------------------------------