Report on Mapping problem of Restriction Enzymes

-By Ayush Sharma (2019101004)

Problem Statement

Write a program to generate a restriction map for a specific RE & compare your result with Mapper. Give RE & Genomic Sequence used.

Report

RE & Genome

Selected RE = NaeI which has recognition site “GCCGGC”
Selected Genomic Sequence = PBR322

```
import re as regex
G_SEQ = "PBR322"
ENZYME = "NaeI"

# enzyme = Nael, cutting index = 2
#             |
# 5' ...GCC | GGC...3'
# 3' ...CGG | CCG...5'
#           |
restriction_endonuclease = "gccggc"
cutting_index = 3
re_length = len(restriction_endonuclease) # genomic sequence = PBR322

GENOMIC_SEQUENCE = "ttctcatgtttgacagcttatcatcgataagctttaatgcggtagtttatcacagttaaattgctaacgcag
DNA_LENGTH = len(GENOMIC_SEQUENCE)

recognition_sites = regex.finditer(restriction_endonuclease, GENOMIC_SEQUENCE)

start_index = []
for site in recognition_sites:
    start_index.append(site.start())
position_cutpoint = []
for i in range(len(start_index)):
    tup = (start_index[i]+1,start_index[i]+cutting_index+1)
    position_cutpoint.append(tup)
FRAGMENTS = len(start_index)+1

print("Genetic Sequence :",G_SEQ)
print("Enzyme :", ENZYME)
print("DNA Length :",DNA_LENGTH)
print("RE :", restriction_endonuclease)
print("Total Fragments :", FRAGMENTS)
print("Frequence :",FRAGMENTS-1)
print()
print("S.No.   Position/Cutpoints   [Assuming indexing 1 in DNA]")
sno = 0
for tup in position_cutpoint:
    sno+=1
    (s, e) = tup
    print(sno,"      ",s,"/",e)
```

Code Explanation

First of all using `regex` python library, I match all the substring in Genomic sequence which matches RE and extracted it’s starting index. Then using those indices I made a list of tuple (start index, cut point). Assumed DNA indexing to be 1. Then printed all the related info i.e. Total fragments, Frequence and positions of recognition site (starting) and their crresponding cutpoints.

Comparing result with online Mapper

As we can see Fragments formed finally in output of the code is 5 with Frequence = 4 which matches with Frequence column on the online Mapper, which is also having value 4 i.e. total number of recognition site for Enzyme “NaeI”. Also, the fragments index is according to the data found in online mapper when run for the same Genomic sequence. Cutsite is same as cutpoints in the code output.

Name	Sequence	Frequence	Positions/Cutsite
AaaI	CGGCCG	1	939/940
AagI	ATCGAT	1	23/25
AasI	GACNNNNNNGTC	2	2162/2169, 2575/2582
AatII	GACGTC	1	4284/4289
AccI	GTMKAC	2	651/653, 2244/2246
NaeI	GCCGGC	4	401/404, 769/772, 929/932, 1283/1286
NarI	GGCGCC	4	413/415, 434/436, 548/550, 1205/1207
NbII	CGATCG	1	3733/3737
NciI	CCSGG	10	170/172, 534/536, 1258/1260, 1484/1486, 1812/1814, 2118/2120, 2153/2155, 2852/2854, 3548/3550, 3899/3901
NcuI	GAAGA	11	464/477, 3123/3136, 738, 1009, 1601, 2352, 3214, 3969, 4047, 4156, 4352
NdaI	GGCGCC	4	413/415, 434/436, 548/550, 1205/1207

```
ayushsharma@ayush-sharma:~/Documents/College IIIT-H Files/sem4/Sc
Genetic Sequence : PBR322
Enzyme : NaeI
DNA Length : 4361
RE : gccggc
Total Fragments : 5
Frequence : 4

S.No.   Position/Cutpoints   [Assuming indexing 1 in DNA]
1       401 / 404
2       769 / 772
3       929 / 932
4       1283 / 1286
ayushsharma@ayush-sharma:~/Documents/College IIIT-H Files/sem4/Sc
```