

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

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

Code Explanation

First of all using `regex` python library I match all the substring in Genomic sequence and extracted it's starting index. Then using those indices I made a list of fragments i.e. parts / substring of the genomic sequence after action of RE. Then printed the total number of fragments and their position and length.

```
# Return an iterator over all non-overlapping restriction_endonuclease matches in the GENOMIC_SEQUENCE.
# For each match, the iterator returns a Match object.
recognition_sites = regex.finditer(restriction_endonuclease, GENOMIC_SEQUENCE)

# Collecting all the start indices.
start_index = []
for site in recognition_sites:
    start_index.append(site.start())

j=0
fragments = []
for i in range(len(start_index)):
    tup = (j, start_index[i]+cutting_index)
    j = start_index[i]+cutting_index+1
    fragments.append(tup)

FRAGMENTS = len(start_index)+1
```

Comparing result with online Mapper

As we can see Total Fragments in Output of the code is 5 and on the online Mapper Frequency column is having value 4 i.e. total number of restriction site for Enzyme “NaeI”. It is in accordance with our findings and also the fragments index is according to the data found in online mapper when run for the same Genomic sequence.

```
ayushsharma@ayush-sharma:~/Documents/College IIII-II Files/sem4/Science2/Assignments/Assignment1/Ayush/02$ python code.py
Genetic Sequence : PBR322
Enzyme : NaeI
DNA Length : 4361
RE : gccggc
Total Fragments : 5
S.No. LOCATION SIZE
1 (0, 482) 483
2 (483, 778) 368
3 (771, 938) 168
4 (931, 1284) 354
5 (1285, 4360) 3876
ayushsharma@ayush-sharma:~/Documents/College IIII-II Files/sem4/Science2/Assignments/Assignment1/Ayush/02$
```

Name	Sequence	Frequency	Positions/Cutsite
Aaai	CGGCCG	1	939/940
AagI	ATCGAT	1	2325
AasI	GACNNNNNGTC	2	2162/2168, 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
NaeI	GCCGGC	4	413/415, 434/436, 548/550, 1205/1207
NciI	CGATCG	1	3733/3737
NciI	CCGG	10	1701/172, 534/536, 1258/1260, 1484/5496, 1813/1814, 2118/2120, 2133/2135, 2852/2854, 3548/3550, 3899/3901
NcoI	GAGA	11	484/477, 512/5136, 738, 1009, 1601, 2362, 3214, 3969, 4047, 4156, 4352
NdeI	GCCGGC	4	413/415, 434/436, 548/550, 1205/1207