

Universidade Federal do Rio de Janeiro

Centro de Ciências da Saúde

Instituto de Biofísica Carlos Chagas Filho

Disciplina: CFB017 - Programação para Biociências

Professor: Dr. Vitor Lima Coelho

TAC 4 - Programação Básica

1 - Motivos de DNA ou proteínas são sequências curtas de DNA/aminoácidos que se repetem um número considerável de vezes no genoma/proteoma que podem ter ou tem comprovadamente uma função biológica.

- a) Escreva um código em Python que leia uma sequência de DNA ou de aminoácidos digitada pelo usuário e procure esse motivo em várias sequências de DNA ou aminoácidos de um arquivo multi-FASTA;
- b) Escreva uma função que conta quantas vezes o motivo se repete (**dica:** use a função **count** de Python);
- c) Você deve escrever uma função genérica que busca tanto sequências de DNA ou de aminoácidos em um arquivo multi-FASTA;
- d) Você pode usar qualquer arquivo multi-FASTA utilizado até agora na disciplina.
- e) Imprima na tela o identificador das sequências que contém este motivo;



IBCCF

INSTITUTO DE BIOFÍSICA
CARLOS CHAGAS FILHO



UFRJ
faz 100
ANOS

1920 | 2020

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Código:

```
# importação de bibliotecas:
import re
from Bio.Seq import Seq
from Bio import SeqIO

# ITEM B: Função que conta os motivos usando count()
def Counting_Motifs(sequence,motif):
    gene = str(line.id)
    mc = sequence.count(motif)
    if mc != 0:

#ITEM D: Printando os identificadores das sequências que possuam o
motivo
        print("Identificador:",gene,"Motivo:",motif,"Count:",mc,sep=" ")
# ITEM A: Lendo sequência do usuário
motif = str(input("Insira seu motivo de DNA ou AA:")).upper()
# ITEM C: re.search para saber se o input é um DNA ou AA:
# o if sozinho sozinho é o mesmo que perguntar se tal condição é
verdadeira
# se o motif tiver APENAS as letras A,C,G ou T, será considerada uma
sequência de DNA
if re.search('^[ACGT]+$ ', motif):
    # DNA:
    print("É uma sequência de DNA.")
    multifasta =
SeqIO.parse(open("/home/carol/Documents/CFB017/TriTrypDB_Annotated_C
DS.fasta","r"), "fasta")
    for line in multifasta:
        sequence = str(line.seq)
        Counting_Motifs(sequence, motif)
# se o motif tiver outra letra que não A,C,G ou, T, será considerada
uma sequencia de AA
else:
    # Aminoácidos
    print("É uma sequência de AA.")
    multifasta =
SeqIO.parse(open("/home/carol/Documents/CFB017/TriTrypDB_Proteins.fa
sta",'r'), "fasta")
    for line in multifasta:
        sequence = str(line.seq)
        Counting_Motifs(sequence, motif)
```



IBCCF

INSTITUTO DE BIOFÍSICA
CARLOS CHAGAS FILHO



UFRJ
faz 100
ANOS

1920 | 2020

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Outputs:

```
In [1]: runfile('/home/carol/Documents/CFB017/TAC 4/TAC4.py', wdir='/home/carol/Documents/CFB017/TAC 4')
```

Insira seu motivo de DNA ou AA:TTTTCAACAC

É uma sequência de DNA.

Identificador: TcCLB.508807.5 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.511427.19 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.508765.30 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.507083.109 Motivo: TTTTCAACAC Count: 2
Identificador: TcCLB.506775.130 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.510227.10 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.507041.30 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.510149.9 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.504137.150 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.504247.20 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.511543.30 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.507093.80 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.506775.100 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.507037.80 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.509769.30 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.510609.90 Motivo: TTTTCAACAC Count: 1
Identificador: TcCLB.505943.30 Motivo: TTTTCAACAC Count: 1

```
In [2]: runfile('/home/carol/Documents/CFB017/TAC 4/TAC4.py', wdir='/home/carol/Documents/CFB017/TAC 4')
```

Insira seu motivo de DNA ou AA:MASEMTCVKSI

É uma sequência de AA.

Identificador: TcCLB.411427.49:mRNA-p1 Motivo: MASEMTCVKSI Count: 1