

Prueba 4 - Proyecto

DTSE1003 - Programación I

Descripción

Este trabajo consiste en aplicar Python a un problema biológico de análisis de secuencias de aminoácidos del virus COVID-19. Los datos de entrada corresponden a una lista de Python que tiene almacenadas diferentes secuencias aminoacídicas.

Cada secuencia está formada por 14 letras. Cada letra es un código que identifica a un aminoácido diferente:

Aminoácido	Código de una letra
Alanina	A
Arginina	R
Asparagina	N
Ácido aspártico	D
Cisteína	C
Glutamina	Q
Ácido glutámico	E
Glicina	G
Histidina	H
Isoleucina	I
Leucina	L
Lisina	K
Metionina	M
Fenilalanina	F
Prolina	P
Serina	S
Treonina	T
Triptófano	W
Tirosina	Y
Valina	V

Un ejemplo de secuencia podría ser `ARTWRYNSTRSAGH`

Para obtener las secuencias ordenadas en una lista, se debe ejecutar la función `obtenerSecuencias()` la cual devuelve una lista de Python.

Tarea

Construir un programa en Python que analice las secuencias para buscar patrones estadísticos en estas secuencias. Su programa deberá entregar como resultado una tabla con esta estadística, tal como se describirá más adelante. Su programa es muy importante ya que permite determinar que aminoácidos son más frecuentes en una secuencia y orienta el desarrollo de nuevas versiones de vacunas.

Deberá tomar como base el archivo `programa.py` el cual ya trae una función incorporada `obtenerSecuencias()` que le permitirá obtener los datos. Usted deberá elaborar el resto del código necesario.

El programa produce como resultado un análisis estadístico por posición. La primera columna contiene el código de una letra de los 20 aminoácidos posibles y las siguientes columnas, la frecuencia de ocurrencia de cada aminoácido en la posición.

La tabla a imprimir debería presentar los datos en el siguiente orden (los datos son referenciales):

A	2	1	0	0	0		
R	0	0	0	0	0		
N	0	0	0	0	0		
D	0	0	0	0	0		
C	0	0	0	0	0		
E	0	0	0	0	0		
Q	0	0	0	0	0		
G	0	0	4	0	0		
H	2	2	0	0	0		
I	0	0	0	0	0	.	.
L	0	0	0	0	0		
K	0	0	0	0	0		
M	0	0	0	0	0		
F	0	1	0	0	0		
P	0	0	0	0	0		
S	0	0	0	1	2		
T	0	0	0	3	2		
W	0	0	0	0	0		
Y	0	0	0	0	0		
V	0	0	0	0	0		

Otra explicación más en detalle sobre como armar la tabla

A modo de ejemplo, supongamos que se tienen las siguientes secuencias

```
ARTNH
SRTNK
TRTNH
ARTNK
```

Por ejemplo la frecuencia de ocurrencia de "A" en la primera posición sería igual a 2, ya que de las cuatro secuencias 2 tienen A en la primera posición. Además, la frecuencia de ocurrencia de "S" sería de 1 y la de "T" igual a 1, también en la primera posición. En términos de la tabla, la información a mostrar sería:

A	2
R	1
N	0
.	
.	
S	1

Una forma fácil de hacer este ejercicio es emplear diccionarios con listas en su interior para almacenar la cantidad de las frecuencias requeridas.

Fecha de entrega

Martes 12 de diciembre, 11:00 am