

Manual de prácticas

Práctica 1. El dogma central de la biología molecular

Nayeli Luis

Características generales

1. Práctica que se entrega en **parejas**.
2. Entregables:
 - Un archivo PDF con respuestas y descripción del programa. Este archivo debe contener los nombres de los integrantes del equipo y el formato de su nombre es el siguiente:
`p1_cuestionario.pdf`
 - Un script en **Python** (`archivo.py`). Este archivo debe tener como comentario los nombres de los integrantes y el formato del nombre es: `p1_dogma_central.py`.

Ambos archivos se entregan en Classroom, y lo deben entregar los dos integrantes del equipo.

Instrucciones

1. Elige una proteína de la siguiente [lista](#), y busca el archivo en texto plano que le corresponde [aquí](#). Este archivo tiene un formato especial que es conocido como **FASTA**. Los archivos con formato **FASTA** son comunes en la bioinformática pues son archivos de texto plano con secuencias de proteínas o aminoácidos y tienen la siguiente estructura:

```
>NZ_CP010822.1 Thermus aquaticus Y51MC23 chromosome, complete genome
GTGGCCTTGACGACGAGGCGGTCTGGCAGCACGTTCTGGAGCACATCCGCCCAACATACCGAGGTGGAGTACCACAC
CTGGTTTGAAGGATCCGCCCTCGGTATCCGGGAAGGGTTTTGGAGCTGGCGGTGCCACCTCTTCGCCCTGGACT
GGATCAAGCGGCACTACGCCCTCGGTATCCAGGAAGGCTTAGGCCCTCTGGGGGCCAAGGTACCCGCTTTGAGCTTTTG
GTGTGCCCGAGGCCGCCAGCGGCTCAGGTGGACATCTTCAGGCCGCTCCCGAGGCCGACCAAGGGAAGTCCAAGCT
```

La primera línea inicia con **>** e indica un encabezado, que es información sobre la secuencia en cuestión. La segunda línea es la secuencia. Considera ésta información para el desarrollo de tu programa.

2. Contesta las siguientes preguntas sobre tu proteína:
 - ¿De qué organismo se trata? Haz una breve descripción del organismo. Te recomiendo que des click en el link que está en la [lista de proteínas](#) para darte una idea.
 - El organismo, ¿Es un eucarionte o un procarionte?
 - ¿La secuencia que se encuentra en el archivo es un gen o el DNA completo del organismo? ¿Por qué?
 - ¿Esta proteína es única en el organismo? Sino es el caso, ¿Qué otros organismos la presentan?
 - ¿Cuál es la función general de la proteína?

Las respuestas a éstas preguntas deben estar en `p1_cuestionario.pdf`.

3. Crea un programa donde apliques los procesos del dogma central de la biología molecular en la secuencia incluida en el archivo de la proteína que elegiste. Considera que esta secuencia va de 5'-3'.
 - El programa debe aceptar como argumento un archivo en formato **FASTA** y debe arrojar como resultado:
 1. La cadena complementaria.
 2. La secuencia transcrita, es decir, el mRNA.
 3. La secuencia de aminoácidos. Es probable que no todos los codones de la cadena de DNA codifiquen para un aminoácido, para éstos codones que el programa arroje un **-** (guión). [Aquí](#) hay un diccionario con el código genético, te dejo un [artículo](#) por sino recuerdas qué es el código genético y recuerda que siempre puedes utilizar [ésta](#) poderosa herramienta de búsqueda.

Notas importantes

- Para éste ejercicio no tomaremos en cuenta que la secuencia pueda tener regiones promotoras u otro tipo de regiones.
- No pueden utilizar ninguna paquetería de Python que ya tenga implementadas las funciones para hacer esto.