



# UAX

UNIVERSIDAD ALFONSO X EL SABIO

# Práctica 1

GRADO EN BIOMEDICINA

*Bioinformática*

[idelhgar@uax.es](mailto:idelhgar@uax.es)

2023

# Práctica 1

**En esta práctica veremos como procesar ciertos archivos en bioinformática empleando Jupyter Notebook y nuestro entorno Ubuntu.**

**Creamos nuestra carpeta de trabajo llamada Practica1.**

**Ejecutamos en ella nuestro entorno en bash y Jupyter Notebook**

**En la carpeta tenemos que tener nuestros tres archivos:**

- **Ecoli\_protein.faa**
- **Ecoli\_CDS.fna**
- **Ecoli\_genome.fna**

# Búsqueda de patrones en archivo fasta

**Empleamos Ecoli\_genome.fna**

- 1. Buscar secuencias de inicio de transcripción (TATAAA o TATAAT) y de fin de transcripción (>AAAAA). ¿Cuántas hay de cada una?**
- 2. Buscar secuencias de unión de factores de transcripción (FT). Las secuencias son ATXXTC, ACAXTT y TTTCXXA. X se refiere a cualquier nucleótido. En una línea puede haber más de una.**
- 3. Busca subsecuencias que empiecen con TACA y acaben con AAAAAT.**

# Búsqueda de información en un multifasta

Empleamos Ecoli\_CDS.fna

1. ¿Cuántas secuencias hay?
2. Extraer los códigos (ej: CP018245.1) de los headers y meterlos en un archivo llamado `ecoli_headers_cds.txt`. ¡¡¡Quitar los “>” en el output!
3. Buscar la secuencia `AGCGGTGCAAAACCAGGTACA`. ¿En qué gen está?
4. Extraer las descripciones de la proteína (`[protein=...]`) y juntarlas con los identificadores en un archivo

# Búsqueda de información en un multifasta

**Empleamos Ecoli\_protein.fna**

- 1. Extraer el identificador sin el “>” e introducirlo en un archivo llamado Ecoli\_proteinID.txt**
- 1. Determinar si la existe un patrón formado por 1 fenilalanina, 1 lisina, 2 alaninas, 1 arginina y 1 serina.**

# Búsqueda de sitios de restricción

**Empleamos Ecoli\_genome.fna**

**EcoRI: GAATTC**

**SmaI: CCCGGC**

- 1. ¿Hay secuencias de corte de las dos enzimas?**
- 2. ¿Cuántos de cada una?**
- 3. ¿En qué números de línea?**

# Edición de un archivo PDB

**Descargar 3C5X.pdb**  
**Traspassarlo a nuestra VM**

- 1. Cambiar las CYS por CYX**
- 2. Cambiar CYX por CYS sólo en el encabezado**
- 3. Eliminar parte del encabezado**
- 4. Eliminar puentes de hidrógeno y puentes de sulfuro**
- 5. Mirar cuántas cadenas tiene nuestra proteína**

