

Ejercicios de comandos

Ejercicio 1

En este ejercicio trabajaremos con el genoma del procarionta *E. coli*. La idea detrás de estos análisis muy simples es adquirir cierta comodidad con la línea de comando y también empezar a visualizar el tipo de archivos con los que vamos a trabajar a lo largo de la parte práctica del curso.

a. En la carpeta 2024 hay un nuevo subdirectorío, MATERIAL_LINUX2 . ¿Qué archivos ven listados ahí? ¿De qué tamaño son?

b. Copien el archivo GCF_000005845.2_ASM584v2_genomic.fna a su directorio personal y naveguen de nuevo a su directorio personal.

¿Cómo cambiaría el problema si en lugar de estar en la carpeta MATERIAL_LINUX2 , estuvieran en su directorio personal y quisieran copiar el archivo del directorio MATERIAL_LINUX2 sin moverse?

c. ¿Cómo pueden chequear sus permisos sobre la copia nueva?

¿Qué comando pueden usar para cambiarlos? ¿Cuáles son las dos versiones del comando? ¿Qué números usamos nosotros para que todos los usuarios tuvieran acceso únicamente para leer (read) el archivo?

d. ¿Qué hace el comando head ? ¿Cómo describirían la salida cuando lo usan para acceder al genoma de *E. coli*?

Ejercicio 2

A continuación seguiremos investigando el uso de la terminal haciendo algunas observaciones sobre los genomas de *E. coli* y *Bacillus subtilis*.

a. Descargue y descomprima el genoma de *B. subtilis* desde el siguiente enlace, en la misma carpeta que su copia del genoma de *E. coli*.

```
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/009/045/GCF_000009045.1_ASM904v1/GCF_000009045.1_ASM904v1_genomic.fna.gz
```

b. ¿Qué largo tienen ambos genomas?

sugerencias

i. Observar la estructura del archivo y determinar qué información no tienen que contar.

ii. Recordar el uso de `|`.

iii. `tr -d 'x' archivo.txt` elimina todas las instancias de la letra 'x' de `archivo.txt`

iv. `\n` es el carácter que los comandos interpretan cuando encuentran un salto de línea en un archivo

c. ¿Cuántas veces aparece la secuencia Shine-Dalgarno (`AGGAGG`) en cada genoma? ¿Qué función cumplía esta secuencia? ¿Qué comandos usarías para redirigir todas las líneas donde aparece a un archivo nuevo?

d. ¿Cuántas veces aparece `ATG` en cada genoma? ¿Y otros codones de inicio como `GTG` y `TTG`? En un problema real, para este análisis es clave otra información, por ejemplo sobre el ORF (Open Reading Frame) para algo al respecto del uso de codones en *E. coli* y *B. subtilis*, pero prueba mirar cuántas veces aparecen los tripletes en cada genoma.

Ejercicio 3

En este ejercicio investigaremos otro tipo de archivo del formato `.gff`. En futuras clases trabajarán más con estos archivos y entenderán mejor su formato, la información que contienen, cómo se generan y para qué se usan en distintos tipos de análisis genómicos, pero con un par de búsquedas en la terminal se van a poder hacer una idea de su propósito.

a. Descargue y descomprima un archivo `.gff` desde el enlace a continuación a su directorio personal. Notar la correspondencia con el genoma de *E. coli*.

```
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF_000005845.2_ASM584v2/GCF_000005845.2_ASM584v2_genomic.gff.gz
```

b. Busquen todas las instancias de `threonine synthase`. ¿Qué tipo de información aparece? ¿Cuántos "features" aparecen asociados a las búsquedas `rRNA`, `polymerase`, y `pseudogene`?

c. ¿Cómo se estructura el archivo `.gff`? ¿Qué información contiene?

d. En la cartilla del Práctico 1 encontrarán algunas búsquedas para hacer en relación al otro archivo que estaba en la carpeta `MATERIAL_LINUX2`, `P68871.gff`. ¿Qué tipo de

molécula es P68871?

sugerencia

Investigar unas cuantas líneas del archivo. Maximicen el tamaño de la ventana de la terminal para ver un poco mejor.

Ejercicio 4 (Tarea)

Al final de la cartilla del Práctico 1, encontrarán una tarea para seguir aplicando lo aprendido. Allí encontrarán indicaciones para descargar y trabajar con un archivo fasta de *Trypanosoma cruzi*, un organismo eucariota que pertenece al grupo de los protistas flagelados, y el agente de la enfermedad de Chagas.

Además de esta relevancia clínica, presentan muchos mecanismos inusuales en su regulación de la expresión génica: por ejemplo, se ejerce casi totalmente a nivel post-transcripcional. Siga las indicaciones de la cartilla del Práctico 1 y observe las diferencias entre lo que encuentra para este organismo eucariota y los genomas procariotas de los ejercicios anteriores.