



**Maestría en Bioinformática
Principios y Lógica de Programación
Docente: Ing. María Gabriela Echeverría**

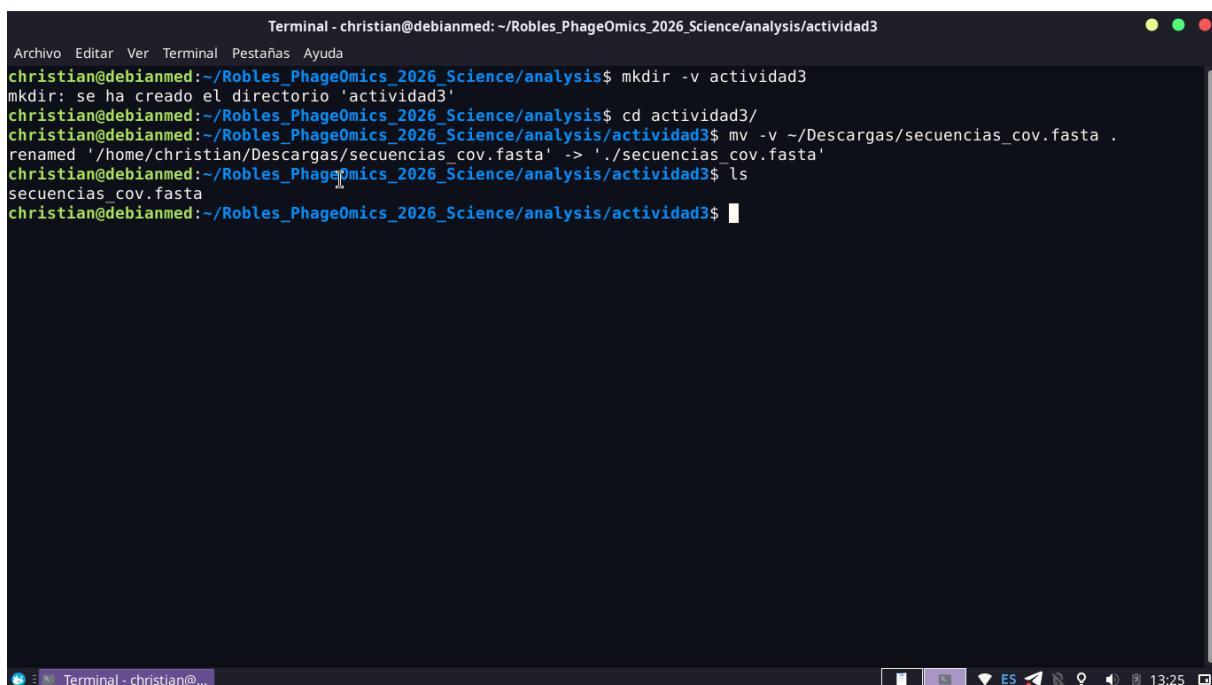
**Actividad 3: Extracción de datos y
ordenamiento de secuencias en arhivos
multifasta**

Estudiante: Christian Robles

1. Introducción

En esta tercera actividad trabajamos con el archivo `secuencias_cov.fasta`, descargado desde el aula virtual, con el propósito de practicar la extracción de información en los encabezados de cada secuencia. A partir de estos datos se generan reportes organizados que incluyen el número de acceso (Accession ID) y la fecha de colecta, ordenados de manera sistemática. La tarea se realizó en **Debian 13 Trixie**, utilizando el entorno de escritorio **Xfce**, **xfce4-terminal** y el gestor de archivos **Thunar**.

2. Creación del directorio de trabajo y movimiento del archivo FASTA



```
Terminal - christian@debianmed: ~/Robles_PhageOomics_2026_Science/analysis/actividad3
Archivo Editar Ver Terminal Pestañas Ayuda
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis$ mkdir -v actividad3
mkdir: se ha creado el directorio 'actividad3'
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis$ cd actividad3/
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$ mv -v ~/Descargas/secuencias_cov.fasta .
renamed '/home/christian/Descargas/secuencias_cov.fasta' -> './secuencias_cov.fasta'
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$ ls
secuencias_cov.fasta
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$
```

Figura 1: Figura 1. Creación de directorios y movimiento de archivos

El primer paso de la Actividad 3 consistió en crear un nuevo directorio dentro de `analysis` llamado `actividad3`, mover el archivo `secuencias_cov.fasta` desde la carpeta de descargas hacia este nuevo directorio y verificar su presencia con `ls`. Todo esto se realizó con los comandos:

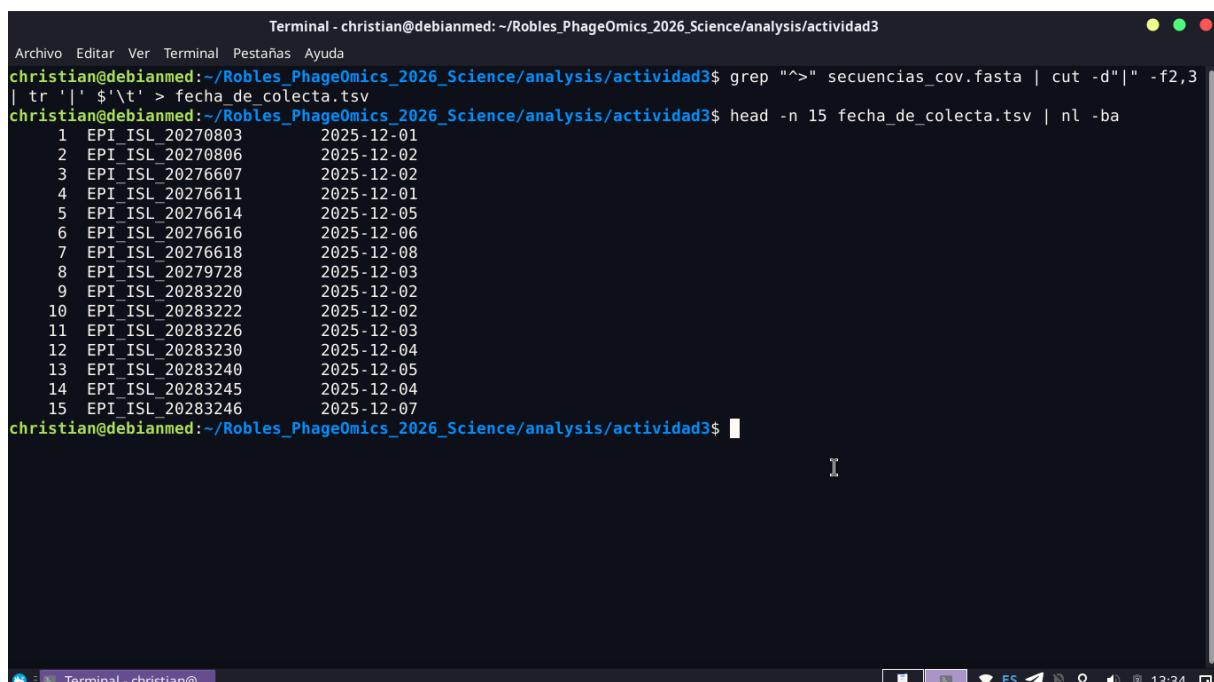
```
mkdir -v actividad3
cd actividad3
mv -v ~/Descargas/secuencias_cov.fasta .
ls
```

Comandos utilizados:

- `mkdir -v actividad3` → crea el directorio `actividad3` y muestra un mensaje de confirmación.

- cd actividad3 → nos mueve dentro del nuevo directorio para trabajar directamente allí.
- mv -v /Descargas/secuencias_cov.fasta . → mueve el archivo desde la carpeta de descargas al directorio actual (.), mostrando un mensaje de lo realizado gracias a la opción -v (verbose).
- ls → lista el contenido del directorio, confirmando que el archivo se encuentra en el directorio actual.
- El archivo secuencias_cov.fasta quedó correctamente ubicado dentro de analysis/actividad3.

3. Extracción de fechas de colecta



```
Terminal - christian@debianmed: ~/Robles_PhageOomics_2026_Science/analysis/actividad3
Archivo Editar Ver Terminal Pestañas Ayuda
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$ grep '^>' secuencias_cov.fasta | cut -d'|' -f2,3 | tr '|'|' $'\t' > fecha_de_colecta.tsv
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$ head -n 15 fecha_de_colecta.tsv | nl -ba
 1 EPI_ISL_20270803      2025-12-01
 2 EPI_ISL_20270806      2025-12-02
 3 EPI_ISL_20276607      2025-12-02
 4 EPI_ISL_20276611      2025-12-01
 5 EPI_ISL_20276614      2025-12-05
 6 EPI_ISL_20276616      2025-12-06
 7 EPI_ISL_20276618      2025-12-08
 8 EPI_ISL_20279728      2025-12-03
 9 EPI_ISL_20283220      2025-12-02
10 EPI_ISL_20283222      2025-12-02
11 EPI_ISL_20283226      2025-12-03
12 EPI_ISL_20283230      2025-12-04
13 EPI_ISL_20283240      2025-12-05
14 EPI_ISL_20283245      2025-12-04
15 EPI_ISL_20283246      2025-12-07
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$
```

Figura 2: Figura 2. Extracción de fechas de colecta mediante los comandos grep, cut y tr

El siguiente paso fue procesar el archivo secuencias_cov.fasta para obtener un archivo tabulado (para mejorar la visibilidad de los datos desde la terminal) con los Accession ID y las fechas de colecta. Esto se logró mediante un encadenamiento de comandos que filtran, cortan y transforman el contenido:

```
grep '^>' secuencias_cov.fasta | cut -d'|' -f2,3 | tr '|'|' $'\t' >
  fecha_de_colecta.tsv

head -n 15 fecha_de_colecta.tsv | nl -ba
```

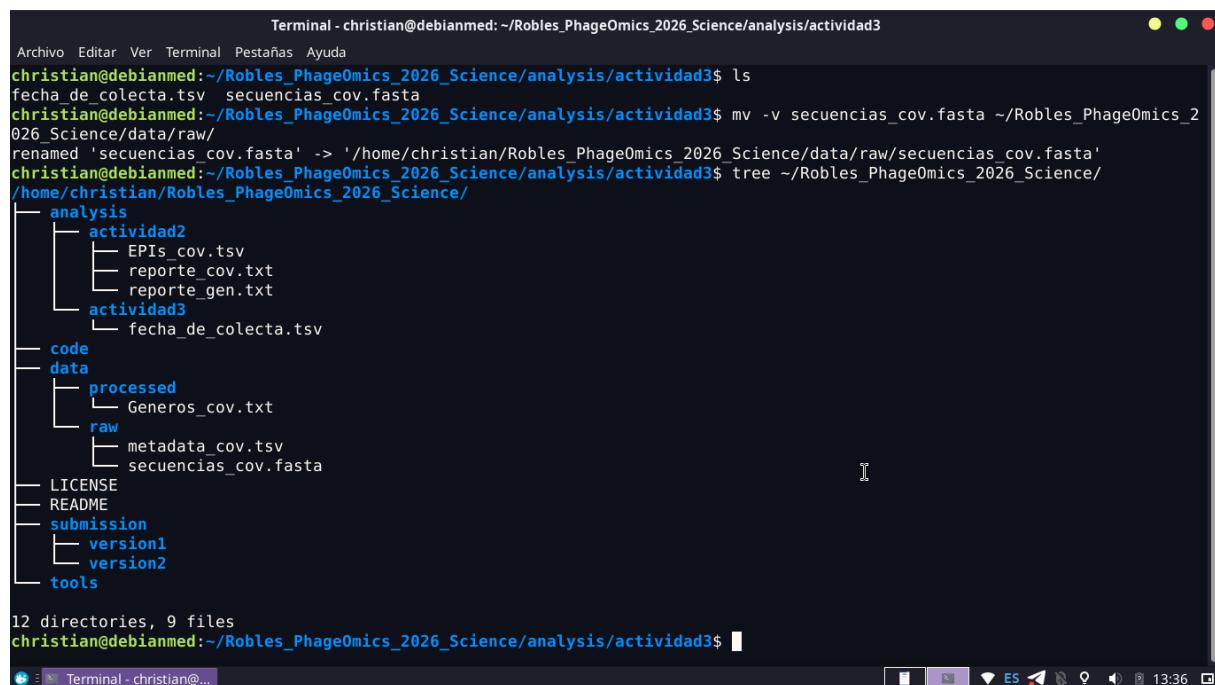
Comandos utilizados:

- grep "^>" → selecciona únicamente las líneas del archivo FASTA que corresponden a encabezados de secuencia (empiezan con el símbolo >).

- `cut -d " | f2 , 3` → divide cada línea usando el carácter | como delimitador y extrae los campos 2 y 3 (ID y fecha de colecta).
- `tr` → reemplaza el delimitador | por tabulaciones, generando un archivo .tsv lo que facilita su visualización desde la terminal.
- `>fecha_de_colecta.tsv` → guarda el resultado en un nuevo archivo.
- `head -n 15 fecha_de_colecta.tsv | nl -ba` → muestra las primeras 15 líneas del archivo y las numera, incluyendo posibles líneas en blanco.

Se decidió encadenar los comandos porque de esta manera, en una sola línea se obtiene el archivo `fecha_de_colecta.tsv` y se revisan sus primeras entradas, lo que asegura su eficiencia y reproducibilidad.

4. Organización del archivo FASTA en carpeta raw



The screenshot shows a terminal window titled 'Terminal - christian@debianmed: ~/Robles_PhageOomics_2026_Science/analysis/actividad3'. The window displays a command-line session and a file tree structure. The command history includes:

```

Archivo Editar Ver Terminal Pestañas Ayuda
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$ ls
fecha_de_colecta.tsv  secuencias_cov.fasta
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$ mv -v secuencias_cov.fasta ~/Robles_PhageOomics_2026_Science/data/raw/
renamed 'secuencias_cov.fasta' -> '/home/christian/Robles_PhageOomics_2026_Science/data/raw/secuencias_cov.fasta'
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$ tree ~/Robles_PhageOomics_2026_Science/
/home/christian/Robles_PhageOomics_2026_Science/
├── analysis
│   ├── actividad2
│   │   ├── EPIS_cov.tsv
│   │   ├── reporte_cov.txt
│   │   └── reporte_gen.txt
│   └── actividad3
│       └── fecha_de_colecta.tsv
├── code
├── data
│   ├── processed
│   │   └── Generos_cov.txt
│   └── raw
│       ├── metadata_cov.tsv
│       └── secuencias_cov.fasta
└── LICENSE
└── README

12 directories, 9 files
christian@debianmed:~/Robles_PhageOomics_2026_Science/analysis/actividad3$ 

```

The terminal window has a dark theme with light-colored text. The title bar shows the application name and path. The command line at the bottom shows the user's prompt and the current directory. The main area displays the command history and the output of the 'tree' command, which shows the hierarchical structure of the project directory.

Figura 3: Figura 3. movimiento de archivos originales a raw y verificación de estructura del proyecto con el comando tree.

Una vez generado el archivo `fecha_de_colecta.tsv`, el siguiente paso fue organizar los datos del proyecto. Para ello se movió el archivo `secuencias_cov.fasta` desde el directorio de análisis hacia la carpeta `data/raw`, que está destinada a almacenar los datos originales sin procesar:

```

mv -v secuencias_cov.fasta
↪ ~/Robles_PhageOomics_2026_Science/data/raw/

```

5. Verificación de la estructura final del proyecto

Finalmente, para comprobar que los archivos quedaron correctamente organizados en sus respectivas carpetas, se utilizó el comando `tree` sobre el directorio principal del proyecto:

```
tree ~/Robles_PhageOomics_2026_Science/
```

El comando `tree` recorre de manera recursiva el directorio indicado y muestra su contenido en forma de árbol. Esto permite visualizar tanto las carpetas como los archivos, y verificar que cada elemento se encuentra en la ubicación correcta.

6. Conclusión

La actividad permitió aplicar comandos de la terminal para extraer y organizar información contenida en los encabezados del archivo `secuencias_cov.fasta`. Mediante los comandos `grep`, `cut` y `tr`. En conjunto, estas herramientas posibilitaron la generación de un reporte con datos clave como el número de acceso y la fecha de colecta, demostrando la eficiencia de la línea de comandos en la manipulación de información biológica.

7. Bibliografía

- Fox, B., & Ramey, C. (2025). *GNU Bash Reference Manual* [Consultado en enero de 2026]. <https://www.gnu.org/software/bash/manual/>
- Newham, C., & Robbins, A. (2005). *Learning the Bash Shell* (3rd). O'Reilly Media.
- Shotts, W. E. (2019). *The Linux Command Line: A Complete Introduction* (2nd). No Starch Press.