

CLASE 9

Introducción a Shell

DBT 792
GENÉTICA Y GENÓMICA EN PRODUCCIÓN ANIMAL

Profesor
Dr. José Gallardo

PLAN DE LA CLASE

Introducción

- Herramientas computacionales para trabajar en proyectos de genómica:
- Shell (Terminal), Linux, Bash, SSH, FTP, Nano
- Conda / Bioconda, SRA toolkit.

Práctica

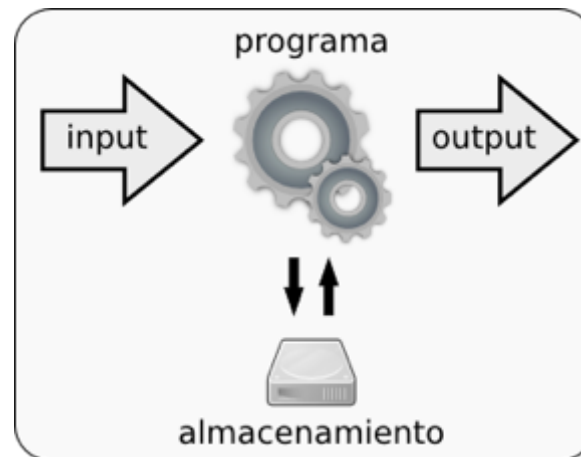
- a) Conectar via SSH a un servidor LINUX.
- b) Familiarizarse con el interprete de comandos Shell.
- c) Instalar conda, bioconda, nano y SRA toolkit.
- d) Descargar Biosamples desde SRA.

Herramientas computacionales para trabajar en un proyecto de genómica

SHELL DE UNIX

Shell de Unix o simplemente Shell: Término usado en informática para referirse a un intérprete de comandos del Sistema Operativo Unix.

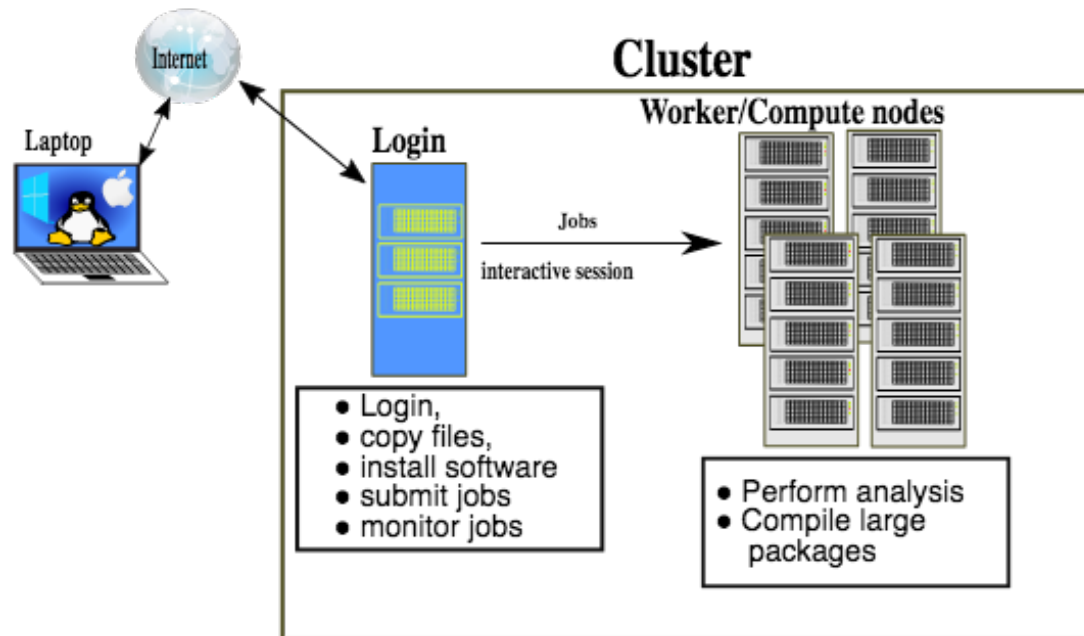
Función: Permite controlar un ordenador y ejecutar un programa (software, paquete o librería) basado en una interfaz de texto.



VENTAJAS DE USAR SHELL

Conectividad y acceso remoto: El uso de Shell facilita la conectividad y el acceso remoto a recursos computacionales desde cualquier parte del mundo.

Multitareas: Es posible ejecutar múltiples tareas como copiar archivos, instalar software, enviar, monitorear y realizar tareas y análisis, almacenar los resultados o datos.

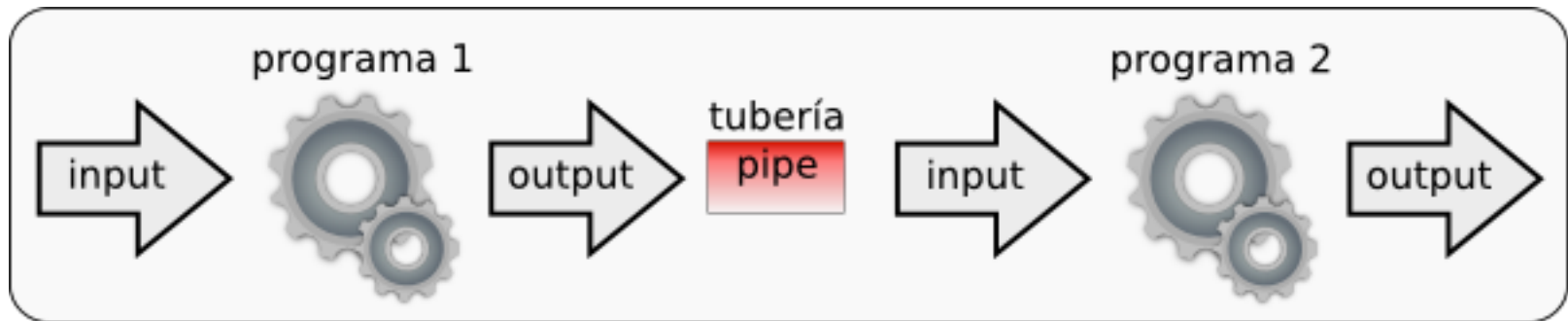


VENTAJAS DE USAR SHELL

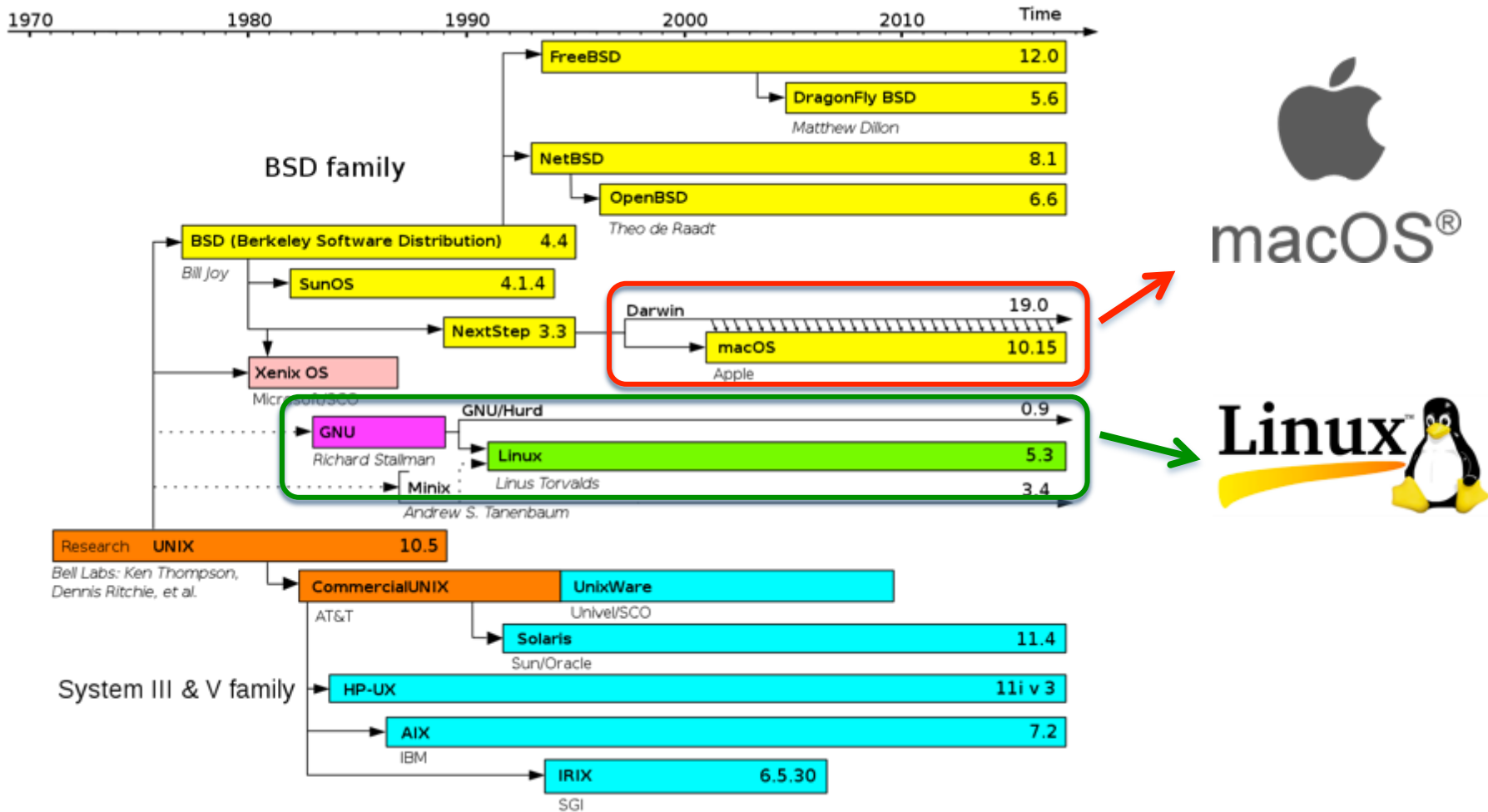
Reproducibilidad. Los procesos realizados mediante la línea de comandos son muy fáciles de documentar puesto que tan sólo debemos guardar el texto que hemos introducido en la pantalla.

Velocidad. Los programas que funcionan en línea de comandos suelen ser extraordinariamente livianos y rápidos, al contrario de software que funcionen con interfaz gráfica.

Transmisión eficiente de flujos de información: Permite transmitir mediante pipe o tuberías flujos de trabajo o procesos, donde el resultado de un proceso es usado por otro programa.



UNIX TIMELINE: LA EVOLUCIÓN A LINUX Y MAC



VERSIONES DE SHELL Y EL TERMINAL

Bourne shell (sh): desarrollado por Stephen Bourne de los Laboratorios Bell de AT&T. en 1977, aun es muy popular. Los archivos ejecutables tiene extensión .sh

Bourne Again shell (bash): Desarrollado en 1989 por Brian Fox para el Proyecto de software libre GNU como un reemplazo de Bourne shell.

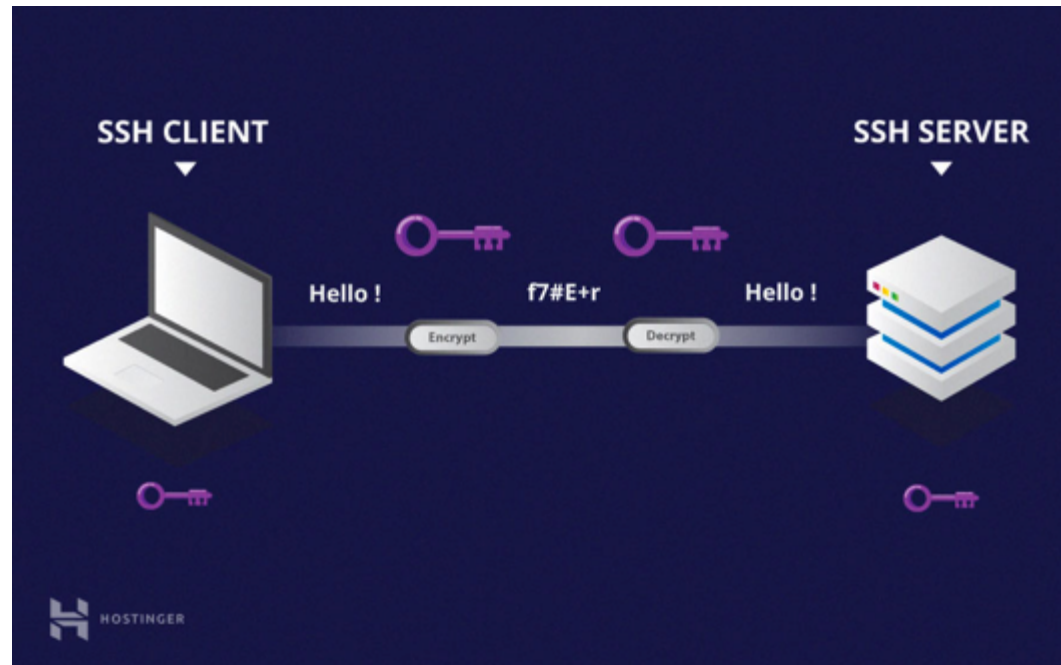
Terminal



Terminal: Para acceder a la shell en MacOS o Linux usamos el terminal. Alternativamente en Windows es posible usar un emulador de la terminal (PuTTY o similar).

PROTOCOLO DE CONECTIVIDAD Y ACCESO REMOTO

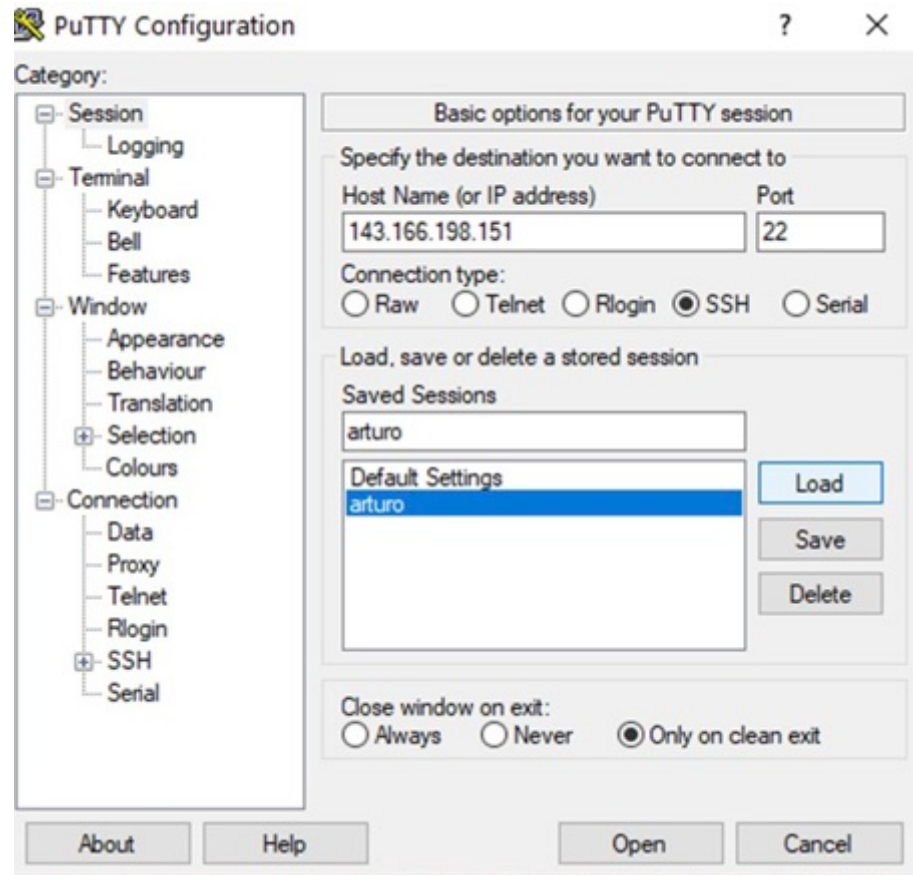
SSH (Secure Shell) ó intérprete de comandos seguro, es un protocolo y programa que sirve para establecer una comunicación entre dos máquinas remotas (cliente/servidor) a través de un canal seguro donde toda la información está cifrada.



SOFTWARE PARA ACCESO REMOTO VIA SSH

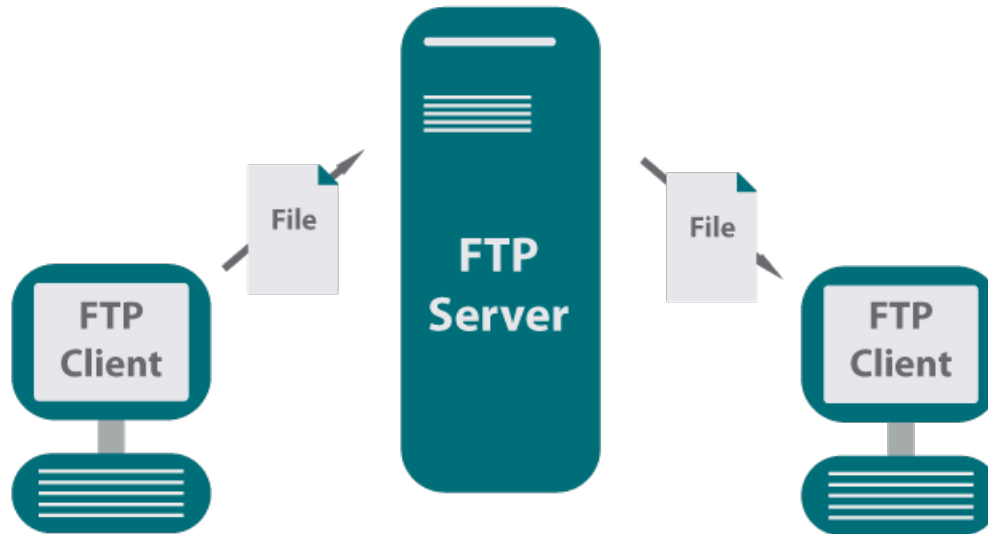
PUTTY

<https://www.putty.org/>



PROTOCOLO DE TRANSFERENCIA DE ARCHIVOS

Protocolo de transferencia de archivos (FTP, por sus siglas en inglés): Es un método simple para transferir archivos de una ubicación en la red a otra.



SOFTWARE PARA TRABAJAR CON FTP

FTP client on windows

<http://filezilla.sourceforge>

Cuidado con los virus



FTP client on windows

<https://winscp.net/eng/docs/lang:es>



FTP client on Mac:

<https://cyberduck.io/>



TFP client on unix:

<http://lftp.yar.ru>



EDITORES DE TEXTO MÁS POPULARES (HPC +)

Un **editor de texto** es un sencillo programa informático que nos permite crear y modificar archivos o scripts (Programas) para el análisis de secuencias NGS.

GNU nano

<https://www.nano-editor.org/>

Es de código abierto y disponible para en macOS, Linux y Windows.



Vim

<https://www.vim.org/>

Vim es un editor de texto altamente configurable,
estable incluido en sistemas UNIX, LINUX y MacOS.



HUELLA DIGITAL DE UN ARCHIVO

MD5 (Message Digest Algorithm 5): Es un algoritmo que se utiliza comprobar la integridad de un archivo luego de su descarga.

¿Qué hace MD5?: En términos simples el algoritmo genera un código de 32 caracteres en formato hexadecimal a partir del contenido de un archivo. Cada carácter puede tomar un valor entre 0 y 9 y entre a y f (16 alternativas para cada carácter). Lo que da 16^{32} códigos diferentes. Cualquier cambio, aunque sea mínimo genera un código totalmente distinto.

¿Cómo se comprueba la integridad?: Es muy sencillo, se compara el código del archivo generado por el creador del archivo con el que obtenemos una vez que el archivo ha sido descargado.

ALTERNATIVAS PARA GENERAR EL CÓDIGO

Windows: MD5summer

<http://www.md5summer.org/>



macOS: md5

```
$ echo "Jose" | md5
```

```
5ebe7595c2e896697367674fc99755e0
```

```
$ echo "Margarita" | md5
```

```
47e6b656af38ea564692aa4523f57e36
```

Linux: md5sum o SHA1sum

CONDA: GESTOR DE PAQUETES Y DEPENDENCIAS

Conda: Gestor de paquetes basado en **Python** que permite administrar la instalación de paquetes y dependencias de varios lenguajes de programación incluyendo Python, R, Ruby, Lua, Scala, Java, JavaScript, C/ C++, FORTRAN, entre otros.

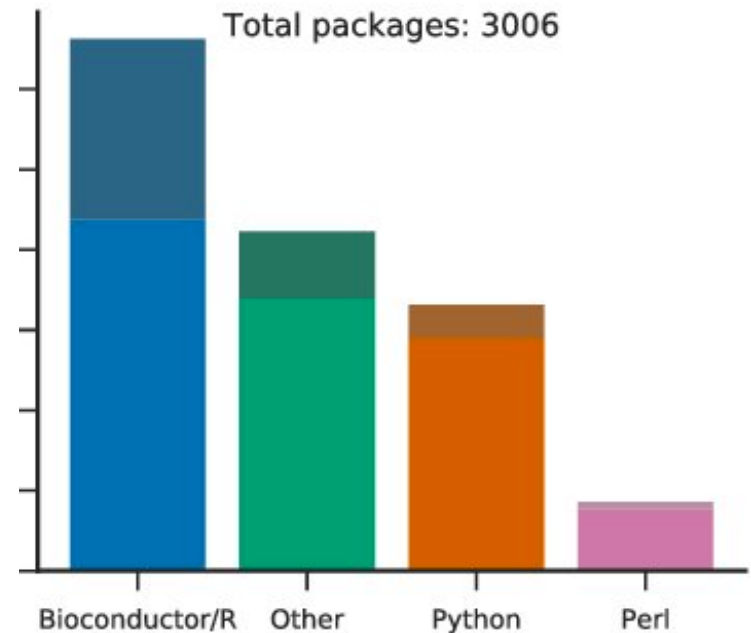
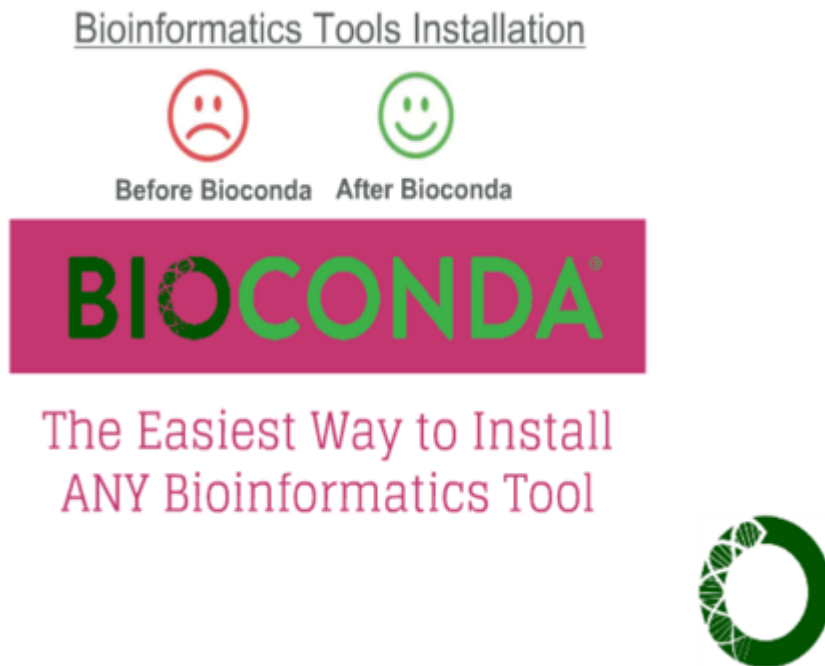
Miniconda: Versión reducida de Conda con lo mínimo necesario para ejecutar las tareas clave (Conda, Python, y otras dependencias menores).

Ventajas:

- 1) Es de código abierto.
- 2) Puede trabajar en macOS, Linux y Windows.
- 3) Permite instalar, correr, y actualizar cientos de software de ciencia de datos y bioinformática de manera sencilla y eficiente mediante **canales**.

BIOCONDA: CANAL DE BIOINFORMÁTICA.

Bioconda: Canal de Conda en el cual están localizados los paquetes especializados en bioinformática. Actualmente, existen más de 7000 paquetes disponibles basados en los lenguajes de programación R, Python, Perl y otros (Fig 1, Grüning et al. 2018).



SRA TOOLKIT

SRA Toolkit: Permite descargar y convertir automáticamente archivos **.sra** en otros formatos usando un interprete de comandos. Es compatible con Linux, Windows y Mac.

Formatos:

fastq, **fasta**, **sff**, **sam** (human-readable bam, aligned or unaligned), otros.



Tools and Software

[Download SRA Toolkit](#)

[SRA Toolkit Documentation](#)

[SRA-BLAST](#)

[SRA Run Browser](#)

[SRA Run Selector](#)

Práctica herramientas computacionales:

Instalar software

SOFTWARE

1.- Instalar software para acceso remoto SSH.

Tu PC es windows: Instala **PuTTY**.

Tienes una MAC o usas Linux: usaremos la terminal.

2.- Instalar software para transferencia de archivos vis FTP.

Tu PC es windows: Instala **WinSCP**.

Tienes una MAC: Instala **Cyberduck**

3.- Instalar editor de textos.

Recomiendo **nano**.

OBJETIVOS DEL TRABAJO PRÁCTICO

Esta práctica de Bioinformática con Shell de Unix tiene como propósito:

- 1) Conectar via SSH a un servidor LINUX.
- 2) Familiarizarse con el interprete de comandos Shell.
- 3) Configurar etapas iniciales de un proyecto de genómica aplicada, incluyendo creación de carpetas y directorios.
- 4) Instalar conda, bioconda, nano y SRA toolkit.
- 5) Descargar Biosamples desde SRA.

CONEXIÓN REMOTA SSH CON UN SERVIDOR USANDO PuTTY

