

Introducción a la Secuenciación de Nueva Generación

Manual de instalación de las herramientas necesarias para el taller

El siguiente manual es una guía básica para la instalación de las herramientas necesarias para crear un pipeline completo de Secuenciación de Nueva Generación (NGS por sus siglas en inglés) durante el taller. Para ello se requiere contar con un ambiente o sistema operativo basado en Unix, ya sea una de las distribuciones para Linux (Ubuntu, Debian, Kali, Arch, etc..) o MacOS. En caso de contar con sistema operativo Windows, se sugiere tener un double boot system con Linux, o instalar WSL.

Atención: SOLO para participantes con sistema operativo Windows

Instalar WSL (Windows Subsystem for Linux):

¿Qué es el Subsistema de Windows para Linux?

El Subsistema de Windows para Linux permite a los desarrolladores ejecutar un entorno de GNU/Linux, incluida la mayoría de herramientas de línea de comandos, utilidades y aplicaciones, directamente en Windows, sin modificar y sin la sobrecarga de una máquina virtual tradicional o una configuración de arranque dual. Se puede hacer lo siguiente:

- Elige tus distribuciones de GNU/Linux favoritas de Microsoft Store.
- Ejecuta herramientas comunes de línea de comandos, como grep, sed, awk.
- Ejecuta scripts de shell de Bash y aplicaciones de línea de comandos de GNU/Linux.
- Instala software adicional mediante el administrador de paquetes de distribución de GNU/Linux.
- Invoca aplicaciones de Windows mediante un shell de línea de comandos de tipo UNIX.
- Invoca aplicaciones de GNU/Linux en Windows.
- Ejecución de aplicaciones gráficas GNU/Linux que están integradas directamente en el escritorio de Windows.
- Uso de la aceleración de GPU para el aprendizaje automático, escenarios de ciencia de datos y mucho más.

Instalación:

Buscar y abrir una ventana de la PowerShell de Windows.



Instalar WSL escribiendo “wsl --install” y presionar Enter.



El comando --install realiza las acciones siguientes:

- Habilita los componentes opcionales de WSL y Plataforma de máquina virtual.
- Descarga e instala el kernel de Linux más reciente.
- Establece WSL 2 como valor predeterminado.
- Descarga e instala la distribución de Ubuntu Linux (es posible que sea necesario reiniciar).

Abrir la terminal de Linux

Una vez completado el proceso de instalación de la distribución de Linux con WSL, abra la distribución (Ubuntu de forma predeterminada) mediante el menú Inicio, o buscando la aplicación en Microsoft Store.

Se le pedirá que cree un nombre de usuario y una contraseña para la distribución de Linux.

Username

- El nombre de usuario y la contraseña son específicos de cada distribución de Linux individual que instala y no tienen relación con su nombre de usuario de Windows.
- Tenga en cuenta que mientras escribe la contraseña, no aparecerá nada en la pantalla. Esto se denomina escritura ciega. No verá lo que está escribiendo, esto es completamente normal.
- Cuando haya creado el nombre de usuario y la contraseña, la cuenta será el usuario predeterminado de la distribución e iniciará sesión automáticamente al inicio.
- Recuerda que esta cuenta se considerará el administrador de Linux y tendrá la capacidad de ejecutar comandos administrativos sudo (es decir, de superusuario).
- Cada distribución de Linux que se ejecuta en WSL tiene sus propias cuentas de usuario y contraseñas de Linux. Tendrás que configurar una cuenta de usuario de Linux cada vez que reinstales, restablezcas o agregues una distribución.

Fuente: <https://learn.microsoft.com/es-es/windows/wsl/setup/environment>.

Youtube: <https://www.youtube.com/watch?v=ElhzKQOR4vw>.

Para TODOS los participantes

Para llevar a cabo el taller de la forma más fluida posible, y evitar al máximo los contratiempos, es recomendable tener instaladas las herramientas que se van a usar. Durante el taller, se explicará el uso de cada una de estas, así como la razón de por qué incluirlas en un pipeline de NGS para determinar su conveniencia a la hora de implementarla en un caso real.

Instalar Trimmomatic

Trimmomatic realiza una variedad de tareas de recorte para Illumina paired-end y datos de un solo extremo. Es una herramienta de preprocesamiento de pares optimizada para los datos de secuenciación de próxima generación (NGS) de Illumina. El software incluye varios pasos de procesamiento para el recorte y filtrado de la lectura. Utiliza una arquitectura basada en tuberías que permite aplicar "pasos" individuales (eliminación de adaptadores, filtrado de calidad, etc.) a cada par de lectura/lectura, en el orden especificado por el usuario.

```
sudo apt install trimmomatic
```

Instalar Samtools

SAMtools es un conjunto de utilidades para interactuar y posprocesar alineaciones cortas de lectura de secuencias de ADN en los formatos SAM (Sequence Alignment/Map), BAM (Binary Alignment/Map) y CRAM, escritas por Heng Li. Estos archivos son generados como salida por alineadores de lectura corta como BWA. Se proporcionan herramientas simples y avanzadas que soportan tareas complejas como la llamada de variantes y la visualización de alineaciones, así como la clasificación, la indexación, la extracción de datos y la conversión de formatos.

```
sudo apt install samtools
```

Instalar BWA

BWA es un paquete de software para mapear secuencias de ADN contra un gran genoma de referencia. Consta de tres algoritmos: BWA-backtrack, BWA-SW y BWA-MEM. El primer algoritmo está diseñado para lecturas de secuencias de Illumina de hasta 100 pb, mientras que los dos restantes para secuencias más largas oscilaron entre 70 pb y unas pocas megabases. BWA-MEM y BWA-SW comparten características similares, como la compatibilidad con lecturas largas y alineación quimérica, pero generalmente se recomienda BWA-MEM, que es la última, ya que es más rápida y precisa. BWA-MEM también tiene un mejor rendimiento que BWA-backtrack para lecturas de Illumina de 70-100 pb.

Para todos los algoritmos, BWA primero necesita construir el índice FM para el genoma de referencia (el comando de índice). Los algoritmos de alineación se invocan con diferentes subcomandos: aln/samse/sampe para BWA-backtrack, bwasm para BWA-SW y mem para el algoritmo BWA-MEM.

```
sudo apt -y install bwa
sudo apt update && sudo apt upgrade
```

Instalar Java

Java es una plataforma informática de lenguaje de programación creada por Sun Microsystems en 1995. Ha evolucionado desde sus humildes comienzos hasta impulsar una gran parte del mundo digital actual, ya que es una plataforma fiable en la que se crean muchos servicios y aplicaciones.

Los nuevos e innovadores productos y servicios digitales diseñados para el futuro también siguen basándose en Java. Aunque la mayoría de aplicaciones Java modernas combinan el tiempo de ejecución y la aplicación de Java, todavía existen algunas aplicaciones e incluso sitios web que no funcionan sin instalar Java para escritorio. El sitio web Java.com está pensado para consumidores que todavía necesitan Java en sus aplicaciones de escritorio, sobre todo las aplicaciones que tienen como destino Java 8.

Algunas de las herramientas que se usarán en el taller fueron creadas en este lenguaje de programación, por lo que será necesario instalar el Java Runtime Environment -JRE- y el Java Development Kit -JDK- para que estas puedan correr sin problema alguno.

```
sudo apt install default-jre
sudo apt update && sudo apt upgrade
sudo apt install default-jdk
```

Actualizar, y verificar las versiones de Java instaladas

```
sudo apt update && sudo apt upgrade
java --version
javac --version
```

Verificar que la variable JAVA_HOME se encuentre entre las variables de ambiente

```
echo $JAVA_HOME
```

En caso de no obtener una salida similar a la anterior cuando se llama la variable JAVA_HOME, esta puede agregarse a las variables de ambiente

así:

```
export JAVA_HOME="/usr/lib/jvm/java-11-openjdk-amd64"
```

Instalar Picard tools

Picard Tools es un conjunto de herramientas creado en Java que corre en bajo línea de comandos para manipular archivos con extensión SAM, BAM, y VCF. La lista completa con una breve descripción se puede obtener ejecutando en la terminal el comando: `PicardCommandLine -h`. Se puede observar la descripción de las opciones y parámetros de todos los comandos ejecutando: `PicardCommandLine COMMAND -h`.

```
sudo apt install -y picard-tools
```

Instalar GATK

GATK (Genome Analysis Toolkit). Es una colección de herramientas que se ejecutan bajo la línea de comandos usada para el análisis de secuencias de datos genómicos enfocado en el descubrimiento de variantes. Se puede usar para crear un pipeline completo o en combinación con otras herramientas.

Descargar GATK:

```
wget "https://github.com/broadinstitute/gatk/releases/download/4.3.0.0/gatk-4.3.0.0.zip"
```

Descomprimir GATK:

```
unzip gatk-4.3.0.0.zip
```

Exportar su ubicación y crear un alias para gatk:

```
export PATH="$PWD/gatk-4.3.0.0/:$PATH"
alias gatk="$PWD/gatk-4.3.0.0/gatk"
```

Verificar que GATK esté funcionando:

```
gatk --help
```