

# Linux for Bioinformatics – Organized Guide

---

## ◆ 1. What is Linux?

- Linux is a **free, open-source operating system**, first released in 1991 under the **GNU GPL license**.
- It powers over **90% of the world's top 500 supercomputers**.
- Like macOS, Linux is based on **UNIX**.
- Popular Linux distributions include:
  - **Ubuntu**
  - Debian
  - Fedora
  - CentOS
  - openSUSE
  - Arch Linux
  - Linux Mint
  - Android (also uses the Linux kernel)

---

## ◆ 2. Why Bioinformaticians Use Linux

- Most bioinformatics tools (e.g., **BLAST**, **CLUSTALW**, **PHRAP**) are **Linux-compatible**.
- Excellent for:
  - Remote computing (via **SSH**)

- Running servers and clusters
  - Multitasking and automation
  - Supporting languages like Python, Perl, R
- 

### ◆ 3. Accessing Linux Remotely

#### Mac

- Use **Terminal** (**Applications > Utilities**)

```
ssh user@hostname
```

#### Windows

- Use **PuTTY** or **MobaXterm** for SSH connections
- Alternative: **Cygwin** for Linux-like environment

#### Enabling GUI Programs Remotely

- Add **-X** to SSH command:

```
ssh -X user@hostname
```

- Install **XQuartz** (Mac), **Xming** or **MobaXterm** (Windows)
- 

### ◆ 4. Command Line Basics

- Use **bash shell** to interact with Linux via commands
- Examples:

```
cd ~      # Home directory  
ls -l     # List files with details
```

```
cp a.txt b.txt # Copy file
rm file.txt # Delete file
```

---

## ◆ 5. Common Linux Commands

Command	Purpose
<code>pwd</code>	Show current directory
<code>cd</code>	Change directory
<code>ls</code>	List files
<code>mkdir</code>	Create directory
<code>rm</code>	Delete file
<code>cp</code>	Copy file
<code>mv</code>	Move or rename file
<code>chmod</code>	Change file permissions
<code>echo</code>	Print text or variables
<code>cat</code> , <code>less</code> , <code>head</code> , <code>tail</code>	Read files

---

## ◆ 6. Output Redirection & Wildcards

- Save output:

```
ls > list.txt # Overwrite
ls >> list.txt # Append
```

- Wildcards:

```
ls *.txt # All .txt files
ls b??? # Files starting with 'b' and 4 letters total
```

---

## ◆ 7. Navigating & System Info

Command	Description
<code>whoami</code>	Shows your username
<code>hostname</code>	Server/machine name
<code>pwd</code>	Current working directory

---

## ◆ 8. File Permissions

- Users: **User (u)**, **Group (g)**, **Others (o)**
- Change permissions:

```
chmod 755 file.sh  
# 7 = rwx (user), 5 = r-x (group/others)
```

---

## ◆ 9. Running Scripts

```
echo "echo Hello \${USER}" > hello.sh  
chmod +x hello.sh  
./hello.sh
```

---

## ◆ 10. File Transfer

- Using **scp**:

```
scp file.txt user@host:path/    # To server  
scp user@host:path/file.txt ./  # From server
```

- Using **wget**:

wget http://example.com/file.txt

---

## ◆ 11. Job Control

Command	Function
Ctrl+C	Stop job
Ctrl+Z	Suspend job
fg	Resume job in foreground
bg	Resume in background
jobs	List jobs
&	Run in background
top	Show running processes

---

## ◆ 12. Useful File Commands

Command	Description
touch	Create empty file
grep	Search in text
cut	Extract fields or columns
diff	Compare two files
sort	Sort contents of a file
wc	Word/line/byte count
history	Show command history
!1234	Repeat command #1234

---

## ◆ 13. Environment Variables

```
env          # View all environment variables
echo $PATH   # Show executable path locations
```

---

## ◆ 14. Linux Exercise (Example Workflow)

1. Download file `Mesculenta_147_gene.gff3.gz`
2. Log into HPC (High-Performance Cluster)
3. Create a directory in scratch:

```
mkdir ~/scratch/LinuxExercise
```

4. Transfer file to this directory using `scp`
5. Unzip and rename:

```
gunzip Mesculenta_147_gene.gff3.gz
mv Mesculenta_147_gene.gff3 Cassavagenes.txt
```

6. Count number of genes:

```
grep "gene" Cassavagenes.txt | wc -l
```

7. Save gene rows:

```
grep "gene" Cassavagenes.txt > Genes.txt
```

8. Count genes in `Genes.txt`:

```
wc -l Genes.txt
```

---

## ◆ 15. Recommended Linux Distro for Bioinformatics

Distro	Notes
Ubuntu LTS	Stable, supported 5 years, large community
Debian	Very stable, slower updates, free-software focused
CentOS	Long-term support, used in scientific computing
Fedora	Cutting-edge features, frequent updates
openSUSE Leap	Stable, easy GUI management tools, good software support

---

## ◆ 16. Installing Bioinformatics Tools

### a) Bowtie2 Installation

- Using **conda**:

```
conda install -c bioconda bowtie2
```

- Using APT/YUM:

```
sudo apt install bowtie2    # Ubuntu
sudo yum install bowtie2    # CentOS
```

- Manual:

```
cd ~/tools/bowtie2/
wget
https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.4.2/bowtie2-2.4.2-sra-linux-x86_64
.zip
unzip bowtie2*.zip
export PATH=$PWD/bowtie2-2.4.2-sra-linux-x86_64:$PATH
```

- Test:

```
bowtie2 --help
```

---

## b) SAMtools Installation

```
wget https://github.com/samtools/samtools/releases/download/1.11/samtools-1.11.tar.bz2
tar -jvxf samtools-1.11.tar.bz2
cd samtools-1.11
make
make prefix=$HOME/tools/samtools install
export PATH=$HOME/tools/samtools/bin:$PATH
```

### Common SAMtools Commands

Task	Command
Convert SAM → BAM	<code>samtools view -b -S file.sam &gt; file.bam</code>
Convert BAM → SAM	<code>samtools view -h file.bam &gt; file.sam</code>
Sort BAM	<code>samtools sort file.bam -o sorted.bam</code>
Count reads	<code>samtools view -c file.bam</code>
Flag stats	<code>samtools flagstat file.bam</code>
Get coverage	<code>samtools mpileup -r 'contig:start-end' file.bam</code>

---

## ◆ 17. Key Resources

- [Linux for Bioinformatics \(PDF\)](#)
  - [Introduction to Linux \(NYU GenCore\)](#)
  - [Bowtie2 on SourceForge](#)
  - [Metagenomics Wiki - SAMtools & more](#)
-