

LINUX WORKSHOP



01 Fundamentals

By Jiajia Li (ANU Biological Data Science Institute)

28/02/2025



**Australian
National
University**



Learning Objectives

- Understand what is Linux and how can we use it in biology research
- Understand the concept of command line interface and terminal
- Understand the syntax of Linux commands
- Understand the Linux file system
- Use commands to go to different locations in Linux
- Use commands to create, delete, copy, move, and rename files and directories
- Use commands to view and print file contents
- Use Linux shortcuts and wild cards





Linux operating system

Linux is a **free** and **open-source** operating system that is widely used in various computing devices including personal computers, servers, mobile devices, embedded systems...

It was created by **Linus Torvalds** in 1991 and is based on the **Unix** operating system.

Popular Linux distributions include Ubuntu, Debian, Fedora, Redhat, and CentOS.



You might have heard “Unix” before, and it is used interchangeably with Linux.

Unix is a family of **multitasking, multiuser** computer operating system originally developed in the 1960s and 1970s at AT&T’s Bell Labs. It was designed to be a versatile and powerful operating system for **mainframe** and later, personal computers.

Unix-like operating systems have since developed and used in various forms, with **Linux** being one of the most well-known derivatives. **MacOS** is also one of Unix-based operating systems.



MacOS



Linux in biology research

Linux is extensively used in biology research especially for larger data processing tasks with the need of using a **server** or **supercomputer**.

There are many biological tools that are specifically designed for Linux or have best performance in Linux, including:

- BWA, Bowtie, SAMtools, GATK for **Sequence Alignment & Variant Calling**
- STAR, Kallisto, Salmon for **RNA-Seq & Expression Analysis**
- SPAdes, Canu, Minimap2 for **Structural Variant & Genome Assembly**
- IQ-Tree, RAXML, MUSCLE for **Phylogenetics & Comparative Genomics**

And more...





Command Line Interface (CLI)

A command line interface is a **text-based interface** used to interact with a computer or software by typing commands into a terminal or command prompt.

In a CLI, users communicate with the computer through **text commands** rather than using graphical user interfaces (GUIs).

```
Welcome to FreeBSD

DataMux v1.9.1 alpha 1 (FreeBSD)
[installed at PS/2 port]
E:\>ver

FreeCom version 0.82 pl 3 XMS_Swap (Dec 10 2003 06:49:21)

E:\>dir

Volume in drive C is FREE003.C95
Volume Serial Number is 8E4F-19E3
Directory of C:\

FREE003.DAT          <DIR>    00-26-04  6:23p
FREE003.DAT          435    00-26-04  6:24p
FREE003.DAT          512    00-26-04  6:23p
FREE003.DAT          93,963  00-26-04  6:24p
FREE003.DAT          881    00-26-04  6:24p
FREE003.DAT          512    00-26-04  6:24p
FREE003.DAT          45,815  04-17-04  9:19p
FREE003.DAT          6 file(s)    142,830 bytes
FREE003.DAT          1 dir(s)    1,064,517,632 bytes free
E:\>
```

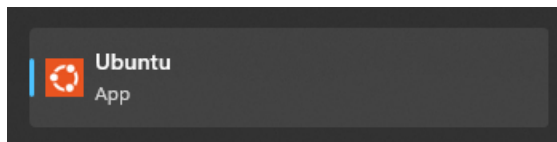
```
root@localhost:~# ping -c 10 wikipedia.org
PING test.peta.wikipedia.org (208.80.152.2) 56(84) bytes of data:
64 bytes from test.peta.wikipedia.org: icmp_seq=1: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=2: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=3: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=4: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=5: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=6: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=7: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=8: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=9: ttl=64: time=0.00 ms
64 bytes from test.peta.wikipedia.org: icmp_seq=10: ttl=64: time=0.00 ms
--- test.peta.wikipedia.org ping statistics ---
10 packets transmitted, 10 received, 0% packet loss, time 0ms
rtt min/avg/max/mdev = 0.00/0.00/0.00/0.00 ms
root@localhost:~# pwd
/root
root@localhost:~# cd /var
root@localhost:~# ls -la
total 72
drwxr-xr-x 18 root root 4096 Sep 14 20:42 .
drwxr-xr-x 23 root root 4096 Sep 14 20:42 ..
drwxr-xr-x  2 root root 4096 May 14 00:15 account
drwxr-xr-x 11 root root 4096 Jul  5 22:25 cache
drwxr-xr-x  3 root root 4096 May 18 16:03 db
drwxr-xr-x  3 root root 4096 May 18 16:03 empty
drwxr-xr-x  2 root root 4096 May 18 16:03 gnu
drwxr-xr-x  2 root root 4096 May 18 16:03 local
drwxr-xr-x  2 root root 4096 Jun  2 18:39 mail
drwxr-xr-x 30 root root 4096 May 18 16:03 lib
drwxr-xr-x  2 root root 4096 May 18 16:03 local
drwxr-xr-x  1 root root 11 May 14 00:12 lock -> ../run/lock
drwxr-xr-x 14 root root 4096 Sep 14 20:42 log
drwxr-xr-x  1 root root 10 Jul  5 22:24 mail -> spool/mail
drwxr-xr-x  2 root root 4096 May 18 16:03 nis
drwxr-xr-x  2 root root 4096 May 18 16:03 opt
drwxr-xr-x  2 root root 4096 May 18 16:03 preserve
drwxr-xr-x  2 root root 4096 Jul  1 22:11 report
drwxr-xr-x  1 root root  6 May 14 00:12 run -> ../run
drwxr-xr-x 14 root root 4096 May 18 16:03 spool
drwxr-xr-x  4 root root 4096 Sep 12 23:59 tmp
drwxr-xr-x  2 root root 4096 May 18 16:03 yp
root@localhost:~# run search.wiki
Loaded plugins: langpacks, presto, refresh-packagekit, remove-with-leaves
refresh-free-updates | 2.7 kB  00:00
refresh-free-updates-primary_db | 2.7 kB  00:00
refresh-nonfree-updates | 2.7 kB  00:00
updates/metalink | 5.9 kB  00:00
updates | 4.7 kB  00:00
updates/primary_db | 73% [=====] 1 62 kB/s 2.6 MB 00:15 ETA
```



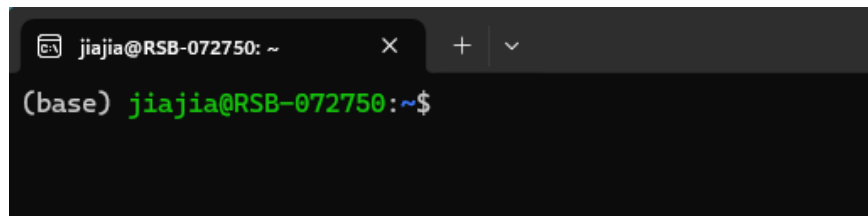


Practice – Open Terminal

Now, please open the terminal on your computer. Windows users to search “Ubuntu” and Mac users to open “Terminal”.



I have an Ubuntu WSL, so my terminal looks like this:





Practice – Open Terminal

Let's learn the different information showed in your CLI:

Name of conda environment

```
jiajia@RSB-072750: ~  
(base) jiajia@RSB-072750:~$
```

Username

Device name

Your location

After the “\$” sign is where you type your command.

To run a command, you just type the command and press enter. For example, run command “pwd”.

```
jiajia@RSB-072750: ~  
(base) jiajia@RSB-072750:~$ pwd  
/home/jiajia  
(base) jiajia@RSB-072750:~$
```





Syntax of Linux commands

In Linux, the commands can be written in a few different ways.

1. The command itself: `(base) jiajia@RSB-072750:~$ ls`

2. Command + Option: `(base) jiajia@RSB-072750:~$ ls -l`

3. Command + Option + Argument: `(base) jiajia@RSB-072750:~$ ls -l /home`

Commands, options, and arguments are separated by a whitespace.

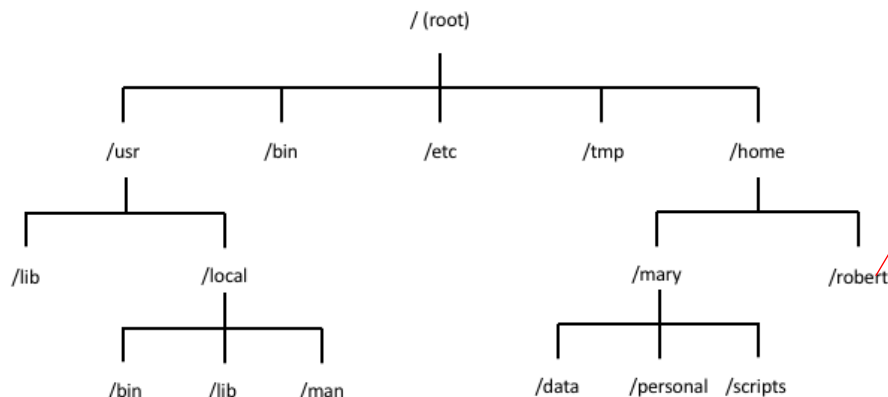
Run these commands and see the result.





Linux file systems

In Linux, there is no Desktop or Start Menu for you to move around. The only places for you to move around is the directories. In Linux, the file system starts from the **root directory “/”** and branches out into subdirectories and files.



When a user logs in, they will be in this location, which also indicates as “~”.

Run ``ls /`` to see what's under your root directory.



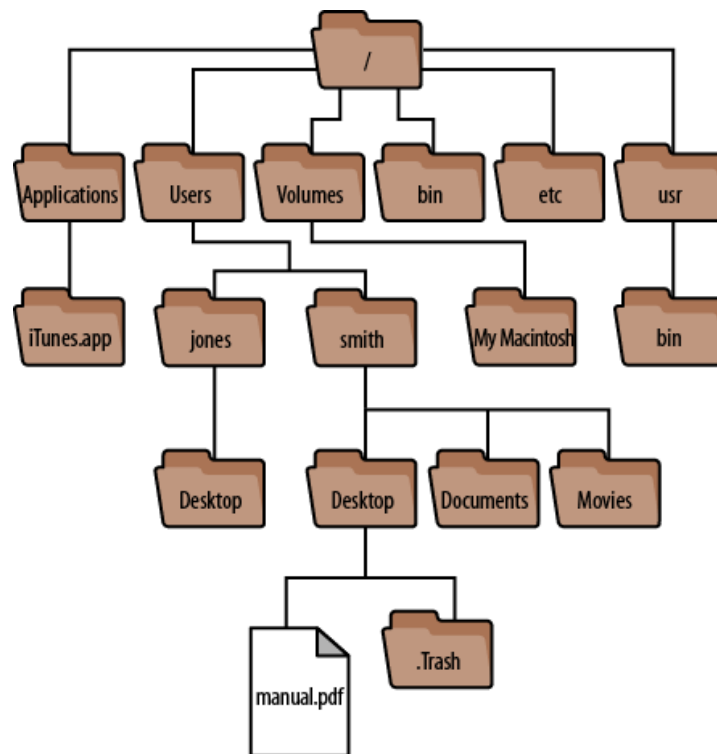


macOS file systems

Similar with macOS, the directories start from **root directory “/”** but macOS has slightly different structures.

When you open a terminal, you will be in “/Users/username”, which also indicated as “~”.

Run ``ls /`` to see what's under your root directory.





Path to file

A path is a string that represents the location of a file or directory within the file system. For example, a path may look like “/home/jiajia/data/sample_1.fastq”.

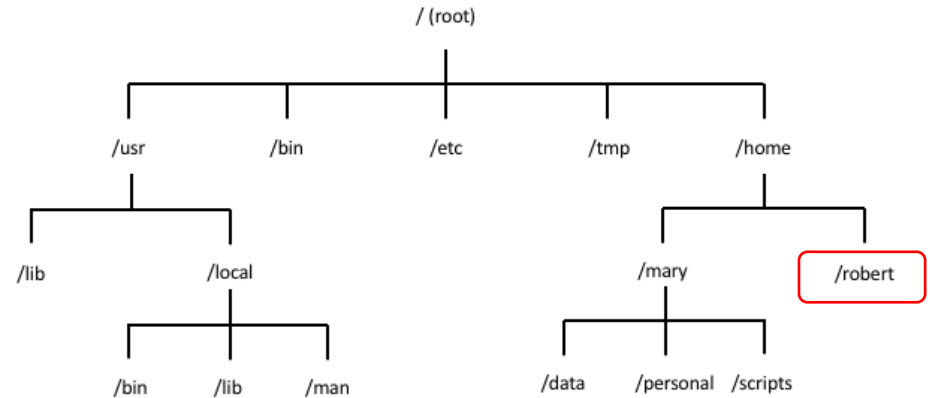
A path can be either **Absolute** or **Relative**. An absolute path starts from the root directory “/”, while a relative path is specified relative to your current directory.

In relative path:

- “.” represents your current directory
- “..” represents the directory above you

Examples:

- ./data/sample_1.fastq
- ../mary/data/sample_1.fastq





Commands to move around in Linux

- ``pwd`` - Get the path of your current directory

```
(base) jiajia@RSB-072750:~$ pwd
/home/jiajia
(base) jiajia@RSB-072750:~$
```

- ``ls`` - List files and directories under your current directory

```
(base) jiajia@RSB-072750:~$ ls
Miniconda3-latest-Linux-x86_64.sh  miniconda3      workshop-RNASeq-Analysis
RNASeq-Home                       variant-calling  workshop-intro-linux-mar-2025
(base) jiajia@RSB-072750:~$
```

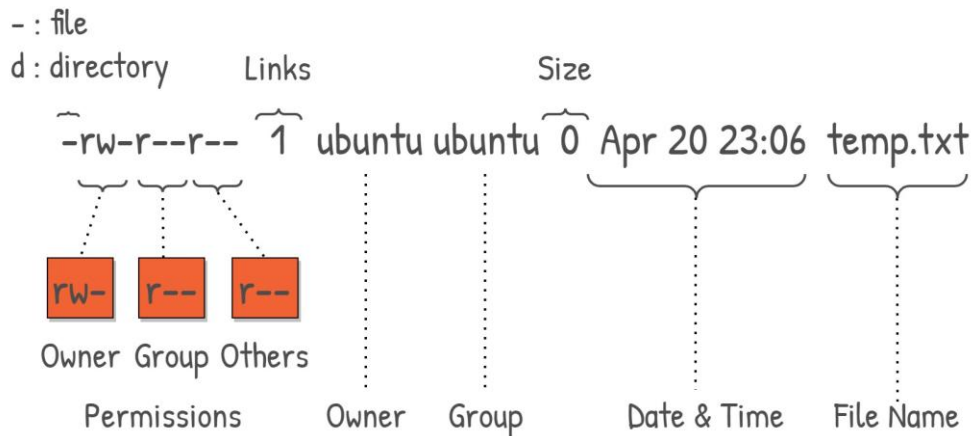
- ``ls`` has many options to list things in different ways
 - ``ls /``: list everything under root directory
 - ``ls -l``: list with long format
 - ``ls -a``: list including hidden files





Long format file information

```
(base) jiajia@RSB-072750:~$ ls -l
total 144344
-rw-r--r-- 1 jiajia jiajia 147784736 Dec 17 10:36 Miniconda3-latest-Linux-x86_64.sh
drwxr-xr-x 2 jiajia jiajia 4096 Jan 22 12:01 RNASeq-Home
drwxr-xr-x 19 jiajia jiajia 4096 Jan 22 12:16 miniconda3
drwxr-xr-x 3 jiajia jiajia 4096 Feb 18 15:07 variant-calling
drwxr-xr-x 3 jiajia jiajia 4096 Jan 22 10:20 workshop-RNASeq-Analysis
drwxr-xr-x 5 jiajia jiajia 4096 Feb 18 15:26 workshop-intro-linux-mar-2025
(base) jiajia@RSB-072750:~$
```





Specify two options together

```
(base) jiajia@RSB-072750:~$ ls -lh
total 141M
-rw-r--r-- 1 jiajia jiajia 141M Dec 17 10:36 Miniconda3-latest-Linux-x86_64.sh
drwxr-xr-x 2 jiajia jiajia 4.0K Jan 22 12:01 RNASeq-Home
drwxr-xr-x 19 jiajia jiajia 4.0K Jan 22 12:16 miniconda3
drwxr-xr-x 3 jiajia jiajia 4.0K Feb 18 15:07 variant-calling
drwxr-xr-x 3 jiajia jiajia 4.0K Jan 22 10:20 workshop-RNASeq-Analysis
drwxr-xr-x 5 jiajia jiajia 4.0K Feb 18 15:26 workshop-intro-linux-mar-2025
(base) jiajia@RSB-072750:~$
```

```
(base) jiajia@RSB-072750:~$ ls -alh
total 142M
drwxr-x-- 12 jiajia jiajia 4.0K Feb 18 15:07 .
drwxr-xr-x 3 root  root  4.0K Nov 18 14:29 ..
-rw----- 1 jiajia jiajia 3.4K Feb 18 17:52 .bash_history
-rw-r--r-- 1 jiajia jiajia 220 Nov 18 14:29 .bash_logout
-rw-r--r-- 1 jiajia jiajia 4.2K Jan 22 12:17 .bashrc
drwx----- 4 jiajia jiajia 4.0K Feb 3 17:07 .cache
drwxr-xr-x 2 jiajia jiajia 4.0K Feb 3 17:07 .conda
-rw-r--r-- 1 jiajia jiajia 116 Feb 3 17:50 .condarc
-rw-r--r-- 1 jiajia jiajia 55 Jan 10 12:45 .gitconfig
drwxr-xr-x 2 jiajia jiajia 4.0K Jan 6 13:41 .landscape
-rw----- 1 jiajia jiajia 20 Jan 10 12:40 .lessht
-rw-r--r-- 1 jiajia jiajia 0 Feb 19 13:43 .motd_shown
-rw-r--r-- 1 jiajia jiajia 807 Nov 18 14:29 .profile
drwx----- 2 jiajia jiajia 4.0K Feb 19 13:54 .ssh
drwxr-xr-x 5 jiajia jiajia 4.0K Jan 6 13:48 .vscode-server
-rw-r--r-- 1 jiajia jiajia 183 Feb 18 12:10 .wget-hsts
-rw-r--r-- 1 jiajia jiajia 141M Dec 17 10:36 Miniconda3-latest-Linux-x86_64.sh
```

`ls -lh` show file size in KB, MB, GB.

`ls -alh` ?





`cd` - change directory

- ``cd /``: go to root directory
- ``cd /home``: go to home directory
- ``cd ..``: go to one directory above
- ``cd ~``: go to your home directory





Create files and directories

- ``mkdir`` - Make Directory. Create a new directory.
- ``mkdir variant-calling`` - Create a new directory called “variant-calling”.

Practice – create a new directory called “variant-calling” under your home directory.

- ``touch`` - Create empty files.

Practice – create a new file called “newfile.txt” in “variant-calling”.





Text editors in Linux

In Windows, there are text editors like Notepad, Microsoft Word, Visual Studio Code...

In Linux, the most common text editors are Nano and Vim. You can use it to edit the content of your file.

- ``nano newfile.txt``
- ``vim newfile.txt``





Nano

```
jiajia@RSB-072750: ~/variant x + v
GNU nano 7.2 newfile.txt

[ New File ]
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^_ Go To Line
```

Once you run the nano command, you will be guided to this interface, that's the nano editor interface. You can start typing.

To save, you can press `ctrl+x`.



[illegible]

For vim, you have to press ``i`` to start editing.

To save, press ``esc`` first and type `:`wq``, then ``enter``.



Read file contents

- ``cat`` - print file contents
- ``cat newfile.txt`` - to print the content of newfile.txt

But sometimes when the file is too big, and we don't want to print out everything all at once. We can use ``less``.

- ``less`` - to view file in a scrollable manner
- ``less SRR2584863_1.fastq.gz`` - to view a FASTQ file
- ``head`` - to print the beginning of a file
- ``head -n 20`` - print the first 20 lines
- ``tail`` - to print the end of a file





Press `space` to scroll down, and press `q` to quit.



Remove files and directories

- ``rmkdir`` remove empty directories
- ``rm`` remove files
- ``rm newfile.txt``
- ``rm -r`` remove directories that has files inside





Move, rename, and copy files and directories

- ``mv`` move files and directories
- ``mv /path/to/source /path/to/destination``
- ``mv /home/jiajia/test.txt /home/jiajia/variant-calling``
- ``mv sample_1.txt ..`` - move file one folder above

Rename file uses the same command:

- ``mv oldname newname``
- ``mv oldfile.txt newfile.txt``
- ``cp`` copy files
- ``cp newfile.txt newfile_copy.txt``
- ``cp -r`` to copy a directories with everything inside





Some useful tricks in Linux

- Getting help with a command:
 - ``man ls``
 - ``ls --help``
- ``history`` get the history of recently used commands
- ``clear`` clear the screen
- Press ``tab`` to auto-complete file names
- Press ``up`` to go to the last used commands
- Press ``ctrl+c`` to terminate running processes
- ``exit`` to exit the terminal





Wild card

Linux provides 2 wildcard characters to represent ambiguous characters or strings in file names.

The question mark `?` represents any single character.

- For example, `ls file?.txt` would list `file1.txt` and `file2.txt` but not `file50.txt`.

The asterisk `*` represents any string of 0 or more characters.

- For example, `ls file*.txt` would list `file.txt`, `file1.txt`, and `file50.txt`, but not `file01.data`.





Input and output redirection

- Input redirection ``<``
- `cat < sample_1.fastq` works the same as `cat sample_1.fastq`.

Normally, we use output redirection ``>`` more in our daily tasks. You can save the output of a command to a file using output redirection.

- `ls -alh > file_info.txt` - save the output of the `ls` command to a file

Similar to output redirection, ``>>`` can append the output to a file.

- `ls -alh >> file_info.txt` append the output after the existing file content.





Pipes `|`

Pipe is another form of redirection; it can redirect the output of the first command to the input of the second command.

- ``ls -l /etc | head -n 5`` show only the first 5 files in /etc

You can pipe as many commands as you like:

- ``ls -l /etc | head -n 5 | tail -n 1`` show only the 5th file in /etc





Downloading and transferring data

- ``curl`` to download files from server using various protocols, including HTTP, FTP, SCP etc.
- ``curl -O``
`https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/005/845/GCF_000005845.2_ASM584v2/GCF_000005845.2_ASM584v2_genomic.fna.gz``
- What does the option “-O” do? Check the help manual.
- ``scp`` secure copy, allows you to transfer files between a remote machine and a local machine.
- ``scp file.txt user@remote_host:/path/to/destination/``
- ``scp file.txt u1133824@dayhoff.rsb.anu.edu.au:~/data``





Linux variables

In Linux, variables are used to store data, such as text strings, numbers, or arrays, with a name that serves as a reference to that data.

- ``variable_name=value``
- `Name="John"`

To access the value of a variable, we use the ``$`` sign followed by the variable name:

- `echo "My name is $name"`

``echo`` is a command to display text or output strings to the terminal.





Loop statements

Loop is a powerful programming tool that enables you to execute a set of commands repeatedly.

`for` loop allows us to iterate over a list of items and perform actions on each item in the list.

Syntax:

```
for thing in list_of_things
do
    command_1
    command_2
    ...
done
```

```
for i in a b c
do
    echo $i
done
```

```
fruits="apple orange banana"

for fruit in $fruits
do
    echo "I like $fruit"
done
```



THANK YOU

Contact Us

Jiajia Li
ANU Biological Data Science Institute

RN Robertson Building, 46 Sullivan's Creek Rd
Canberra ACT 2600

E jiajia.li1@anu.edu.au
W <https://bdsi.anu.edu.au/>



Australian
National
University