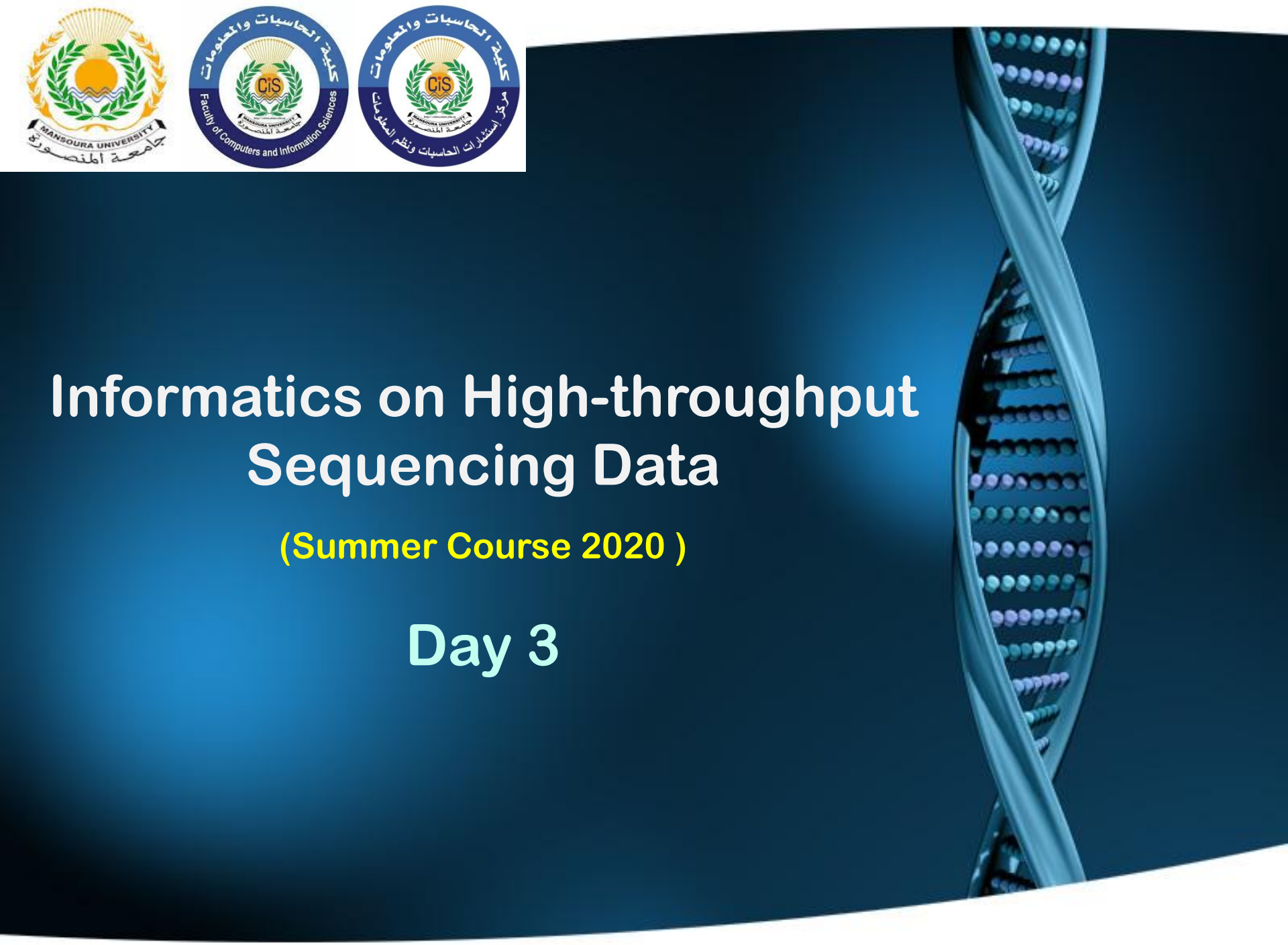
Informatics on High-throughput Sequencing Data

(Summer Course 2020)

Day 3



Agenda

- Unix-based systems.
- Why Linux!
- Let's start!
- Linux Commands for:
  - Files & Directories.
  - System.
  - Process Management.
  - Networking.
  - Compression.
  - Searching.
- Piping output.
- Wildcard character.
- Redirecting output.
- Stream Editor (Sed).
- Linux tools for text files processing.
- Shell Scripting

Ownership of Linux files

User

✓A user is the owner of the file. By default, the person who created a file becomes its owner

Group

✓A user- group can contain multiple users. All users belonging to a group will have the same access permissions to the file.

Other

✓Any other user who has access to a file. This person has neither created the file, nor he belongs to a user group who could own the file. Practically, it means everybody else.

Permission system

Read

✓This permission give you the authority to open and read a file. Read permission on a directory gives you the ability to lists its content.

Write

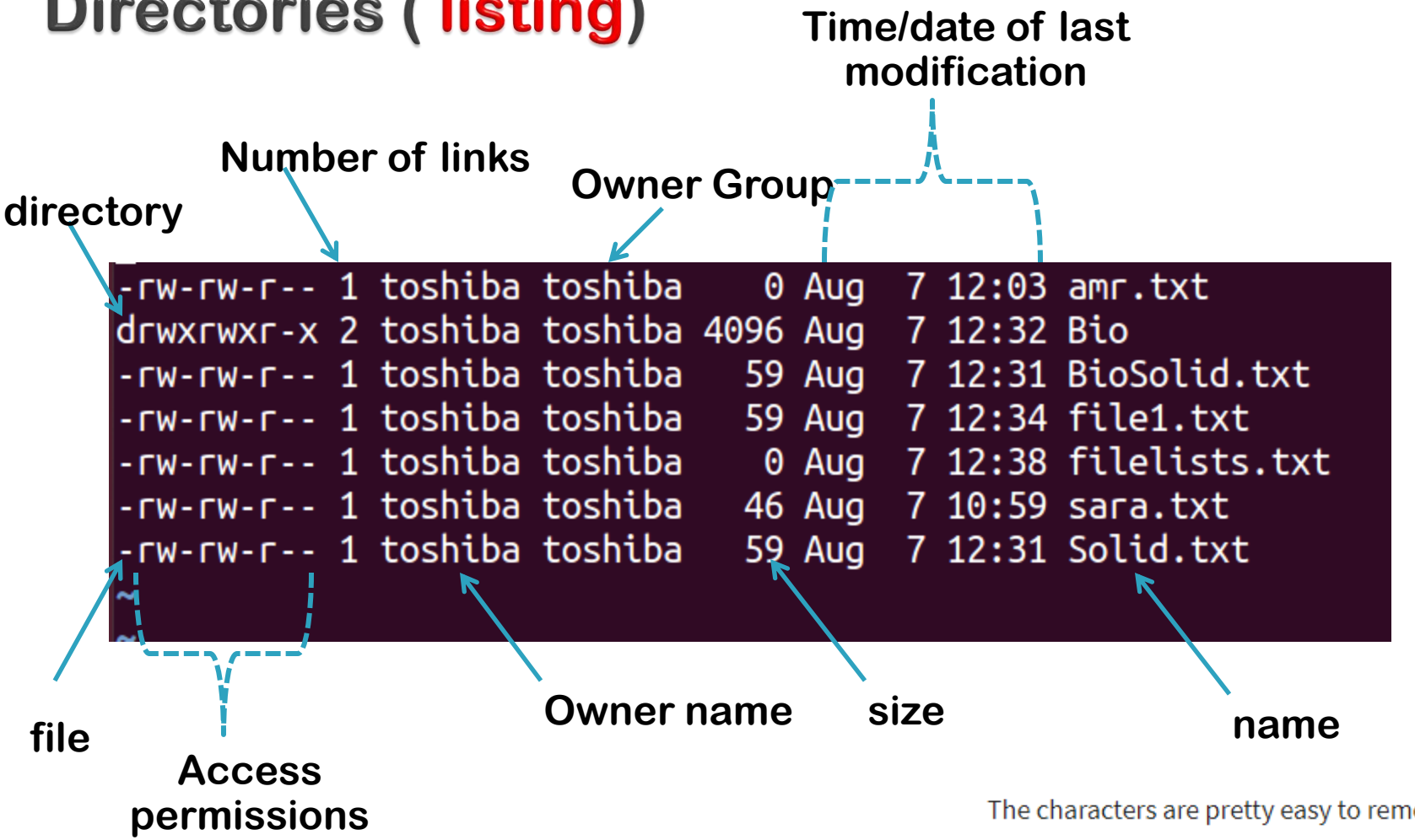
✓The write permission gives you the authority to modify the contents of a file. The write permission on a directory gives you the authority to add, remove and rename files stored in the directory.

Execute

✓In Windows, an executable program usually has an extension ".exe" and which you can easily run. In Unix/Linux, you cannot run a program unless the execute permission is set. If the execute permission is not set, you might still be able to see/modify the program code(provided read & write permissions are set), but not run it.

https://www.guru99.com/file-permissions.html

Directories (listing)



The characters are pretty easy to remember.

r = read permission
w = write permission
x = execute permission
- = no permission

Note: total is the number of blocks taken up by the files

Changing file/directory Permissions

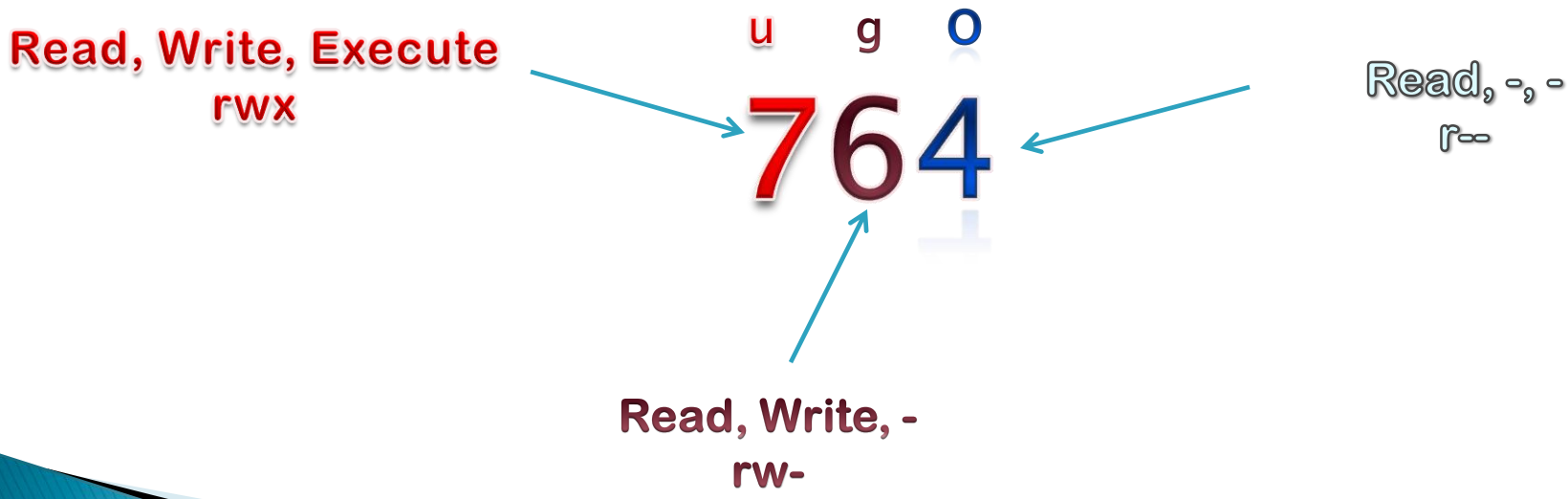
- We can use the 'chmod' command which stands for 'change mode'.

Number	Permission Type	Symbol
0	No Permission	---
1	Execute	--x
2	Write	-w-
3	Execute + Write	-wx
4	Read	r--
5	Read + Execute	r-x
6	Read +Write	rw-
7	Read + Write +Execute	rwx

https://www.guru99.com/file-permissions.html

Changing file/directory Permissions

```
[guest@centos6 Bio1]$ ls -l f
-rw-rw-r--. 1 guest guest 0 Jul 20 06:07 f
[guest@centos6 Bio1]$ chmod 764 f
[guest@centos6 Bio1]$ ls -l f
-rwxrw-r--. 1 guest guest 0 Jul 20 06:07 f
```



Changing file/directory Permissions

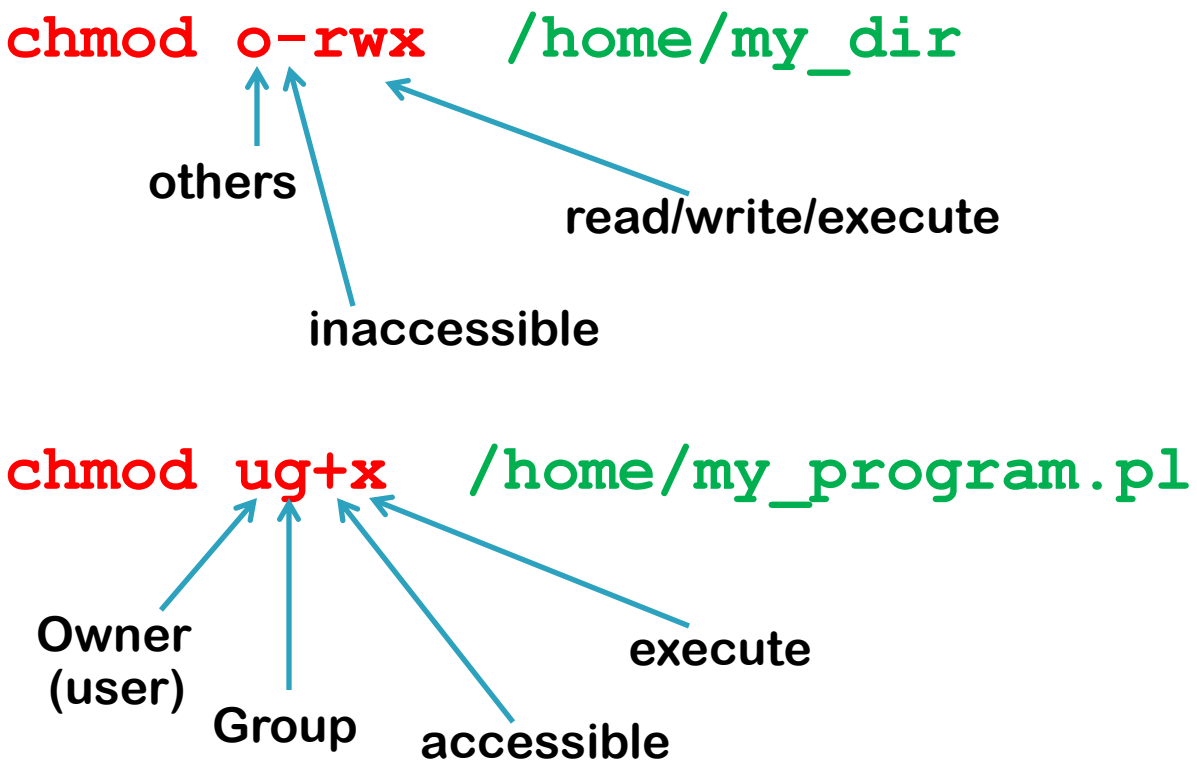
- We can use the '**chmod**' command which stands for 'change mode'.

Operator	Description
+	Adds a permission to a file or directory
-	Removes the permission
=	Sets the permission and overrides the permissions set earlier.

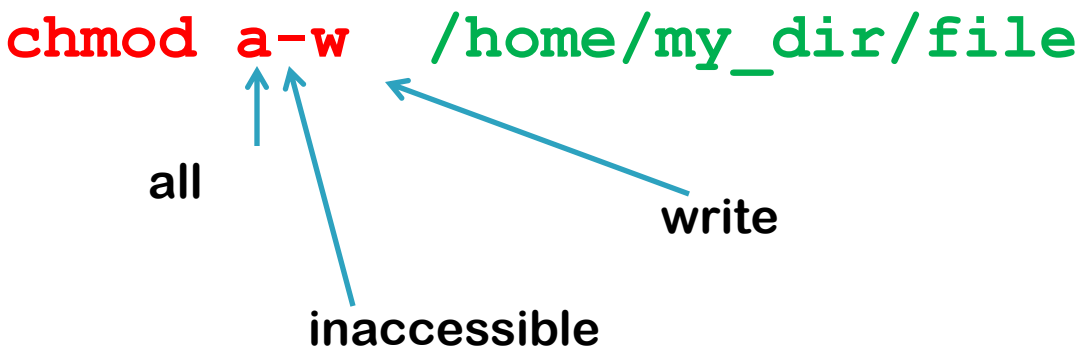
User Denotations	
u	user/owner
g	group
o	other
a	all

https://www.guru99.com/file-permissions.html

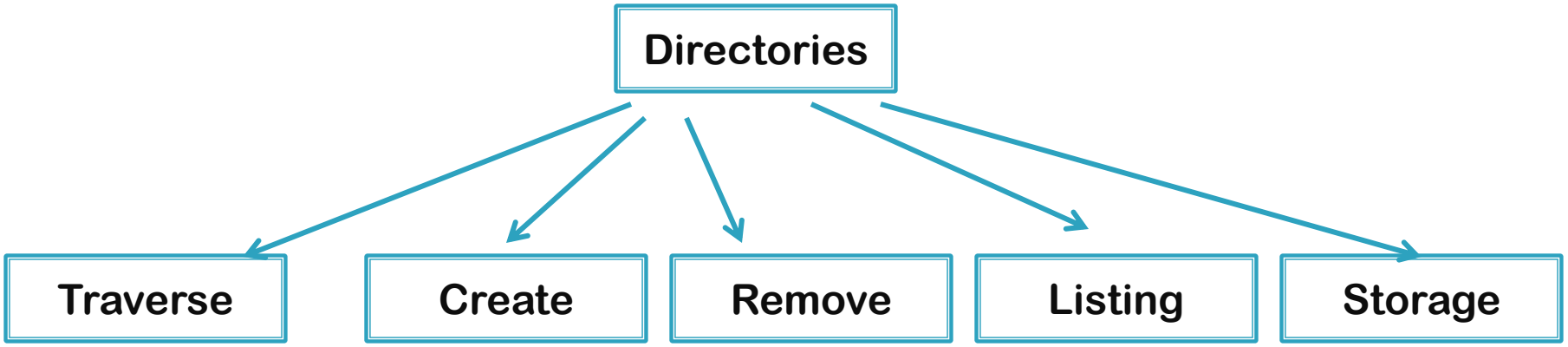
Directories and Files (Permissions)



Directories and Files (Permissions)



Working with directories




```
pwd
cd ./
cd ../
cd /home/bio
```

```
mkdir bio
```

```
rmdir bio
rm -rf bio
rm -ri bio
```

```
ls
ls -al
ls -alt *.txt
ls -alS
```

```
du -hs .
```

Note: rmdir will fail if directory is not empty.

Notes

- There are three main wildcards in Linux:
  - An asterisk (*) – matches one or more occurrences of any character, including no character.
  - Question mark (?) – represents or matches a single occurrence of any character.
  - Bracketed characters ([ ]) – matches any occurrence of character enclosed in the square brackets.

Try & Explain what you see ?

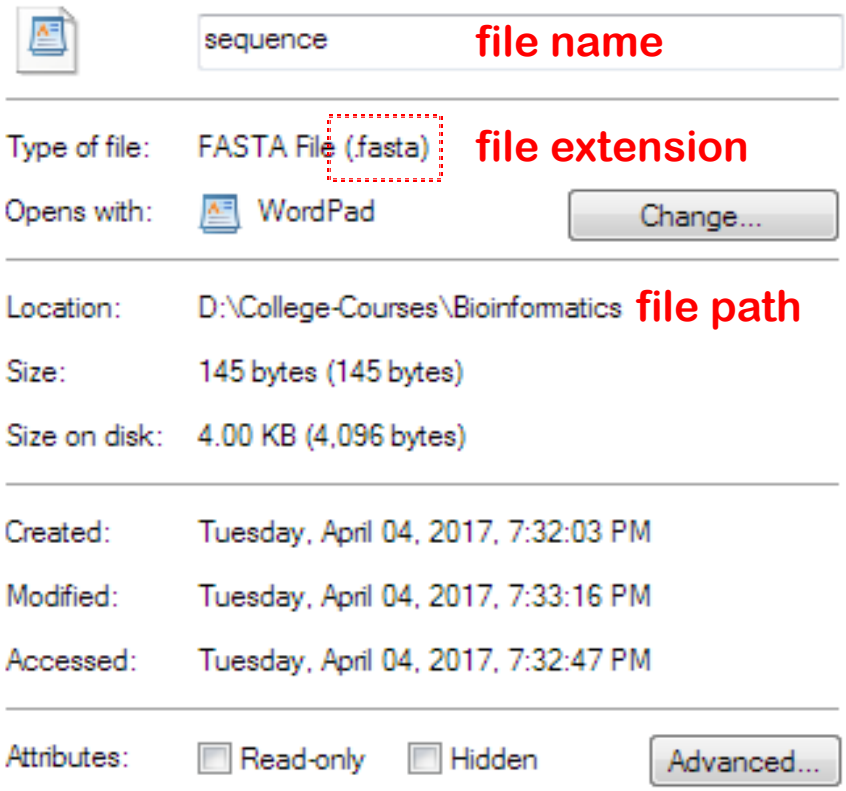
- man ls
- ls apple.*
- ls ?pple.genome
- ls [a-z]*.genome
- ls p*.genome
- ls {apple, peach}.genome
- mkdir apple, cp apple.* apple
- mv pear.genome pear

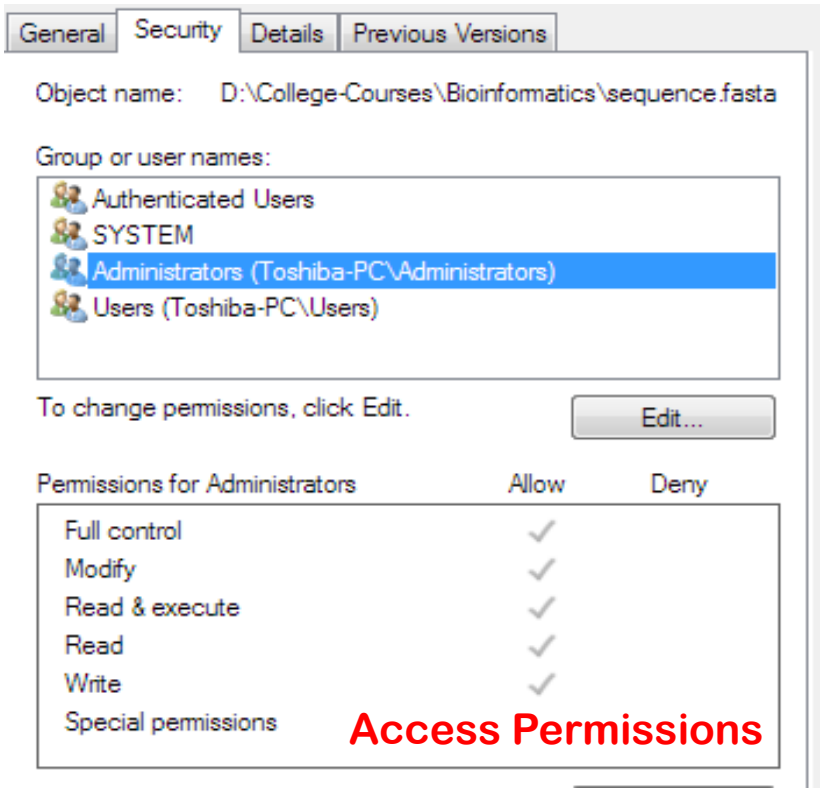
Note! (flowcell, lane, run, read, library)

- --------- a single sequence produced from a sequencer.
- ---------- a collection of DNA fragments that have been prepared for sequencing.
- ---------- a chip on which DNA is loaded and provided to the sequencer.
- --------- one portion of a flowcell. Usually used for technical replicates or different samples.
- ---------- an entire sequencing reaction from start to finish.

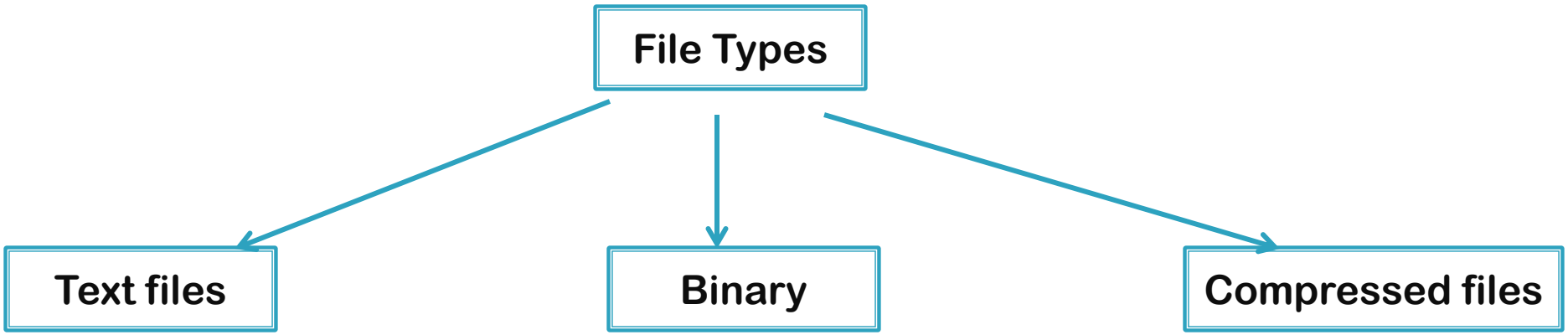
https://learn.gencore.bio.nyu.edu/ngs-file-formats/

Files


sequence
file name
Type of file:
FASTA File (.fasta)
file extension
Opens with:
WordPad
Change...
Location:
D:\College-Courses\Bioinformatics
file path
Size:
145 bytes (145 bytes)
Size on disk:
4.00 KB (4,096 bytes)
Created:
Tuesday, April 04, 2017, 7:32:03 PM
Modified:
Tuesday, April 04, 2017, 7:33:16 PM
Accessed:
Tuesday, April 04, 2017, 7:32:47 PM
Attributes:
Read-only
Hidden
Advanced...


General
Security
Details
Previous Versions
Object name:
D:\College-Courses\Bioinformatics\sequence.fasta
Group or user names:
Authenticated Users
SYSTEM
Administrators (Toshiba-PC\Administrators)
Users (Toshiba-PC\Users)
To change permissions, click Edit.
Edit...
Permissions for Administrators
Allow
Deny
Full control
Modify
Read & execute
Read
Write
Special permissions
Access Permissions

Working with Files



- Text documents (i.e. README files)
- Data in text format (i.e. FASTA, FASTQ)
- Scripts
  - Shell *.sh
  - Perl *.pl
  - Python *.py

- Executables (i.e. samtools, bowtie)
- Data in binary format (i.e. BAM, index files for BWA)

- Usually *.gz, *.zip, *.bz2
- often text files re-formatted to save space on disk

Files (create, edit & view)

`less file` — View file with page navigation.

`head file` — Output the first 10 lines of file.

`tail file` — Output the last 10 lines of file.

`more myfile.txt`

Viewing the contents of a file, one page at a time. To advance a page press Space. To return to the command prompt, enter "q".

`vim file` — Edit file.

Files (create, edit & view)

`touch file` — Used to create, change and modify timestamps of a file.

`cat file` — Used to display the content of text files and to combine several files to one file.

`cat file1 file2`

Output the contents of file1, then file2.

`cat file1 > file2`

Copy file1 to file2.

Files (copy, move & find)

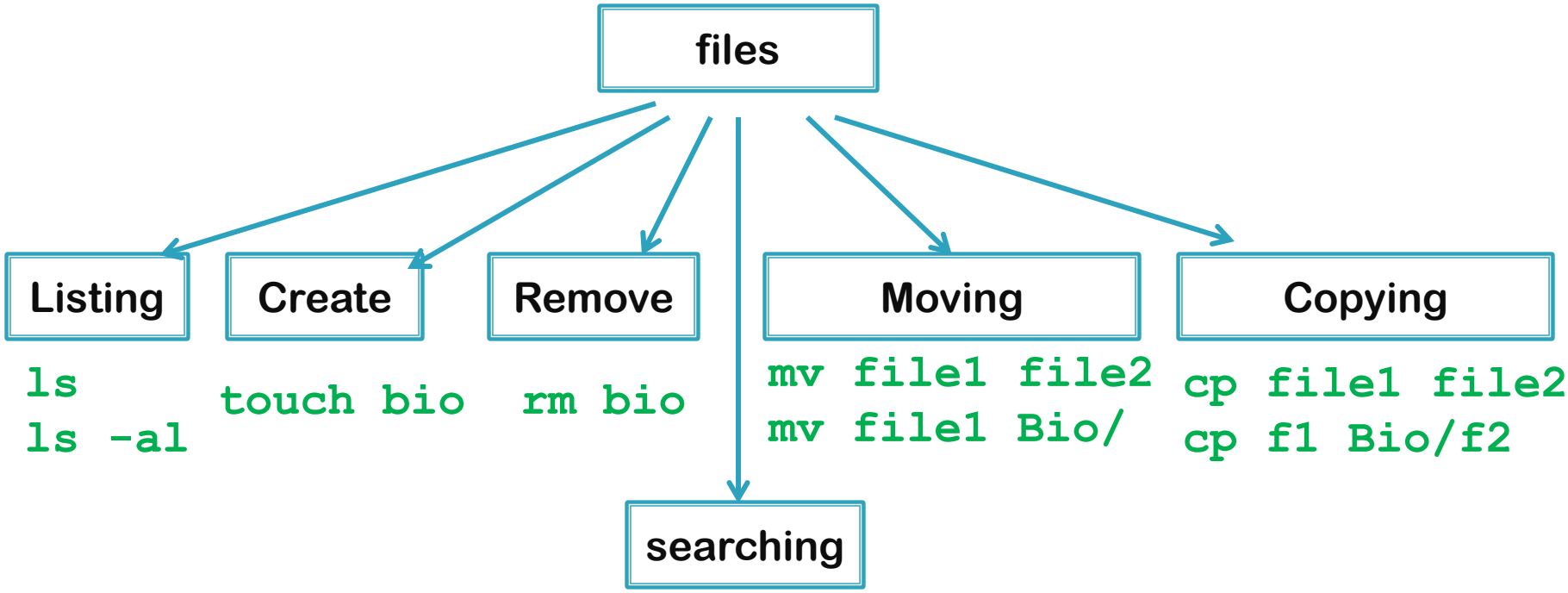
```
cp file1 file2
```
Copy file1 to file2

```
mv file1 reads
```
Rename file1 to reads

```
mv file1 ~/home/Bio
```
Move file1 to Bio directory

```
find . -name file1.bam
```
Search for `file1.bam` starting from the current working directory.

Working with files



Thanks!

// | ?