LIÈGE université
GIGA

GIGA mass storage hands-on

GIGA doctoral school

6th October 2022

Alice Mayer, PhD
GIGA bioinformatic team
bioinfo.giga@uliege.be

1. Preliminary remarks

- all procedures will be done on GIGA mass storage (usually restricted to GIGA members and their collaborators)
- for the purpose of this training, we have created a "friend" user for the GIGA/ULG members who don't have their ULG ID yet.
- other server will have different structure, entry point and/or connection procedure but usually same principles

SAMBA connection
(Finder, Windows explorer)

2.1 SAMBA connection instructions

- Windows:
  https://gitlab.uliege.be/giga-bioinfo/user-guides-wiki/-/wikis/mass-storage/mass-storage-connection#windows
- Mac:
  https://gitlab.uliege.be/giga-bioinfo/user-guides-wiki/-/wikis/mass-storage/mass-storage-connection#mac-os-x
- Linux:
  https://gitlab.uliege.be/giga-bioinfo/user-guides-wiki/-/wikis/mass-storage/mass-storage-connection#linux

2.2 SAMBA connection practical

Space exploration and small file transfer: your $HOME

- your entry point is your HOME (only you can read/write there)
- you can create folders and/or transfer files these
  - Very important naming conventions : https://gitlab.uliege.be/giga-bioinfo/user-guides-wiki/-/wikis/mass-storage/mass-storage-management-practices#mandatory-rules
  - create a training folder in your home (for exemple DS_training)
  - transfer (drag and drop) a small text file from your computer into it ### See naming conventions !!!!
- the _SHARE_ folder is a link to shared resources (research, platforms, resources and administration)

2.2 SAMBA connection practical

Shared space exploration

- _SHARE_/Research/… if you are GIGA member, you can go down the tree to find your lab (you may have access or not depending of what your PI asked for you)
- _SHARE_/Platforms/GEN/BIOINFO/TRAINING (folder we created for this training => temporarily allowed to read and write => if you delete something it disappear for everybody!)
- _SHARE_/Resources (you can read but not write, if you can't read a file, we probably made a mistake in setting the permissions => let us know)
- go to _SHARE_/Resources/Genomes/Scripts => you can open and read (most) files but not add a new file in the folder or modify/delete one of them

2.3 important considerations

- SAMBA useful to
  - see what's in a folder
  - read/update small files (typically smaller than 1Gb)
  - transfer small files (from or to your computer)

- best practices:
  - if working on a document, save it on your computer and only transfer the final version to mass storage
  - naming conventions: https://gitlab.uliege.be/giga-bioinfo/user-guides-wiki/-/wikis/mass-storage/mass-storage-management-practices#mandatory-rules

- WARNING: when you delete a file/folder on the mass storage, data will be deleted for everybody !!!! Don't delete files or folders without being sure nobody else need it. Before to delete a folder/file, check you are not unintentionally deleting shared files (ex: if you delete the _SHARE_ folder in your home, you will delete all your lab's data !!!!!!!!)

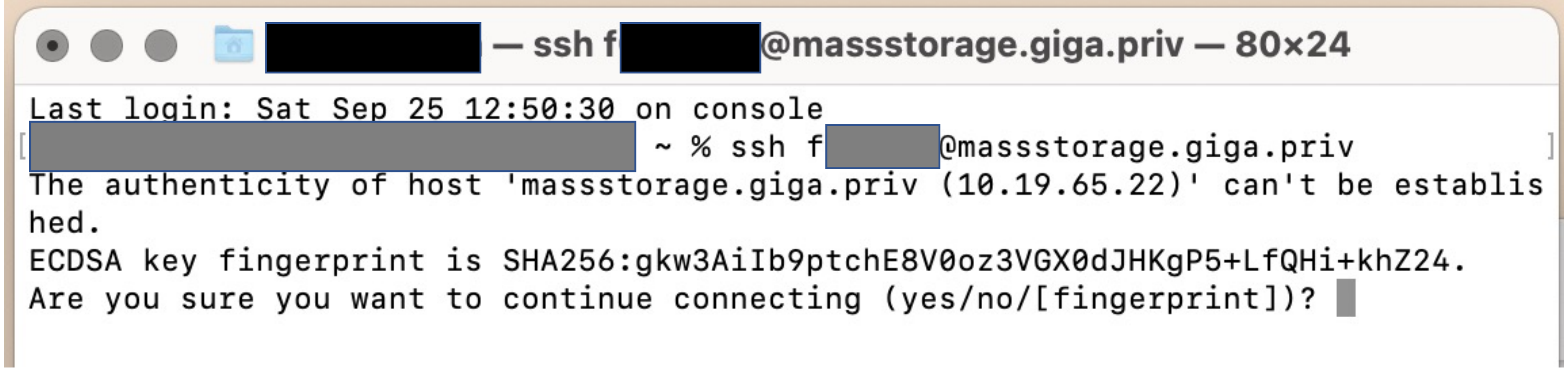
  If you delete a file/folder by mistake, contact the bioinformatic team or the UDIMED as soon as possible (if the file/folder was backed up, we can restore it during 28 days!)

SSH connection

3.1 SSH connection instruction

- open command line terminal
  - MAC or LINUX: open Terminal
  - Windows: look for PowerShell in start menu or install and run MobaXterm https://mobaxterm.mobatek.net/
- type "ssh u123456@massstorage.giga.priv"
- If you see a message about ECDSA key fingerprint (see next slide), answer "yes"
- Enter password when prompted (nothing will appear on screen while typing)

Message that could appear at first connection



```
Last login: Sat Sep 25 12:50:30 on console
[redacted] ~ % ssh f[redacted]@massstorage.giga.priv
The authenticity of host 'massstorage.giga.priv (10.19.65.22)' can't be establis
hed.
ECDSA key fingerprint is SHA256:gkw3AiIb9ptchE8V0oz3VGX0dJHKgP5+LfQHi+khZ24.
Are you sure you want to continue connecting (yes/no/[fingerprint])?
```

Screen once logged

(X11 and bash error messages are specific to my configuration)

```
[alice@M[redacted] ~ % ssh u[redacted]@massstorage.giga.priv
[u[redacted]@massstorage.giga.priv's password:
X11 forwarding request failed on channel 0
Last login: Mon Oct  4 01:00:18 2021 from 10.39.7.185
Welcome to

      /\/\   __ _ ___ ___  / _\ |_ ___  _ __ __ _  __ _  ___
     /    \ / _` / __/ __| \ \| __/ _ \| '__/ _` |/ _` |/ _ \
    / /\/\ \ (_| \__ \__ \ _\ \ || (_) | | | (_| | (_| |  __/
    \/    \/\__,_|___/___/ \__/\__\___/|_|  \__,_|\__, |\___|
                                                  |___/

                You are logged in on NASGW2
                In case of problem, contact the Helpdesk
                   Ticket     : https://sam.segi.uliege.be/
                   Phone      : 04/366.49.99
                   E-mail     : helpdesk@segi.ulg.ac.be

    --> For more information about the GIGA cluster and mass storage:
        https://gitlab.uliege.be/giga-bioinfo/user-guides-wiki/-/wikis/cluster/cluster-home
-bash: warning: setlocale: LC_CTYPE: cannot change locale (UTF-8): No such file or directory
-bash: module: command not found
[u[redacted]@nasgw2 ~]$
```

3.2 what's SSH ?

- Encrypted network protocol
- linux system, use bash as command interpreter

Google is you best friend

ex: "linux how to do xxxx" or "bash how to do xxxx"

3.3 SSH connection practical

- your entry point = your $HOME (the same as with SAMBA connection)
- it's a linux system and bash interface => when looking for how to do stuff, google "linux how to do xxxx" or "bash how to do xxxx"
- basic commands
  - list directory content with "`ls -lh <path>`"
  - move around with "`cd <path>`" and "cd .." to go one folder up (parent folder)
  - to go back to home, use "`cd $HOME`"
  - print working directory with "`pwd`" or "`realpath ./`"
  - read a text file with "`less <path/to/file.txt>`" (type "q" to close it)
- move around and test different commands
- see supplemental slides with other bash commands

File transfer

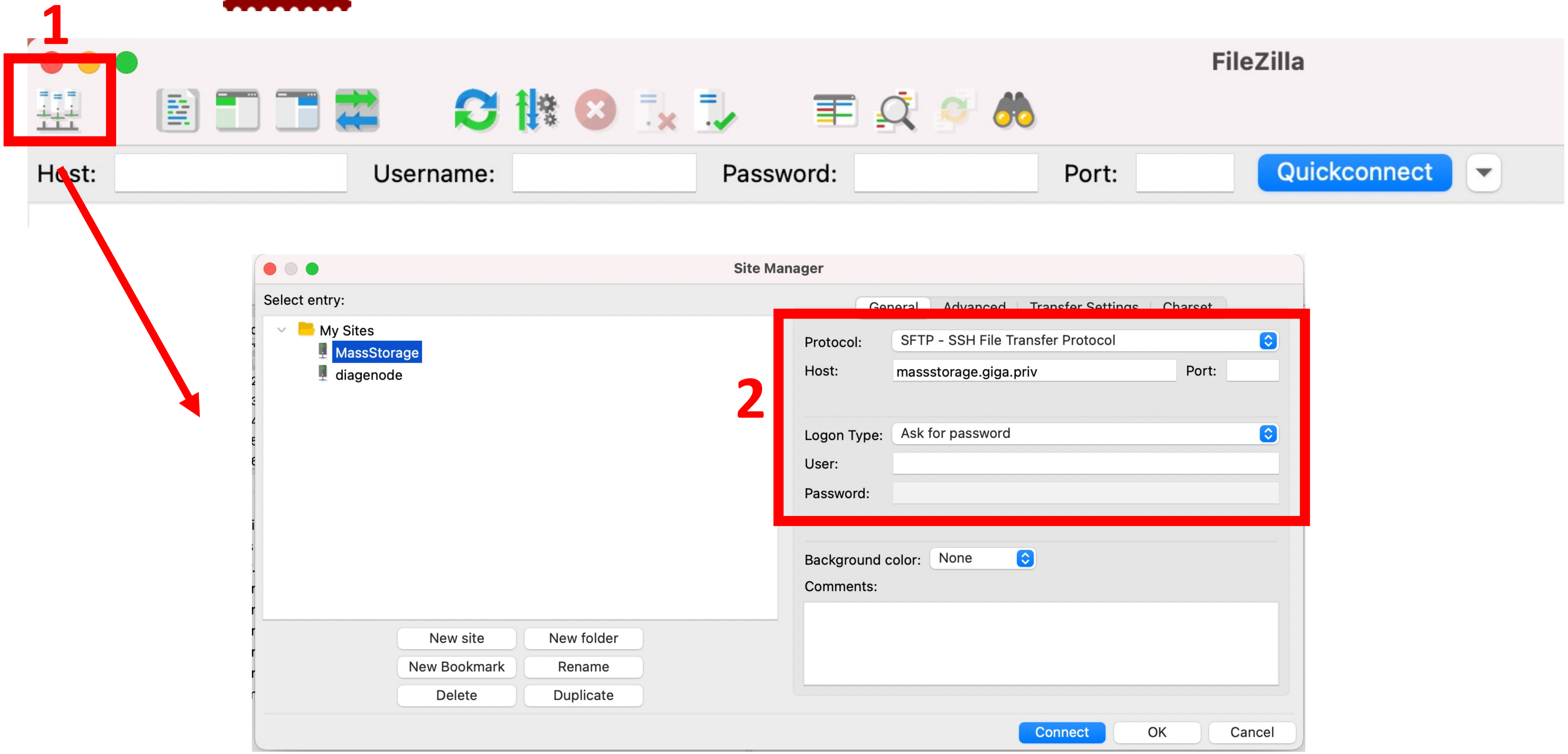
LIÈGE université
GIGA

4. File transfer from/to mass storage

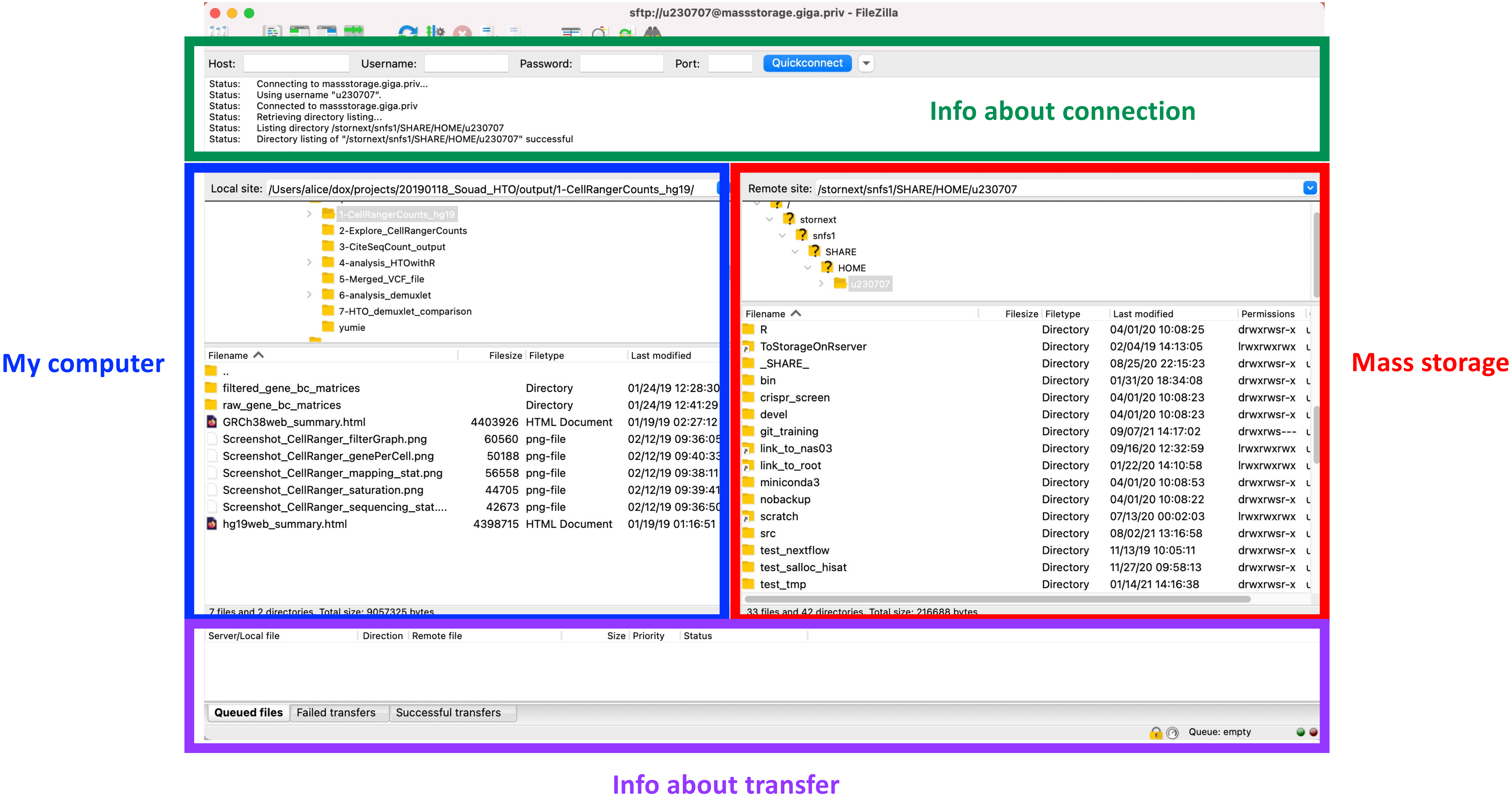
- SAMBA (drag and drop)
  - **Warning:** if you lose your connection the transfer is aborted and some files may be partially transferred (without warning!!!) => always check integrity afterward

- FileZilla (secure protocol, graphical interface),
  - **Warning:** if you lose your connection the transfer is aborted and some files may be partially transferred (warning will appear in graphical interface) => always check messages in transfer window and always better to also check integrity afterward
  - https://filezilla-project.org/
  - Host = sftp://massstorage.giga.priv

FileZilla



Protocol: SFTP - SSH File Transfer Protocol
Host: massstorage.giga.priv
Port:
Logon Type: Ask for password
Background color: None



sftp://u230707@massstorage.giga.priv - FileZilla

```
Status:	Connecting to massstorage.giga.priv...
Status:	Using username "u230707".
Status:	Connected to massstorage.giga.priv
Status:	Retrieving directory listing...
Status:	Listing directory /stornext/snfs1/SHARE/HOME/u230707
Status:	Directory listing of "/stornext/snfs1/SHARE/HOME/u230707" successful
```

Local site: /Users/alice/dox/projects/20190118_Souad_HTO/output/1-CellRangerCounts_hg19/

- 1-CellRangerCounts_hg19
- 2-Explore_CellRangerCounts
- 3-CiteSeqCount_output
- 4-analysis_HTOwithR
- 5-Merged_VCF_file
- 6-analysis_demuxlet
- 7-HTO_demuxlet_comparison
- yumie

Filename	Filesize	Filetype	Last modified
..			
filtered_gene_bc_matrices		Directory	01/24/19 12:28:30
raw_gene_bc_matrices		Directory	01/24/19 12:41:29
GRCh38web_summary.html	4403926	HTML Document	01/19/19 02:27:12
Screenshot_CellRanger_filterGraph.png	60560	png-file	02/12/19 09:36:05
Screenshot_CellRanger_genePerCell.png	50188	png-file	02/12/19 09:40:33
Screenshot_CellRanger_mapping_stat.png	56558	png-file	02/12/19 09:38:11
Screenshot_CellRanger_saturation.png	44705	png-file	02/12/19 09:39:41
Screenshot_CellRanger_sequencing_stat....	42673	png-file	02/12/19 09:36:50
hg19web_summary.html	4398715	HTML Document	01/19/19 01:16:51

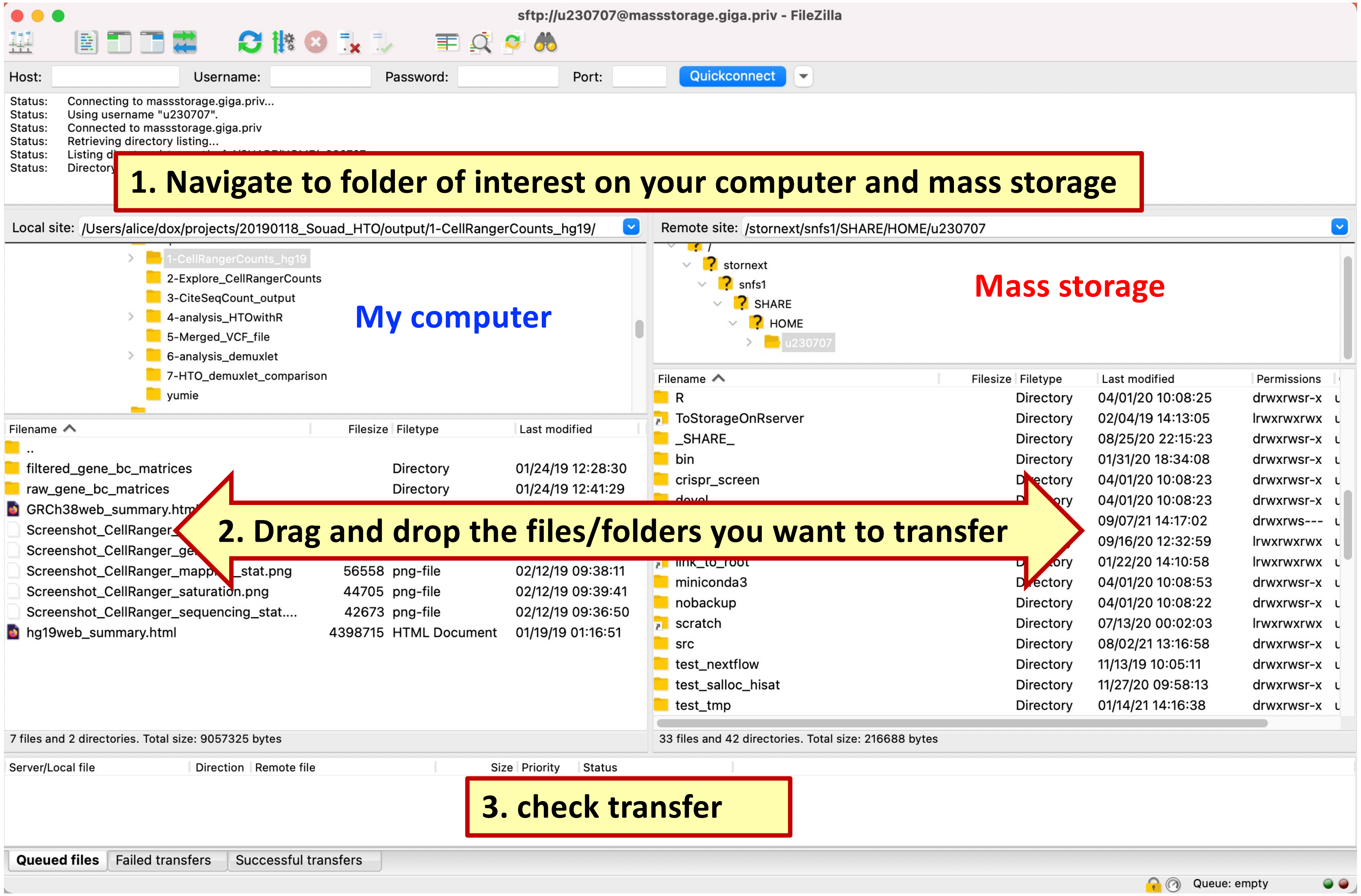
7 files and 2 directories. Total size: 9057325 bytes

Remote site: /stornext/snfs1/SHARE/HOME/u230707

- /
  - stornext
    - snfs1
      - SHARE
        - HOME
          - u230707

Filename	Filesize	Filetype	Last modified	Permissions
R		Directory	04/01/20 10:08:25	drwxrwsr-x
ToStorageOnRserver		Directory	02/04/19 14:13:05	lrwxrwxrwx
SHARE		Directory	08/25/20 22:15:23	drwxrwsr-x
bin		Directory	01/31/20 18:34:08	drwxrwsr-x
crispr_screen		Directory	04/01/20 10:08:23	drwxrwsr-x
devel		Directory	04/01/20 10:08:23	drwxrwsr-x
git_training		Directory	09/07/21 14:17:02	drwxrws---
link_to_nas03		Directory	09/16/20 12:32:59	lrwxrwxrwx
link_to_root		Directory	01/22/20 14:10:58	lrwxrwxrwx
miniconda3		Directory	04/01/20 10:08:53	drwxrwsr-x
nobackup		Directory	04/01/20 10:08:22	drwxrwsr-x
scratch		Directory	07/13/20 00:02:03	lrwxrwxrwx
src		Directory	08/02/21 13:16:58	drwxrwsr-x
test_nextflow		Directory	11/13/19 10:05:11	drwxrwsr-x
test_salloc_hisat		Directory	11/27/20 09:58:13	drwxrwsr-x
test_tmp		Directory	01/14/21 14:16:38	drwxrwsr-x

33 files and 42 directories. Total size: 216688 bytes



```
Status:	Connecting to massstorage.giga.priv...
Status:	Using username "u230707".
Status:	Connected to massstorage.giga.priv
Status:	Retrieving directory listing...
Status:	Listing d
Status:	Directory
```

1. Navigate to folder of interest on your computer and mass storage

Local site: /Users/alice/dox/projects/20190118_Souad_HTO/output/1-CellRangerCounts_hg19/

- 1-CellRangerCounts_hg19
- 2-Explore_CellRangerCounts
- 3-CiteSeqCount_output
- 4-analysis_HTOwithR
- 5-Merged_VCF_file
- 6-analysis_demuxlet
- 7-HTO_demuxlet_comparison
- yumie

My computer

Filename	Filesize	Filetype	Last modified
..			
filtered_gene_bc_matrices		Directory	01/24/19 12:28:30
raw_gene_bc_matrices		Directory	01/24/19 12:41:29
GRCh38web_summary.htm			
Screenshot_CellRanger_			
Screenshot_CellRanger_ge			
Screenshot_CellRanger_mapp _stat.png	56558	png-file	02/12/19 09:38:11
Screenshot_CellRanger_saturation.png	44705	png-file	02/12/19 09:39:41
Screenshot_CellRanger_sequencing_stat....	42673	png-file	02/12/19 09:36:50
hg19web_summary.html	4398715	HTML Document	01/19/19 01:16:51

7 files and 2 directories. Total size: 9057325 bytes

Remote site: /stornext/snfs1/SHARE/HOME/u230707

- /
  - stornext
    - snfs1
      - SHARE
        - HOME
          - u230707

Mass storage

Filename	Filesize	Filetype	Last modified	Permissions
R		Directory	04/01/20 10:08:25	drwxrwsr-x
ToStorageOnRserver		Directory	02/04/19 14:13:05	lrwxrwxrwx
SHARE		Directory	08/25/20 22:15:23	drwxrwsr-x
bin		Directory	01/31/20 18:34:08	drwxrwsr-x
crispr_screen		D ectory	04/01/20 10:08:23	drwxrwsr-x
devel		D ory	04/01/20 10:08:23	drwxrwsr-x
			09/07/21 14:17:02	drwxrws---
			09/16/20 12:32:59	lrwxrwxrwx
link_to_root		D ory	01/22/20 14:10:58	lrwxrwxrwx
miniconda3		D ectory	04/01/20 10:08:53	drwxrwsr-x
nobackup		Directory	04/01/20 10:08:22	drwxrwsr-x
scratch		Directory	07/13/20 00:02:03	lrwxrwxrwx
src		Directory	08/02/21 13:16:58	drwxrwsr-x
test_nextflow		Directory	11/13/19 10:05:11	drwxrwsr-x
test_salloc_hisat		Directory	11/27/20 09:58:13	drwxrwsr-x
test_tmp		Directory	01/14/21 14:16:38	drwxrwsr-x

33 files and 42 directories. Total size: 216688 bytes

2. Drag and drop the files/folders you want to transfer

Server/Local file	Direction	Remote file	Size	Priority	Status

3. check transfer

4. File transfer from/to mass storage

- rsync (command line terminal):
  - instruction to transfer to/from the GIGA mass storage https://gitlab.uliege.be/giga-bioinfo/user-guides-wiki/-/wikis/faq/file_transfer
  - on Windows computer may require installation of a software such as
    - Ubuntu for Windows
    - MobaXterm (recommended)
  - a tiny bit more complicated but a lot more powerful and safe !!!! If a transfer is interrupted, rsync can resume it from where it stopped.
- Windows: "synchronize it"
  - https://www.grigsoft.com/wndsync.htm
  - https://en.wikipedia.org/wiki/Synchronize_It

SSH versus SAMBA

What's the difference between SAMBA and SSH connection ?

- **Via a SAMBA connection, any command you run is running on your computer**
  - if you work on a file on the mass storage from your computer via a SAMBA connection, the file will be transferred to your computer memory (via the network), processed, then the output is sent back to the mass storage => can be very inefficient (especially over VPN)
  - Best practice: transfer file(s) to your computer, works on it there and transfer back the output afterwards
- **Via a ssh connection, any command is run on the mass storage, using the mass storage CPUs/RAM**
  - will be a lot faster (ex: to move file from one folder to another one or to compress files)
  - it's a shared limited resource => don't perform analyses there. Run analysis either on your computer or on another server (ex: GIGA cluster)
- **Note:** if you transfer files between 2 remote servers (mass storage and CECI cluster for example) using FileZilla or "synchronize it", data will go through your computer and that can get very slow (especially over WIFI and/or VPN) => you should ssh to one of the server and use rsync (or an other tool depending of the type of server)

LIÈGE université
GIGA

How to get help ?

5. How to get help ?

- linux commands

Getting help

`$ man <command>`	display a command's manual page
`$ info <command>`	display a command's info entry
`$ whatis <command>`	display a brief description of a command

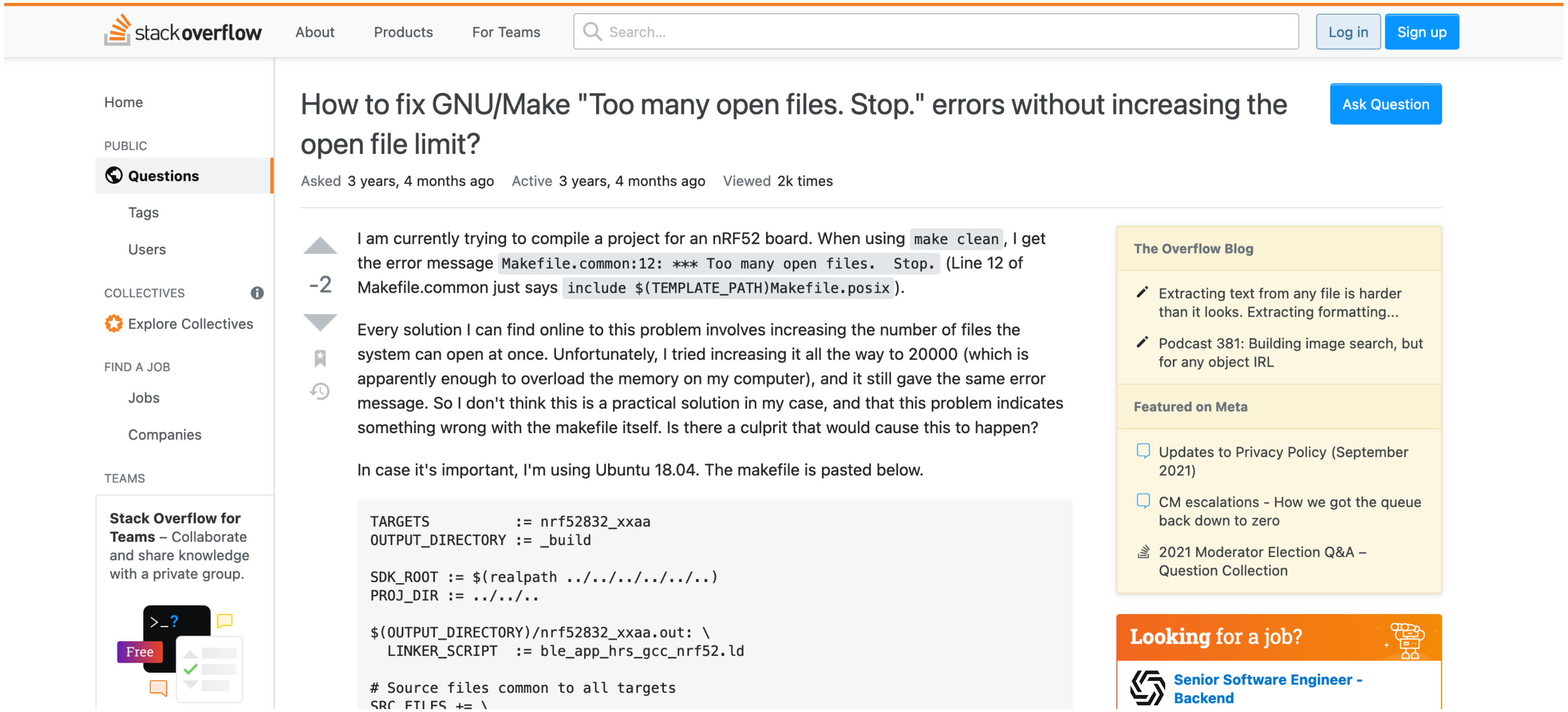
Helpful websites

Google

ex: "linux how to do xxxx" or "bash how to do xxxx"

Helpful websites

- Lots of websites with good explanations about commands, options, etc
- Forum to ask questions and understand error messages



How to fix GNU/Make "Too many open files. Stop." errors without increasing the open file limit?

Asked 3 years, 4 months ago Active 3 years, 4 months ago Viewed 2k times

I am currently trying to compile a project for an nRF52 board. When using `make clean`, I get the error message `Makefile.common:12: *** Too many open files.  Stop.` (Line 12 of Makefile.common just says `include $(TEMPLATE_PATH)Makefile.posix`).

Every solution I can find online to this problem involves increasing the number of files the system can open at once. Unfortunately, I tried increasing it all the way to 20000 (which is apparently enough to overload the memory on my computer), and it still gave the same error message. So I don't think this is a practical solution in my case, and that this problem indicates something wrong with the makefile itself. Is there a culprit that would cause this to happen?

In case it's important, I'm using Ubuntu 18.04. The makefile is pasted below.

```
TARGETS          := nrf52832_xxaa
OUTPUT_DIRECTORY := _build

SDK_ROOT := $(realpath ../../../../../..)
PROJ_DIR := ../../..

$(OUTPUT_DIRECTORY)/nrf52832_xxaa.out: \
  LINKER_SCRIPT  := ble_app_hrs_gcc_nrf52.ld

# Source files common to all targets
SRC_FILES += \
```

The Overflow Blog

- Extracting text from any file is harder than it looks. Extracting formatting...
- Podcast 381: Building image search, but for any object IRL

Featured on Meta

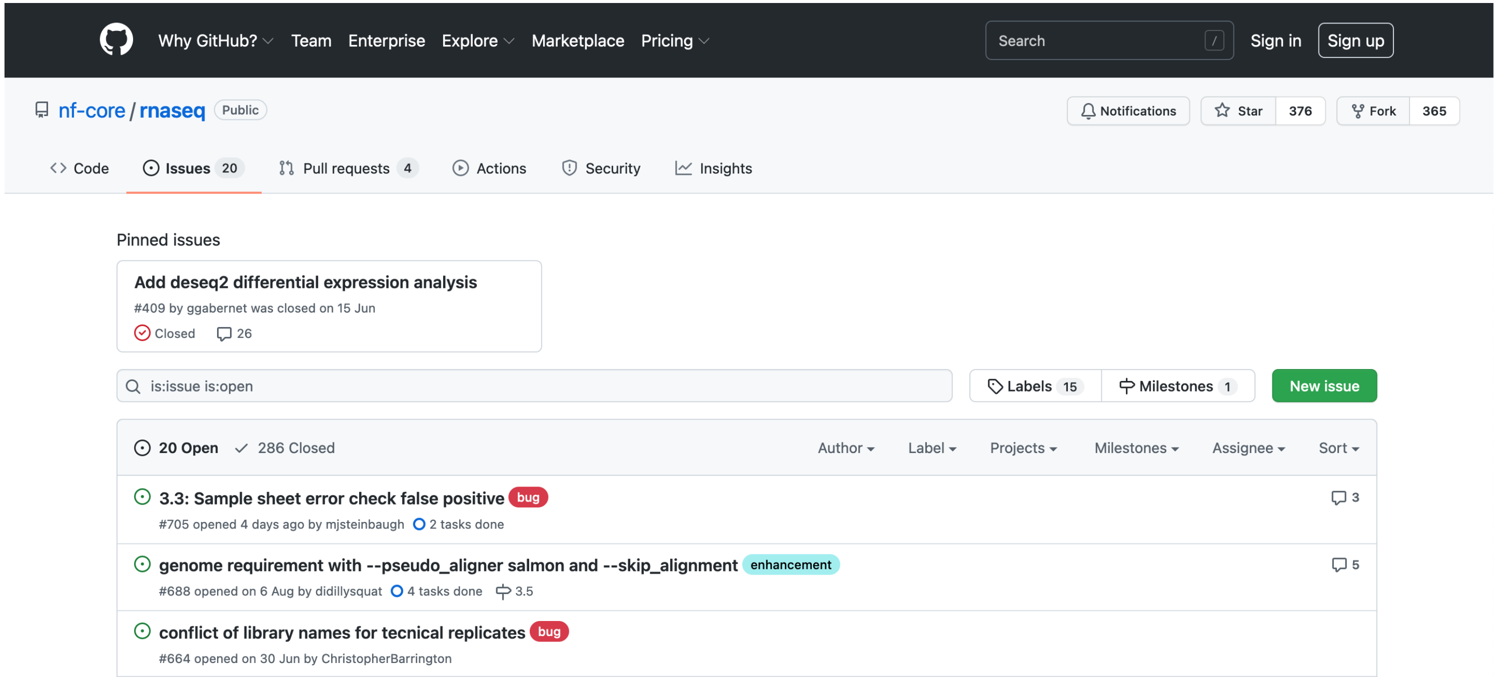
- Updates to Privacy Policy (September 2021)
- CM escalations - How we got the queue back down to zero
- 2021 Moderator Election Q&A – Question Collection

Looking for a job?

Senior Software Engineer - Backend

Helpful websites

- Lots of websites with good explanations about commands, options, etc
- Forum to ask questions and understand error messages
- If the program you use has a GitHub page



Pinned issues

Add deseq2 differential expression analysis
#409 by ggabernet was closed on 15 Jun
Closed 26

is:issue is:open

20 Open ✓ 286 Closed

3.3: Sample sheet error check false positive bug
#705 opened 4 days ago by mjsteinbaugh 2 tasks done 3

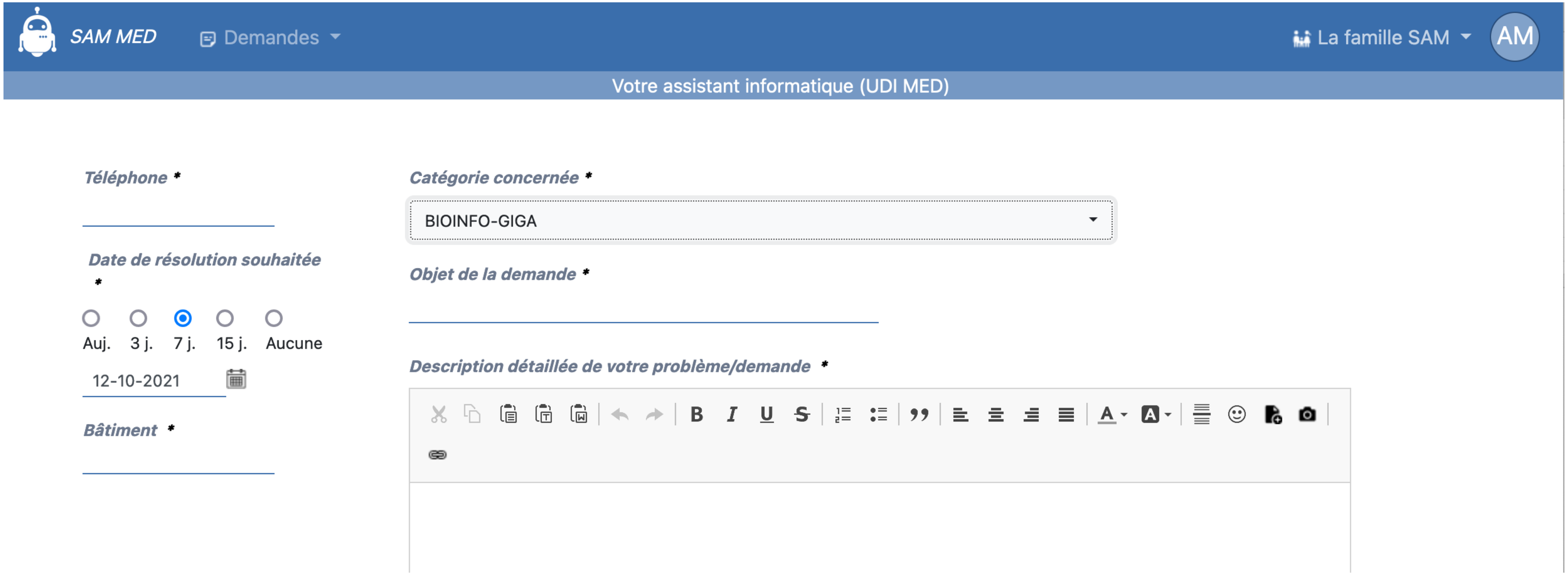
genome requirement with --pseudo_aligner salmon and --skip_alignment enhancement
#688 opened on 6 Aug by didillysquat 4 tasks done 3.5 5

conflict of library names for tecnical replicates bug
#664 opened on 30 Jun by ChristopherBarrington

GIGA bioinfo team

- When ?
  - If you didn't find an answer on internet
  - If you didn't understand the info
  - If you are not sure about what you understood
  - If you need help to implement the solution
- How ?
  - https://sam.med.uliege.be/
    - change "Catégorie concernée" to BIOINFO-GIGA (see next slide)
    - preferred route for solving problems (can be shared with UDI, SEGI, etc)
  - bioinfo.giga@uliege.be

Create a request in SAM



Téléphone *

Date de résolution souhaitée *

○ Auj. ○ 3 j. ● 7 j. ○ 15 j. ○ Aucune

12-10-2021

Bâtiment *

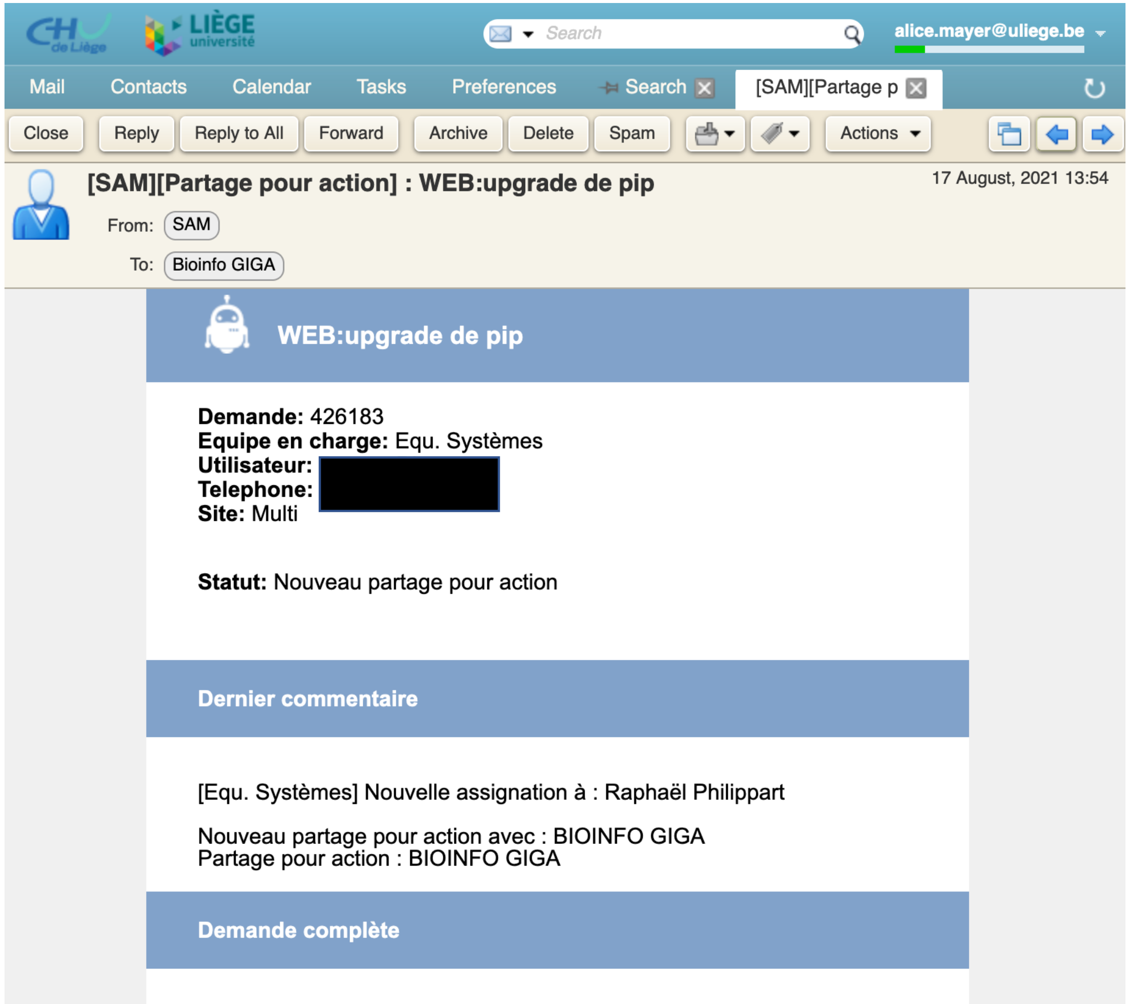
Catégorie concernée *

BIOINFO-GIGA

Objet de la demande *

Description détaillée de votre problème/demande *

You will receive answer by email



[SAM][Partage pour action] : WEB:upgrade de pip

17 August, 2021 13:54

From: SAM

To: Bioinfo GIGA

WEB:upgrade de pip

Demande: 426183
Equipe en charge: Equ. Systèmes
Utilisateur: [redacted]
Telephone:
Site: Multi

Statut: Nouveau partage pour action

Dernier commentaire

[Equ. Systèmes] Nouvelle assignation à : Raphaël Philippart

Nouveau partage pour action avec : BIOINFO GIGA
Partage pour action : BIOINFO GIGA

Demande complète

To reply, click on link **at the bottom of the email**

Cliquez ici pour accéder directement à la demande

Cliquez ici pour accéder à SAM

Ceci est un message automatique de SAM

Copyright © 2021 SEGI -Université de Liège

Important consideration

6. Important consideration

You can run basic commands and manage your data (move, compress, etc) while connected to the mass storage but **it's forbidden to run any analysis there !!!!!**

LIÈGE université
GIGA

Thank you for your attention !
Questions ?

GIGA bioinformatic team

bioinfo.giga@uliege.be

LIÈGE université
GIGA

Supplemental info

From Oxford Biomedical Data Science Training Programme (2018)

The Linux Shell

- Command line interface (CLI) – command interpreter
- Takes commands from the keyboard and gives them to the operating system to perform
- Most widely used is bash (Bourne Again Shell)
- Terminal emulator – program that lets you interact with the shell
- Environment variables
  - variables in the shell that can be accessed by any program run in the shell
  - Configure the path – places to look for executable programs
  - Can be user defined and modified
  - Can be accessed using $ notation

```
$ echo $PATH
```

From Oxford Biomedical Data Science Training Programme

Bash shortcuts

- CTRL-c Abort current command
- CTRL-z Pause current foreground process
- CTRL-l Clear the screen
- CTRL-a Go to start of line
- CTRL-e Go to end of line
- CTRL-u Cut from start of line
- CTRL-k Cut to end of line
- CTRL-r Search history
- CTRL-d Logout (also exit)
- Up arrow – access previous commands
- Tab – autocomplete (will prompt if ambiguous)

From Oxford Biomedical Data Science Training Programme

Linux Streams

- Streams are mechanism to move data from one place to another
- Standard streams
  - Standard in (stdin): the default place where commands listen for information Standard outout (stdout): the default place for output to go
  - Standard error (stderr): the default place for error output to go
- Unless redirected stdout and stderr both print to the terminal
- Pipes (|) connect the standard output of one command to the standard input of another

From Oxford Biomedical Data Science Training Programme

Redirecting input, output & error

`$ command1 < file1`	input file1 to command1
`$ command1 < (command2)`	output of command 2 as file input to command1
`$ command1 > file1`	standard output of command1 to file1
`$ command1 >> file1`	append standard output of command1 to file1
`$ command1 2> file1`	error output of command1 to file1
`$ command1 1>&2`	standard output to same place as standard error
`$ command1 2>&1`	standard error to same place as standard output
`$ command1 > /dev/null`	discard standard output of command1

From Oxford Biomedical Data Science Training Programme

Combining commands

`$ command1; command2`	run cmd1 then cmd2
`$ command1 && command2`	run cmd2 if cmd1 is successful
`$ command1 || command2`	run cmd2 if cmd1 is not successful
`$ command1 | command2`	Pipe stdout from cmd1 to stdin cmd2
`$ command1 |& command2`	Pipe stderr from cmd1 to stdin cmd2

From Oxford Biomedical Data Science Training Programme

Loops

```
$ for i in {1..5}; do COMMAND; done
$ for (i=1;i<=10;i+=2); do COMMAND; done
$ for i in *.txt; do COMMAND; done
$ for x in *.bed; do cat $x | wc -l; done
```

Glob for files

Use filename as variable

From Oxford Biomedical Data Science Training Programme

Managing Linux Processes

- Any Linux command creates a process
- Every process has a process ID (PID)
- Processes can run in the foreground or background
- The priority index of a process is called Nice
  - It can vary between 20 to -19 (default 0)
  - The lower the ‘Nice’ index the higher the priority

```
$ bg
```
To run a stopped process in background

```
$ fg
```
To bring a background process to the foreground

```
$ ps
```
Give the status of processes running for a user

```
$ ps <PID>
```
Gives the status of a particular process

```
$ nice
```
Starts a process with a given priority

```
$ renice
```
Changes priority of an already running process

```
$ kill <PID>
```
Kills a process

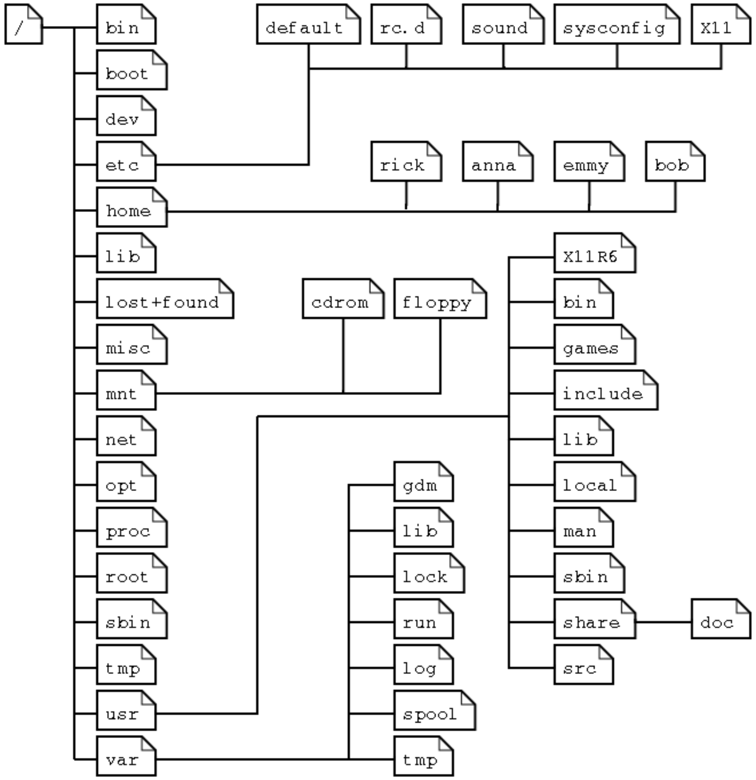
```
$ killall <name>
```
Kills all processes by name

From Oxford Biomedical Data Science Training Programme

The Linux file system

- In Linux everything is a file (or a process)
- A directory is just a file containing names of other files
- Tree structure
  - Finder (OSX)
  - Windows Explorer

Redhat file system layout



From Oxford Biomedical Data Science Training Programme

Navigating the file structure

`$ pwd` print working directory

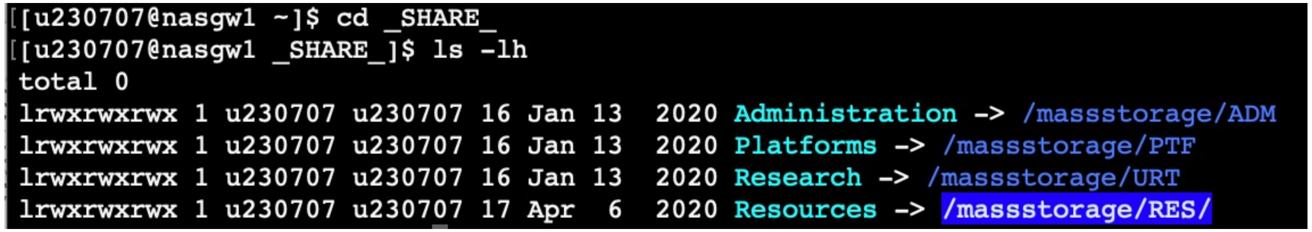
```
[u230707@nasgw1 ~]$ pwd
/stornext/snfs1/SHARE/HOME/u230707
```

`$ ls` list directory contents

`_SHARE_`

`$ cd _SHARE_` change directory

`$ ls`

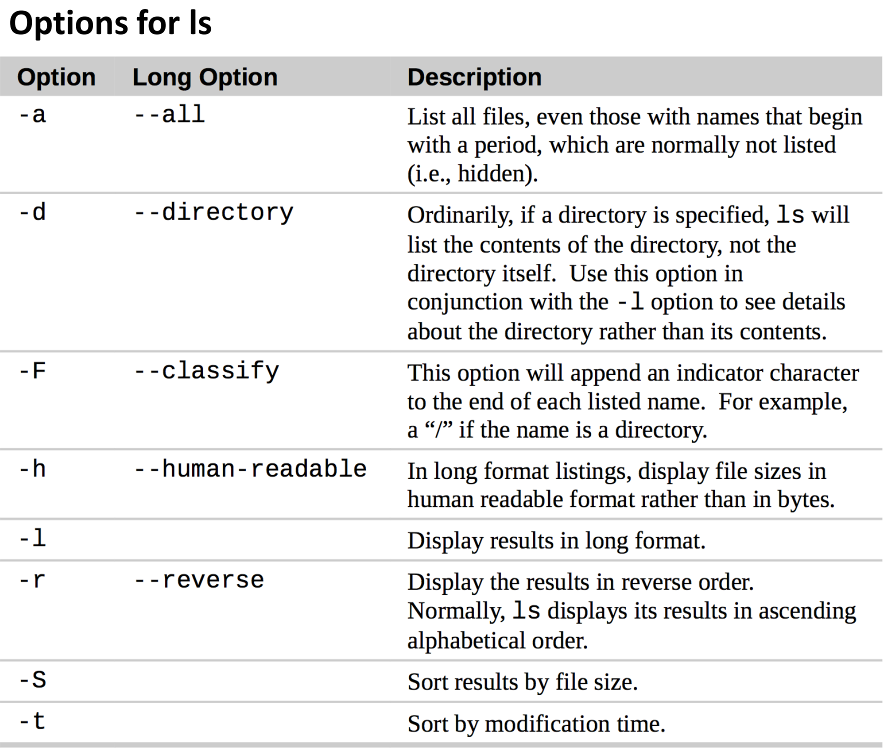


```
[u230707@nasgw1 ~]$ cd _SHARE_
[u230707@nasgw1 _SHARE_]$ ls -lh
total 0
lrwxrwxrwx 1 u230707 u230707 16 Jan 13  2020 Administration -> /massstorage/ADM
lrwxrwxrwx 1 u230707 u230707 16 Jan 13  2020 Platforms -> /massstorage/PTF
lrwxrwxrwx 1 u230707 u230707 16 Jan 13  2020 Research -> /massstorage/URT
lrwxrwxrwx 1 u230707 u230707 17 Apr  6  2020 Resources -> /massstorage/RES/
```

Adapted from Oxford Biomedical Data Science Training Programme

Changing how you view files

Alias: ll = 'ls -l'



Options for ls

Option	Long Option	Description
`-a`	`--all`	List all files, even those with names that begin with a period, which are normally not listed (i.e., hidden).
`-d`	`--directory`	Ordinarily, if a directory is specified, `ls` will list the contents of the directory, not the directory itself. Use this option in conjunction with the `-l` option to see details about the directory rather than its contents.
`-F`	`--classify`	This option will append an indicator character to the end of each listed name. For example, a "/" if the name is a directory.
`-h`	`--human-readable`	In long format listings, display file sizes in human readable format rather than in bytes.
`-l`		Display results in long format.
`-r`	`--reverse`	Display the results in reverse order. Normally, `ls` displays its results in ascending alphabetical order.
`-S`		Sort results by file size.
`-t`		Sort by modification time.

Adapted from Oxford Biomedical Data Science Training Programme

Viewing text files

`$ file <filename>`	print a brief description of the file's content
`$ cat <filename>`	print the file contents to standard out
`$ head -n 20 <filename>`	print the **first** n lines to standard out
`$ tail -n 20 <filename>`	print the **last** n lines to standard out
`$ tail –f <filename>`	Follow changes to the end of the file on standard out
`$ more <filename>`	Display output in the terminal one page at a time
`$ less <filename>`	Like more but allows backward movement
`$ vim <filename>`	Open file in vim terminal text editor
`$ emacs <filename>`	Open file in emacs terminal text editor
`$ nano <filename>`	Open file in nano terminal text editor

From Oxford Biomedical Data Science Training Programme

Creating files and directories

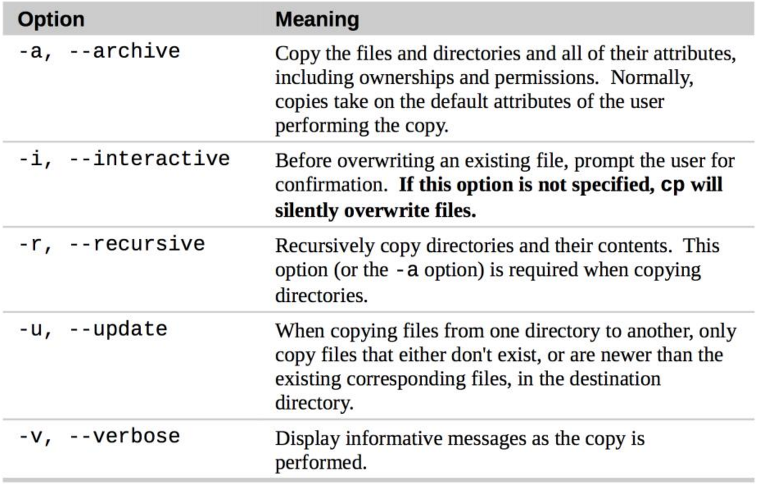
`$ touch <filename>`	Create new empty file (can create several at a time)
`$ vim <filename>`	Open new file in vim editor
`$ emacs <filename>`	Open new file in emacs editor
`$ nano <filename>`	Open new file in nano editor
`$ mkdir <dir1>`	Create new directory
`$ mkdir <dir1> <dir2>`	Create multiple new directories

From Oxford Biomedical Data Science Training Programme

Copying, moving and renaming files

`$ cp file1 file2`	copy file to new file
`$ cp file1 dir1/`	copy file to new directory
`$ mv file1 file2`	Rename file
`$ mv file1 dir1/`	move file to new directory

cp options



Option	Meaning
`-a, --archive`	Copy the files and directories and all of their attributes, including ownerships and permissions. Normally, copies take on the default attributes of the user performing the copy.
`-i, --interactive`	Before overwriting an existing file, prompt the user for confirmation. **If this option is not specified, cp will silently overwrite files.**
`-r, --recursive`	Recursively copy directories and their contents. This option (or the `-a` option) is required when copying directories.
`-u, --update`	When copying files from one directory to another, only copy files that either don't exist, or are newer than the existing corresponding files, in the destination directory.
`-v, --verbose`	Display informative messages as the copy is performed.

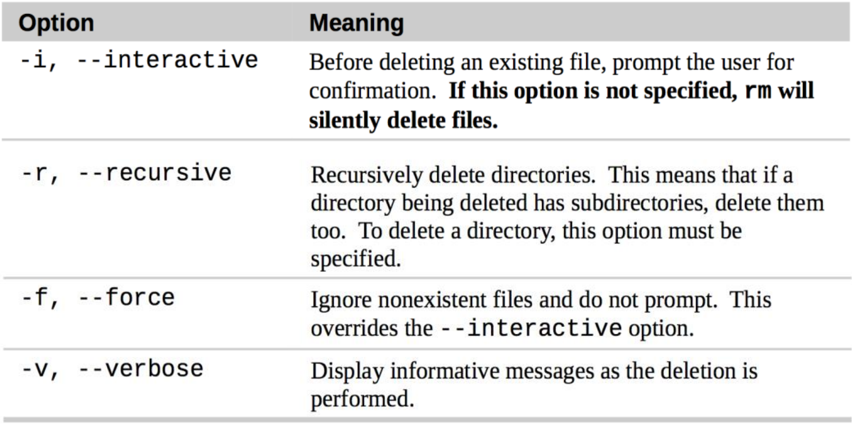
`$ rename 's/old/new/' <files>`	Rename multiple files using pattern matching
`$ rename 's/perl/pl/' *.perl`	

From Oxford Biomedical Data Science Training Programme

Deleting files

`$ rm file1`	remove file
`$ rm *.perl`	remove multiple files
`$ rm -r dir1`	remove directory & contents
`$ rmdir dir1`	remove empty directory

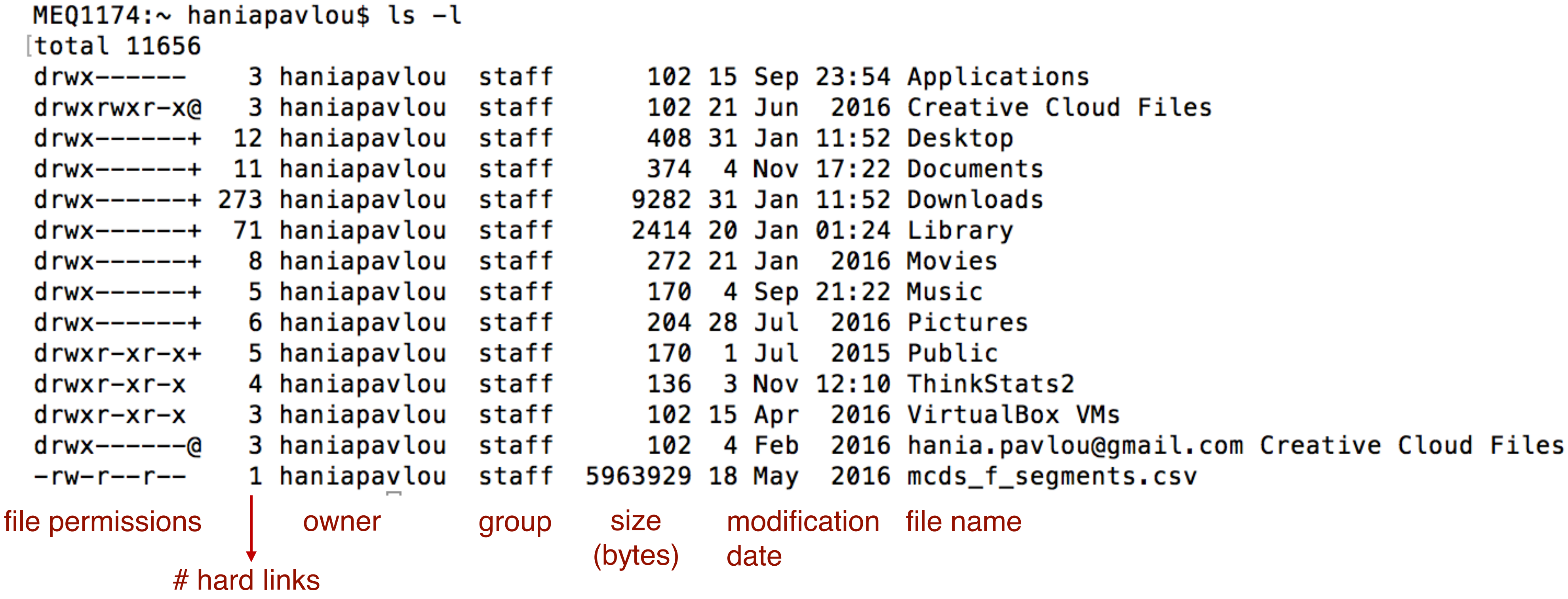
rm options



Option	Meaning
`-i, --interactive`	Before deleting an existing file, prompt the user for confirmation. **If this option is not specified, `rm` will silently delete files.**
`-r, --recursive`	Recursively delete directories. This means that if a directory being deleted has subdirectories, delete them too. To delete a directory, this option must be specified.
`-f, --force`	Ignore nonexistent files and do not prompt. This overrides the `--interactive` option.
`-v, --verbose`	Display informative messages as the deletion is performed.

From Oxford Biomedical Data Science Training Programme

File Properties



```
MEQ1174:~ haniapavlou$ ls -l
[total 11656
drwx------     3 haniapavlou  staff      102 15 Sep 23:54 Applications
drwxrwxr-x@    3 haniapavlou  staff      102 21 Jun  2016 Creative Cloud Files
drwx------+   12 haniapavlou  staff      408 31 Jan 11:52 Desktop
drwx------+   11 haniapavlou  staff      374  4 Nov 17:22 Documents
drwx------+  273 haniapavlou  staff     9282 31 Jan 11:52 Downloads
drwx------+   71 haniapavlou  staff     2414 20 Jan 01:24 Library
drwx------+    8 haniapavlou  staff      272 21 Jan  2016 Movies
drwx------+    5 haniapavlou  staff      170  4 Sep 21:22 Music
drwx------+    6 haniapavlou  staff      204 28 Jul  2016 Pictures
drwxr-xr-x+    5 haniapavlou  staff      170  1 Jul  2015 Public
drwxr-xr-x     4 haniapavlou  staff      136  3 Nov 12:10 ThinkStats2
drwxr-xr-x     3 haniapavlou  staff      102 15 Apr  2016 VirtualBox VMs
drwx------@    3 haniapavlou  staff      102  4 Feb  2016 hania.pavlou@gmail.com Creative Cloud Files
-rw-r--r--     1 haniapavlou  staff  5963929 18 May  2016 mcds_f_segments.csv
```

From Oxford Biomedical Data Science Training Programme

File Permissions - Owner

- Linux is a multi-user operating system
- Access to read, write and execute each file / directory is controlled
- Each file is owned by exactly one user
- You can see the different users on your system by typing:

```
$ cat /etc/passwd
```

- The first field on each line is the name of a unique user

From Oxford Biomedical Data Science Training Programme

File Permissions - Groups

- A group is a collection of one or more users
- Each user can be a member of multiple groups
- A file can be owned by exactly one group
- To see the groups that your user currently belongs to, type:

```
$ groups
```

- To show all of the groups currently available on your system, type:

```
$ cat /etc/group
```

- The last category that you can assign permissions for is the "other" category.
- ‘Other’ = any user that is not the file owner and is not a member of the group that owns the file
- This allows you to set permissions that will apply to anyone outside of the two control groups

From Oxford Biomedical Data Science Training Programme

Types of Permissions

- Each permissions category (owner, group and other) can be assigned permissions that restrict their ability to read, write, or execute a file
- For a regular file
  - read permissions are required to read the contents of a file
  - write permissions are necessary to modify it
  - execute permissions are needed to run the file as a script or an application
- For directories:
  - read permissions are necessary to list the contents of a directory
  - write permissions are required to modify the contents of a directory
  - execute permissions allow a user to change directories into the directory
- Linux represents these types of permissions using two separate symbolic notations:
  - Alphabetic
  - Octal

From Oxford Biomedical Data Science Training Programme

Alphabetic notation

```
$ ls -l
```

```
drwxr-xr-x 3 root root     4096 Apr 26  2012 dir1
-rw-r----- 1 root root     2981 Apr 26  2012 file1
```

- = regular file
l = link
d = directory

user group other

r = read permissions
w = write permissions
x = execute permissions
- = no access

From Oxford Biomedical Data Science Training Programme