Intro to Linux – Part 1

June 19, 2020

BPHL-SEbioinformatics@flhealth.gov

Outline

Linux

File structure

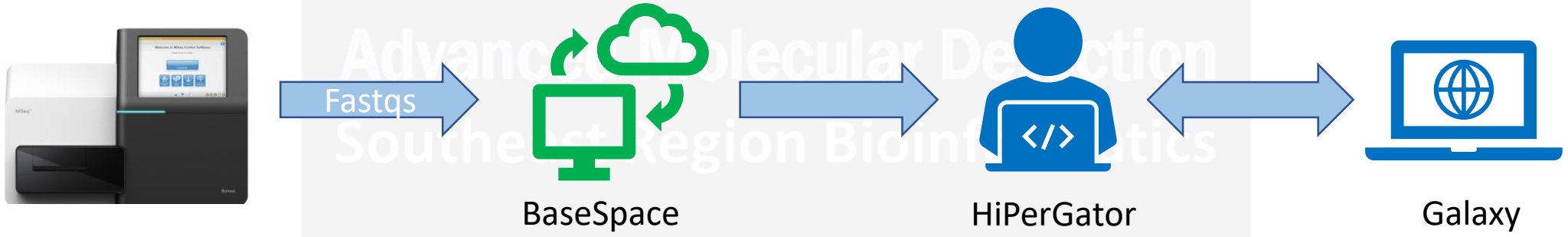
Logging into HiPerGator

Navigating HiPerGator directories

Tip and tricks

Why learn Linux?

- Most bioinformatics tools/pipelines run via command-line
  - Execute scripts in HiPerGator
- Download fastq files from NCBI's Sequence Read Archive (SRA)
- Transfer data directly from BaseSpace to Galaxy



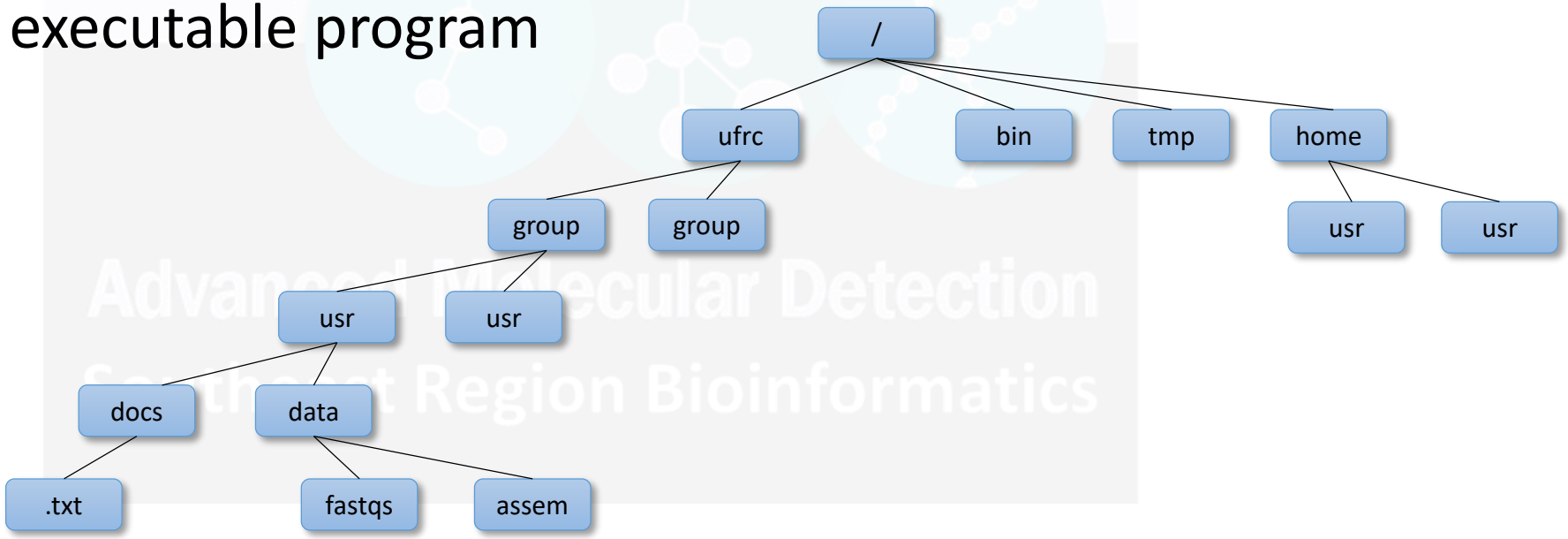
Advanced Molecular Detection
Southeast Region Bioinformatics
BPHL-SEbioinformatics@flhealth.gov

Linux

- Operating system
  - Unix-based
  - Macs are unix-based
- Why is it important in bioinformatics?
  - Open-source tools
  - Many servers or HPCs (high-performance computing clusters) are Linux-based
- Can work in the command-line
  - Terminal
  - Bash shell language

File Structure

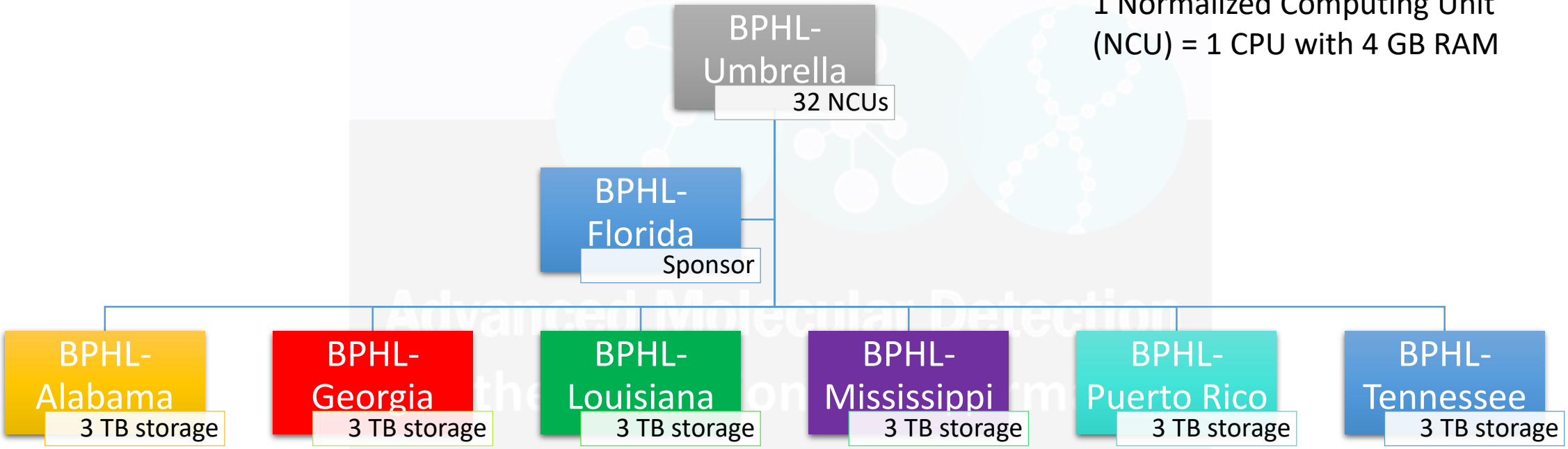
- Everything in Linux is either a file or a process
- A directory is a “file”
- A process = executable program



*Note: /ufrc will change to /blue in the coming weeks

HiPerGator – State groups

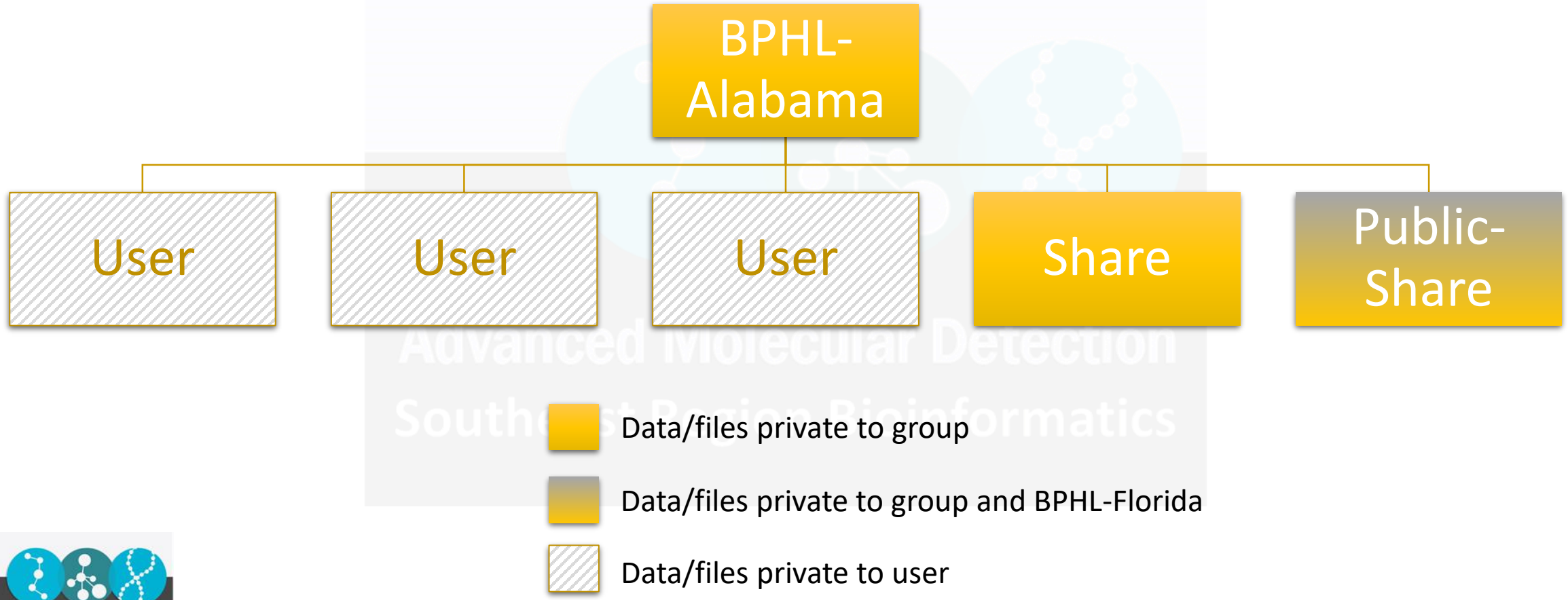
1 Normalized Computing Unit (NCU) = 1 CPU with 4 GB RAM



Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

HiPerGator – State groups

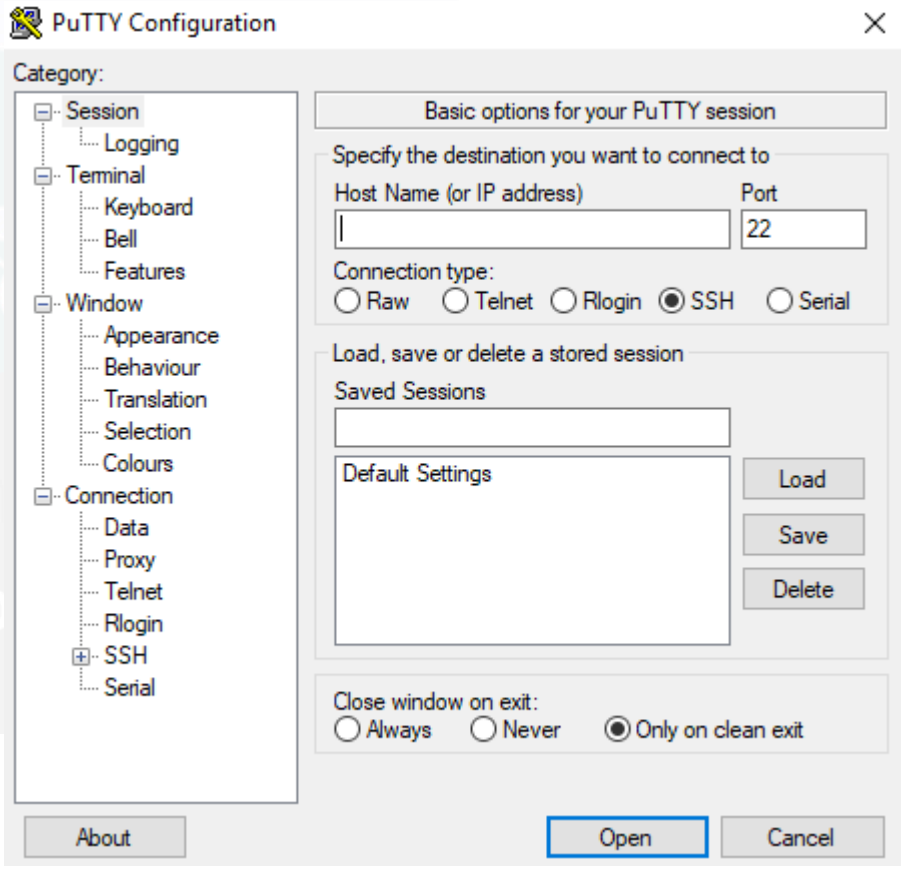


BPHL-SEbioinformatics@flhealth.gov

Logging into HiPerGator

- From Windows
  - Use Putty

Double-click

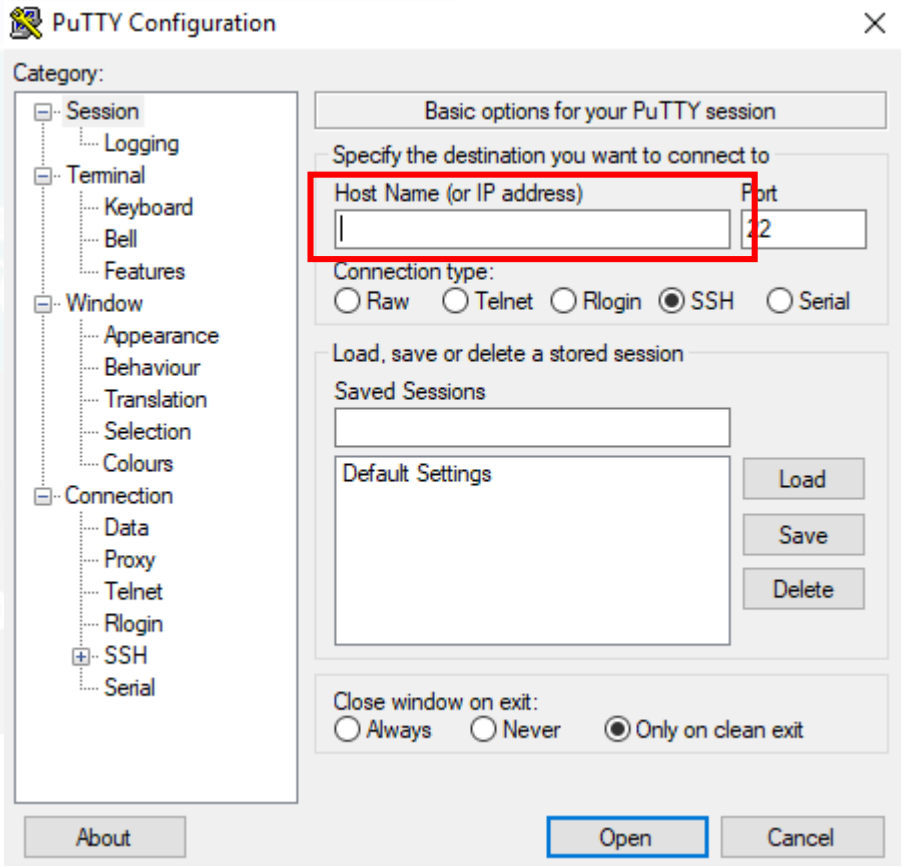


Advanced Molecular Detection
Southeast Region Bioinformatics

BPHL-SEbioinformatics@flhealth.gov

Logging into HiPerGator

- Host Name: hpg.rc.ufl.edu
- Click “Open” or press “Enter”

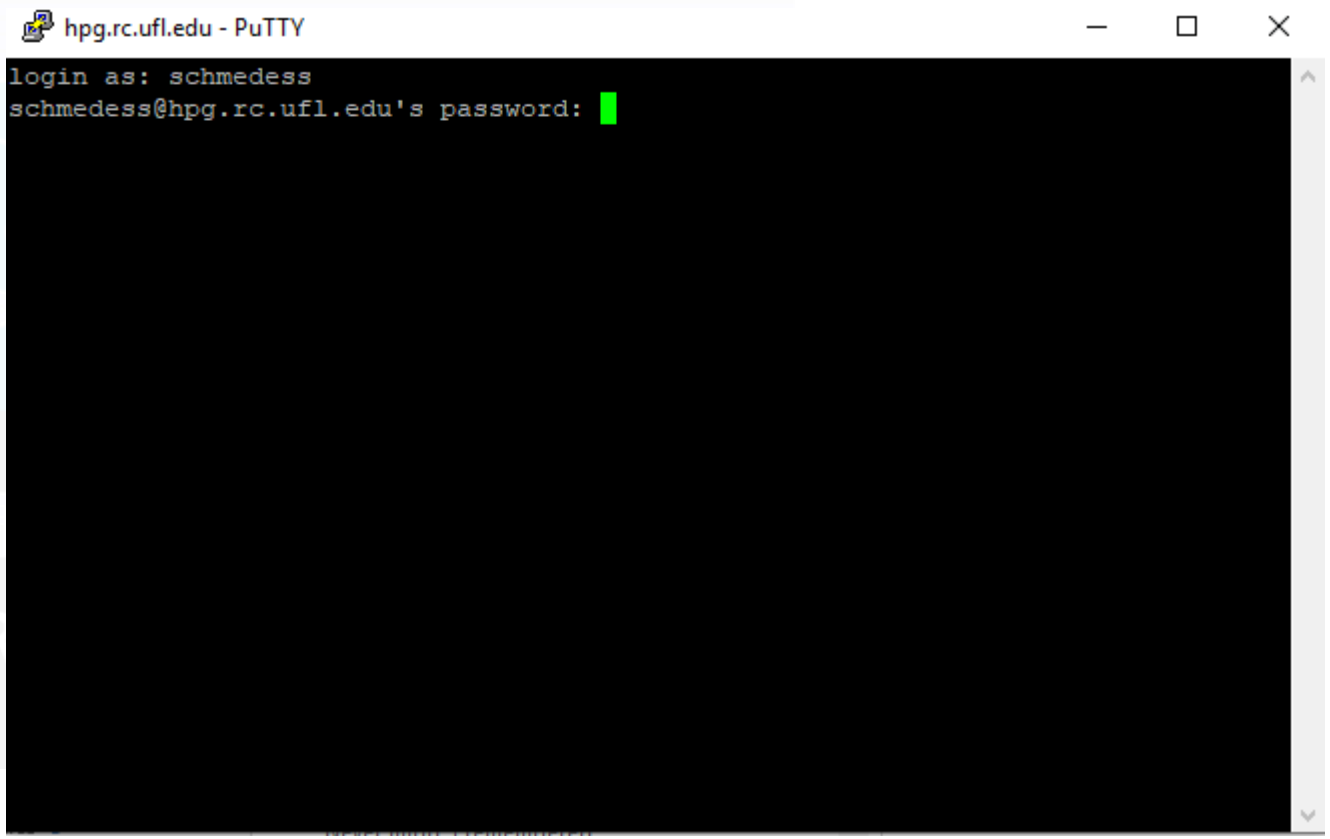


Logging into HiPerGator

- Type your GatorLink ID/HiPerGator Username
- Press “Enter”

Logging into HiPerGator

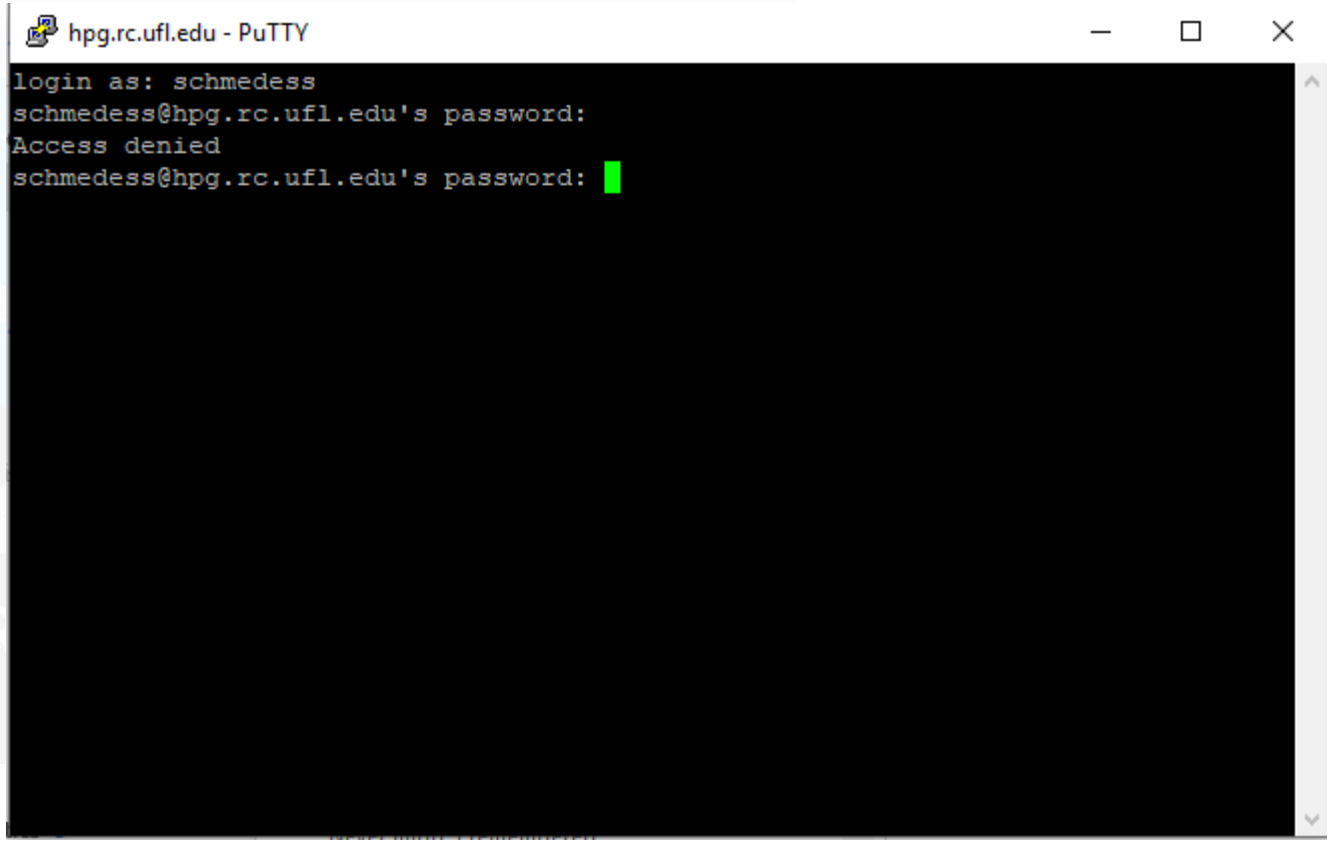
- Type your GatorLink ID/HiPerGator password
- Press “Enter”
- *Note: The password will not display as you type*



Advanced Molecular Detection
Southeast Region Bioinformatics

Logging into HiPerGator

- If you mistype your password, this screen will display
- Simply, re-type your password and try again



Advanced Molecular Detection
Southeast Region Bioinformatics

Logging into HiPerGator

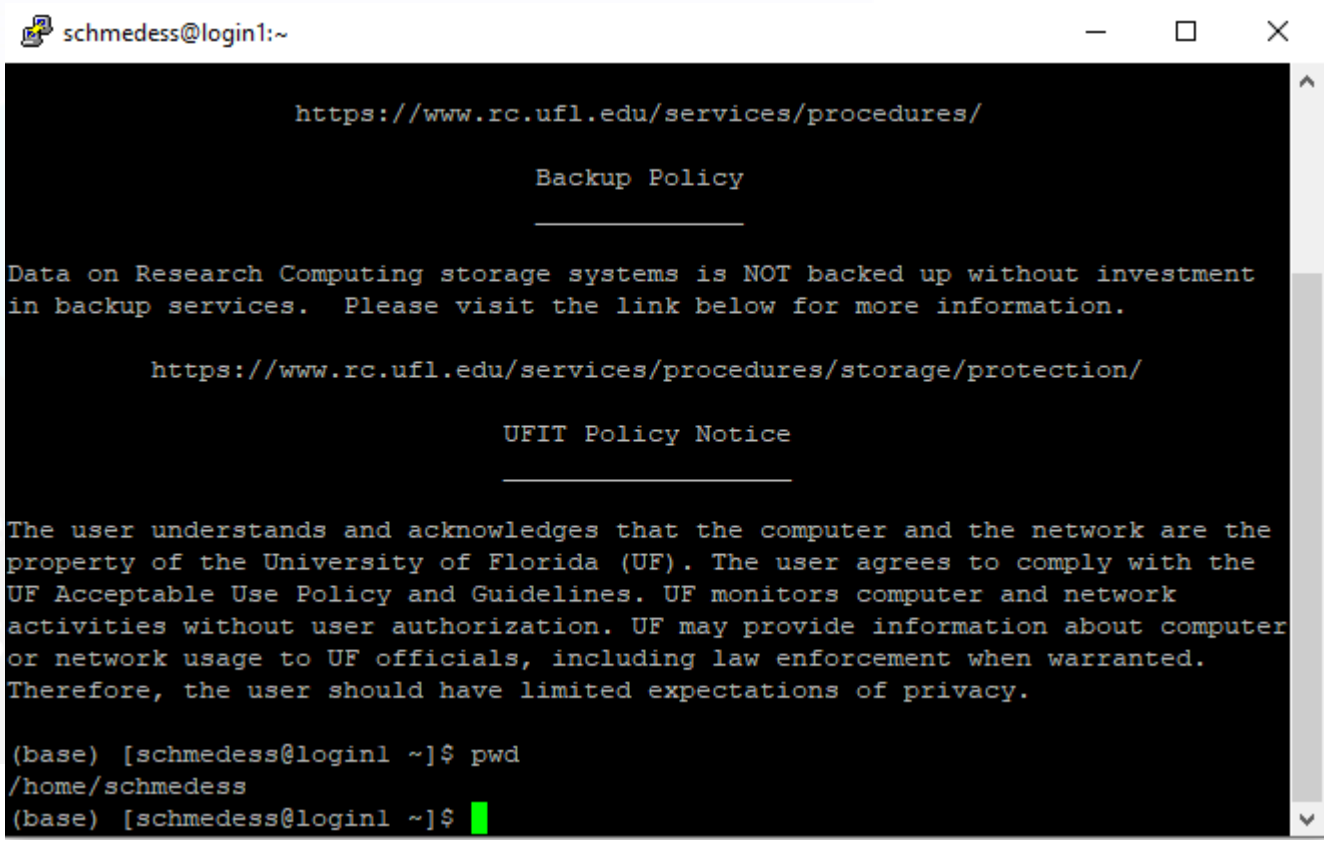
- You are logged in!

Navigating the file structure

- Where are we?

pwd

(print working directory)



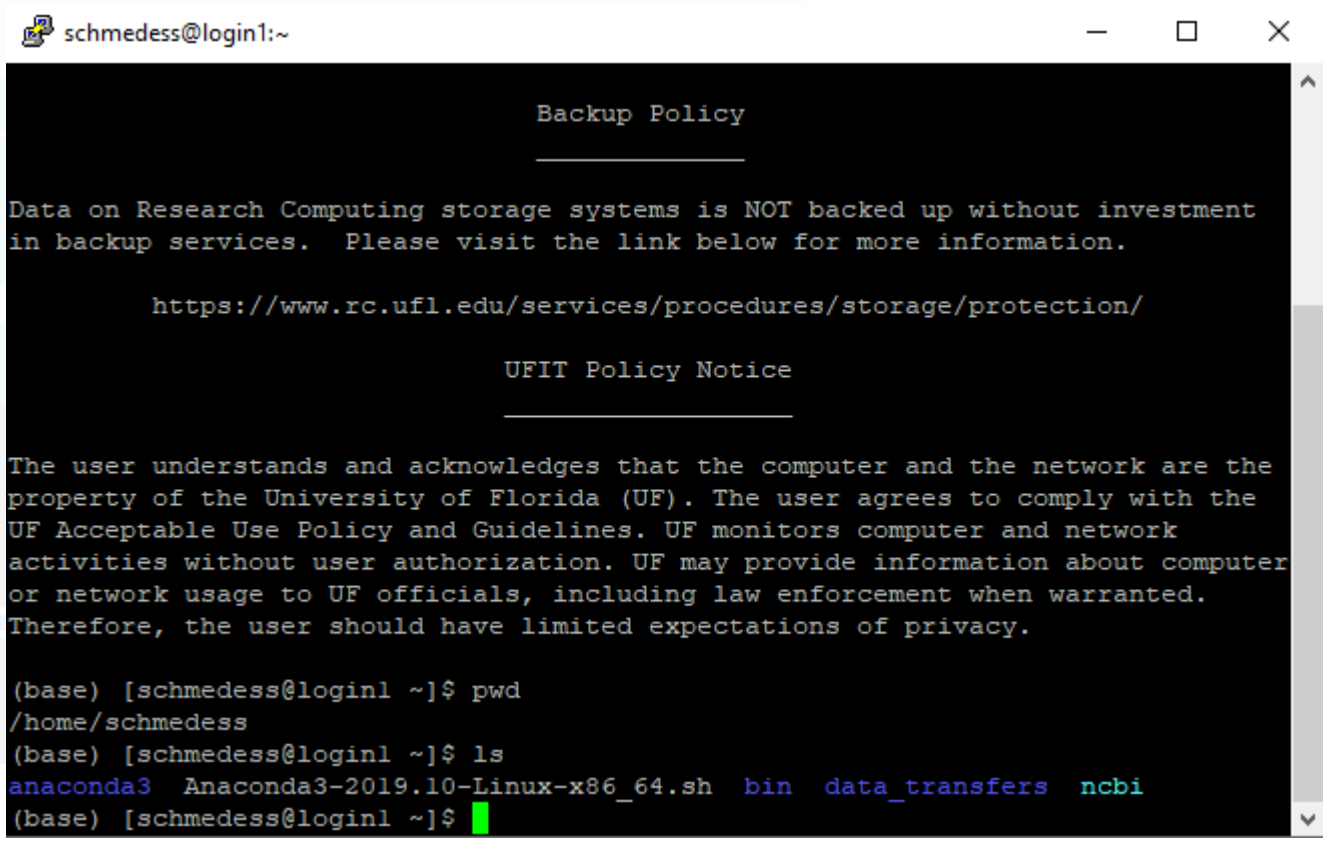
Advanced Molecular Detection
Southeast Region Bioinformatics

Navigating the file structure

- What files are in my current working directory?

ls

(list)



Advanced Molecular Detection
Southeast Region Bioinformatics

Navigating the file structure

- Other information about my files

ls -althr --color=auto

d (directory)

- (file)

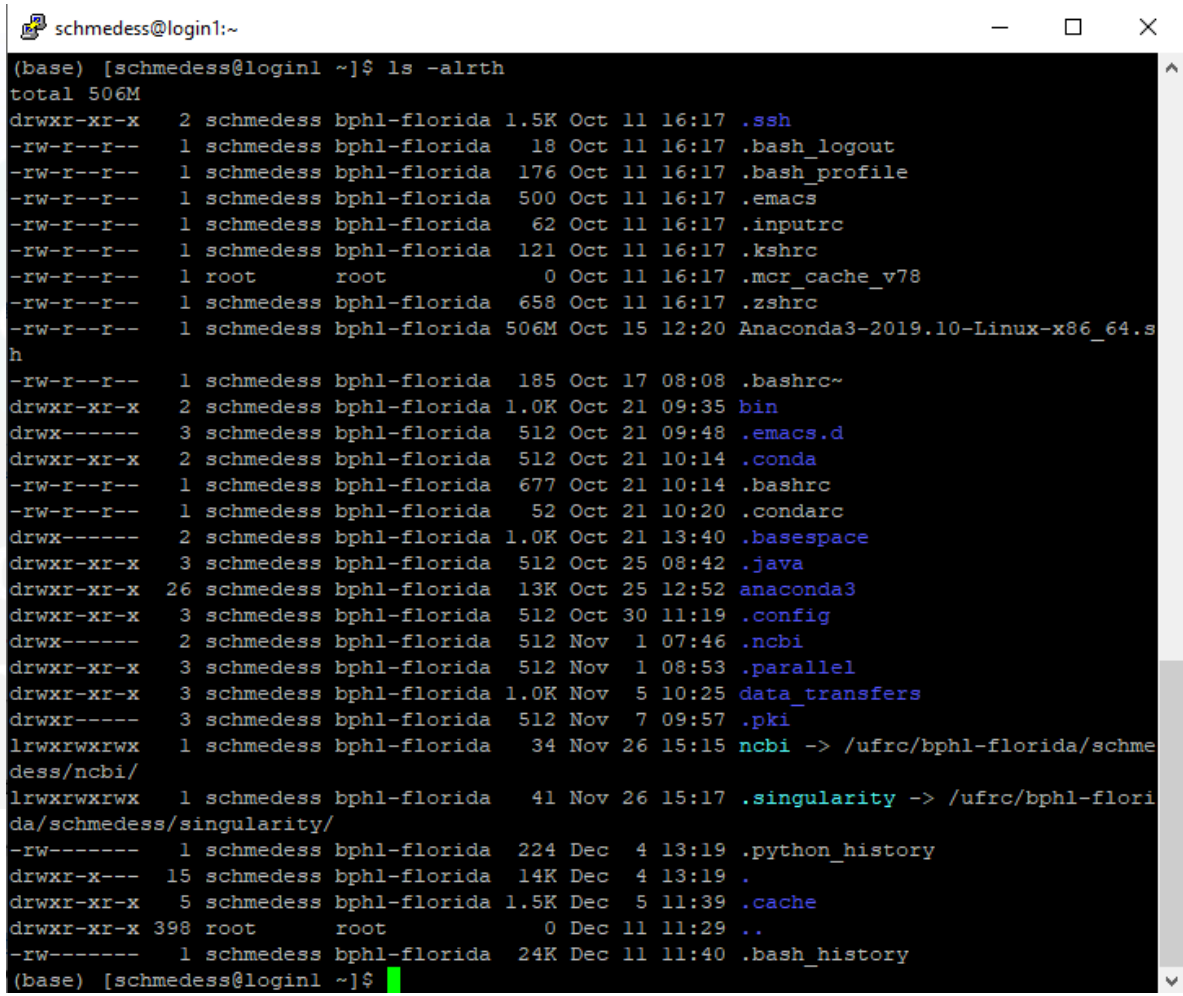
l (link)

. (current directory)

.. (parent directory)

.filename (hidden file)

-rw-r--r-- (file permissions)

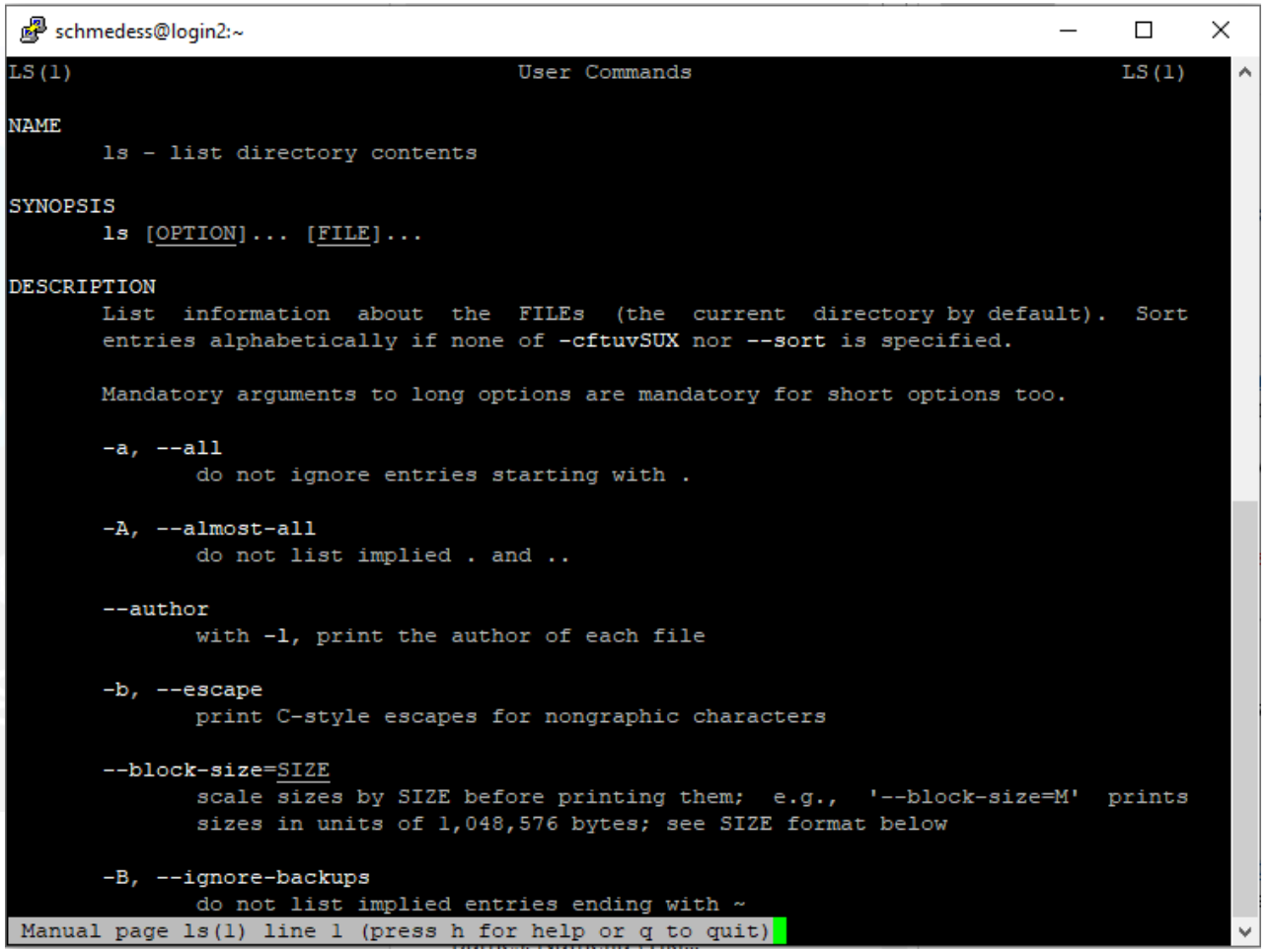


Advanced Molecular Detection
Southeast Region Bioinformatics

What does –althr mean?

man ls

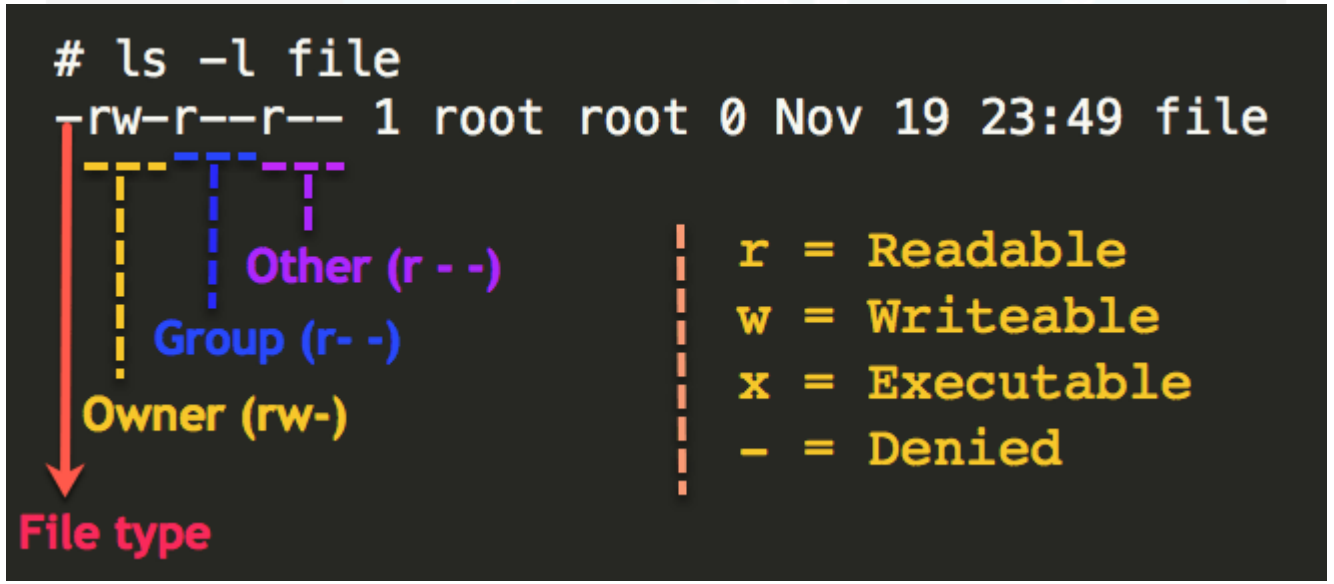
- Flags/options
- Command man ls
- A = all (including hidden files)
- H= “human readable”, prints sizes in 1K, 1 g, 1, M format)
- L = long listing format
- T = show time
- R= reverse order while sorting, most recent at bottom



Advanced Molecular Detection
Southeast Region Bioinformatics

File Permissions

```
drwxr-xr-x   2 schmedess bphl-florida 1.0K Oct 21 09:35 bin
drwx------   3 schmedess bphl-florida  512 Oct 21 09:48 .emacs.d
drwxr-xr-x   2 schmedess bphl-florida  512 Oct 21 10:14 .conda
-rw-r--r--   1 schmedess bphl-florida  677 Oct 21 10:14 .bashrc
-rw-r--r--   1 schmedess bphl-florida   52 Oct 21 10:20 .condarc
```



https://www.thegeekdiary.com/understanding-basic-file-permissions-and-ownership-in-linux/

Advanced Molecular Detection
Southeast Region Bioinformatics

File Permissions

chmod <permissions> <filename>

(change file mode)

Example:

chmod 770 report.txt

-rwxrwx--- report.txt

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

Navigating the file structure

- cd <directory name>
  - change directory
- cd /absolute/path/to/directory/ or relative/path/
- cd ~ or cd
  - Go to your home directory
- cd .. or cd ../
  - Go up one directory in the tree (to parent directory)
- cd ../…/
  - Go up two directories in the file tree (“grandparent directory”)

Change from home to /ufrc

- cd /ufrc/bphl-<state>/<user>
- /home/<user>/
  - Very limited storage – **do NOT store your data here!!!!!**
  - Not high performance – **do NOT run jobs from here!!!!!**
  - cd will take you home
- /ufrc/bphl-<state>/<user>
  - This is where you do the bulk of your work.
  - This is where your data input and output from “jobs” go.
  - cd /ufrc/bphl-<state>/<user> will take you to your user directory in your group

*Note: /ufrc will change to /blue in the coming weeks

More bash commands

- mkdir <name of directory>
  - (make a new directory)
- cp <path/to/file> <path/to/new/location>
  - (copy a file to a new location)
  - cp /path/to/file .
    - (copy file to current working directory)
  - Example:
    - cp sample_*.fastq.gz data/
      - (copy read 1 and read 2 fastq files for sample at the same time to data/)

More bash commands

- touch <filename>
  - (Create new/empty file)
- rm <filename>
  - (remove file)
- rm –r <directory name>
  - (remove directory and all contents)

Looking at files

- cat <file> or <file> <file>
  - “concatenate and print”
  - zcat (gz files)
- more <file>
  - Use Enter to scroll through file
  - q to quit
- less <file>
  - Use up and down arrow to scroll through file
  - q to quit

Looking at files

- head <file>
  - Print first 10 lines
  - head -n <number of lines other than 10>
- tail <file>
  - Print last 10 lines
  - tail -n <number of lines other than 10>

Looking at files

wc – “word count”

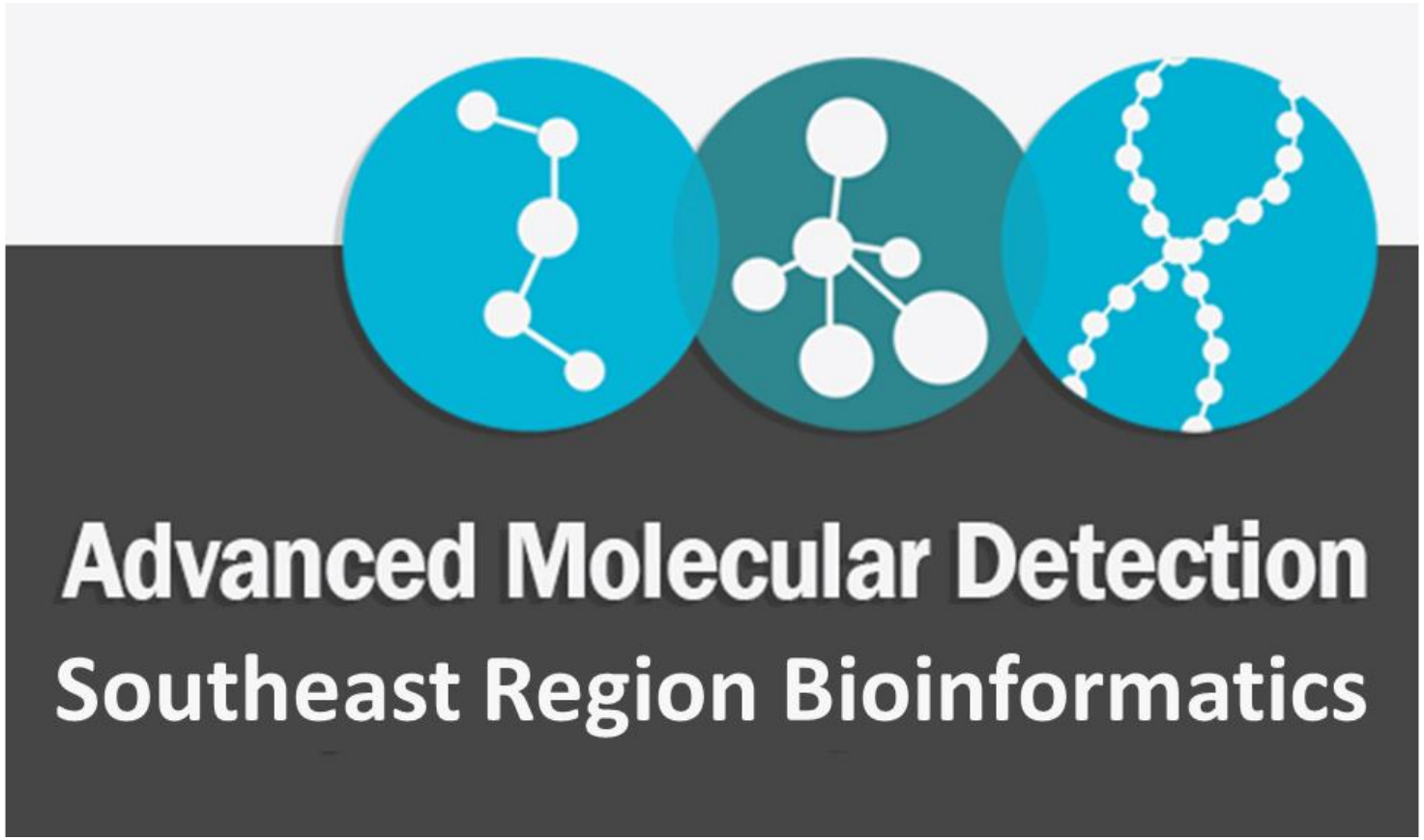
- wc –l <file>
  - Line count
- wc –w <file>
  - Word count
- wc –m <file>
  - Character count

Tricks and Tips

- *wild card
- Tab = auto-complete
- Up arrow = scroll through previous commands
- Control-A = moves cursor to beginning of command prompt line
- Control-E = moves cursor to end of command prompt line
- Alt-F = moves cursor one word forward at a time in command line
- Alt-B = moves cursor one word back at a time in command line
- Control-C = kills a process
- clear = clears terminal

Future Webinars

- Working with the BaseSpace Command Line Interface (CLI)
- Working in Galaxy – including data transfers from HiPerGator
- SARS-CoV-2 Sequencing



Questions???

BPHL-SEbioinformatics@flhealth.gov

Sarah Schmedes, PhD
BRR Lead, Southeast Region

Taj Azarian, PhD
Bioinformatician
Assistant Professor, UCF

Jason Blanton, PhD
Molecular Administrator
State Sequencing Coordinator