HPRC Short Course

Introduction to UNIX

DIVISION OF RESEARCH
TEXAS A&M UNIVERSITY

For More Help...

Website:	hprc.tamu.edu
Email:	help@hprc.tamu.edu
Telephone:	(979) 845-0219
Visit us in person:	Henderson Hall, Room 114A

Help us, help you -- we need more info

- Which Cluster
- UserID/NetID
- Job id(s) if any
- Location of your jobfile, input/output files
- Application used if any
- Module(s) loaded if any
- Error messages
- Steps you have taken, so we can reproduce the problem

Logging in to the system

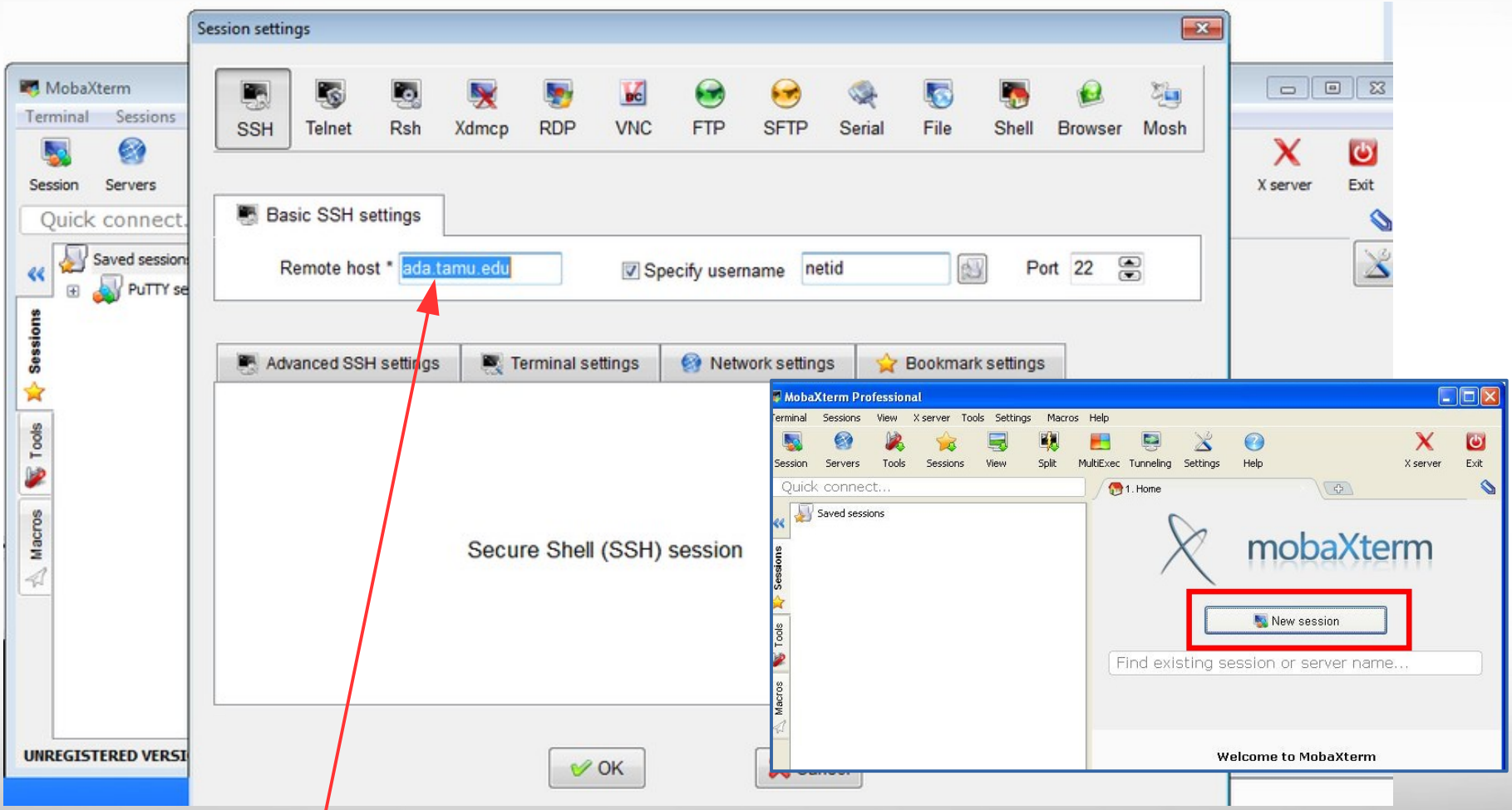
- SSH (secure shell)
  - The only program allowed for remote access; encrypted communication; freely available for Linux/Unix and Mac OS X hosts;
- For Microsoft Windows PCs, use *MobaXterm*
    - **https://hprc.tamu.edu/wiki/HPRC:MobaXterm**
      - You are able to view images and use GUI applications with MobaXterm
  - or *Putty*
    - **https://hprc.tamu.edu/wiki/HPRC:Access#Using_PuTTY**
      - You can not view images or use GUI applications with PuTTY

Your Login Password

- Both state of Texas law and TAMU regulations prohibit the sharing and/or illegal use of computer passwords and accounts;
- Don’t write down passwords;
- Don’t choose easy to guess/crack passwords;
- Change passwords frequently

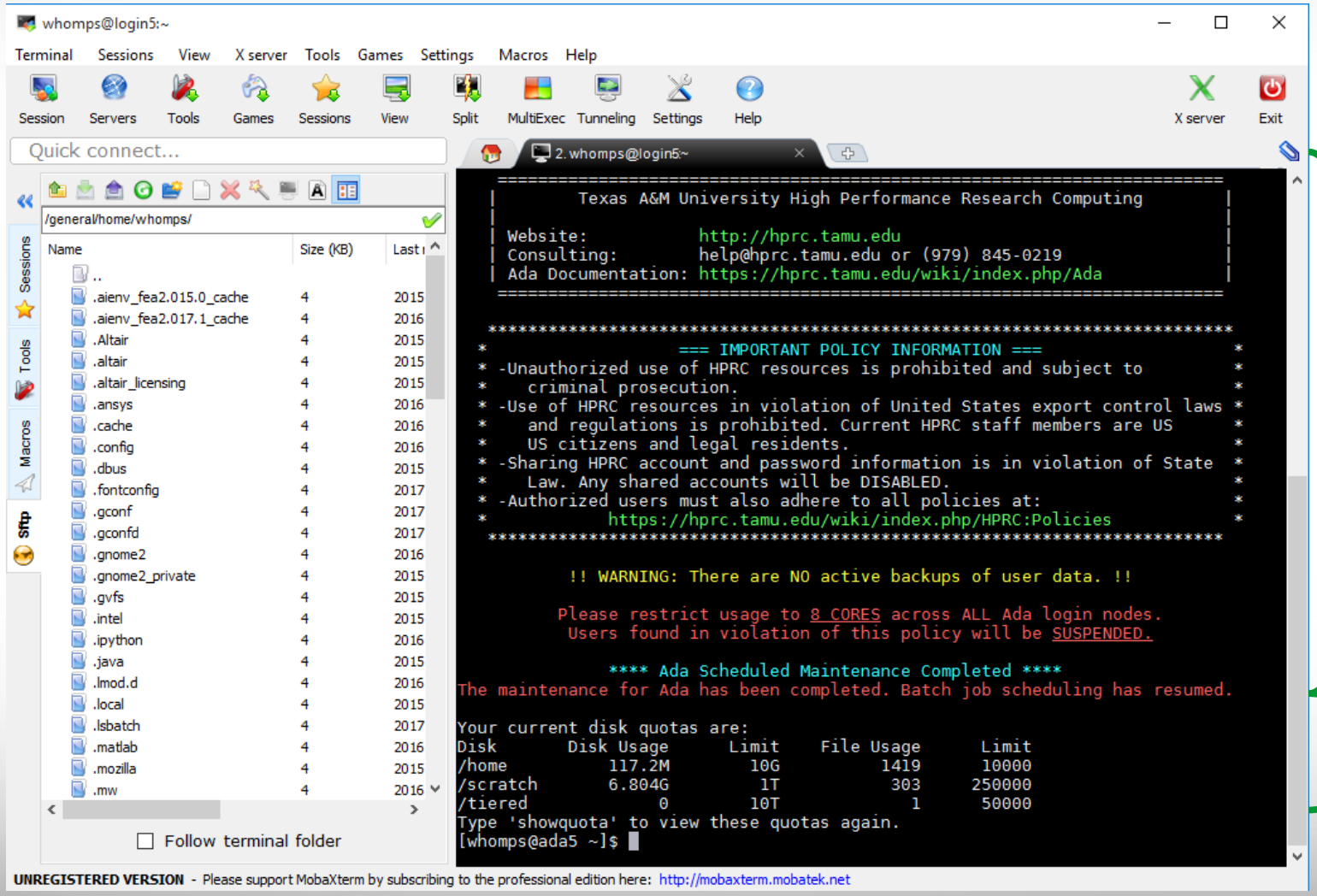
Using SSH - MobaXterm (on Windows)

https://hprc.tamu.edu/wiki/HPRC:MobaXterm



Use **ada.tamu.edu** as Remote host name.

Using SSH - MobaXterm (on Windows)



message of the day

your quotas

Using SSH (on a Linux/Unix Client)

https://hprc.tamu.edu/wiki/Ada:Access

You may see something like the following the first time you connect to the remote machine from your local machine:

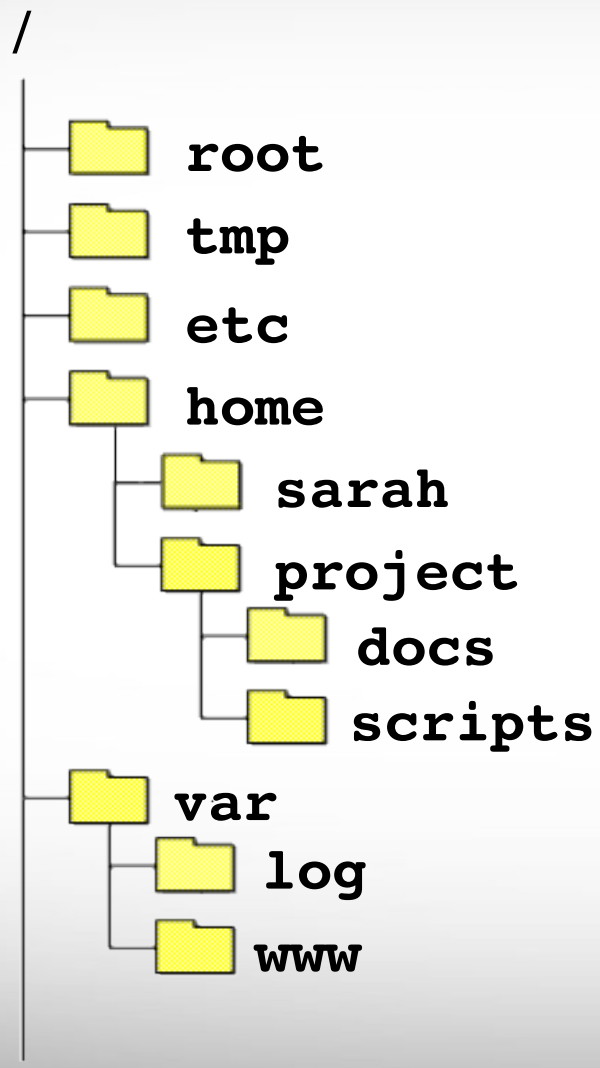
```
ssh user_NetID@ada.tamu.edu
```

```
Host key not found from the list of known hosts.
Are you sure you want to continue connecting (yes/no)?
```

Type yes, hit enter and you will then see the following:

```
Host 'ada.tamu.edu' added to the list of known hosts.
user_NetID@ada.tamu.edu's password:
```

Finding your way around the UNIX directory structure




```
/
/root
/tmp
/etc
/home
/home/sarah
/home/project
/home/project/docs
/home/project/scripts
/var
/var/log
/var/www
```

Where Am I?

- **pwd** command (print current/working directory)

UNIX commands in green

```
pwd
```

output in blue

```
/home/user_NetID
```

UNIX Commands Have Options

Leave a space between the command and the options

double dash means there is a single option which is usually a descriptive word

```
ls --all
```

`--all` show all files including hidden files which begin with `.`

single dash means each character is an option

```
ls -a -l
```

`-a` show all files including hidden
`-l` show file details

most options can be combined behind one dash

```
ls -al
```

`-a` show all files including hidden
`-l` show file details

`.`	current working directory
`..`	back one directory

Search for UNIX Commands Options

Search the manual page for the UNIX command ***ls***

```
man ls
```

Page up	Move up one page
Page down	Move down one page
Spacebar	Move down one page
Mouse scroll wheel	Move up and down

`/all`	search the man page for the text 'all'
n	search forward for next found match
N	search backwards next found match
g	go to first line
G	Go to last line
q	quit

File and directory names

Commonly used:	
A-Z	
a-z	
0-9	
.	
-	dash
_	underscore

Avoid using:	
	spaces
()	parenthesis
“ '	quotes
?	Question mark
$	Dollar sign
*	Asterisk
\	back slash
/	forward slash
:	colon

- Avoid spaces.
- File and directory names are case sensitive
- Avoid spaces in the file name ("*my data file*" vs "*my_data_file.txt*").

Common Directory Commands

- ***mkdir*** command to make a new directory:

```
mkdir directory_name
```

- ***cd*** to change to another directory:

```
cd directory_name
```

- ***rmdir*** to remove an empty directory:

```
rmdir directory_name
```

Changing Directories: the *cd* cmd

- To switch to the parent directory of the current directory:

```
cd ..
```

- Return to home directory

```
cd
cd ~
cd ~/
```

See directory structure

```
tree
```

Exercise:

```
cd
mkdir temp
mkdir temp/hg19
cd temp
pwd
cd hg19
pwd
```

```
cd ../..
pwd
cd -
cd ..
pwd
cd
pwd
tree
```

return to previous pwd

Absolute vs. Relative Pathname

```
/
/root
/tmp
/etc
/home
/home/sarah
/home/sarah/project
/home/sarah/docs/README
/var
/var/log
/var/www
```

If you are in the project directory

```
pwd
```

```
/home/sarah/project
```

The relative path to the README file is `../docs/README`

```
ls ../docs/README
```

The absolute path to the README file `/home/sarah/docs/README`

```
ls /home/sarah/docs/README
```

gedit text editor

From the UNIX terminal command line enter this command to start gedit and edit a file called my_favorite_foods.txt

```
gedit my_favorite_foods.txt &
```

the **&** will detach gedit from the terminal so you can continue to use the terminal

Naming files:

use all lower case characters
separate words with an underscore
make the file name very descriptive of what is in the file even if its rather long
common file extensions:

.pl	a Perl script
.py	a Python script
.gz	a file that has been compressed (zipped) to reduce file size (.zip)
.txt	a generic text file
.tsv	tab separated values (columns are separated by a tab)
.csv	comma separated values
.jar	a Java ARchive file
.png	image file in png format
.tar.gz	compressed tar file (sometimes .tgz)

Use the tab key to complete a file name

Type the first few characters of the file name

```
ls -l my
```

then hit the tab key to auto complete the file name

```
ls -l my_favorite_foods.txt
```

then hit enter to see the command results which will show the file size in human readable format

If the tab key did not complete the file name then either the file does not exist or there are two or more files that begin with the same characters in which case you need to hit tab twice then type a few more characters and hit tab again to complete.

Count lines in a file

```
wc my_favorite_foods.txt
```

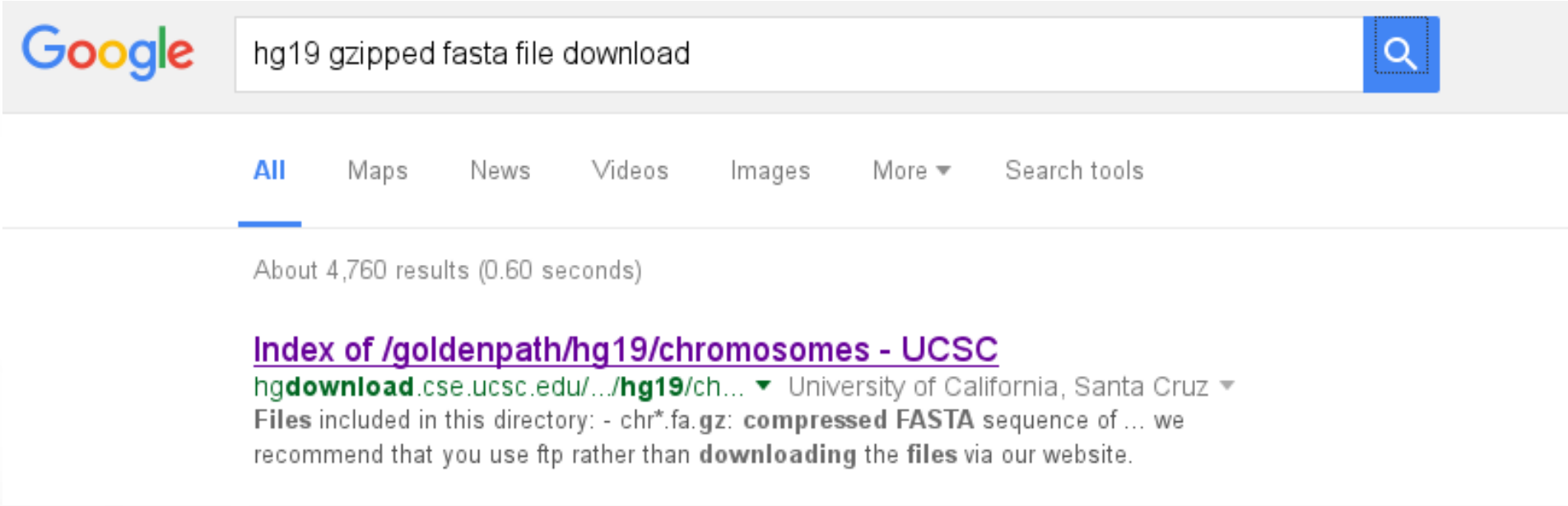
```
 29 109 876 my_favorite_foods.txt
```

What does the output mean?

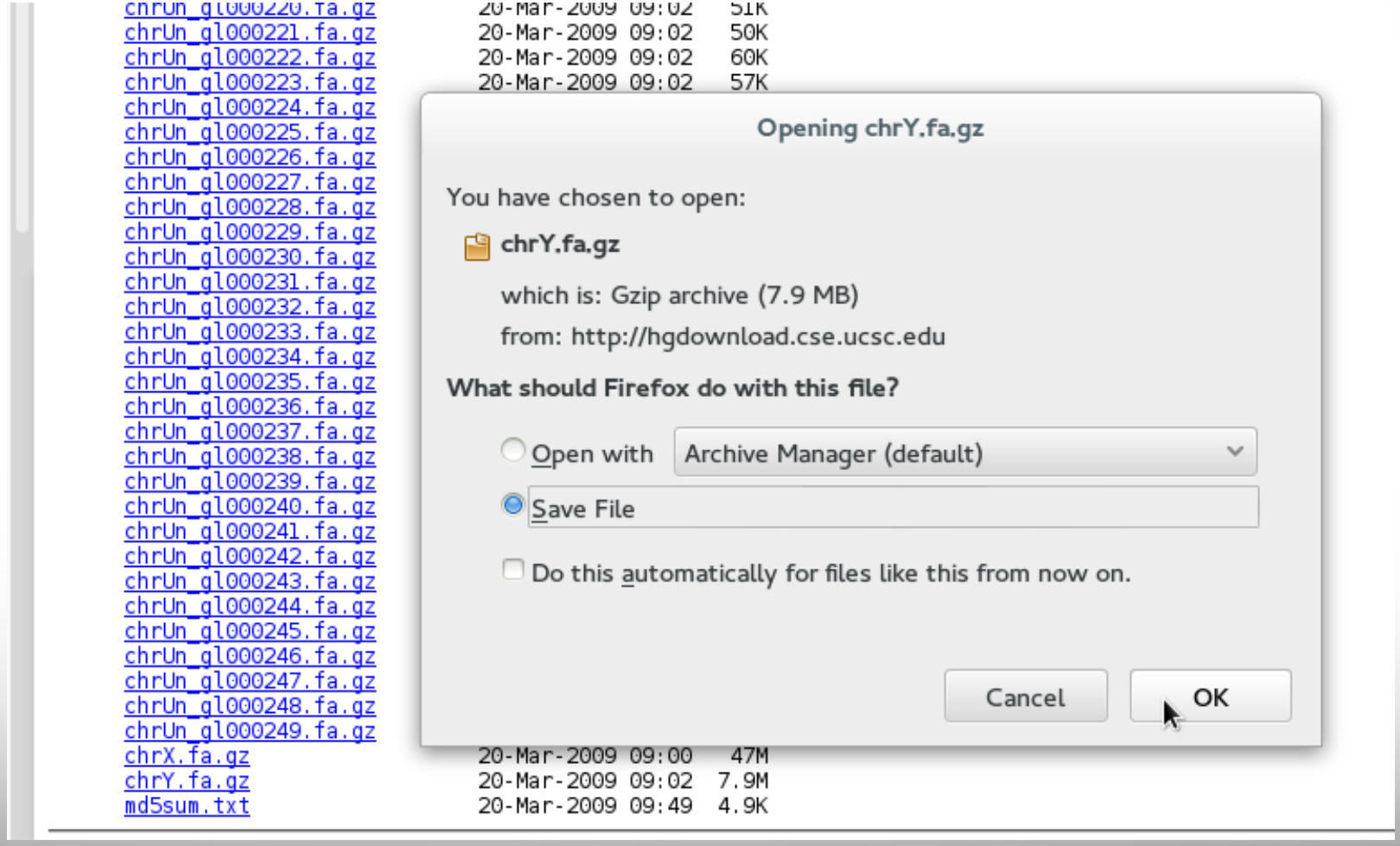
Use the man page for wc to find out.

How do you just print the newline counts?

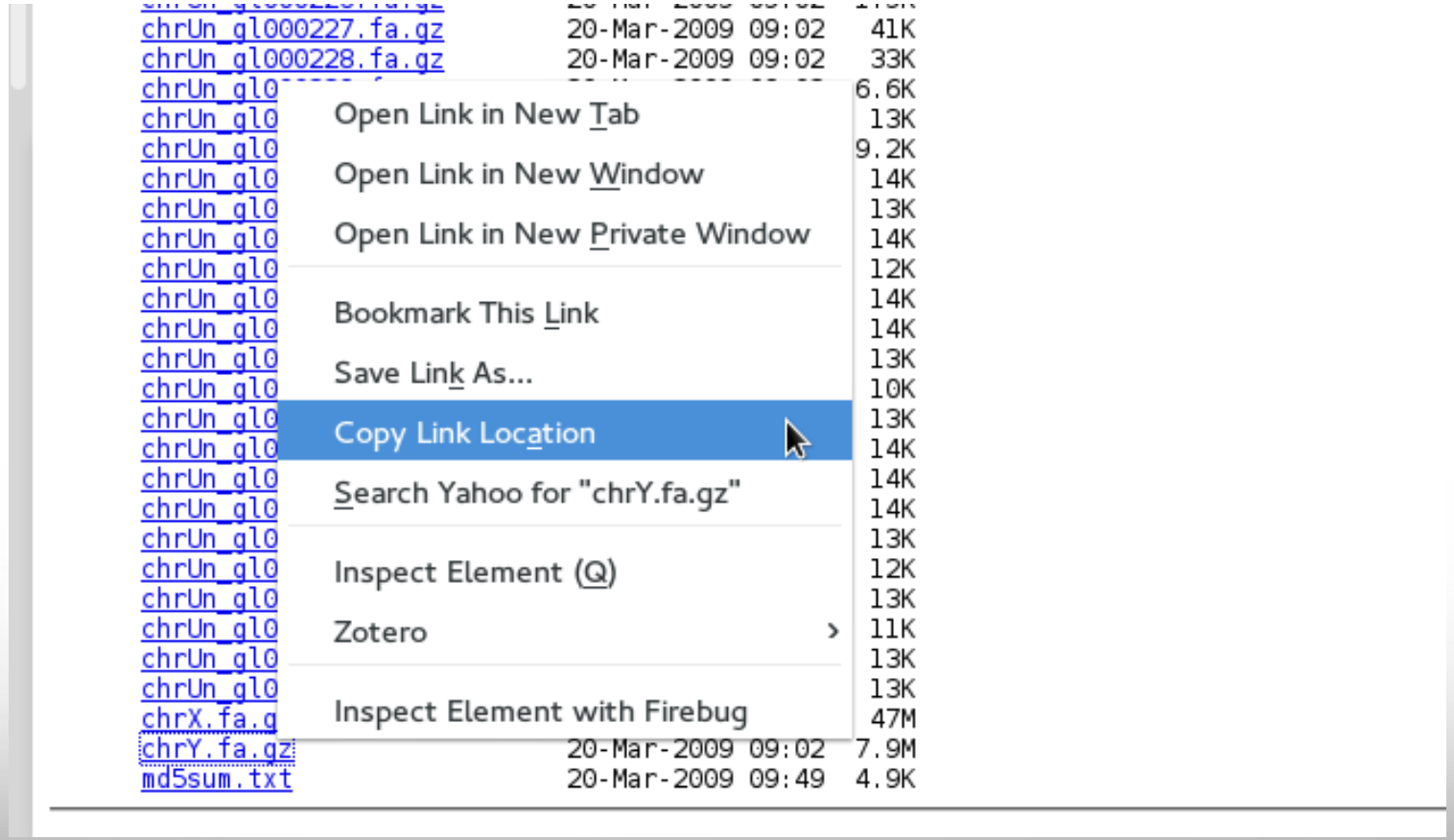
Download a File from the Web to your pwd



Don't Left Click and download a file to your desktop



Right Click and Copy the URL so you can download it directly to your UNIX computer



Copy a file directly to your UNIX directory

Use the **wget** command to get a file from a URL

Type **wget** then a space then click the middle mouse button to paste the URL
(you can also use 'Shift + Insert' to paste into the UNIX terminal)

```
cd ~/temp/hg19
```

```
wget http://hgdownload.cse.ucsc.edu/goldenpath/hg19/chromosomes/chrY.fa.gz
```

```
wget http://hgdownload.cse.ucsc.edu/goldenpath/hg19/chromosomes/md5sum.txt
```

List the directory contents to see the file with details (-l)
and human readable file sizes (-h)

```
ls -lh
```

Copying and renaming Files

Use the tab key to help prevent typos when typing filenames and directories

1a.
```
cp ch                                   (then hit tab)
```

UNIX will complete the file name for you

1b.
```
cp chrY.fa.gz
```

Make a copy of the ***chrY.fa.gz*** file called ***chrY_copy.fa.gz***

1c.
```
cp chrY.fa.gz chrY_copy.fa.gz
```

Rename the ***chrY_copy.fa.gz*** file to ***chrY_hg19.fa.gz***

```
mv chrY_copy.fa.gz chrY_hg19.fa.gz
```

Deleting Files: the *rm* cmd

```
rm [options] [file_name]
```

- Commonly used options
  - -i prompt user before any deletion

Use the wildcard * to list all files ending with gz

Exercise:

```
ls *gz

rm -i chrY.fa.gz

ls
```

Displaying File Contents

Check the file size before attempting to open with a text editor

```
ls -lh md5sum.txt
```

```
cat md5sum.txt
```

- **cat** prints all the contents of a file(s) to the screen.
- The **more** command, and its improved version **less**, display an text file one page at a time.
  - Hit space bar for next page
  - Type **q** to quit
  - Use **zmore** or **zless** for compressed files (.gz)

```
more md5sum.txt
less md5sum.txt
zmore chrY_hg19.fa.gz
```

Displaying File Contents

- Use head and tail commands to see first and last 10 lines of a file respectively

```
head md5sum.txt

tail md5sum.txt
```

- **head** and **tail** are not for compressed files (.gz)
- There is not a zhead or ztail command
- Use **zcat** together with the **head** or **tail** command

```
zcat chrY_hg19.fa.gz | head

zcat chrY_hg19.fa.gz | tail
```

'*grep*' – Search pattern(s) in files

```
grep [options] PATTERN [files ...]
```

```
grep chrX md5sum.txt
```

Count the number of lines that match pattern

```
grep -c random md5sum.txt
```

Search multiple matches

```
grep -e chrX -e chrY md5sum.txt
```

Exclude a pattern; show non-matching lines

```
grep -v random md5sum.txt
```

Use **zgrep** for compressed files (.gz)

```
zgrep chr chrY_hg19.fa.gz
```

Piping output of UNIX commands

Use the pipe character **|** to send results to another command

Search the output of md5sum.txt for the string 'chrX'

```
cat md5sum.txt | grep chrX
```

Use the standard output redirect operator **>** to **create** a new file

```
grep chrX md5sum.txt > chr_xy.txt
```

Use the standard output redirect operator **>>** to **append** to a file

```
grep chrY md5sum.txt >> chr_xy.txt
```

History of your commands

- Your commands are saved to a file in your home directory (.bash_history)
- You can use the up/down arrows to scroll through your previous commands
- Type history to see all your commands

```
history

history | tail
```

See last 10 commands

- Search your history commands using | and grep

```
history | grep wget
```

Types of File: the *file* cmd

```
file [name]
```

- Displays a brief description of the contents or other type information for a file or related object.

```
file md5sum.txt
```

```
md5sum.txt: ASCII English text
```

```
file chrY_hg19.fa.gz
```

```
chrY_hg19.fa.gz: gzip compressed data,
 was "chrY.fa", last modified ...
```

CRLF line terminators

- Windows editors such as Notepad will add hidden return/line feed characters that can cause problems with some applications

```
file my_file.txt
```

```
my_file.txt: ASCII English text, with CRLF line terminators
```

```
dos2unix my_file.txt
file my_file.txt
```

```
my_file.txt: ASCII English text
```

Editing an ASCII file

- There are many editors available under Linux.
- Text mode
  - nano (simple)
  - vi or vim (more advanced)
  - emacs (more advanced)
- Graphic mode (require X11)
  - gedit
  - xemacs / gvim
- Be aware of text file edited under Windows (CR/LF vs LF). Use **`dos2unix`** to convert a DOS/Windows edited text file to UNIX format.

```
dos2unix my_dos_file.txt
```

The UNIX Filesystem

File Attributes: A look with *ls*

```
[user_NetID@ada ~]$ ls -l
total 37216
drwx------    7 user_NetID    user_NetID         121 Sep  9 10:41 abaqus_files
-rw-------    1 user_NetID    user_NetID        2252 Aug 24 10:47 fluent-unique.txt
-rw-------    1 user_NetID    user_NetID    13393007 Aug 24 10:40 fluent-use1.txt
-rw-------    1 user_NetID    user_NetID         533 Aug 24 11:23 fluent.users
drwxr-xr-x    3 user_NetID    user_NetID          17 May  7 16:56 man
-rw-------    1 user_NetID    user_NetID    24627200 Sep  9 10:49 myHomeDir.tar
lrwxrwxrwx    1 root     root            21 May 28 16:11 README -> /usr/local/etc/README
-rwx------    1 user_NetID    user_NetID         162 Sep  7 12:20 spiros-ex1.bash
-rwx--x--x    1 user_NetID    user_NetID          82 Aug 24 10:51 split.pl
drwxr-xr-x    2 user_NetID    user_NetID           6 May  5 11:32 verifyOLD
```

file permissions

hard link count

user name

group name

file size in bytes

file modification date

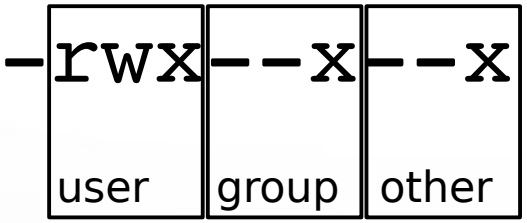
file modification tim

file name

File Ownership and Permissions

```
-rwx--x--x    1 user_NetID    staff          82 Aug 24 10:51 split.pl
```

permissions user group



- There are 3 sets of permissions for each file
  - 1st set - user (the owner)
  - 2nd set - group (to which file owner belongs)
  - 3rd set - other (all other users)
- The r indicates read permission
- The w indicated write permission
- The x indicated execute permission

Directory Permissions

```
drwx------    7 user_NetID   staff         121 Sep  9 10:41 abaqus_files
```

- The meanings of the permission bits for a directory are slightly different than for regular files:
  - ***r*** permission means the user can list the directory's contents
  - ***w*** permission means the user can add or delete files from the directory
  - ***x*** permission means the user can cd into the directory; it also means the user can execute programs stored in it
- Notice that if the file is a directory, the leading bit before the permissions is set to ***d***, indicating directory.

Changing Attributes: the ***chmod*** cmd

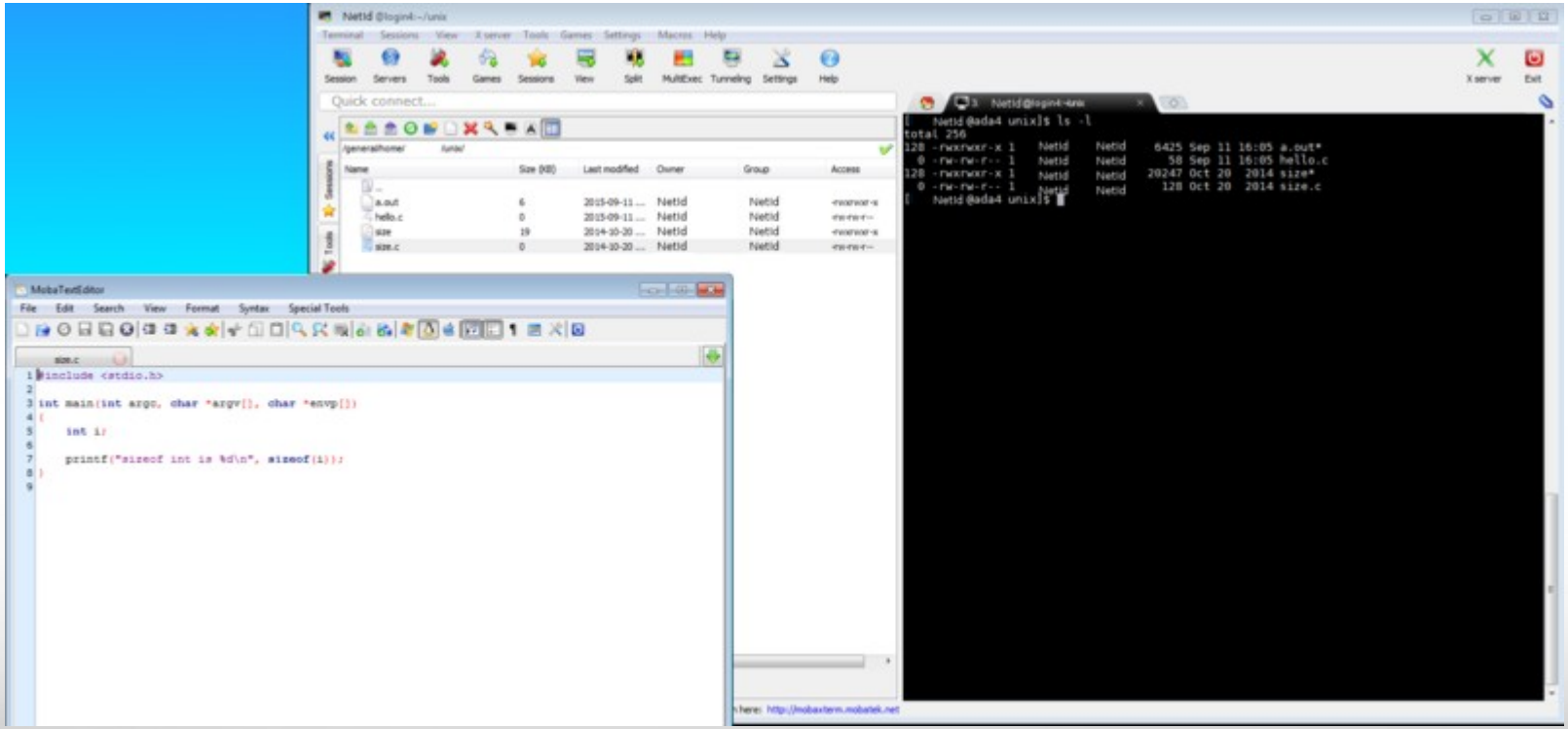
```
chmod [options] [permission mode] [target_file]
```

```
chmod 755 chr_xy.txt     ( the permissions will be set to -rwxr-xr-x )

chmod o-x chr_xy.txt     ( the permissions will change to -rwxr-xr-- )

chmod ug-x chr_xy.txt    ( the permissions will change to -rw-r--r-- )

chmod g+w chr_xy.txt     ( the permissions will change to -rw-rw-r-- )
```

```
u = user                    r = read
g = group                   w = write
o = other                   x = execute
-x = remove executable permissions
+x = enable executable permissions
```

Transfer data between Windows hosts with MobaXterm

- On a Windows system, you can use MobaXterm to transfer files to/from an HPRC cluster



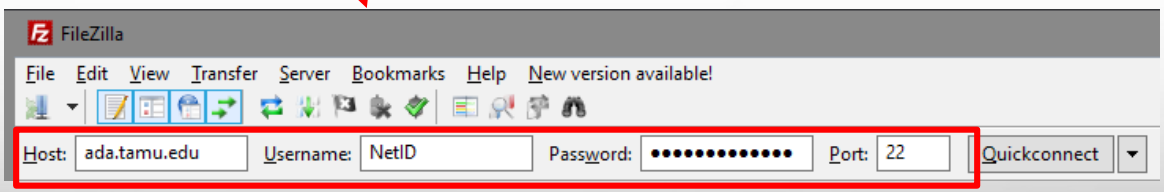
https://hprc.tamu.edu/wiki/HPRC:MobaXterm

File Transfers Using FileZilla

The FileZilla Client:

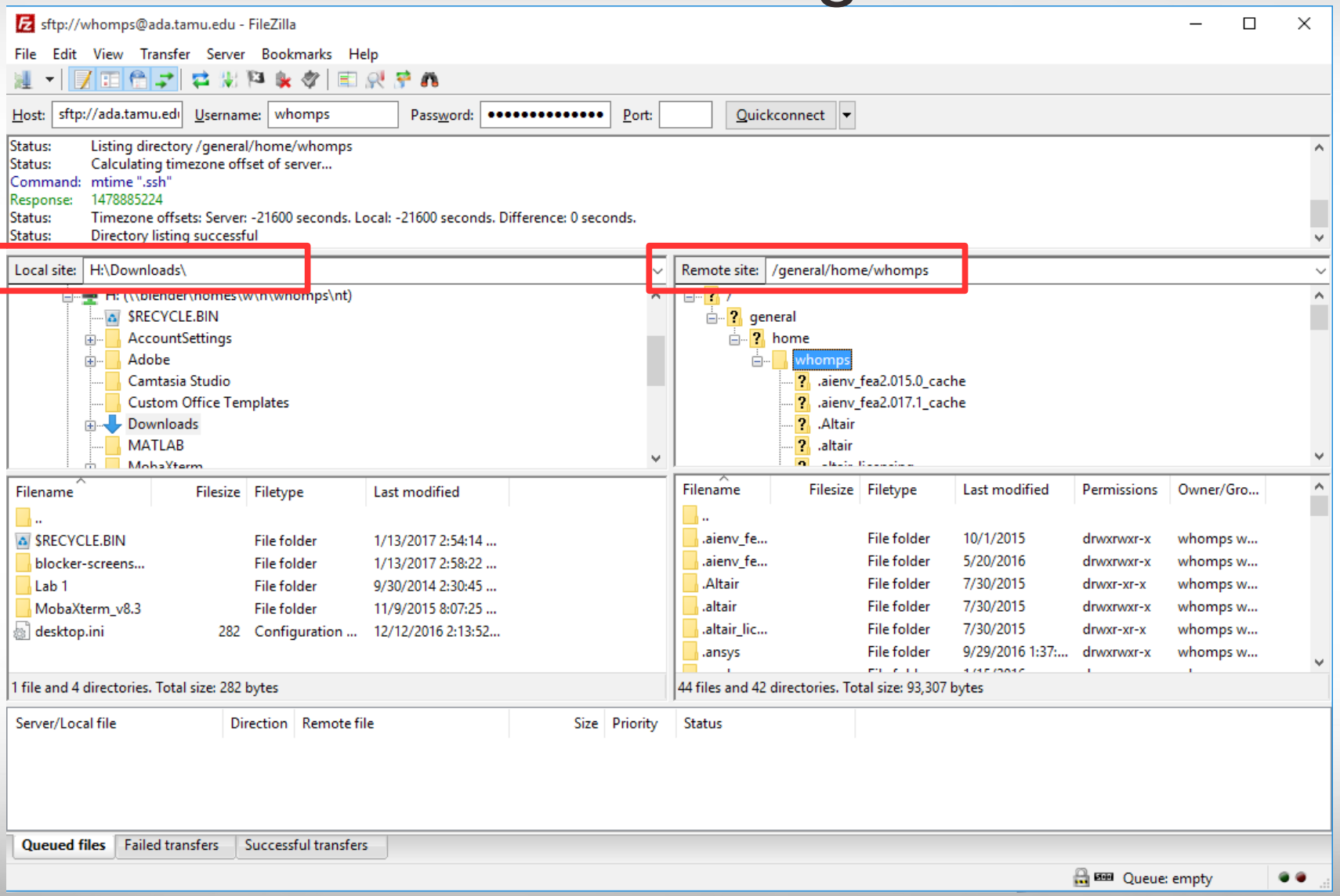
1) Available on Windows, OS X, and UNIX/Linux
2) Allows permissions to be preserved or implied
3) Easy to use without previous experience
4) Can drag and drop files between computers

Download from:
https://filezilla-project.org



Connect with remote login

File Transfers Using FileZilla



Transfer data between UNIX hosts with *scp*

```
scp [[user@]host1:]filename1 [[user@]host2:]filename2
```

Copy a file **from** your UNIX desktop **to** your ada home directory

```
scp myfile1 user@ada.tamu.edu:
```

Copy & rename file **from** your UNIX desktop **to** dir1 in your $HOME

```
scp myfile1 user@ada.tamu.edu:dir1/remote_myfile1
```

Copy a file **to** your UNIX desktop **from** your ada home directory

```
scp user@ada.tamu.edu:myfile2 ./

scp user@ada.tamu.edu:myfile2 local_myfile2

scp -r user@ada.tamu.edu:dir3 local_dir/
```

Bash Environment Variables

HOME	pathname of current user's home directory
PATH	the search path for commands. It is a colon separated list of directories in which the shell looks for commands.
SCRATCH	pathname of current user's scratch directory

```
echo $HOME
```

```
/home/user_NetID
```

```
echo $SCRATCH
```

```
/scratch/user/user_NetID
```

```
cd $SCRATCH
ls $HOME
```

The Search Path

- The shell uses the PATH environment variable to locate commands typed at the command line
- The value of PATH is a colon separated list of full directory names.
- The PATH is searched from left to right. If the command is not found in any of the listed directories, the shell returns an error message
- If multiple commands with the same name exist in more than one location, the first instance found according to the PATH variable will be executed.

```
echo $PATH
```

```
PATH=/opt/TurboVNC/bin:/software/tamusc/local/bin:
/software/lsf/9.1/linux2.6-glibc2.3-x86_64/etc:
/software/lsf/9.1/linux2.6-glibc2.3-x86_64/bin:
/usr/local/bin:/bin:/usr/bin:/usr/local/sbin:/usr/sbin:
/sbin:/usr/lpp/mmfs/bin:/opt/ibutils/bin:/home/user_NetID/bin
```

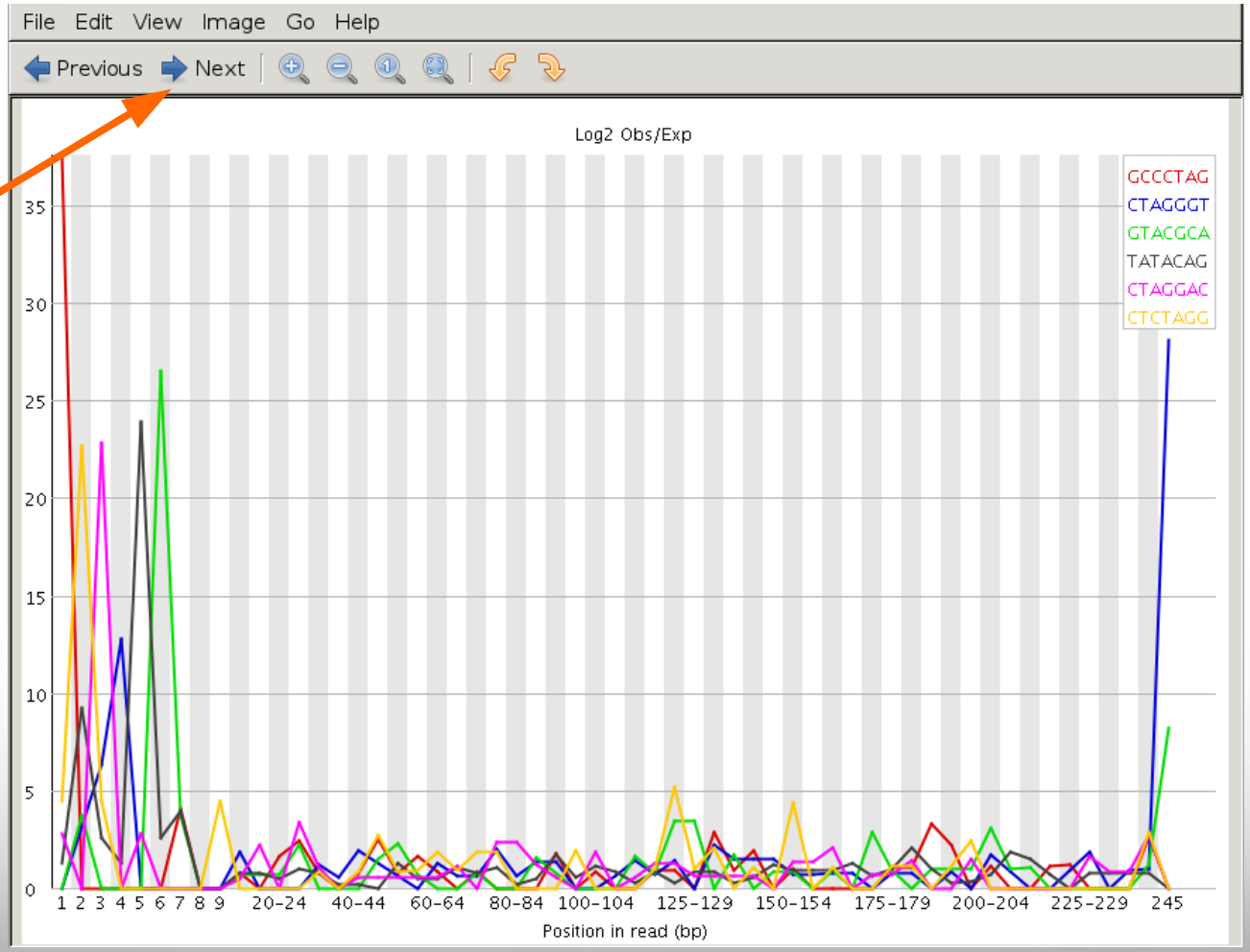
add a directory to the PATH for the current UNIX session

```
export PATH=$PATH:/home/user_NetID/bin
```

Viewing image files with Eye of Gnome image viewer

```
eog kmer_profiles.png
```

Click to see next image in current directory



The '*find*' Command

```
find [target dir] [expression]
```

`find ./ -name "*.txt"`	Anything in ./ ending in .txt
`find $SCRATCH -mtime -2 -type f`	Files Modified within last 2 days
`find $SCRATCH -mtime +5`	Modified more than 5 days ago
`find /tmp -user user_NetID`	Owned by user_NetID

`.` is the same as `./` which means current directory
`$SCRATCH` is your **`/scratch/user/user_NetID`** directory

Command Aliases

- Set a shortcut command or alias for the **grep** command

```
alias grep='grep --color=auto'
```

- Try the new alias

```
grep chrX md5sum.txt
```

- To save alias for each login session, add the alias command to your ~/.bashrc file
  - After editing the ~/.bashrc file, you will need to run the source command for the current session

```
source ~/.bashrc
```

```
# .bashrc

# Source global definitions
if [ -f /etc/bashrc ]; then
    . /etc/bashrc
fi

# User specific aliases and functions
alias grep='grep --color=auto'
```

The '*tar*' Command

```
tar [options] [tar file] [file or dir name]
```

- Used to "package" multiple files (along with directories if any) into one file suffixed with a *.tar* suffix by convention.
- Commonly used options
  - **x** extract files from a tar
  - **c** create a new tar
  - **t** list the contents of a tar
  - **v** verbosely list files processed
  - **f** use the specified tar file
  - **z** the tar file is compressed

The '*tar*' Command - examples

```
cd
```

go to your home directory

Package the temp directory into a file called my_hg19.tar

```
tar -cvf my_hg19.tar temp
```

Package the current directory into a compressed file

```
tar -cvzf my_hg19.tar.gz temp
```

Show the contents of the compressed tar file

```
tar -tzf my_hg19.tar.gz
```

Change the name of your original temp directory

```
mv temp temp_old
```

Extract all contents from the compressed tar file

```
tar -xvzf my_hg19.tgz
```

The '***tar***' Command

- Be careful when extracting files (overwriting old files).
- Where files are extracted depends on how they were packaged.
- Always a good idea to check Table of Contents (***-t*** option) before extraction.

Redirection Operators

<	redirects input (use this with bsub on Ada)
>	redirects output
>>	appends output
<<	input from *here document*

2>	redirects error
&>	redirects output and error
>&	redirects output and error
2>&1	redirects error to where output is going
1>&2	redirects output to where error is going

References

Here are some slides from TACC and LSU on the similar subject.

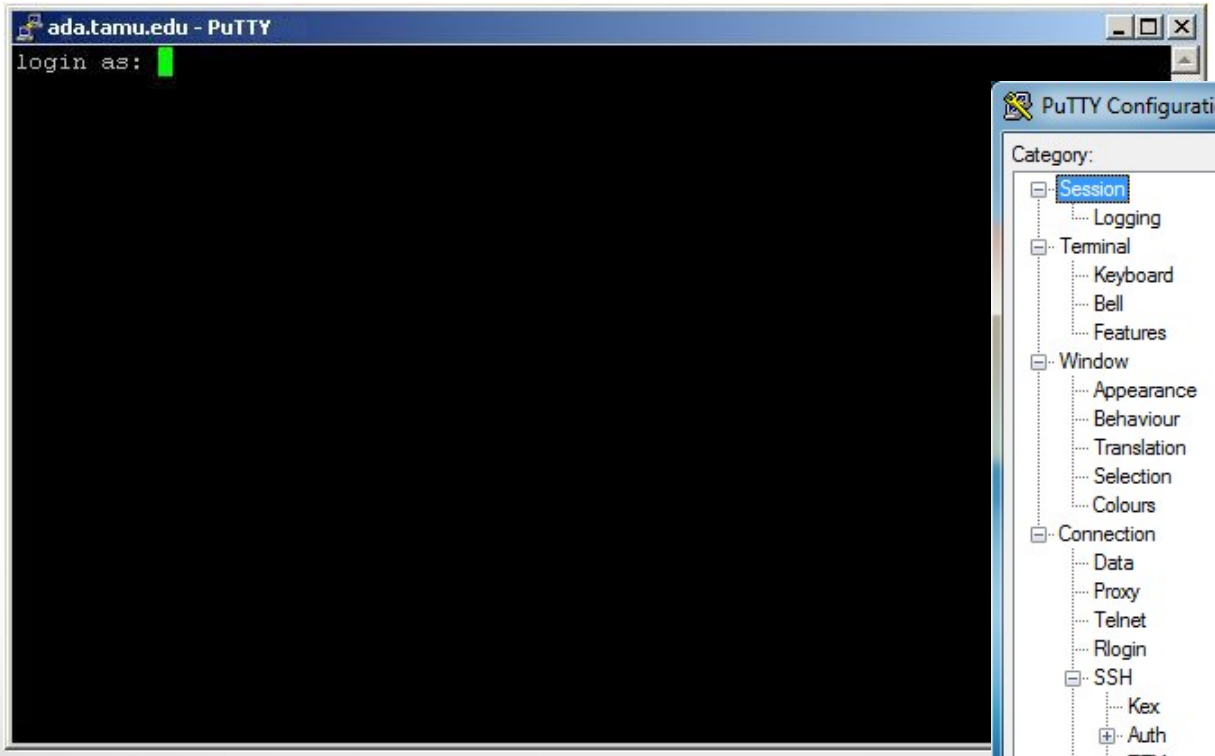
- Linux/Unix Basics for HPC: October 9, 2014 (with video) [TACC]
  https://portal.tacc.utexas.edu/-/linux-unix-basics-for-hpc
- Express Linux Tutorial: Learn Basic Commands in an Hour [TACC]
  https://portal.tacc.utexas.edu/c/document_library/get_file?uuid=ed6c16e9-bcbc-4b70-9311-5273b09508b8&groupId=13601

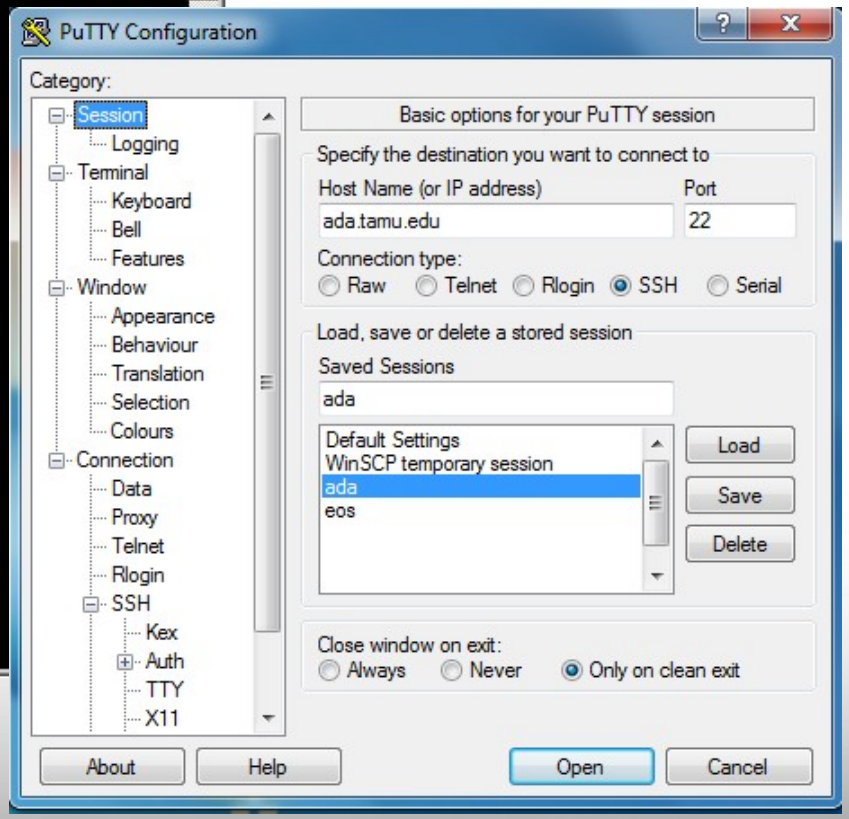
```
exit                    # exit the terminal session
                        # can also use Ctrl+d to detach session
```

Additional Slides

Using SSH - Putty (on Windows)

https://hprc.tamu.edu/wiki/HPRC:Access:Windows#Using_PuTTY





Use **ada.tamu.edu** as host name.