

Baseline Skills and Setup



Objectives

- Basic understanding of Virtual Machines and the terminal
- Review basic Linux commands
- Become familiar with Discovery Environment and VICE at CyVerse



Outline

- VMs & terminals
- File system navigation
- Creating, copying, moving, deleting files
- Networking commands
- Software installation & compression commands



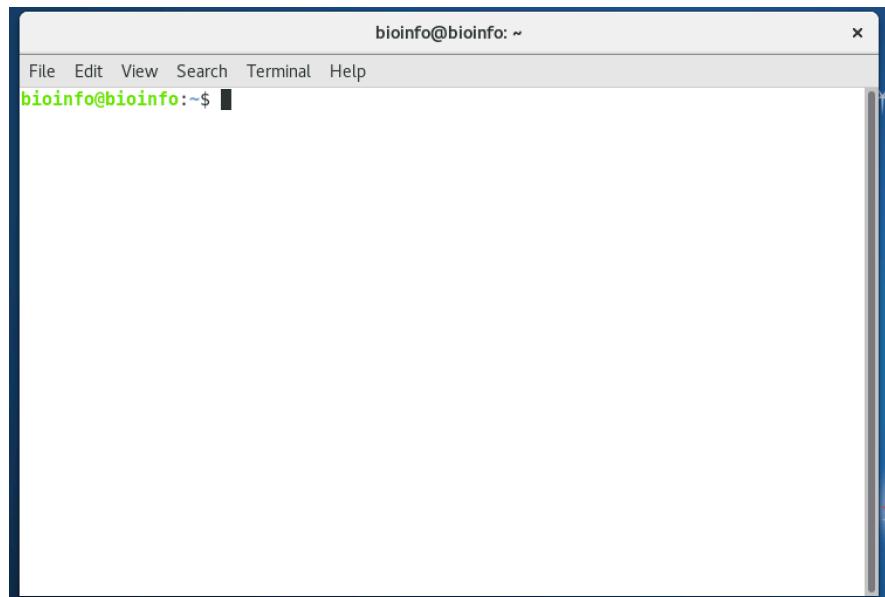
What is a Virtual Machine?

- A VM is an operating system which has been installed inside a simulated environment.



What is a Terminal?

- A terminal is a textual interface for interacting with a computer.
- Using the terminal, one can issue powerful and concise command-line instructions for the computer to follow.



Why use the command-line (terminal)?

- Most software for biological data analysis is used through UNIX command-line operations.
- Most of the servers for biological data analysis use Linux/ Unix as their operating system.
- Data analysis on calculation servers are much faster since we can use more CPUs and RAM than in a PC or laptop (*e.g.* BTI's "Boyce" server has 64 cores and 1TB RAM).
- Large NGS data files can not be opened or loaded in most graphical software and web sites.



We will use the CyVerse platform
today

About CyVerse

Vision: Transforming Science through Data-Driven Discovery

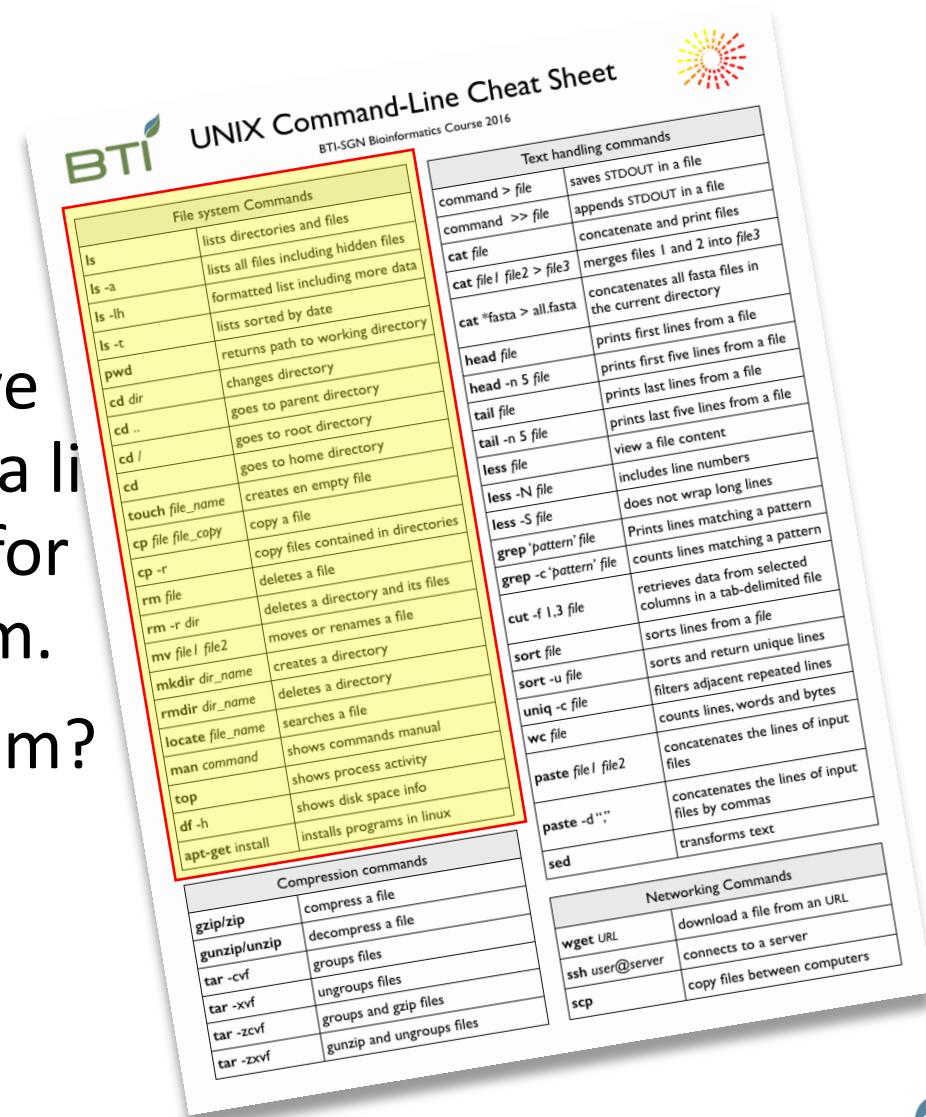
Mission: To design, deploy, and expand a national Cyberinfrastructure for Life Sciences research, and to train scientists in its use.

CyVerse provides life scientists with powerful computational infrastructure to handle huge datasets and complex analyses, thus enabling data-driven discovery. Our extensible platforms provide data storage, bioinformatics tools, data visualization, interactive analyses, cloud services, APIs, and more.

CyVerse is funded by the National Science Foundation's Directorate for Biological Sciences. We are a dynamic virtual organization, led by the University of Arizona in partnership with the Texas Advanced Computing Center and Cold Spring Harbor Laboratory, fulfilling a broad mission to enable data-driven, collaborative research.

Command-line File System Navigation

- The cheat sheet you have been provided contains a list of common commands for navigating the file system.
- But what *is* the file system?



The graphic features a title "UNIX Command-Line Cheat Sheet" with a subtitle "BTI-SGN Bioinformatics Course 2016" and a BTI logo. It includes a small sunburst icon. The cheat sheet is organized into several sections:

- File system Commands** (highlighted with a red border):

ls	lists directories and files
ls -a	lists all files including hidden files
ls -lh	formatted list including more data
ls -t	lists sorted by date
pwd	returns path to working directory
cd dir	changes directory
cd ..	goes to parent directory
cd /	goes to root directory
cd	goes to home directory
touch file_name	creates an empty file
cp file file_copy	copy a file
cp -r	copy files contained in directories
rm file	deletes a file
rm -r dir	deletes a directory and its files
mv file1 file2	moves or renames a file
mkdir dir_name	creates a directory
rmdir dir_name	deletes a directory
locate file_name	searches a file
man command	shows commands manual
top	shows process activity
df -h	shows disk space info
apt-get install	installs programs in linux
- Text handling commands**:

command > file	saves STDOUT in a file
command >> file	appends STDOUT in a file
cat file	concatenate and print files
cat file1 file2 > file3	merges files 1 and 2 into file3
cat *fasta > all.fasta	concatenates all fasta files in the current directory
head file	prints first lines from a file
head -n 5 file	prints first five lines from a file
tail file	prints last lines from a file
tail -n 5 file	prints last five lines from a file
less file	view a file content
less -N file	includes line numbers
less -S file	does not wrap long lines
grep 'pattern' file	Prints lines matching a pattern
grep -c 'pattern' file	counts lines matching a pattern
cut -f 1,3 file	retrieves data from selected columns in a tab-delimited file
sort file	sorts lines from a file
sort -u file	sorts and return unique lines
uniq -c file	filters adjacent repeated lines
wc file	counts lines, words and bytes
paste file1 file2	concatenates the lines of input files
paste -d ":" file	concatenates the lines of input files by commas
sed	transforms text
- Compression commands**:

gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files
- Networking Commands**:

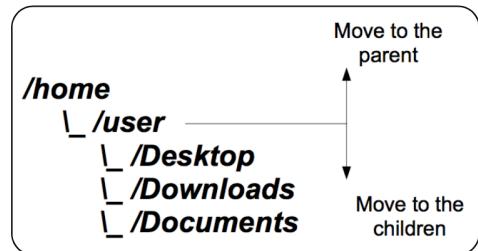
wget URL	download a file from an URL
ssh user@server	connects to a server
scp	copy files between computers



<https://github.com/bcbc-group/Botany2022NMGWorkshop>



The File System



```
bioinfo@bioinfo:~$ ls -lh
total 32K
drwxr-xr-x  2 bioinfo bioinfo 4.0K Mar  5 16:12 Data
drwxr-xr-x  2 bioinfo bioinfo 4.0K May 29 11:03 Desktop
drwxr-xr-x  2 bioinfo bioinfo 4.0K May 29 11:25 Documents
drwxr-xr-x  2 bioinfo bioinfo 4.0K May 29 11:24 Downloads
drwxr-xr-x 17 bioinfo bioinfo 4.0K Mar  5 16:17 Programs
drwxr-xr-x  2 bioinfo bioinfo 4.0K Mar  5 12:47 Public
drwxr-xr-x  3 bioinfo bioinfo 4.0K Mar  5 15:10 R
drwxr-xr-x  2 bioinfo bioinfo 4.0K Mar  5 16:12 Scripts
bioinfo@bioinfo:~$
```



Home and Root Directories

```
Noes-MacBook-Pro:~ Noe$ ls -lht
total 0
drwx-----+ 29 Noe  staff  986B May 31 11:24 Desktop
drwx-----@  8 Noe  staff  272B May 31 08:26 Dropbox
drwx-----+ 54 Noe  staff  1.8K May 30 16:01 Downloads
drwx-----+  8 Noe  staff  272B May 28 21:06 Pictures
drwxr-xr-x  18 Noe  staff  612B May 17 11:12 BTI
drwxr-xr-x   5 Noe  staff  170B May  8 11:44 programs
drwx-----+ 15 Noe  staff  510B Apr 10 08:33 Documents
drwxr-xr-x   6 Noe  staff  204B Mar 18 09:22 VirtualBox VMs
drwxr-xr-x   8 Noe  staff  272B Mar 14 19:26 py_devel
drwx-----@ 51 Noe  staff  1.7K Mar 11 15:08 Library
```

Home directory

/home/bioinfo
/home/noe
/home/noe/Desktop

```
noe@debian-virtualbox:~$ ls -l /
total 108
drwxr-xr-x  2 root root  4096 Sep 26  2012 bin
drwxr-xr-x  3 root root  4096 Nov  9  2012 boot
drwxr-xr-x  15 root root  3140 May 31 12:46 dev
drwxr-xr-x 130 root root 12288 May 31 12:45 etc
drwxr-xr-x  5 root root  4096 Feb 28 13:54 export
drwxr-xr-x  4 root root  4096 Nov  7  2012 home
lrwxrwxrwx  1 root root   30 Sep 26  2012 initrd.img
drwxr-xr-x 12 root root 12288 Nov  9  2012 lib
drwxr-xr-x  2 root root 12288 Nov  9  2012 lib32
lrwxrwxrwx  1 root root   4 Sep 26  2012 lib64 -> /1
drwxr-xr-x  2 root root 16384 Sep 26  2012 lost+found
drwxr-xr-x  3 root root  4096 Sep 26  2012 media
drwxr-xr-x  2 root root  4096 May  1  2012 mnt
drwxr-xr-x  2 root root  4096 Sep 26  2012 opt
dr-xr-xr-x 134 root root     0 May 31 12:45 proc
drwxr-xr-x  10 root root  4096 Nov 15  2012 root
drwxr-xr-x  2 root root  4096 Nov  9  2012 sbin
drwxr-xr-x  2 root root  4096 Jul 21  2010 selinux
drwxr-xr-x  2 root root  4096 Sep 26  2012 srv
drwxr-xr-x  13 root root     0 May 31 12:45 sys
drwxrwxrwt  11 root root  4096 May 31 19:56 tmp
drwxr-xr-x  11 root root  4096 Sep 26  2012 usr
drwxr-xr-x  14 root root  4096 Sep 26  2012 var
```

Root directory

/bin, /lib, /usr code and code libraries
/var logs and other data
/home user directories
/tmp temporary files
/etc configuration information
/proc special file system in Linux



pwd, ls, and cd

- **pwd** — where am I?
- **ls** — what is inside the current directory?
- **cd** — change my current directory.

Exercise 2:

- Change into the “/opt/” directory and list the files found there.



Wildcards, Shortcuts, and Command History

```
ls *txt
```

list all txt files in current directory

```
ls P*s
```

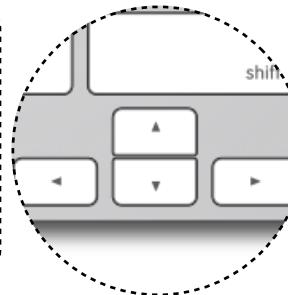
list files starting with P and ending with s,
e.g.: Pictures, Photos, Programs ...

ctrl-c stop process

ctrl-a go to beginning of line

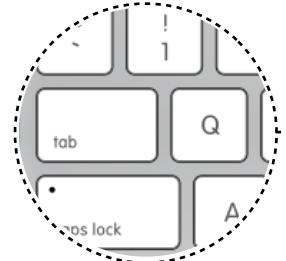
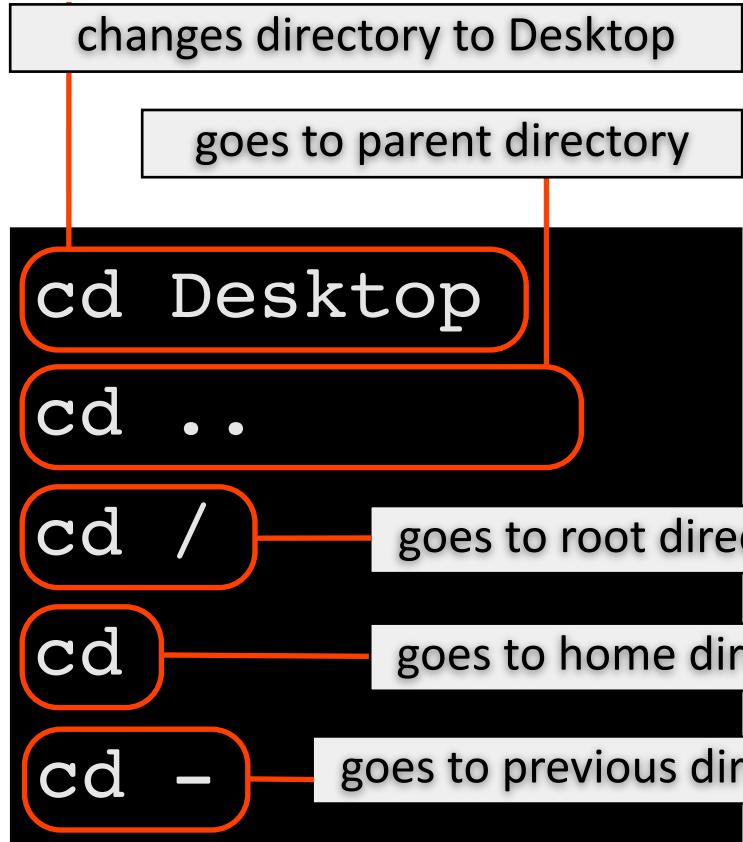
ctrl-e go to end of line

ctrl-r search in command history

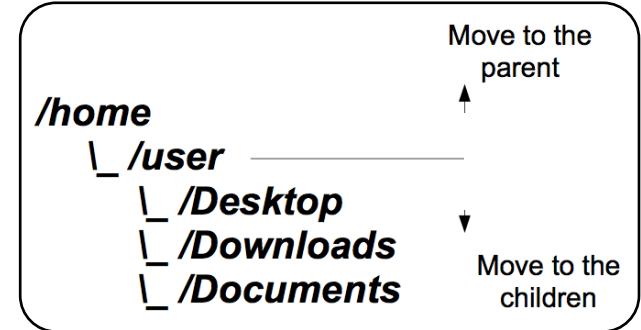


Use up and down
arrows to navigate
the command
history

cd changes directory



Use tab key to
autocomplete names



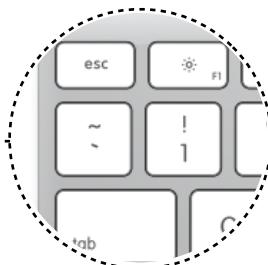
Absolute and relative paths

list files in Desktop using an absolute path

```
ls /home/user/Desktop
```

```
ls Desktop/
```

```
ls ~/Desktop
```



list files in Desktop using your home as a reference

list files in Desktop using a relative path (from your home: /home/bioinfo)

Absolute and relative paths

Absolute paths do not depend on where you are

```
ls /home/bioinfo/Desktop
```

```
ls ~/Desktop
```

~/ is equivalent to /home/bioinfo/

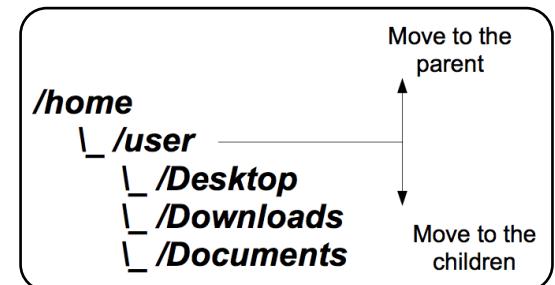
Absolute and relative paths

goes to *Desktop* from when you are in your home (/home/bioinfo)

```
cd Desktop/
```

```
ls .. /Documents
```

list files from *Documents* when you are in *Desktop*



Create, copy, move and delete files

creates an empty file called tmp_file.txt

```
touch tmp_file.txt
```

Tip: name files in lower case
and with underscores
instead of using spaces



Create, copy, move and delete files

creates an empty file called tmp_file.txt

copies tmp_file.txt in file_copy.txt

Tip: name files in lower case
and with underscores
instead of using spaces

```
touch tmp_file.txt
```

```
cp tmp_file.txt file_copy.txt
```



Create, copy, move and delete files

creates an empty file called tmp_file.txt

copies tmp_file.txt in file_copy.txt

```
touch tmp_file.txt
```

```
cp tmp_file.txt file_copy.txt
```

```
mv file1.txt file2.txt
```

Tip: name files in lower case
and with underscores
instead of using spaces



Create, copy, move and delete files

creates an empty file called tmp_file.txt

copies tmp_file.txt in file_copy.txt

```
touch tmp_file.txt
```

```
cp tmp_file.txt file_copy.txt
```

```
mv file1.txt file2.txt
```

```
rm file.txt
```

Tip: name files in lower case
and with underscores
instead of using spaces



Create, copy and delete directories

creates an empty directory called *dir_name*

```
mkdir dir_name
```



Music



Pictures



programs

Create, copy and delete directories

creates an empty directory called *dir_name*

deletes *dir_name* directory if it is empty

`mkdir dir_name`

`rmdir dir_name`



Music



Pictures



programs

Create, copy and delete directories

creates an empty directory called *dir_name*

deletes *dir_name* directory if it is empty

```
mkdir dir_name
```

```
rmdir dir_name
```

```
rm -r dir_name
```



Music



Pictures



programs

Create, copy and delete directories

creates an empty directory called *dir_name*

deletes *dir_name* directory if it is empty

```
mkdir dir_name
```

```
rmdir dir_name
```

```
rm -r dir_name
```

delete *dir_name* and its files

```
cp -r dir_name dir_copy
```

copy *dir_name* and its files in a new folder



Music



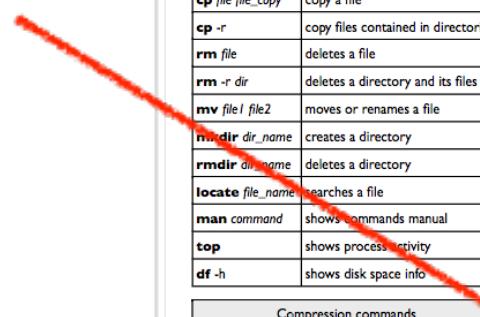
Pictures



programs

Networking Commands

● Networking commands



File system Commands	
ls	lists directories and files
ls -a	lists all files including hidden files
ls -lh	formatted list including more data
ls -t	lists sorted by date
pwd	returns path to working directory
cd dir	changes directory
cd ..	goes to parent directory
cd /	goes to root directory
cd	goes to home directory
touch file_name	creates an empty file
cp file file_copy	copy a file
cp -r	copy files contained in directories
rm file	deletes a file
rm -r dir	deletes a directory and its files
mv file1 file2	moves or renames a file
mkdir dir_name	creates a directory
rmdir dir_name	deletes a directory
locate file_name	searches a file
man command	shows commands manual
top	shows process activity
df -h	shows disk space info
Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxf	gunzip and ungroups files
Text handling commands	
command > file	saves STDOUT in a file
command >> file	appends STDOUT in a file
cat file	concatenate and print files
cat file1 file2 > file3	merges files 1 and 2 into file3
cat *fasta > all.fasta	concatenates all fasta files in the current directory
head file	prints first lines from a file
head -n 5 file	prints first five lines from a file
tail file	prints last lines from a file
tail -n 5 file	prints last five lines from a file
less file	view a file
less -N file	includes line numbers
less -S file	wraps long lines
grep 'pattern' file	Prints lines matching a pattern
grep -c 'pattern' file	counts lines matching a pattern
cut -f 1,3 file	retrieves data from selected columns in a tab-delimited file
sort file	sorts lines from a file
sort -u file	sorts and return unique lines
uniq -c file	filters adjacent repeated lines
wc file	counts lines, words and bytes
paste file1 file2	concatenates the lines of input files
paste -d ","	concatenates the lines of input files by commas
sed	transforms text
Networking Commands	
wget URL	download a file from an URL
ssh user@server	connects to a server
scp	copy files between computers
apt-get install	installs applications in linux



Networking Commands

connects your terminal to your account in a server

```
ssh user_name@server_address
```

```
wget https://btiscience.org/wp-content/uploads/BCBClogo.png
```

```
scp afp@boyce.sgn.cornell.edu:/home/afp/file.txt .
```

copy *file.txt* from [your home in the server] to [the current directory in your computer]

Tip: use the command `pwd` to get the path for `cp` and `scp`





SEARCH & DOWNLOAD CENTER

GETTING STARTED TUTORIAL

[CLOSE AND DON'T SHOW AGAIN](#)

My Services

**CoGe**

Online system for quick and easy retrieval and comparison of genomic information and sequences.

[LAUNCH](#)**Discovery Environment**

Use hundreds of bioinformatics apps and manage data in the CyVerse Data Store from a simple web...

[LAUNCH](#)**Atmosphere**

Cloud computing platform for CyVerse

[LAUNCH](#)**Data Commons**

A unified system for managing and sharing your data across CyVerse's tools and services

[LAUNCH](#)

CyVerse Discovery Environment

The screenshot displays the CyVerse Discovery Environment interface. The top navigation bar includes a logo for 'DISCOVERY ENVIRONMENT', a search bar, and various user icons (Data, Help, Cart, Notifications, and a user profile). A sidebar on the left contains icons for Home, Data, Projects, Groups, Settings, and Help. The main content area shows a list of 'Community Data' projects, each with a checkbox, a folder icon, and the project name. The projects listed are: Astrolabe, BioViz, Botany2020NMGWorkshop, Brassica_dwarf, CAP_TCN, CHEESEHEAD, EuPathDB-Data, G-OnRamp_hubs, GoreLab, KBase_staging, Legume_Federation, and NAM. A 'Customize Columns' button is located at the top right of the list. At the bottom, there are navigation buttons for page navigation and a page size selector set to 100.

	Name ↑	Actions
<input type="checkbox"/>	Astrolabe	⋮
<input type="checkbox"/>	BioViz	⋮
<input type="checkbox"/>	Botany2020NMGWorkshop	⋮
<input type="checkbox"/>	Brassica_dwarf	⋮
<input type="checkbox"/>	CAP_TCN	⋮
<input type="checkbox"/>	CHEESEHEAD	⋮
<input type="checkbox"/>	EuPathDB-Data	⋮
<input type="checkbox"/>	G-OnRamp_hubs	⋮
<input type="checkbox"/>	GoreLab	⋮
<input type="checkbox"/>	KBase_staging	⋮
<input type="checkbox"/>	Legume_Federation	⋮
<input type="checkbox"/>	NAM	⋮



CyVerse Discovery Environment

The screenshot shows the CyVerse Discovery Environment interface. The top navigation bar includes a logo, "DISCOVERY ENVIRONMENT", a search bar, and various user icons. Below the header, a sidebar on the left provides navigation links for "Community Data", "Data", "Data Types", "Search", "Help", and "Logout". The main content area displays a list of "Community Data" items, each with a checkbox, a folder icon, and a name. The items listed are: Astrolabe, BioViz, Botany2020NMGWorkshop, Brassica_dwarf, CAP_TCN, CHEESEHEAD, EuPathDB-Data, G-OnRamp_hubs, GoreLab, KBase_staging, Legume_Federation, and NAM. A "Refresh" button, "Details" button, and "Add To Bag" button are located at the top right of the list. A "Customize Columns" button is also present. A context menu is open over the "Astrolabe" item, listing options: Details, Metadata, Copy Link, Copy Path, and Move to Trash.

Name	Action
Astrolabe	Details
BioViz	Metadata
Botany2020NMGWorkshop	Copy Link
Brassica_dwarf	Copy Path
CAP_TCN	Move to Trash
CHEESEHEAD	
EuPathDB-Data	
G-OnRamp_hubs	
GoreLab	
KBase_staging	
Legume_Federation	
NAM	



iCommands

https://learning.cyverse.org/projects/data_store_guide/en/latest/step2.html

```
$ iinit
One or more fields in your iRODS environment file (.irodsEnv) are
missing; please enter them.
Enter the host name (DNS) of the server to connect to: data.cyverse.org
Enter the port number: 1247
Enter your irods user name: #your_cyverse_username
Enter your irods zone: iplant
Those values will be added to your environment file (for use by
other i-commands) if the login succeeds.

Enter your current iRODS password: #your_cyverse_password
```

```
cd /scratch
```

```
iget -rPT /iplant/home/shared/Botany2020NMGWorkshop
```



Compression commands

Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files



UNIX Command-Line Cheat Sheet

BTI-SGN Bioinformatics Course 2014



File system Commands	
ls	lists directories and files
ls -a	lists all files including hidden files
ls -lh	formatted list including more data
ls -t	lists sorted by date
pwd	returns path to working directory
cd dir	changes directory
cd ..	goes to parent directory
cd /	goes to root directory
cd	goes to home directory
touch file_name	creates an empty file
cp file1 file2	copy a file
cp -r	copy files contained in directories
rm file	deletes a file
rm -r dir	deletes a directory and its files
mv file1 file2	moves or renames a file
mkdir dir_name	creates a directory
rmdir dir_name	deletes a directory
locate file_name	searches a file
man command	shows commands manual
top	shows process activity
df -h	shows disk space info
Text handling commands	
command > file	saves STDOUT in a file
command >> file	appends STDOUT in a file
cat file	concatenate and print files
cat file1 file2 > file3	merges file1 and 2 into file3
cat *fasta > all.fasta	concatenates all fasta files in the current directory
head file	prints first lines from a file
head -n 5 file	prints first five lines from a file
tail file	prints last lines from a file
tail -n 5 file	prints last five lines from a file
less file	view a file
less -N file	includes line numbers
less -S file	wraps long lines
grep 'pattern' file	Prints lines matching a pattern
grep -c 'pattern' file	counts lines matching a pattern
cut -f 1,3 file	retrieves data from selected columns in a tab-delimited file
sort file	sorts lines from a file
sort -u file	sorts and return unique lines
uniq -c file	filters adjacent repeated lines
wc file	counts lines, words and bytes
paste file1 file2	concatenates the lines of input files
paste -d ","	concatenates the lines of input files by commas
sed	transforms text

Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files

Networking Commands	
wget URL	download a file from an URL
ssh user@server	connects to a server
scp	copy files between computers
apt-get install	installs applications in linux



Compression commands

Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files

group and compress files

```
tar -zcvf file.tar.gz f1 f2
```

```
tar -zxvf file.tar.gz
```

decompress and ungroup a tar.gz file



UNIX Command-Line Cheat Sheet

BTI-SGN Bioinformatics Course 2014



File system Commands	
ls	lists directories and files
ls -a	lists all files including hidden files
ls -lh	formatted list including more data
ls -t	lists sorted by date
pwd	returns path to working directory
cd dir	changes directory
cd ..	goes to parent directory
cd /	goes to root directory
cd	goes to home directory
touch file_name	creates an empty file
cp file_file_copy	copy a file
cp -r	copy files contained in directories
rm file	deletes a file
rm -r dir	deletes a directory and its files
mv file1 file2	moves or renames a file
mkdir dir_name	creates a directory
rmdir dir_name	deletes a directory
locate file_name	searches a file
man command	shows commands manual
top	shows process activity
df -h	shows disk space info

Compression commands	
gzip/zip	compress a file
gunzip/unzip	decompress a file
tar -cvf	groups files
tar -xvf	ungroups files
tar -zcvf	groups and gzip files
tar -zxvf	gunzip and ungroups files

Networking Commands	
wget URL	download a file from an URL
ssh user@server	connects to a server
scp	copy files between computers
apt-get install	installs applications in linux

files, directories or wildcards



Commands to install software

ubuntu®

debian
GNU/Linux

```
aptitude search blast
```

```
sudo aptitude install blast2
```

```
sudo apt-get install pbzip2
```

call the command with super user permissions

installs *pbzip2* in your computer



Installing MaSuRCA

```
cd /scratch  
git clone https://github.com/alekseyzimin/  
masurca.git  
cd masurca  
mv MaSuRCA-3.4.2.tar.gz /scratch  
cd ..  
tar -xvf MaSuRCA-3.4.2.tar.gz  
cd MaSuRCA-3.4.2  
../install.sh
```



Outline

- VMs & terminals
- File system navigation
- Creating, copying, moving, deleting files
- Networking commands
- Software installation & compression commands

