

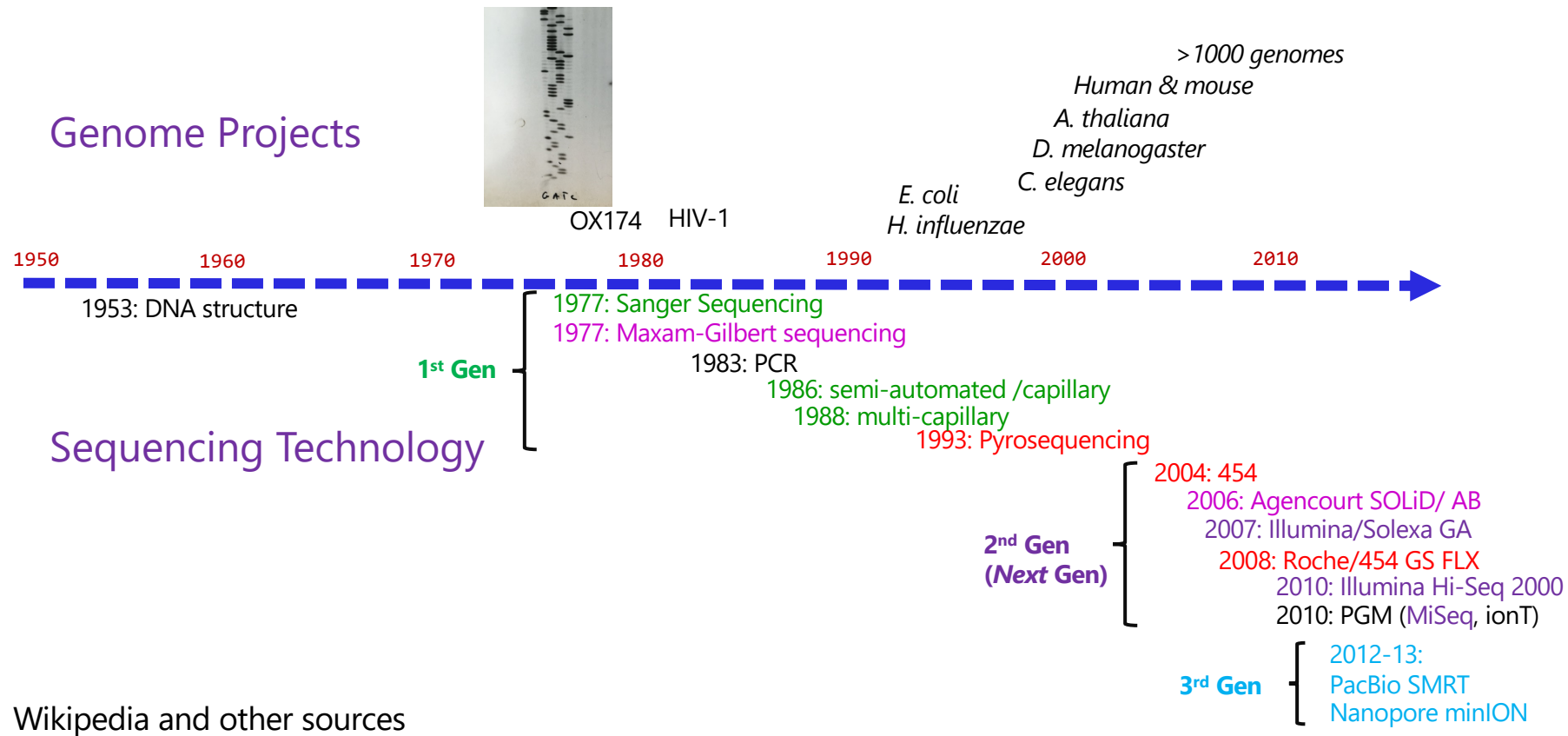
Introduction to Basic Unix

Adapted from: Andrew Severin
Virtual Research Support Core, Iowa State University

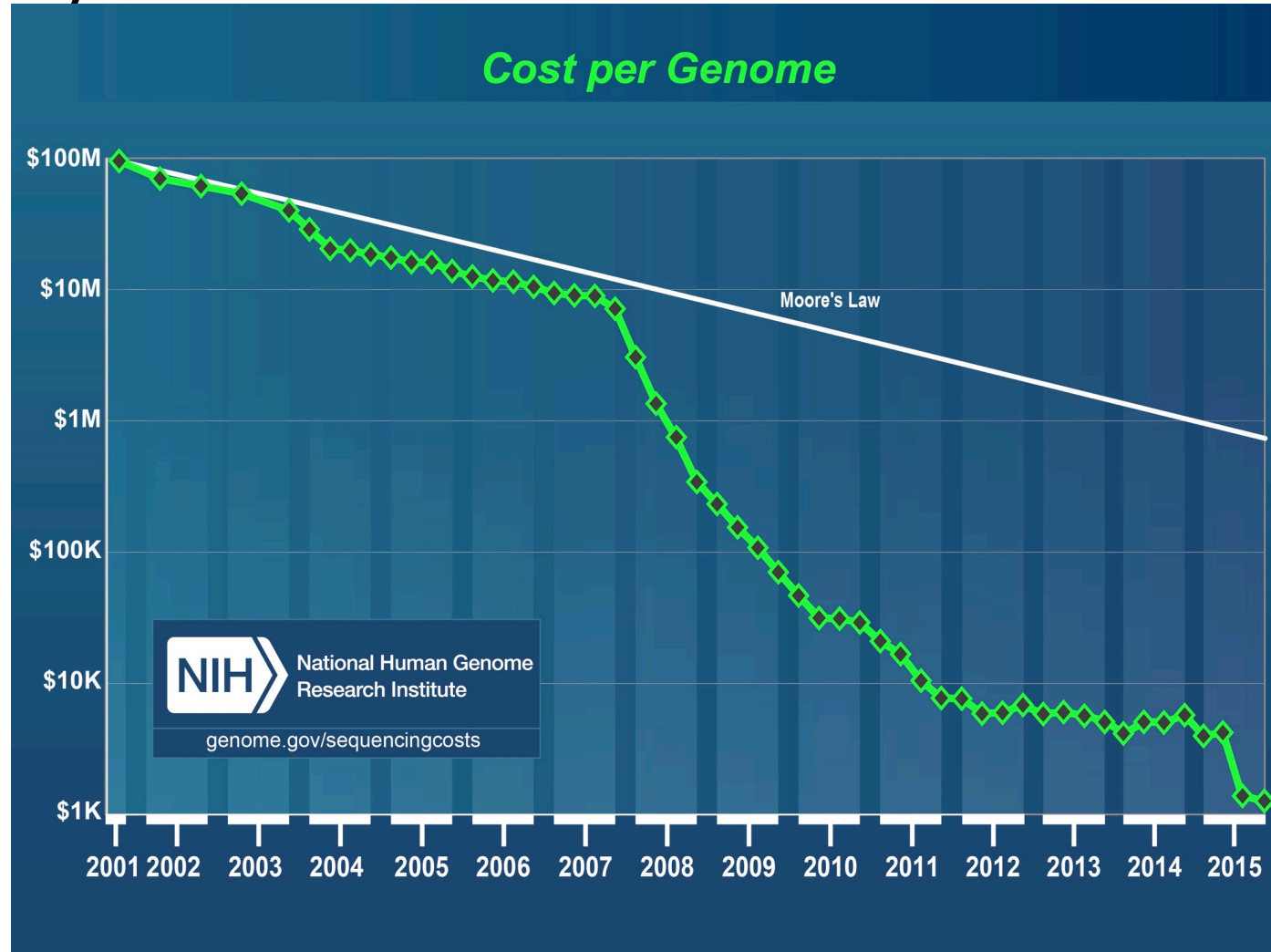
Download this Power point at <https://tinyurl.com/ydzd2gen>

<https://isugenomics.github.io/bioinformatics-workbook/>

Why learn UNIX?



Why do you need these skills?



What is Unix?

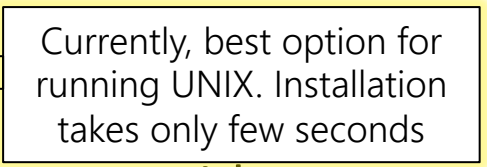
- Widely used multiuser operating system
- Linux: free version of UNIX-like operating system
 - Red Hat Enterprise Linux, Ubuntu, and CentOS
- Used on high-end workstations, database servers, web servers and managing shared resources
- Standard features include:
 - Security, reliability, scalability
 - supports multi-user (as in 100s)

Anyone can learn UNIX

- It is just another way of operating your computer
- No more difficult than learning Word, Excel or Powerpoint
- Biggest difference
 - In Linux: You type the command to execute
 - In Windows and Mac: You use your mouse to execute a command
 - Remember. In the Terminal, “Don’t touch the mouse”

How to get started?

- **Mac OS:** Native application called “Terminal” allows you to run UNIX commands. No installation necessary, just search for the program and start using it!
- **Linux OS:** Again, pre-installed in the OS. Search for either Terminal/Console program and can start running UNIX commands.
- **Windows OS:** Requires emulators for running UNIX commands. Popular options include:
 - Git BASH: msysgit.github.io
 - Cygwin : www.cygwin.com
 - Or ssh into the Purdue cluster computer with putty



Currently, best option for running UNIX. Installation takes only few seconds

Some tips before starting

- Write down the commands on a sheet of paper as you move through the exercise
- Typing first few letters of command/file and then pressing TAB will auto complete the word. Pressing TAB-TAB, displays all matching commands/files for the letters you typed.
- Use arrow keys to cycle through your command history (up and down).
- Commands are case sensitive (eg. Cat ≠ cat)
- Do not use space or special characters for naming files/folders
- Be careful before you delete, there is no recycle bin.
- Be familiar with the syntax and options for the commands

Some tips before starting

- Cancel a command
 - “CTRL+C” (“Ctrl” key and letter “c” key pressed together)
 - “command+C”
C for Cancel!
- Many programs use “q” or “Esc” to quit/exit
- Double check your commands: chances are, most of the times you don’t see an error, instead, you get wrong results.
- To clear the screen: Ctrl+L
- The command reset will reset your terminal

Purdue cluster computing

- <https://www.rcac.purdue.edu/compute/>
- We use scholar and bell

Computational Resources

ITaP maintains many different resources for computation. Here is a brief introduction to current computational resources. More information and detailed documentation are available for each resource listed below.



interconnect.

Weber

Weber is Purdue's new specialty Community Cluster for restricted and controlled data. Applications and research which is restricted by ITAR or DFARS 800-171. Weber was built in August 2019. Weber currently consists of HP compute nodes with two 10-core Intel Xeon-E5 "Haswell" processors (20 cores per node) and 64 GB of memory. All nodes have 56 Gbps EDR Infiniband



Gilbreth

Gilbreth is a new type of addition to Purdue's Community Clusters, designed specifically for applications which are able to take advantage of GPU accelerators. While applications must be specially-crafted to use GPUs, a GPU-enabled application can often run many times faster than the same application could on general-purpose CPUs. Due to the increased cost of GPU-equipped nodes, Gilbreth is being offered with some new purchase options to allow for shared access at a lower price point than the full cost of a node.

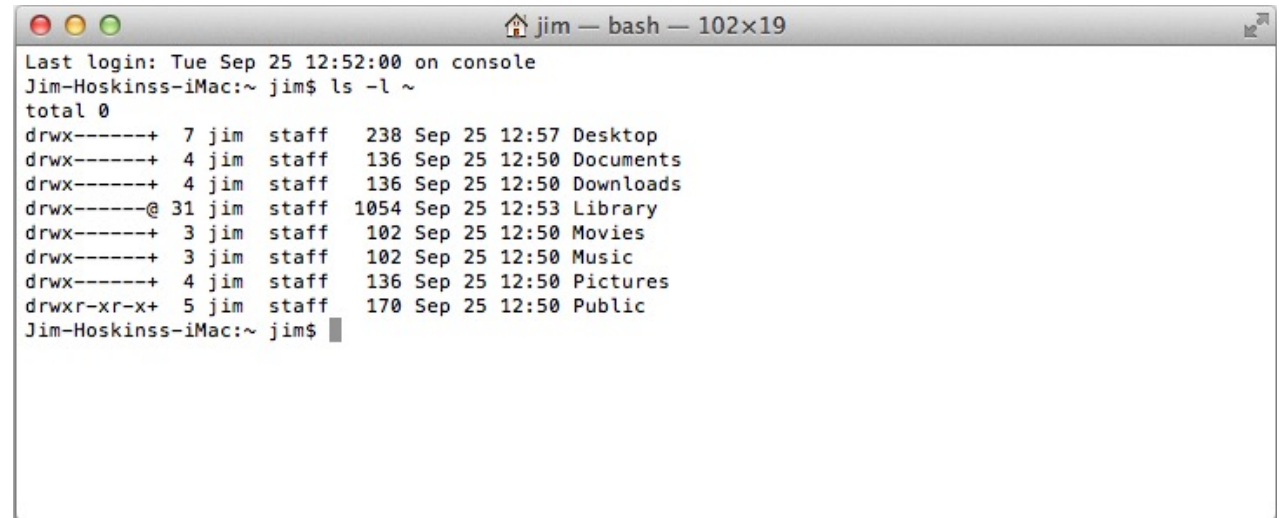


Scholar

The Scholar cluster is open to Purdue instructors from any field whose classes include

Logging into Purdue Cluster computing

On a mac: Terminal

A screenshot of a macOS Terminal window. The title bar at the top reads 'jim — bash — 102x19'. The window contains the following text:

```
Last login: Tue Sep 25 12:52:00 on console
Jim-Hoskinss-iMac:~ jim$ ls -l ~
total 0
drwx-----+  7 jim  staff   238 Sep 25 12:57 Desktop
drwx-----+  4 jim  staff   136 Sep 25 12:50 Documents
drwx-----+  4 jim  staff   136 Sep 25 12:50 Downloads
drwx-----@ 31 jim  staff  1054 Sep 25 12:53 Library
drwx-----+  3 jim  staff   102 Sep 25 12:50 Movies
drwx-----+  3 jim  staff   102 Sep 25 12:50 Music
drwx-----+  4 jim  staff   136 Sep 25 12:50 Pictures
drwxr-xr-x+  5 jim  staff   170 Sep 25 12:50 Public
Jim-Hoskinss-iMac:~ jim$
```

Logging into Snyder on a mac

ssh – short for “secure shell”
– opens a secure terminal on a remote server

Example:

```
$ ssh username@bell.rcac.purdue.edu
```

```
$ ssh username@scholar.rcac.purdue.edu
```

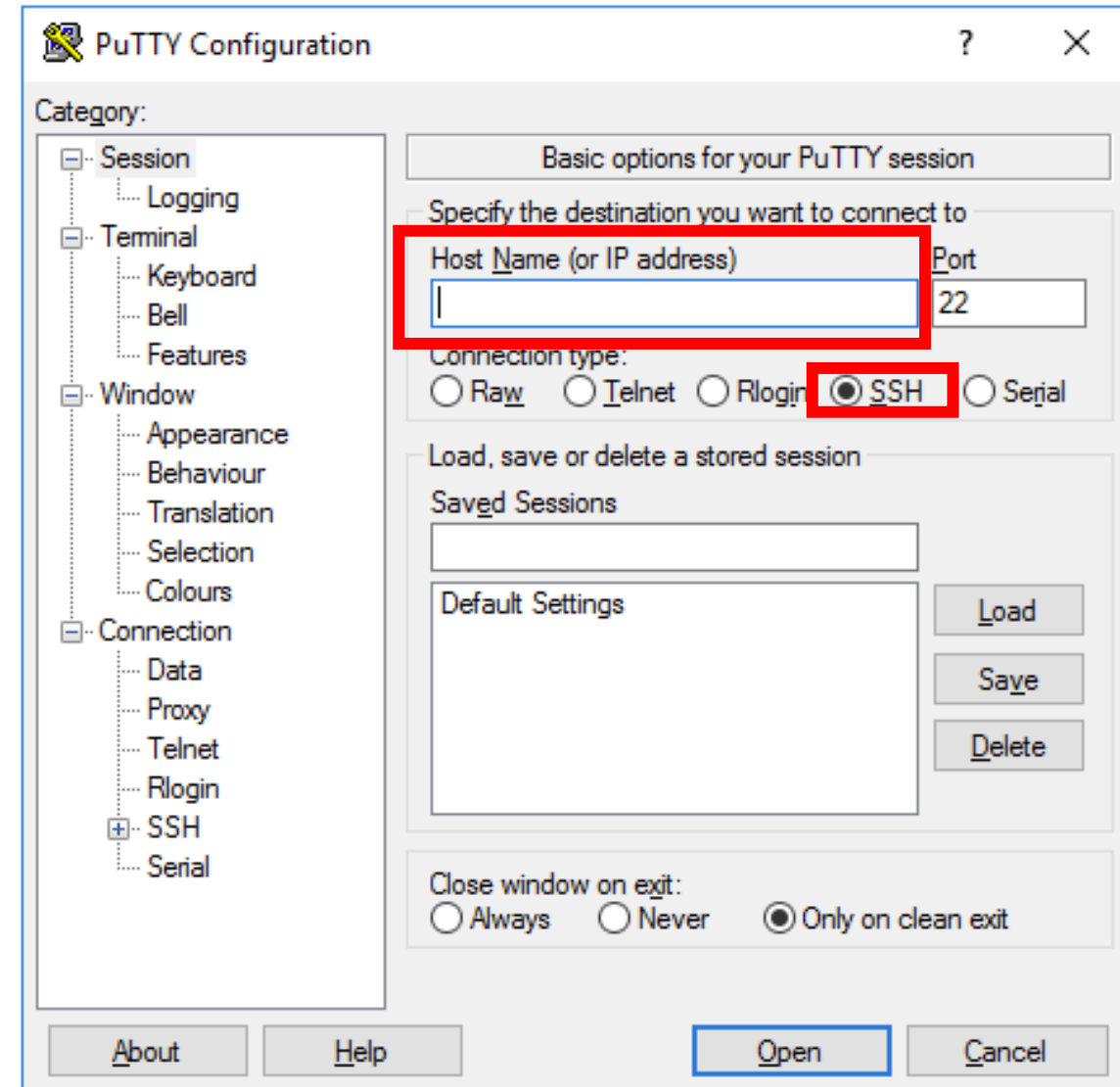
Here is my login command

```
$ ssh john2185@bell.rcac.purdue.edu
```

```
$ ssh john2185@scholar.rcac.purdue.edu
```

Logging into Snyder on Windows machine

- Open up putty
- Put `bell.rcac.purdue.edu` or `scholar.rcac.purdue.edu` in the Host name space
- Make sure ssh is selected
- This will open a new window and they will ask for user name and password.



Terminal

```
Last login: Fri Dec 28 13:16:34 on ttys000
1151POUL105-02L:~ john2185$ ssh john2185@snyder.rcac.purdue.edu
Password:
Last login: Fri Dec 28 13:18:04 2018 from crtn-ansc-desk30148.agriculture.purdue.edu

=====
Welcome to the Snyder Cluster
=====
Snyder consists of:

Nodes:
Snyder-A    ppn=20    256 GB memory
Snyder-B    ppn=20    512 GB memory
Snyder-C    ppn=20    512 GB memory
Snyder-D    ppn=20    1 TB memory
Snyder-E    ppn=24    384 GB memory

Scratch Storage:
Quota: 100 TB / 1,000k files
Path: $CLUSTER_SCRATCH
Type command: "myquota"

Queues:
Type command: "qlist"

Software:
Type command: "module avail" or "module spider"

User guide:
www.rcac.purdue.edu/knowledge/snyder

Help:
www.rcac.purdue.edu/help

News:
www.rcac.purdue.edu/news/Snyder

=====
*****

WARNING: You are connecting from an off-campus IP address.

; February 11, 2019 you will be required to use the VPN
arKey to access the Community Clusters from off campus

See the news posting for more information:
https://www.rcac.purdue.edu/news/1234

*****

john2185@snyder-e002:/scratch/snyder/j/john2185 $ ls -l somefi
ls: cannot access somefile: No such file or directory
john2185@snyder-e002:/scratch/snyder/j/john2185 $
```

prompt

command

argument

standard out

standard error

Getting started on the Bell system

Start an interactive session

```
$ sinteractive -t 2:00:00 -A microbiome -n6
```

-t is for time hours:min:sec

-A is the queue name

-n is the number of CPUs requested

Get set up

```
$ cp /depot/microbiome/data/data-shell.zip $RCAC_SCRATCH
```

```
$ cd $RCAC_SCRATCH
```

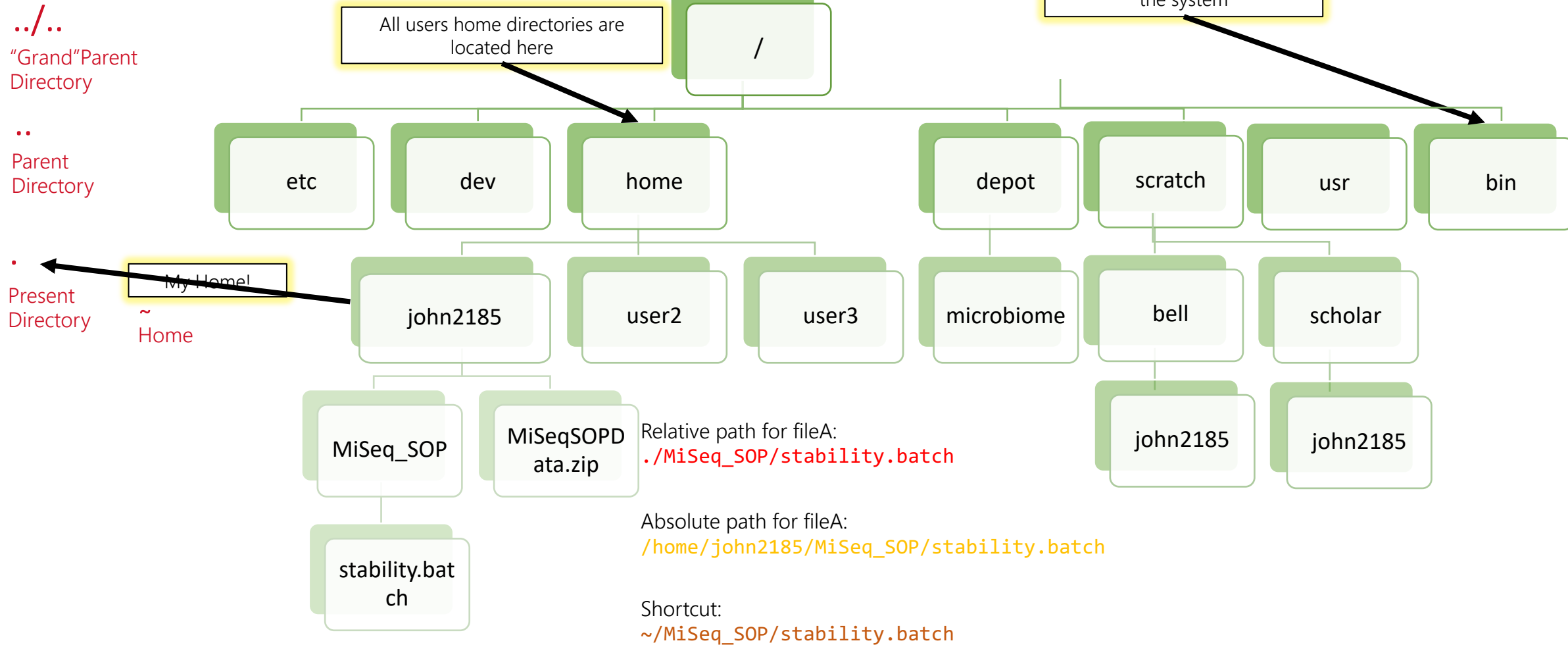
```
$ unzip data-shell.zip
```

pwd: path of working directory

```
$ pwd
```

```
/scratch/bell/john2185
```

Organization



Changing Directories

cd – Change to a directory

Example:

```
$ cd $RCAC_SCRATCH
```

```
$ pwd
```

Pro-tip: try to press tab after typing a few letters of a directory

Challenge: Try to get to your scratch directory without
\$RCAC_SCRATCH

Education

A Quick Guide to Organizing Computational Biology Projects

William Stafford Noble^{1,2*}

1 Department of Genome Sciences, School of Medicine, University of Washington, Seattle, Washington, United States of America, **2** Department of Computer Science and Engineering, University of Washington, Seattle, Washington, United States of America

Two Guiding Principles:

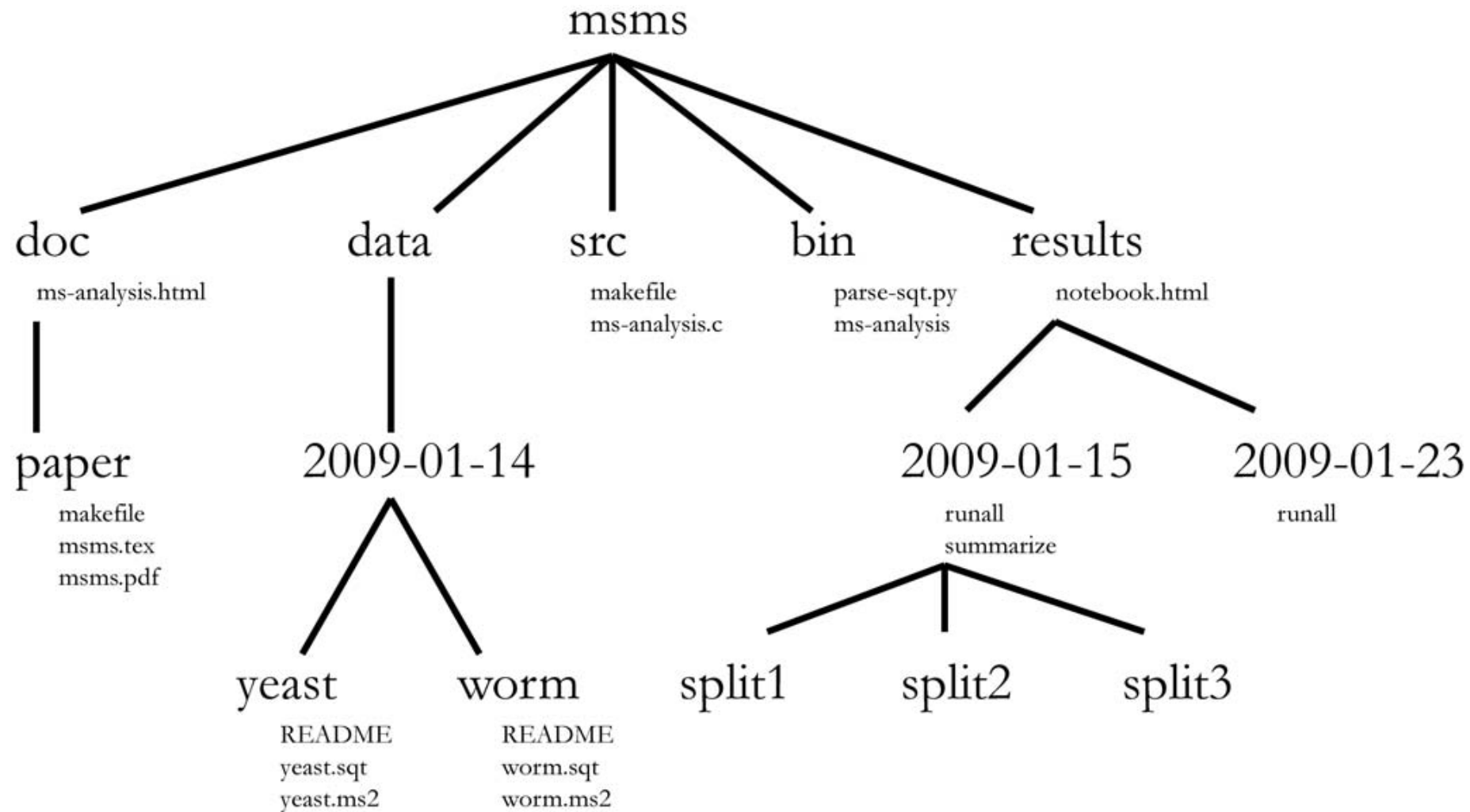
Principle #1

The core guiding principle is simple: Someone unfamiliar with your project should be able to look at your computer files and understand in detail what you did and why. This “someone” could be any of a variety of people: someone who read your

Principle #2

Murphy’s Law: Everything you do, you will probably have to do over again. Inevitably, you will discover some flaw in

Directory organization



Practical advice

- Lab notebook
 - First organize by date completed
 - Second organize by topic
- Record every operation
- Add comments to your code
- Avoid editing intermediate files
- Use file names in your scripts
- Use relative paths within your project folder

Unix tutorial

- Go to:

<http://swcarpentry.github.io/shell-novice/>

The path to your directory is:

/depot/microbiome/data/ANSC595/<your_name>

And complete the first 3 modules

1. Introducing the shell
2. Navigating Files and Directories
3. Working with Files and Directories

Answer the questions listed at the top of each module and submit on BrightSpace.

Class tutorials

For much of the remaining lessons go to

<https://github.com/john2929/ANSC595/wiki>

For our class materials.