Getting started

This class will emphasize UNIX skills to support doing genomics and evolutionary analysis with bioinformatics tools. There are many many tutorials and workshops out there. Many are available for free and linked here

- Data Carpentry and Software Carpentry part of The Carpentries
- Data Intensive Biology training like Shell Genomics
- Getting Started with Genomics Tools
- Happy Belly Bioinformatics

Logging into Cluster

You will need to have a terminal to get onto the system. On Mac that is called 'Terminal'. On Windows MobaXTerm is the best tool. Choose the 'Free' and 'Portable Version'.

Existing documentation and tutorial available at http://hpcc.ucr.edu/ for using the HPCC.

See the Linux Basics

To login to the cluster we need to use ssh client. This allows secure communication with the cluster. The UCR cluster is accessed using the host `cluster.hpcc.ucr.edu`.

```
$ ssh -X USERNAME@cluster.hpcc.ucr.edu
```

This will initiate a UNIX session running on the cluster 'head node' by connecting through a secure connection. There are multiple machines which serve as this login node where we can stage our analysis to run on the worker nodes that are on the cluster so you may see different names like 'pelican', 'pigeon' when you log in each time. Much more detail on the setup of the cluster and resources available at http://hpcc.ucr.edu.

You should now see a message as well as a prompt:

```
-------------------------------------------------------------------------------
 University of California, Riverside - HPCC (High-Performance Computing Center)
-------------------------------------------------------------------------------

More information about HPCC and how to use the resources provided can
be found at http://hpcc.ucr.edu/manuals_linux-cluster_intro.html

Please send all questions and support requests to support@hpcc.ucr.edu

Note: The default version of R is now 3.6.0
```

```
----------------------------------------------------------------------------
```

`username@pelican:~$`

Let's setup some initial things. Make a SSH folder so we can copy thing over.

```
[hpcc] $ mkdir ~/.ssh
[hpcc] $ chmod 700 ~/.ssh # sets the permission for this folder
```

You want to change your password on the cluster to something only you know. This is good practice since your password was emailed to you. It will prompt you for your current password and a new one.

`[hpcc] $ passwd`

You can log off by typing * `exit` * ˆD (Control-D)

The prompt on our system by default will start with the name of the computer as well as the current directory you are logged into. This prompt like most things on the system can be customized.

`hostname:[directory]$`

The `-X` option tells the system to forward your X11 connection which is necessary for running interactive graphics (eg showing an image, running a graphical editor program like emacs)

The command line interface (CLI)

The command line provides ability to interact with the filesystem (files and folders) and run programs. A collection of UNIX utilities.

Directories and files

ls - list the files and folders in a directory. Options include `-l` to list with details (long). `-t` list ordered by time created (time). These can be combined as `ls -ls` or `ls -l -s`. Specify a folder to list other than the current directory with another argument `ls -l data`.

mkdir Create a directory. Give the `-p` option to create an necessary sub folders and also to not give warnings if a folder already exists.

```
$ mkdir test
$ mkdir Alpha/Beta/Zeta # will give error
$ mkdir -p Alpha/Beta/Zeta # will not give error
```

rmdir Remove a folder. Only works if folder is empty

rm Remove a file or folder. This is command to be careful with. To delete a folder that contains many folders.

more See the contents of a text file, one page at a time. Go to the next page with 'space'. Can search for a specific text with slash (**/**).

```
$ more Gene_list.txt
```

less See the contents of a text file, one page at a time. Less has *more* options than **more** with arrows which will let you navigate up and down pages and a search option - use the slash (**/**) and then type in a search text it will highlight all the options.

```
$ less Gene_list.txt
```

head see the first lines in a file. By default this is 10 lines. But you can specify as many as you want with **-n LINES** option. Useful to get the beginning of a report or see what is the header in a spreadsheet file.

```
$ head -n 15 FILE.txt
```

tail see the last lines in a file. By running this command can see the last 10 lines by default. Can specify number of lines with **-n LINES** option. Useful when looking at a log-file and want to see the last reported messages.

```
$ tail -n 12 FILE.txt
```

echo Prints out messages.

Logging in with SSH keys

Let's make it simpler to login to the cluster without having to type our password all the time.

Setup your SSH access on your computer (your laptop). Generate RSA key pair on your computer. It will ask you for a password. You get to pick any password here you can remember. This will be a different password from your cluster one.

```
[your laptop] $ ssh-keygen -t rsa
```

See the new files created

```
[your laptop] $ ls ~/.ssh/
[your laptop] $ id_rsa id_rsa.pub
```

Let's also configure how our ssh works by customizing the ssh config. We need to edit a text file. On OSX this can be with **vi**, **emacs**, **nano**

```
nano ~/.ssh/config
```

"'text ForwardX11 yes ForwardX11Trusted yes ForwardAgent yes

Host hpcc Hostname cluster.hpcc.ucr.edu User YOURHPCCUSERNAME ServerAliveInterval 10 "'

This will open up a window for editing. You should add this content to your file. Here's a file you can copy onto the server too. `scp ssh-config_example.txt YOURHPCCUSERNAME@cluster.hpcc.ucr.edu:.ssh`

Web Access with Jupyter

We will setup access to the cluster and support a notebook interfaceto the cluster.

```
[your laptop] $ ssh cluster.hpcc.ucr.edu
sbatch -p short /bigdata/gen220/shared/login/submit_jupyter.sh
```

Practice steps.

1. Generate a new direction