**Homework 1 - Linux Refresh**

In this homework I will be asking you to review the purpose of the shell and how it works, common shell commands and demonstrate scripting skills.

For the class download this file [shell-lesson-data.zip](https://swcarpentry.github.io/shell-novice/data/shell-lesson-data.zip)

For the homework. You will need this file [Hutia\_DNA.fasta](https://www.dropbox.com/s/07y610c8zt63nt5/Hutia_DNA.fasta?dl=0).

**General concepts to refresh and lessons to review if you need them:**

[Shell Commands](http://swcarpentry.github.io/shell-novice/01-intro/index.html)

[Working with Files and Directories](http://swcarpentry.github.io/shell-novice/03-create/index.html)

[Pipes and Filters](http://swcarpentry.github.io/shell-novice/04-pipefilter/index.html)

[Loops](http://swcarpentry.github.io/shell-novice/05-loop/index.html)

[Shell Scripts](http://swcarpentry.github.io/shell-novice/06-script/index.html)

[Documentation](https://blog.submain.com/code-documentation-the-complete-beginners-guide/) -specifically in-line-comments

**Remember these Common commands as you work through this exercise**

ls mkdir cd pwd

cp mv rm rmdir

cat less head tail

grep wc uniq top

man sort wc \*

| . .. /

**Part I: Questions**

1. What is a Graphical User Interface (GUI) and how does that differ from the Command Line Interface (CLI)?
   1. GUI- shows you everything in a user-friendly manner, allowing you to click on what you want to do and has drop down menus.
   2. CLI- is just a shell that allows you to enter code and run programs. There is nothing to show you want to enter, it is up to you
2. What does the shell do?

Run programs

1. What is the command used for listing things in a directory?

Ls (I prefer ll on Ubuntu on Windows)

1. What is the command used for changing directories?

cd

1. What command would you use to get your current working directory?

Cd . (unclear what you are asking, current directory is ., to get to home would be cd )

1. How do you get the manual for these commands?

Type man in front of the command (I also use the --help as an argument sometimes too)

1. What does the shell prompt look like?

$

1. How would you list things in a directory in chronological order?

Ls -t

1. Name two ways to get to the ‘home’ or ‘root’ directory.

I am a little confused by this as well. My root directory is just / and my home is /home/ckhui, maybe it is different for Windows. To get to root, I type cd /. To get to home, I type cd or cd ~

1. What is the difference between an absolute path and a relative path?

Absolute path tells you the entire name of the path and is accessible anywhere you are. Relative path changes relative to where you are. For example .. is the folder containing your folder, which changes on where you are.

1. What are the two relative path directories we talked about and what do they mean?

Current folder .

Folder containing your current folder ..

1. Why shouldn’t you put spaces in filenames?

Unless you properly negate the spaces in your codes, it will think each space is a different argument.

1. Name a way to have multiple readable words in a filename without spaces?

Use\_this\_

1. What is the program nano, what does it do? Do you use nano? If not what do you use?

It is a text editor in the terminal. I use nano frequently, I have tried vim but I am very bad at it.

1. \* is a form of a wildcard. What does it mean?

Zero or more characters

1. The following code would match what? ls \*.txt

List all files ending in .txt

1. Why do we need to be careful with the mv command?

It will replace a file with the same name

1. Why do we need to be careful with the rm command?

It can delete files permanently

1. What is the difference between > and >>?

> means put into this, and >> means add to this (first one will create new or replace everything, second one add to existing files)

1. What does head do?

Shows the top 10 lines of a file by default

1. What is the purpose of | (pipe)?\

To string together commands, feed one into another

1. Interpret the following command:

cat huge\_file.fasta | uniq | head -n 5 >proteins.fasta

print the top 5 unique lines of huge\_file.fasta into protein.fasta

1. What is a loop and when would you use it?

I loop run a command or commands through multiple items. You would use a loop if you want to do the same thing to multiple files or items, such as print out the name of multiple files.

1. In a directory with the following files. What would the following loop do?
2. $ for filename in \*.txt; do
3. > cat ${filename} >>fruits.txt
4. > done

This would print each filename of all the .txt files in the current folder and add it to the end of the file called fruits.txt

1. What is a shell script?

A small program

1. What are the benefits of writing and running a script over typing the code in?

If it is something you will use over and over, for example analyzing a specific type of data you collect, then it will save you (or someone else) lots of time over writing out the whole code each time.

1. When looking at a script what does # mean? Why would you use one?

It is a comment line and not read by the code. This way you can write in anything you want to help explain the code without affecting it.

1. What does grep stand for and what does it do?

Global regular expression print. It will pull out any specific thing you are looking for, such as all lines with ‘GAGA’

1. What does find do?

It locates files based on specific criteria

**Part II - Writing Code**

As part of this exercise please open the shell and practice moving around into different files. Best practice is to try a little bit every day. Try to challenge yourself by not using the mouse. For each of the questions below type your code and the result from the shell prompt.

1. Open the shell and change directories to the desktop and list the contents, paste your code here.

cd /mnt/c/Users/Cassa/Desktop/

drwxrwxrwx 1 ckhui ckhui 512 Jan 20 15:44 ./

drwxrwxrwx 1 ckhui ckhui 512 Dec 30 09:41 ../

drwxrwxrwx 1 ckhui ckhui 512 Nov 15 12:37 BioInfo/

-rwxrwxrwx 1 ckhui ckhui 2394 Aug 4 12:03 'Cassandra - Chrome.lnk'\*

-rwxrwxrwx 1 ckhui ckhui 2371 Jan 20 15:44 'GitHub Desktop.lnk'\*

-rwxrwxrwx 1 ckhui ckhui 1887 Jan 12 07:27 'Google Docs.lnk'\*

-rwxrwxrwx 1 ckhui ckhui 1899 Jan 12 07:27 'Google Sheets.lnk'\*

-rwxrwxrwx 1 ckhui ckhui 1899 Jan 12 07:27 'Google Slides.lnk'\*

-rwxrwxrwx 1 ckhui ckhui 282 Aug 2 17:44 desktop.ini\*

1. Draw the file structure from the root directory into one of the folders on the desktop. You can submit a pdf with a drawing. Select three locations and type the commands for changing directories to those locations use a combination of absolute and relative paths.
2. On your computer create a directory on your desktop. Create two files without opening them that have file extensions .txt. Then use wildcards to list the files in that directory. Paste you code here.

mkdir BIOL792\_2

cd BIOL792\_2

touch 1.txt 2.txt

ll \*.txt

1. Download the file [Hutia\_DNA.fasta](https://www.dropbox.com/s/07y610c8zt63nt5/Hutia_DNA.fasta?dl=0) from the Github repo. Make a new directory on the desktop. Move the file into that directory. List the contents of the directory. How many lines are in this file? Fasta files are a type of data file that holds DNA sequences. They are all formatted the same. The first line starts with a > and a name followed by a line break, and then DNA sequence followed by a line break. The next line starts with an > and another name followed by a line break and the DNA sequence. With that in mind, how could you use linux commands to determine how many sequences are in this file? How many sequences are in this file?

mkdir HW1

mv Hutia\_DNA.fasta HW1/

cd HW1/

ll

wc -l Hutia\_DNA.fasta

2382004 Hutia\_DNA.fasta

grep ">" Hutia\_DNA.fasta | wc -l

1191002

1. Build a single line of code that would take the first 100 sequences of this file and put them in a new file called ‘MyOutputFile.fasta’. Use a combination of wc, sort and uniq in a single line of code to tell me how many unique lines are in this file. Paste your commands and the answer here.

head -n 100 Hutia\_DNA.fasta > MyOutputFile.fasta

sort MyOutputFile.fasta | uniq | wc -l

100

(all lines are unique)

1. Write code that would create a loop to copy all the files in one directory ending in .txt to another directory. Paste the code here.

#!/usr/bin/sh

######

# to copy files from current folder to HW1 folder on Desktop

# Usage: sh cpfiles.sh

#

######

for file in \*.txt

do

#copy all files to designated location

cp $file /mnt/c/Users/Cassa/Desktop/HW1

done

1. Write a bash script with in-line documentation (hint #) to show how to run a made up python program (script) called 'find\_taxa.py' on set of files ending in ‘.fasta’. Show in the script how to run the program on each of those files. What is the name of your script? Type the script here.

#!/usr/bin/bash

###########

# This is to run find\_taxa.py on all .fasta files in current folder

# Usage: bash run\_find\_taxa.py

###############

bash find\_taxa.py \*.fasta

1. From the Hutia\_DNA.fasta file tell me how on how many lines do we find the pattern ‘GAGA’. What was the code used to find this?

grep 'GAGA' Hutia\_DNA.fasta | wc -l

304806