Day 2 Introduction to Unix/Linux Worksheet
TAs: Samuel Hunter and Ariel Eraso

1. Things to do before we begin
   a. Make sure your terminal (MobaXterm or Ubuntu for PC) is comfortable to work with
      i. Make window large
      ii. Customize your terminal (font, text colour and background)
         1. For Mac: press Command + Plus (increase text size) or Command + Minus (decrease text size). Further options are available in the **Terminal** dropdown menu, under **Preferences**
         2. For Windows: In MobaXterm hit Session in top left. From here click SSH in new window to open new session settings. Clicking on Terminal settings will present you with A Terminal font settings . Once done hit okay and enter login info here: Remote host * [ ] and hit okay to save changes.
   b. Find a cheat sheet for unix/linux
      i. https://files.fosswire.com/2007/08/fwunixref.pdf
      ii. http://www.mathcs.emory.edu/~valerie/courses/fall10/155/resources/unix_cheatsheet.html
   c. Find vim cheat sheet
      i. https://vim.rtorr.com/
      ii. https://www.cs.cmu.edu/~15131/f17/topics/vim/vim-cheatsheet.pdf
2. Log into the server using your username and the server information
   a. `ssh <username>@IP/server` (example `ssh rutendo@bonsai`)
3. Finding your bearings in unix
   a. Print your working directory (command `$ pwd`)
   b. To move through directories use `$ cd` (command to navigate to your home `$ cd ~`)
4. Creating directories
   a. Make a directory called "workshop-day2"

```
-bash-4.2$ mkdir workshop-day2
```

   b. Make a "bedfiles" directory in your "workshop-day2" directory

```
-bash-4.2$ mkdir workshop-day2/bedfiles
```

   c. Make more directories in your "workshop-day2" directory
      i. "bin"
      ii. "scripts"
      iii. "data"

```
[-bash-4.2$ cd workshop-day2/
[[rutendos@ip-172-31-12-93 workshop-day2]$ mkdir scripts bin data
```

5. List contents of your directory

ls	ls -l	ls -a	ls -h
List contents in a directory	Generate a detailed list of contents in a directory	Display all file including hidden files	Make information human readable

    a. Generate a list of directory contents

```
[rutendos@ip-172-31-12-93 workshop-day2]$ ls
bedfiles  bin  data  scripts
```

    b. Generate a detailed list of directory contents

```
[rutendos@ip-172-31-12-93 workshop-day2]$ ls -l
total 16
drwxrwxr-x 2 rutendos rutendos 6144 Jul  7 10:38 bedfiles
drwxrwxr-x 2 rutendos rutendos 6144 Jul  7 10:40 bin
drwxrwxr-x 2 rutendos rutendos 6144 Jul  7 10:40 data
drwxrwxr-x 2 rutendos rutendos 6144 Jul  7 10:40 scripts
```

    c. List directories using relative paths

```
[rutendos@ip-172-31-12-93 workshop-day2]$ ls .
bedfiles  bin  data  scripts
[rutendos@ip-172-31-12-93 workshop-day2]$ ls ..
workshop-day2
[rutendos@ip-172-31-12-93 workshop-day2]$ ls ../..
daramirez  hynesgra  jrdemasi  maallen3  magruca  mgruca13
[rutendos@ip-172-31-12-93 workshop-day2]$ ls ~
workshop-day2
[rutendos@ip-172-31-12-93 workshop-day2]$
```

6. Check where you are
    a. `$ pwd`

```
[rutendos@ip-172-31-12-93 workshop-day2]$ pwd
/Users/rutendos/workshop-day2
```

7. Moving files
    a. Copy chr1_bedfiles.tar.gz file into the "bedfiles" directory

```
$ cp /scratch/Workshop/SR2021/day2/bedfiles/chr1_bedfiles.zip bedfiles/
```

    b. Change name for "chr1_bedfiles.tar.gz" to "chr1bedfiles.tar.gz"

```
[[rutendos@ip-172-31-12-93 workshop-day2]$ cd bedfiles/
[[rutendos@ip-172-31-12-93 bedfiles]$ mv chr1_bedfiles.zip chr1bedfiles.zip
[rutendos@ip-172-31-12-93 bedfiles]$
```

8. Decompress the files
   a. Decompress the zipped file "chr1bedfiles.tar.gz"

```
[rutendos@ip-172-31-12-93 bedfiles]$ unzip chr1bedfiles.zip
Archive:  chr1bedfiles.zip
  inflating: ENCFF050DNG.chr1.bed
  inflating: ENCFF084PQ0.chr1.bed
  inflating: ENCFF635XTH.chr1.bed
  inflating: ENCFF837ENN.chr1.bed
  inflating: ENCFF857QTH.chr1.bed
```

9. Explore the files in the directory
   a. Get a detailed list (including hidden files) of files in the "bedfiles" directory and make
      the list is human readable

```
[[rutendos@ip-172-31-12-93 bedfiles]$ ls -lah
total 15M
drwxrwxr-x 2 rutendos rutendos 6.0K Jul  7 11:00 .
drwxrwxr-x 6 rutendos rutendos 6.0K Jul  7 10:40 ..
-rw-r--r-- 1 rutendos rutendos 3.5M Jul  7 10:52 chr1bedfiles.zip
-rw-rw-r-- 1 rutendos rutendos 1.1M Jun 28 15:40 ENCFF050DNG.chr1.bed
-rw-rw-r-- 1 rutendos rutendos 1.9M Jun 28 15:40 ENCFF084PQ0.chr1.bed
-rw-rw-r-- 1 rutendos rutendos 1.4M Jun 28 15:40 ENCFF635XTH.chr1.bed
-rw-rw-r-- 1 rutendos rutendos 3.4M Jun 28 15:40 ENCFF837ENN.chr1.bed
-rw-rw-r-- 1 rutendos rutendos 2.9M Jun 28 15:40 ENCFF857QTH.chr1.bed
```

   b. View the first 20 lines of the ENCFF050DNG.chr1.bed file (this will show you the first
      n lines of a file)

```
-bash-4.2$ head -20 ENCFF050DNG.chr1.bed
chr1    234741661    234750175    Rank_9   403    .    10.57151       40.39918       37.28195
chr1    16160808     16164611     Rank_10  403    .    9.05042 40.33364       37.35299
chr1    161368080    161371005    Rank_21  379    .    11.07757       37.93226       34.94309
chr1    245132770    245135764    Rank_28  367    .    8.70089 36.74874       33.80583
chr1    59247400     59252652     Rank_43  346    .    8.24778 34.66340       31.72422
chr1    242010975    242011616    Rank_47  344    .    8.72231 34.43312       31.57939
chr1    245025125    245029064    Rank_53  338    .    8.40179 33.86665       30.98131
chr1    26945955     26948801     Rank_69  323    .    8.44152 32.36346       29.54492
chr1    243416537    243420024    Rank_71  322    .    8.99760 32.29551       29.46327
chr1    113160256    113163148    Rank_80  315    .    8.45791 31.59736       28.81804
chr1    150549459    150553721    Rank_83  314    .    9.11023 31.48746       28.69695
chr1    51442087     51444900     Rank_88  309    .    9.76360 30.98812       28.17010
chr1    249131278    249134036    Rank_90  309    .    7.37405 30.92263       28.16320
chr1    10269750     10272875     Rank_98  303    .    8.29112 30.35679       27.59029
chr1    46151106     46154352     Rank_104      301    .    8.22899 30.17658       27.43262
chr1    93643659     93647626     Rank_118      295    .    7.65178 29.58767       26.91262
chr1    78443815     78445923     Rank_126      293    .    7.91636 29.39694       26.68419
chr1    62900986     62904106     Rank_143      288    .    7.68062 28.83642       26.12630
chr1    244996279    245000172    Rank_168      282    .    9.78096 28.22511       25.51217
chr1    36395693     36397642     Rank_169      282    .    8.38384 28.21035       25.45159
```

   c. View the file using each of the commands
      i.   tail
      ii.   more
      iii. less
      iv.   cat (this will print the whole file, more or less is better)

head	**tail**	**more**	**less**	**cat**
Print the first 10 lines of a file	Print the last 10 lines of a file	Filter for paging through text one screenful at a time.	Allows backward movement in the file as well as forward movement (opposite of more)	Concatenate files and print on the standard output
`-n <int>` specify how many lines to be printed	`-n <int>` specify how many lines to be printed	`SPACE` to scroll `q` to quit	`SPACE` or `^V` or `f` or `^F` or `arrows` to scroll `q` to quit	Can concatenate multiple files

    d. Scroll through ENCFF837ENN.chr1.bed with `more`

       ● To scroll with more use the <SPACE> bar
       ● More only allows scrolling in one direction

    e. Scroll through ENCFF837ENN.chr1.bed with `less`

- `less` is more in that it has more functionality and flexibility than `more`
- To score with less you can use arrow keys

10. File manipulation

cut	cut -f	cut -d
Remove sections from each line of files	Select specific fields in a file	Use DELIM instead of TAB for field delimiter

a. Select the first 3 columns of the ENCFF837ENN.chr1.bed bedfile.

11. Search files

find	grep	grep -i	grep -v	grep -c	grep -w
Search for	Search for	Ignore case	Select lines	Count lines	Select only

files in a directory hierarchy	PATTERNS in each file	(upper versus lower case)	that do not match PATTERN (inverse match)	that match PATTERN	those lines containing matches that form whole words.

a. Search for the ENCFF050DNG.chr1.bed file in the bedfiles directory

```
[rutendos@ip-172-31-12-93 ~]$ find -iname ENCFF050DNG.chr1.bed
./workshop-day2/bedfiles/ENCFF050DNG.chr1.bed
```

- ● This is very helpful for searching where a file is located

b. Search for lines with the character '151' in ENCFF050DNG.chr1.bed

```
rutendosigauke — rutendos@ip-172-31-12-93:~/workshop-day2/bedfiles — ssh rutendos@3.15.81.98 — 106×23
[[rutendos@ip-172-31-12-93 bedfiles]$ cat ENCFF050DNG.chr1.bed | grep -w '151'
chr1    160311944       160315047       Rank_2634    151    .    5.88029 15.17147        12.93766
chr1    156443033       156483004       Rank_2640    151    .    5.28292 15.15764        12.89305
chr1    153698824       153703698       Rank_2649    151    .    4.63789 15.13737        12.98069
chr1    11332580        11334232        Rank_2650    151    .    5.79466 15.13692        12.84683
chr1    212207948       212209443       Rank_2669    151    .    5.53319 15.10618        12.80687
chr1    8874582 8878692 Rank_2671       151    .    5.55842 15.10148        12.85965
```

12. Counting lines/features in unix

wc	wc -l	uniq -c	grep -c
Print newline, word, and byte counts for each file	Print the newline counts	Prefix lines by the number of occurrences	Count lines that match PATTERN

```
[rutendos@ip-172-31-12-93 data]$ wc SRR4090090_chr1.bedGraph
  4160947  12482841 100338164 SRR4090090_chr1.bedGraph
[rutendos@ip-172-31-12-93 data]$ wc -l SRR4090090_chr1.bedGraph
4160947 SRR4090090_chr1.bedGraph
[rutendos@ip-172-31-12-93 data]$ grep -c 'chr1' SRR4090090_chr1.bedGraph
4160947
```

13. Sorting files
   a.

sort	sort -n	sort -k
Sorts lines of text files	Numeric sort	sort via a key, key defined as location and type

```
[rutendos@ip-172-31-12-93 bedfiles]$ sort ENCFF050DNG.chr1.bed | head
chr1    100014682       100016094       Rank_32110      34      .       3.07636 3.46227 1.74507
chr1    10002601        10004373        Rank_7870       90      .       4.35943 9.05913 6.97220
chr1    100045653       100046535       Rank_110101     16      .       1.99599 1.66912 0.22234
chr1    100055083       100057012       Rank_28624      37      .       3.23170 3.73726 1.96705
chr1    100064610       100066587       Rank_67204      22      .       2.40614 2.25161 0.66713
chr1    10006485        10006980        Rank_141882     13      .       1.89740 1.37608 0.08605
chr1    100068791       100069860       Rank_22834      43      .       3.68465 4.39127 2.57169
chr1    100096332       100096501       Rank_127196     15      .       1.96071 1.52614 0.10746
chr1    100101667       100102685       Rank_88145      19      .       2.28319 1.92411 0.38577
chr1    10010334        10011733        Rank_37652      31      .       2.84001 3.13169 1.41958
[rutendos@ip-172-31-12-93 bedfiles]$ sort -k1,1 -k2,2n ENCFF050DNG.chr1.bed | head
chr1    10048   10288   Rank_45352      28      .       2.24384 2.81052 1.16802
chr1    566741  566895  Rank_100426     17      .       2.05015 1.77373 0.28207
chr1    567461  567697  Rank_40653      29      .       1.86131 2.99347 1.30290
chr1    569809  570042  Rank_46482      27      .       1.80564 2.77234 1.11190
chr1    713170  714757  Rank_25024      41      .       2.89645 4.11096 2.33642
chr1    729739  729996  Rank_127263     15      .       2.02760 1.52558 0.12187
chr1    750107  750247  Rank_46301      27      .       2.84820 2.77775 1.03236
chr1    752177  756117  Rank_22910      43      .       3.45087 4.37894 2.64764
chr1    758008  759100  Rank_159163     11      .       1.67400 1.12286 0.00000
chr1    760404  761440  Rank_124670     15      .       1.99384 1.55038 0.13306
```

14. Standard out and standard error
   a. `>` Write standard out to a file
   b. `>>` Append standard out to a file
   c. `2>` Write standard error to a file
15. Running multiple commands at the sequentially
   a. Multiple commands can be run using pipes "|"
      i. Open a file with cat, get the first and third column while grabbing lines with a "FEATURE" and writing the standard output to OUTFILE.txt

      cat FILE | cut -f 1,3 | grep "FEATURE" > OUTFILE.txt

   b. Example of running multiple commands on a bedgraph file
      i. Copy `/scratch/Workshop/SR2021/day2/bedfiles/SRR4090090_chr21_2 2.bedGraph.gz` to your `bedfiles` directory
      ii. Decompress the file (hint: check the file extension)

      iii. Open the file and filter chr21 coordinates from the bedgraph and write the file with only columns 1,2 and 3 to your data folder
```
[rutendos@ip-172-31-12-93 ~]$ cd /Users/rutendos/workshop-day2/bedfiles/
[rutendos@ip-172-31-12-93 bedfiles]$ head SRR4090090_chr1.bedGraph
chr1    10059   10068   2
chr1    10068   10071   4
chr1    10071   10082   6
chr1    10082   10083   9
chr1    10083   10089   15
chr1    10089   10110   17
chr1    10110   10114   15
chr1    10114   10115   18
chr1    10115   10118   17
chr1    10118   10119   20
[rutendos@ip-172-31-12-93 bedfiles]$ cat SRR4090090_chr1.bedGraph | cut -f 1,2,3 | grep 'chr1' >]
../data/SRR4090090_chr1.bedGraph
```

16. Using vim
   a. Create a file called vim_example.txt in your home directory
      i. **touch vim_example.txt**
   b. Open the file with vim

      i.   **`vim vim_example.txt`**

Vim has 2 operating modes:

Command mode (default) (Vim commands only)

    1. Insert mode (editing)

`i/I`   ➢ insert

`o/O`   ➢ open new line

`a/A`   ➢ append

    2. Switch back to command mode

`esc`

To save (in command mode):

:w

To save the file and exit vim (in command mode):

:wq

To exit without saving the file (in command mode):

:q!

a. Other helpful/useful unix hints
   i. **`Ctrl + r:`** search your command history
   ii. **`Up/down arrow keys`**: scroll through your command history, one at a time
   iii. **`Ctrl + a`**: Move your cursor to the front of the line
   iv. **`Ctrl + e`**: Move your cursor to the end of the line
   v. **`Tab`**: Auto-completes file name, if there is only one matching file. If there are multiple files matching your pattern, press Tab again to see a list of matching files
b. Other helpful vim hints
```
In Command Mode:
gg – Move to first line of the file
G – Move to last line of the file
V – Highlight lines
y – Yank (copy) highlighted lines
d – Cut highlighted lines
p – Paste cut/copied lines
```

c. Editing the SRR4090090_chr21_22.bedGraph
   i. See slides