

NGS Bioinformatics

Practical assignment Linux_Session1 MEMORANDUM

Module topic: Linux

Contact session title: Introduction to the command line

Trainer: Amel Ghouila and Sunir Panj

Introduction to the command line

Introduction

The aim of this assignment is to practice the command lines we covered during this first session of the module.

Tools used in this session

You'll be using only your terminal.

Please note

- **Hand-in information** please upload your completed assignment to the Vula 'Assignments' tab. Take note of the final hand-in date for each assignment, which will be indicated on Vula.

Instructions

To begin, open a new terminal window . For each of the questions, type the command you used. Remember to always check the content of directories and files to understand content and formats before starting to use them and extract information.

1. What is your current directory? What's it's absolute path?

Home directory : `home/manager`

2. How can you check your current working directory?

`pwd`

3. Move to the unix directory then complete the exercise below



```
cd ./course_data/unix
```

```
Or cd /home/manager/course_data/unix
```

4. Create a directory Assignment under Session1 using a single command (remember Session1 is under /unix/practical/)

```
mkdir ./practical/Session1/Assignment
```

5. Move to the directory you just created

```
cd ./practical/Session1/Assignment
```

6. Copy the file PccAS_v3.gff3 located under ~/course_data/rna_seq_pathogen/data to Assignment. Write 2 different possible commands to do this.

```
cp ~/course_data/rna_seq_pathogen/data/PccAS_v3.gff3
```

7. How can you check that the file has been properly copied?

use command `ls` or `less PccAS_v3.gff3` (or any other command to read file content)

8. What's the size of the file PccAS_v3.gff3? Type both the command you used to get the information and the size

```
ls -lh
```

File size: 2.1M

9. How many lines does PccAS_v3.gff3 contain?

```
wc -l PccAS_v3.gff3
```

Number of lines: 29614

10. Display the first 15 lines of PccAS_v3.gff3

```
head -n15 PccAS_v3.gff3
```

11. Display all the lines containing details of all genes contained in PccAS_v3.gff3?

Please note genes are a type of feature. More details about gff3 format here:

<https://learn.gencore.bio.nyu.edu/ngs-file-formats/gff3-format/>

```
grep gene PccAS_v3.gff3
```



12. Create a new file excluding all the gene features and name this file
PccAS_v3_withoutgenes.gff3

```
grep -v gene PccAS_v3.gff3 > PccAS_v3_withoutgenes.gff3
```

13. Rename the resulting file PccAS_v3_withoutgenes.gff3 to question_12_results.gff3

```
mv PccAS_v3_withoutgenes.gff3 question_12_results.gff3
```

14. How many CDS does PccAS_v3.gff3 contain?

- Write 2 separate commands to do this
- Combine 2 commands using |
- Write one single command

```
a) grep CDS PccAS_v3.gff3 > PccAS_v3_CDS.gff3
```

```
wc -l PccAS_v3_CDS.gff3
```

```
b) grep CDS PccAS_v3.gff3 | wc -l
```

```
c) grep -c CDS PccAS_v3.gff3
```

15. Extract all information sequence ID: PccAS_01_v3 and copy it to a file
PccAS_01_v3.gff3

```
grep PccAS_01_v3 PccAS_v3.gff3 > PccAS_01_v3.gff3
```

16. Write a command to display the names of files ending with .gff3 under the directory
Assignment using wildcards

```
find -name "*.gff3"
```

17. Create a subdirectory Genomics under Assignment

```
mkdir Genomics
```

18. Download Plasmodium falciparum fasta file available at:

http://plasmodb.org/common/downloads/release-9.0/Pfalciparum/fasta/PlasmoDB-9.0_Pfalciparum_BarcodeIsolates.fasta

```
wget http://plasmodb.org/common/downloads/release-9.0/Pfalciparum/fasta/PlasmoDB-9.0\_Pfalciparum\_BarcodeIsolates.fasta
```



19. How many lines does the file contain?

`wc -l PlasmoDB-9.0_Pfalciparum_BarcodeIsolates.fasta`

20. Save all the commands you type to a file named Assignment1_commands

`history > Assignment1_commands`

Part 1: participant's answer

<type your answers here. For most of the questions, you need to enter the command you used >

