

Module 01 Activity

Module 01: Introduction to the Linux command line

Introduction

The purpose of this assignment is to help you practice the Linux command-line skills introduced during the first session of this module. It is designed to reinforce your ability to navigate the file system, inspect files, and apply common terminal commands independently.

Please note

For this assignment, please complete all your work in your **personal scratch space**, located at:

```
/scratch/${USER}
```

This is your temporary working area on the Explorer HPC system.

For each of the questions, **type the command you used**. Remember to always check the content of directories and files to understand the content and formats before starting to use them and extracting information.

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

2. How can you check your current working directory?

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

4. Create a directory called activity1

5. Move to the directory you just created

6. Copy the file PccAS_v3.gff3 located under [/courses/BINF6310.202610/data/module04](#) to activity1. Write 2 different possible commands to do this.

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

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

9. How many lines does PccAS_v3.gff3 contain?

10. Display the first 15 lines of PccAS_v3.gff3. (command and output)

11. Display all the lines containing details of all genes contained in PccAS_v3.gff3? (command only).

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/>

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

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

14. How many CDS does PccAS_v3.gff3 contain?

- a. Write 2 separate commands to do this
- b. Combine 2 commands using |
- c. Write one single command

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

16. Write a command to display the names of files ending with .gff3 under the directory activity1 using wildcards

17. Create a subdirectory bioinformatics under activity1

18. Download Plasmodium falciparum fasta file available at:

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

19. How many lines does the file contain?

20. Save all the commands you type to a file named activity1_commands (hint – ‘history’)
