**NGS Bioinformatics**

**Practical assignment Linux\_Session1**

**Module topic: Linux**

**Contact session title: Introduction to the command line**

**Trainer: Amel Ghouila**

**Participant:** <*write your name here>*

**Date:** <*write today’s date here*>

**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?
2. How can you check your current working directory?
3. Move to the unix directory then complete the exercise below
4. Create a directory Assignment under Session1 (remember Session1 is under /unix/practical/)
5. Move to the directory you just created
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.
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
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/

1. Create a new file excluding all the gene features and name this file PccAS\_v3\_withoutgenes.gff3
2. Rename the resulting file PccAS\_v3\_withoutgenes.gff3 to question\_12\_results.gff3
3. How many CDS does PccAS\_v3.gff3 contain?
   1. Write 2 separate commands to do this
   2. Combine 2 commands using |
   3. Write one single command
4. Extract all information sequence ID: PccAS\_01\_v3 and copy it to a file PccAS\_01\_v3.gff3
5. Write a command to display the names of files ending with .gff3 under the directory Assignment using wildcards
6. Create a subdirectory Genomics under Assignment
7. Download Plasmodium falciparum fasta file available at: <http://plasmodb.org/common/downloads/release-9.0/Pfalciparum/fasta/PlasmoDB-9.0_Pfalciparum_BarcodeIsolates.fasta>
8. How many lines does the file contain?
9. Save all the commands you type to a file named Assignement1\_commands

**Part 1: participant’s answer**

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

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_