Stage 3

Project 1: BASh Basic

Participants who contributed significantly (slack handle alone): @MernaSalem @Sarani

1. Login to your coding workspace
2. Create a folder titled your name

Sarani:

```
$ mkdir Sarani
```

Merna:

```
$ mkdir MernaSalem
```

3. Create another new directory titled biocomputing and change to that directory with one line of command

```
$  mkdir biocomputing ; cd biocomputing
```

4. Download these 3 files:
   a. https://raw.githubusercontent.com/josoga2/dataset-repos/main/wildtype.fna
   b. https://raw.githubusercontent.com/josoga2/dataset-repos/main/wildtype.gbk
   c. https://raw.githubusercontent.com/josoga2/dataset-repos/main/wildtype.gbk

```
$ wget https://raw.githubusercontent.com/josoga2/dataset-repos/main/wildtype.fna > wildtype.fna ; wget https://raw.githubusercontent.com/josoga2/dataset-repos/main/wildtype.gbk > wildtype.gbk ; wget https://raw.githubusercontent.com/josoga2/dataset-repos/main/wildtype.gbk > wildtype.gbk
```

5. OH! You made a mistake. You have to move the .fna file to the folder titled your name directly. (Do this with one command. Hint: See our cheatsheet)

```
$ mv wildtype.fna /home/sarani_admin/Sarani
```

```
$ mv wildtype.fna /home/merna_admin/MernaSalem
```

6. OH No! The gbk file is a duplicate, they are actually the same thing. Please delete it.

```
$ rm wildtype.gbk
```

7. The .fna file is actually from a bacteria, and it should definitely have a TATA (tata) box for initiating gene transcription. The molecular biologist is trying to understand the implication of dual TATA sequences. The files got mixed up and we are not sure which is wildtype and which is mutant. The mutant should have “tatatata” while the normal should have just “tata”. Can you confirm if the file is mutant or wild type

```
$ grep -i tatatata wildtype.fna
```

   The file is mutant type, as this command shows the presence of the tatatata pattern in it..

8. If it is mutant, print all the lines that show it is a mutant into a new file

```
$ grep -i tatatata wildtype.fna.1 > mutant_lines.fna
```

9. What is your favorite gene? (In any organism). Each team member should pick a unique gene different from every other person

Sarani: LOC103165089 protein LEG1 homolog [Ornithorhynchus anatinus (platypus)]
Merna: X91251.1 H.sapiens mRNA for BDNF protein

10. Download the fasta format of the gene from NCBI Nucleotide

```
$  wget
"https://www.ncbi.nlm.nih.gov/sviewer/viewer.fcgi?id=NM_00132377
6.1&db=nuccore&report=fasta&retmode=text" -O sequence.fasta
```

Merna:

```
$ wget
"https://www.ncbi.nlm.nih.gov/sviewer/viewer.fcgi?id=X91251.1&db
=nuccore&report=fasta&retmode=text" -O sequence.fasta
```

11. How many lines are in the FASTA file (with the exception of the header)

Sarani: 20
Merna: 20

```
$ echo $(( $(wc -l < sequence.fasta) - 2 ))
```

12. How many times does A occur

Sarani: 324 times
Merna: 361

```
$ sed 1d sequence.fasta | grep -o -i 'A' | wc -l
```

13. How many times does G occur

Sarani: 305 times
Merna: 358

```
$ sed 1d sequence.fasta | grep -o -i 'G' | wc -l
```

14. How many times does C occur

Sarani: 389 times
Merna: 315

```
$ sed 1d sequence.fasta | grep -o -i 'C' | wc -l
```

15. How many times does T occur

Sarani: 317 times
Merna: 301

```
$ sed 1d sequence.fasta | grep -o -i 'T' | wc -l
```

16. Calculate the %GC content of your gene

Sarani: 51%
Merna: 50%

```
$ sudo apt install bc

gc_content=$(sed 1d sequence.fasta | grep -o -i '[GC]' | wc -l)

total_bases=$(sed 1d sequence.fasta | grep -o -i '[AGTC]' | wc
-l)

gc_percentage=$(echo "scale=2; $gc_content / $total_bases * 100"
| bc)
```

```
echo "GC content: $gc_percentage%"
```

17. Create a nucleotide (.fasta) file title your name

Sarani:

```
$ cp sequence.fasta Sarani.fasta
```

Merna:

```
$ cp sequence.fasta Merna.fasta
```

18. "`echo`" the following into the file using `>>:` `the number of A, G, T and C in the file you created above.`

Sarani:

```
$ num_A=$(sed 1d sequence.fasta | grep -o -i 'A' | wc -l)

num_G=$(sed 1d sequence.fasta | grep -o -i 'G' | wc -l)

num_T=$(sed 1d sequence.fasta | grep -o -i 'T' | wc -l)

num_C=$(sed 1d sequence.fasta | grep -o -i 'C' | wc -l)

echo "Number of A: $num_A" >> Sarani.fasta

echo "Number of G: $num_G" >> Sarani.fasta

echo "Number of T: $num_T" >> Sarani.fasta

echo "Number of C: $num_C" >> Sarani.fasta
```

Merna:

```
$ num_A=$(sed 1d Merna.fasta | grep -o -i 'A' | wc -l)
num_G=$(sed 1d Merna.fasta | grep -o -i 'G' | wc -l)
num_T=$(sed 1d Merna.fasta | grep -o -i 'T' | wc -l)
num_C=$(sed 1d Merna.fasta | grep -o -i 'C' | wc -l)
echo "Number of A: $num_A" >> Merna.fasta
echo "Number of G: $num_G" >> Merna.fasta
echo "Number of T: $num_T" >> Merna.fasta
echo "Number of C: $num_C" >> Merna.fasta
```

19. Upload the file to your team's github repo in a folder called /output

Sarani:

```
$ git clone https://github.com/Sarani-NS/HackBio
cd HackBio/
cd Task3
mkdir -p output
cd ../
mv /c/Users/sante/Sarani/Sarani.fasta ./output/
git add output/Sarani.fasta
git commit -m "Ajout du fichier Sarani.fasta"
git push origin main

https://github.com/Sarani-NS/HackBio/tree/main/Task3/output
```

Merna:

```
$
https://github.com/Sarani-NS/HackBio/blob/main/Task3/output/Mern
a.fasta
```

20. Save all the codes you have used in this project in a file named yourname.sh Upload all the codes you have used to your team's github repo in a folder called /script

Sarani:

```
$ nano Sarani.sh
mkdir -p ./script
mv /c/Users/sante/Sarani/HackBio/Task3/Sarani.sh ./script/
cd ../
git add Task3/script/Sarani.sh
git commit -m "Added the script file Sarani.sh"
git push origin main

https://github.com/Sarani-NS/HackBio/tree/main/Task3/script
```

Merna:

```
$
https://github.com/Sarani-NS/HackBio/blob/main/Task3/script/Mern
a.sh
```

21. Clear your terminal space and print all the commands you have used today.

```
$ clear && history
```

22. List the files in the two folders and share a screenshot of your terminal below

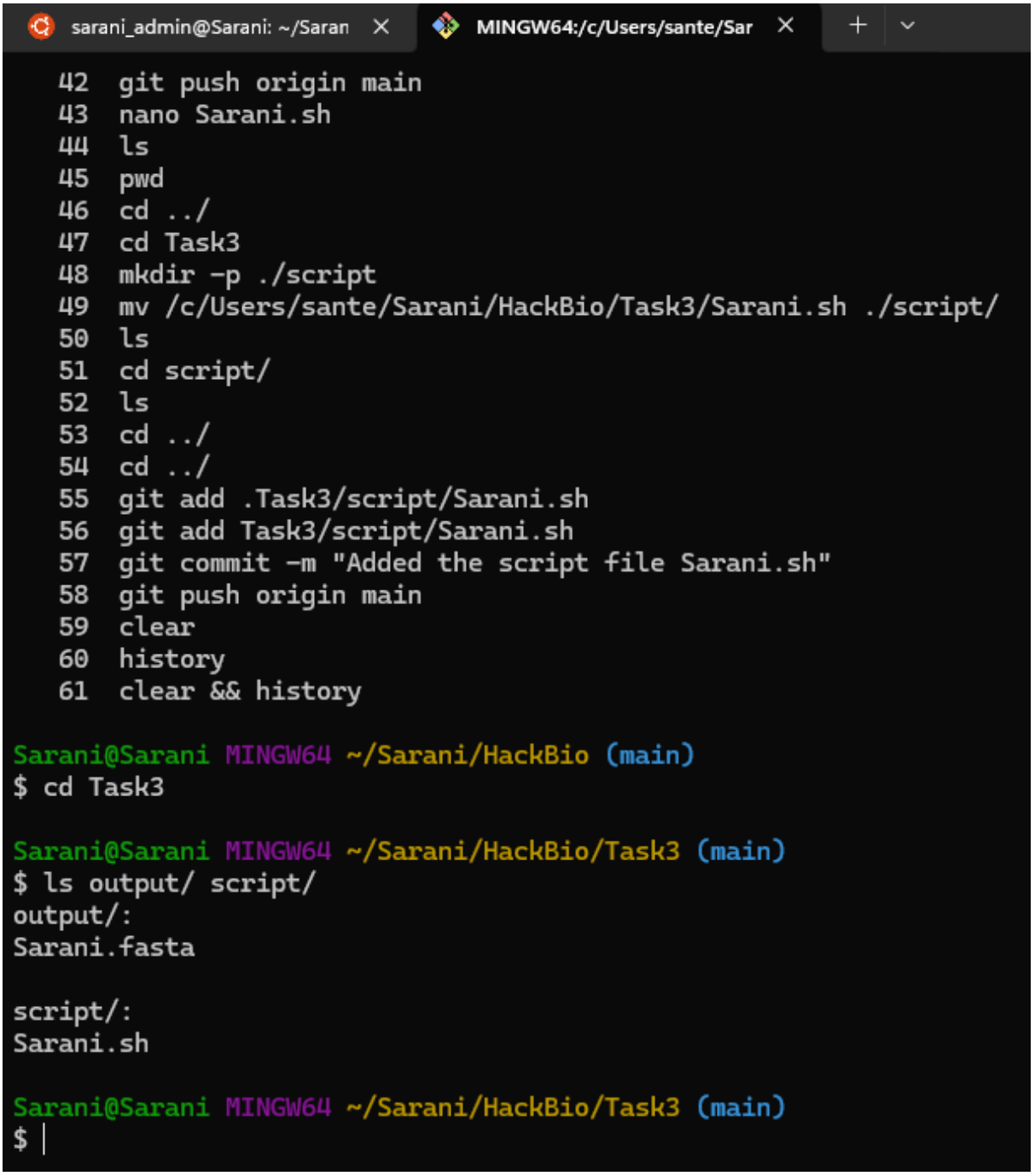
```
$ ls output/ script/
```

```
Sarani@Sarani MINGW64 ~/Sarani/HackBio/Task3 (main)
$ ls output/ script/
output/:
Sarani.fasta

script/:
Sarani.sh
```

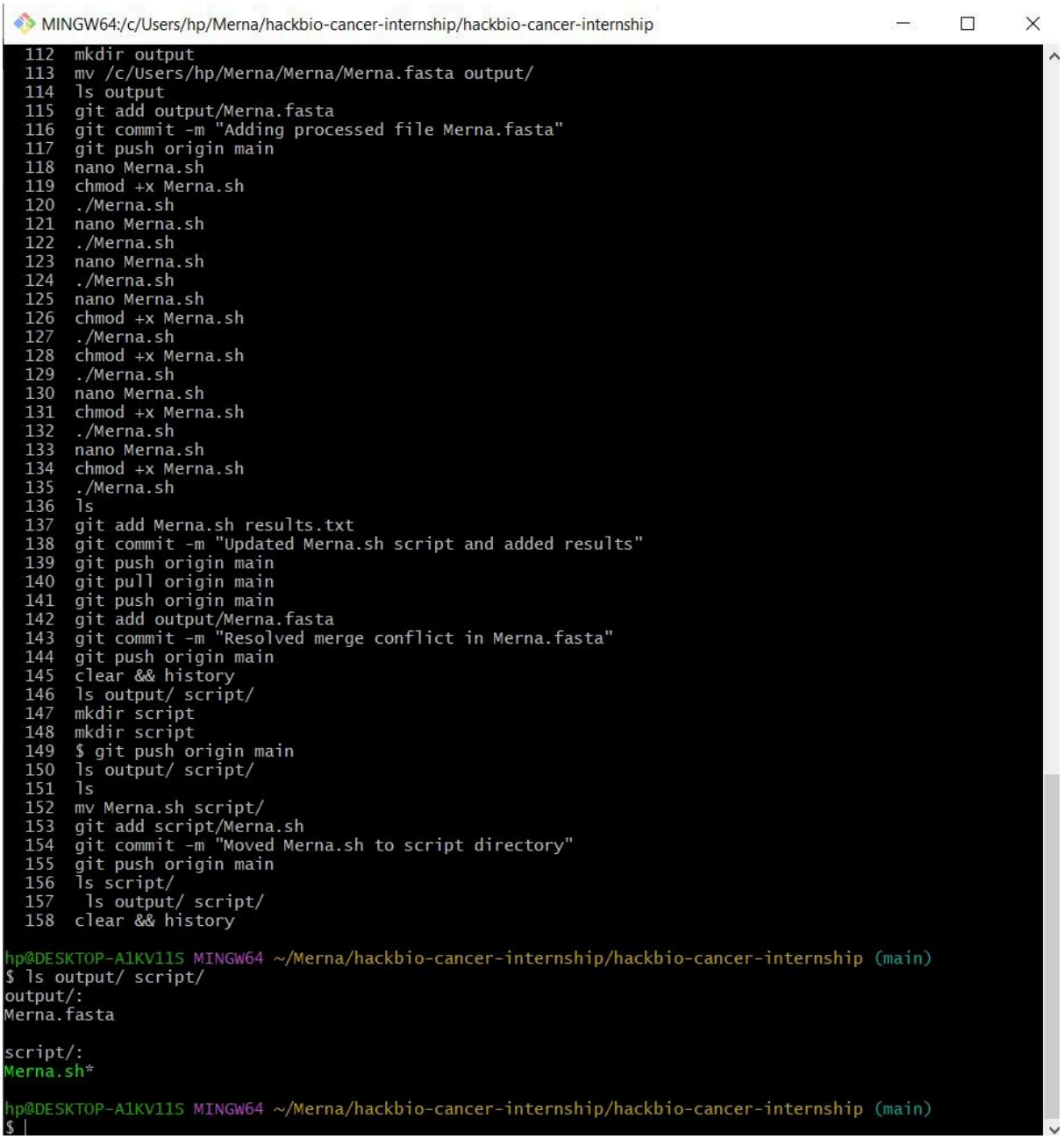
Merna:

```
hp@DESKTOP-A1KV11S MINGW64 ~/Merna/hackbio-cancer-internship/hackbio-cancer-internship (main)
$ ls output/ script/
output/:
Merna.fasta

script/:
Merna.sh*

hp@DESKTOP-A1KV11S MINGW64 ~/Merna/hackbio-cancer-internship/hackbio-cancer-internship (main)
$
```

23. Take a screenshot of your terminal screen currently and paste it below


sarani_admin@Sarani: ~/Saran
MINGW64:/c/Users/sante/Sar
42 git push origin main
43 nano Sarani.sh
44 ls
45 pwd
46 cd ../
47 cd Task3
48 mkdir -p ./script
49 mv /c/Users/sante/Sarani/HackBio/Task3/Sarani.sh ./script/
50 ls
51 cd script/
52 ls
53 cd ../
54 cd ../
55 git add .Task3/script/Sarani.sh
56 git add Task3/script/Sarani.sh
57 git commit -m "Added the script file Sarani.sh"
58 git push origin main
59 clear
60 history
61 clear && history
Sarani@Sarani MINGW64 ~/Sarani/HackBio (main)
$ cd Task3
Sarani@Sarani MINGW64 ~/Sarani/HackBio/Task3 (main)
$ ls output/ script/
output/:
Sarani.fasta
script/:
Sarani.sh
Sarani@Sarani MINGW64 ~/Sarani/HackBio/Task3 (main)
$

Merna:


MINGW64:/c/Users/hp/Merna/hackbio-cancer-internship/hackbio-cancer-internship
112 mkdir output
113 mv /c/Users/hp/Merna/Merna/Merna.fasta output/
114 ls output
115 git add output/Merna.fasta
116 git commit -m "Adding processed file Merna.fasta"
117 git push origin main
118 nano Merna.sh
119 chmod +x Merna.sh
120 ./Merna.sh
121 nano Merna.sh
122 ./Merna.sh
123 nano Merna.sh
124 ./Merna.sh
125 nano Merna.sh
126 chmod +x Merna.sh
127 ./Merna.sh
128 chmod +x Merna.sh
129 ./Merna.sh
130 nano Merna.sh
131 chmod +x Merna.sh
132 ./Merna.sh
133 nano Merna.sh
134 chmod +x Merna.sh
135 ./Merna.sh
136 ls
137 git add Merna.sh results.txt
138 git commit -m "Updated Merna.sh script and added results"
139 git push origin main
140 git pull origin main
141 git push origin main
142 git add output/Merna.fasta
143 git commit -m "Resolved merge conflict in Merna.fasta"
144 git push origin main
145 clear && history
146 ls output/ script/
147 mkdir script
148 mkdir script
149 $ git push origin main
150 ls output/ script/
151 ls
152 mv Merna.sh script/
153 git add script/Merna.sh
154 git commit -m "Moved Merna.sh to script directory"
155 git push origin main
156 ls script/
157 ls output/ script/
158 clear && history
hp@DESKTOP-A1KV11S MINGW64 ~/Merna/hackbio-cancer-internship/hackbio-cancer-internship (main)
$ ls output/ script/
output/:
Merna.fasta
script/:
Merna.sh*
hp@DESKTOP-A1KV11S MINGW64 ~/Merna/hackbio-cancer-internship/hackbio-cancer-internship (main)
$

Project 2: Installing Bioinformatics Softwares on the terminal

1. Activate your base conda environment

```
$ conda activate
```

2. Create a conda environment names funtools

```
$ conda create --name funtools
```

3. Activate the funtools environment

```
$ conda activate funtools
```

4. Install Figlet using conda

```
$ conda config --add channels conda-forge
conda install pyfiglet
```

Used an alternative found on github, as the original package wasn't installable using conda (even after trying on the official anaconda website, on the right):

```
PackagesNotFoundError: The following packages are no
t available from current channels:

  - figlet

Current channels:

  - https://repo.anaconda.com/pkgs/main/win-64
  - https://repo.anaconda.com/pkgs/main/noarch
  - https://repo.anaconda.com/pkgs/r/win-64
  - https://repo.anaconda.com/pkgs/r/noarch
  - https://repo.anaconda.com/pkgs/msys2/win-64
  - https://repo.anaconda.com/pkgs/msys2/noarch
```

```
PackagesNotFoundError: The following packages are no
t available from current channels:

  - tsnyder::figlet

Current channels:

  - https://repo.anaconda.com/pkgs/main/win-64
  - https://repo.anaconda.com/pkgs/main/noarch
  - https://repo.anaconda.com/pkgs/r/win-64
  - https://repo.anaconda.com/pkgs/r/noarch
  - https://repo.anaconda.com/pkgs/msys2/win-64
  - https://repo.anaconda.com/pkgs/msys2/noarch
```

5. Run the following command figlet {your name}. Put a screenshot of what you see below 😀

6. Install bwa through the bioconda channel

```
$ conda config --add channels bioconda
conda install bwa
```

7. Install blast through the bioconda channel

```
$ conda install -c bioconda blast
```

8. Install samtools through the bioconda channel

```
$ conda install -c bioconda samtools
```

9. Install bedtools through the bioconda channel

```
$ conda install -c bioconda bedtools
```

10. Install spades.py through the bioconda channel

```
$ conda install -c bioconda spades.py
```

11. Install bcftools through the bioconda channel

```
$ conda install -c bioconda bcftools
```

12. Install fastp through the bioconda channel

```
$ conda install -c bioconda fastp
```

13. Install multiqc through the bioconda channel

```
$  conda install -c bioconda multiqc
```