Stage 1

Project 1: BASh Basic

You are to achieve this short story with the command line alone.
Create your copy of the file and enter your command in the terminal space ($) below each action.
Your Team Name:
Participants who contributed significantly (slack handle alone):

N/B: The story here is fictional and the files are just hypothetical. Please don't use it for any serious research work.
Please copy exactly what worked. Do not paraphrase. A single mismatch makes you loose your point.
To submit this project, make this document open using the 🔒**share icon** at the top right corner. Copy the link and submit it on HackBio platform (See submission Guide).

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

```
$ mkdir Sopuruchi_Mba
```

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
$ wget https://raw.githubusercontent.com/josoga2/dataset-repos/main/wildtype.gbk
$ wget https://raw.githubusercontent.com/josoga2/dataset-repos/main/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/sopuruchilizbeth/Sopuruchi_Mba
```

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

```
$ rm wildtype.gbk1
```

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

```
$ if grep -q "tatatata" wildtype.fna; then echo "mutant"; else
echo "wildtype"; fi
```

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

```
$ grep "tatatata" wildtype.fna > mutant.fna
```

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

```
$ PDCD1
```

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

```
$ wget
"https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nu
cleotide&id=NM_005018.3 &rettype=fasta&retmode=text" -O
pdcd1gene.fasta
```

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

```
$ grep -c -v "^>" pdcd1gene.fasta
```

12. How many times does A occur

```
$ grep -v ">" pdcd1gene.fasta | tr -d '\n' | awk -F "A"
'{count+=NF-1} END {print count}'
```

13. How many times does G occur

```
$ grep -v ">" pdcd1gene.fasta | tr -d '\n' | awk -F "G"
'{count+=NF-1} END {print count}'
```

14. How many times does C occur

```
$ grep -v ">" pdcd1gene.fasta | tr -d '\n' | awk -F "C"
'{count+=NF-1} END {print count}'
```

15. How many times does T occur

```
$ grep -v ">" pdcd1gene.fasta | tr -d '\n' | awk -F "T"
'{count+=NF-1} END {print count}’
```

16. Calculate the %GC content of your gene

```
$ awk '!/^>/{n+=length($0); g+=gsub(/[GgCc]/,"") } END{print g/n
* 100 "%"}' pdcd1gene.fasta
```

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

```
$ touch sopuruchi.fasta
```

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

```
$ echo 'A count is: 382' >> sopuruchi.fasta && echo 'G count is:
651' >> sopuruchi.fasta && echo 'T count is: 351' >>
sopuruchi.fasta && echo 'C count is: 713' >> sopuruchi.fasta
```

```
$ link
https://github.com/Sopuruchilizbeth/Hackbio_Internship-/tree/mai
n/Stage%201/Output
```

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

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`

```
$ link
https://github.com/Sopuruchilizbeth/Hackbio_Internship-/tree/mai
n/Stage%201/Script%20here
```

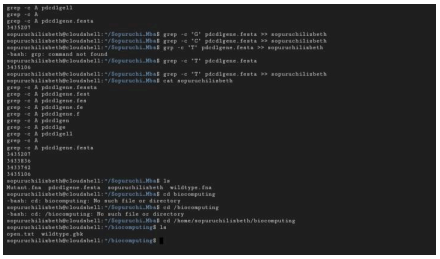
21. Clear your terminal space and print all the commands you have used today.

```
$ clear
$ history
```

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

```
$ ls
```

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



Project 2: Installing Bioinformatics Softwares on the terminal

N/B: You need to install and setup your conda environment with either anaconda or miniconda.

Please copy exactly what worked. Do not paraphrase. A single mismatch makes you loose your point.

1. Activate your base conda environment

```
$ mkdir -p ~/miniconda3
$  cd miniconda3
$  wget
https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
$  conda activate
```

2. Create a conda environment names funtools

```
$ conda create -n funtools
```

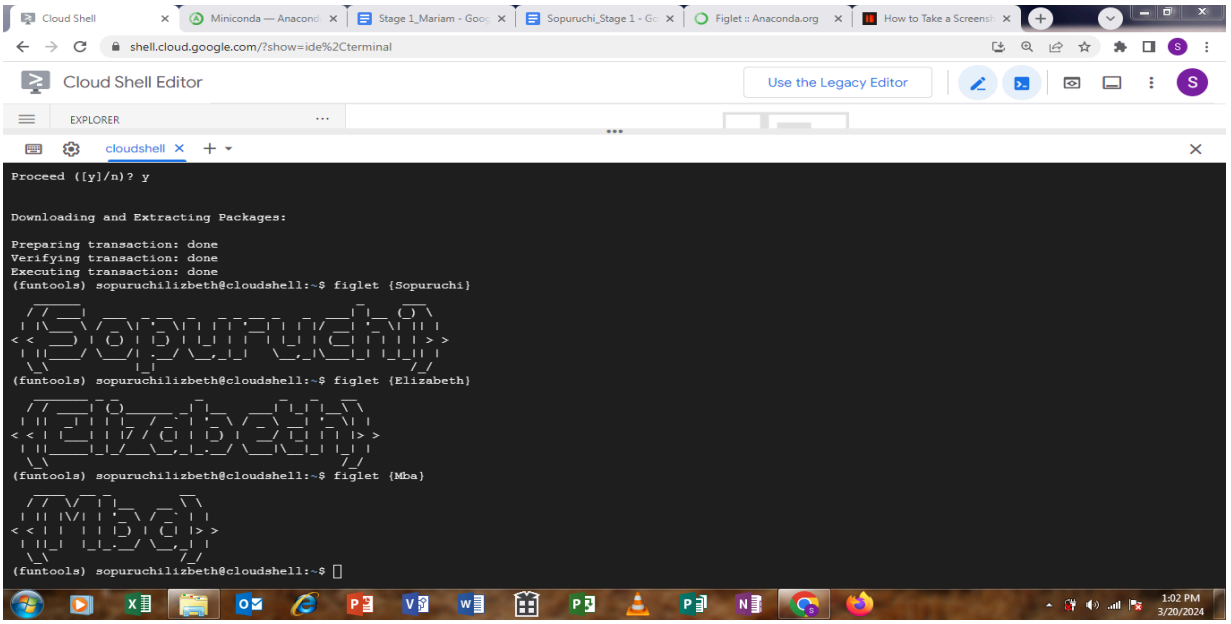
3. Activate the funtools environment

```
$ conda activate funtools
```

4. Install Figlet using conda

```
$ conda install tsnyder::figlet
```

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



6. Install bwa through the bioconda channel

```
$ conda install bioconda::bwa
```

7. Install blast through the bioconda channel

```
$ conda install bioconda::blast
```

8. Install samtools through the bioconda channel

```
$ conda install bioconda::samtools
```

9. Install bedtools through the bioconda channel

```
$  conda install bioconda/label/cf201901::bedtools
```

10. Install spades.py through the bioconda channel

```
$  conda install bioconda/label/cf201901::spades
```

11. Install bcftools through the bioconda channel

```
$ conda install bioconda/label/cf201901::bcftools
```

12. Install fastp through the bioconda channel

```
$ conda install fastp
```

13. Install multiqc through the bioconda channel

```
$ pip install multiqc
```

To submit this project, make this document open using the 🔒**share icon** at the top right corner. Copy the link and submit it on HackBio platform.

Finally, everyone in your team should be ready to discuss your code submission with everyone.

Learning Resources:

The Official learning resource for this internship is **HackBio's Genomics Course**. Sign up to enjoy uninterrupted and synchronized flow of bioinformatics knowledge. If you have access to the course already, everything you need for the internship is already provided in the course.

However, we have plans for you if you are unable to purchase the course. We have gathered some resources for you to help you learn and navigate the internship better.

Stage 1

- How to Access the terminal for the purpose of stage 1 and 2
- Introduction to BASh
- How to setup and use Conda
- An article on bioconda usage from HackBio
- The rest is practice! practice!! practice!!!