Project 1: Bash Basic

#!/bin/bash

# --- Project 1: Bash Basics ---

# 1. Print your name

echo afifah zahwah

 # 2. Create a folder titled your name

mkdir afifahzahwah

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

mkdir biocomputing && cd biocomputing

# 4. Download these 3 files:

# The last URL is a duplicate, which will be named wildtype.gbk.1

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. Move the .fna file to the folder titled your name

mv wildtype.fna ../afifahzahwah/

# 6. Delete the duplicate gbk file

rm wildtype.gbk.1

# 7. Confirm if the .fna file is mutant or wild type (tatatata vs tata)

echo "--- Checking for Mutant 'tatatata' ---"

mutant\_count=$(grep -o "tatatata" wildtype.fna | wc -l)

echo "Count for 'tatatata': $mutant\_count"

echo "--- Checking for Wild Type 'tata' ---"

wildtype\_count=$(grep -o "tata" wildtype.fna | wc -l)

echo "Count for 'tata': $wildtype\_count"

# 8. If mutant, print all matching lines into a new file

if [ "$mutant\_count" -gt 0 ]; then

    echo "Conclusion: This is a mutant."

    echo "Saving matching lines to mutant\_lines.txt."

    grep "tatatata" wildtype.fna > mutant\_lines.txt

else

    echo "Conclusion: This is a wild type. No 'tatatata' sequences found."

fi

# Move back to the parent directory to access the biocomputing folder

cd ..

# 9. Count number of lines (excluding header) in the .gbk file

echo "Counting lines in .gbk file (excluding header)..."

 header\_lines=$(grep -n "ORIGIN" biocomputing/wildtype.gbk | cut -d: -f1)

total\_lines=$(wc -l < biocomputing/wildtype.gbk)

sequence\_lines=$((total\_lines - header\_lines - 1)) # -1 to exclude the // line

echo "Number of sequence lines: $sequence\_lines"

# 10. Print the sequence length of the .gbk file.

echo "--- Sequence length ---"

grep "LOCUS" biocomputing/wildtype.gbk | awk '{print $3}'

# 11. Print the source organism of the .gbk file.

grep "SOURCE" biocomputing/wildtype.gbk | awk '{$1=$2=""; print $0}' | xargs

# 12. List all the gene names of the .gbk file.

echo "--- Gene names ---"

grep '/gene=' biocomputing/wildtype.gbk

# 13. Clear your terminal space and print all commands used today

clear

echo "mkdir, cd, wget, rm, grep, wc, if, >, cut, awk, xargs, clear, ls"

# 14. List the files in the two folders

ls afifahzawah

ls biocomputing

Project 2: Installing Bioinformatics Software on the Terminal

# 1. Activate your base conda environment

conda activate

# 2. Create a conda environment named funtools

conda create -n funtools

# 3. Activate the funtools environment

conda activate funtools

# 4. Install Figlet using conda

conda install -c conda-forge figlet

# 5. Run figlet <your name>

figlet AFIFAH\_ZAHWAH

# 6. Install bwa through the bioconda channel

conda install -c bioconda 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

# 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# 1. Activate your base conda environment