

**Indian Institute of Technology Gandhinagar**  
**BE623 Biocomputing**  
**Sem1 2025-2026**

**Lab session –2**

**Shell scripting basics, variables, export,  
conditionals and loops**

- Based on previous understanding of previous lab session, create a folder named “Lab\_session2” in your home folder
- Download the files uploaded on the Google Classroom in this folder.
- Navigate to this folder using terminal

**Important Instructions -**

**Introduction to vi**

vi- Text editor in most linux based systems

Different modes in vi- **Command mode, Insert mode, Escape mode/Last line mode**

Action	Command
Enter insert mode	i
Save & exit	:wq
Exit without saving	:q!
Delete current line	dd
Copy current line	yy
Paste after cursor	p
Move to start of file	gg
Move to end of file	G
Move to the start of line	^
Move to the end of line	\$

**1) Open vi:**

- vi testfile.txt
- Type “Hello World” (Hint- press i first).
- Save and exit (Hint- press escape then type :wq and enter).

- Reopen the file to verify contents. (Hint- vi filename)

## 2) Basic Linux commands

- head sequence.fasta
- head -n 5 sequence.fasta
- tail sequence.fasta
- tail -n 5 sequence.fasta
- wc -w sequence.fasta
- wc -l sequence.fasta

## 3) less command to view files

```
less <filename>
```

Command	Action
Page Up or b	Scroll back one page
Page Down or space	Scroll forward one page
Up arrow	Scroll up one line
Down arrow	Scroll down one line
G	Move to the end of the text file
1G or g	Move to the beginning of the text file
/characters	Search forward to the next occurrence of <i>characters</i>
n	Search for the next occurrence of the previous search
h	Display help screen
q	Quit less

## 4) Copy “protein.fasta” file to current folder

- grep TDVNT protein.fasta
- grep -o TDVNT protein.fasta
- grep -v TDVNT protein.fasta
- grep -c ">" sequence.fasta

### 5) **echo** - Display text or variable values

```
echo "Hello World"

echo $HOME      # print home directory
echo $USER      # print your username
echo *.fasta    # print all the fasta files in the folder
echo * and ls difference can be explained here.
```

### 6) **export** - Make a variable available to other programs or shells

```
myvar="bioinfo"
export myvar
bash -c 'echo $myvar' # printing
printenv #printing all the environment variable.
```

### 7) **Variables with file operations**

```
file="sequence5.fasta"
echo "Number of sequences in $file:"
grep -c ">" $file
```

### 8) **Creating bash script file (filename.sh)**

[vi sequence\\_reader\\_1.sh](#)

```
#!/bin/bash

file="sequence5.fasta"
echo "Number of sequences in $file:"
grep -c ">" $file
```

### 9) **Using 'if' statement**

[vi sequence\\_line\\_counter.sh](#)

```
#!/bin/bash

if [ -f "sequence1.fasta" ]; then
    wc -l sequence1.fasta
else
    echo "File sequence1.fasta not found."
fi
```

### 10) Processing files with 'for' loop

[vi for\\_sequence\\_reader.sh](#)

```
#!/bin/bash

for file in *.fasta; do
    echo "Processing: $file"
    echo "Sequences: $(grep -c ">" "$file")"
    echo "File size (characters): $(wc -c < "$file")"
done
```

### 11) Processing files with 'while' loop

[vi while\\_sequence\\_reader.sh](#)

```
#!/bin/bash

ls *.fasta | while read -r file; do
    echo "Processing: $file"
    echo "Sequences: $(grep -c ">" "$file")"
    echo "File size (characters): $(wc -c < "$file")"
done
```

## 12) Grep with regular expressions

`grep "^>" sequence5.fasta`      # Match headers (start with >)

`grep "A$" protein.fasta`      # Lines ending with A

`grep "A.G" protein.fasta`      # A, then any char, then G

`grep "AL[DS]E" sequence5.fasta`      # AL, then D or S, then E

`grep "A[^T]G" sequence5.fasta`      # A, then NOT T, then G

`grep "A\{2\}" sequence5.fasta`      # AA (three As)

`grep -E "RR|KK" sequence5.fasta.`      # Search for lines containing 'RR' or 'KK' in the file sequence5.fasta

`grep "K" sequence5.fasta`      # list all the lines with K including header

`grep -v "^>" sequence5.fasta | grep "K"`      # list all the lines with K that exclude header

`grep -v "^>" sequence5.fasta | grep "C.*C.*C.*C"`      # Match sequences with cysteine-rich regions (≥4 Cys)