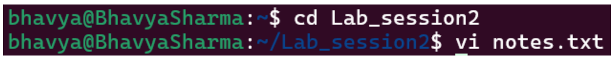
Indian Institute of Technology Gandhinagar

BE623 Biocomputing
Sem1 2025-2026
Lab Assignment –2

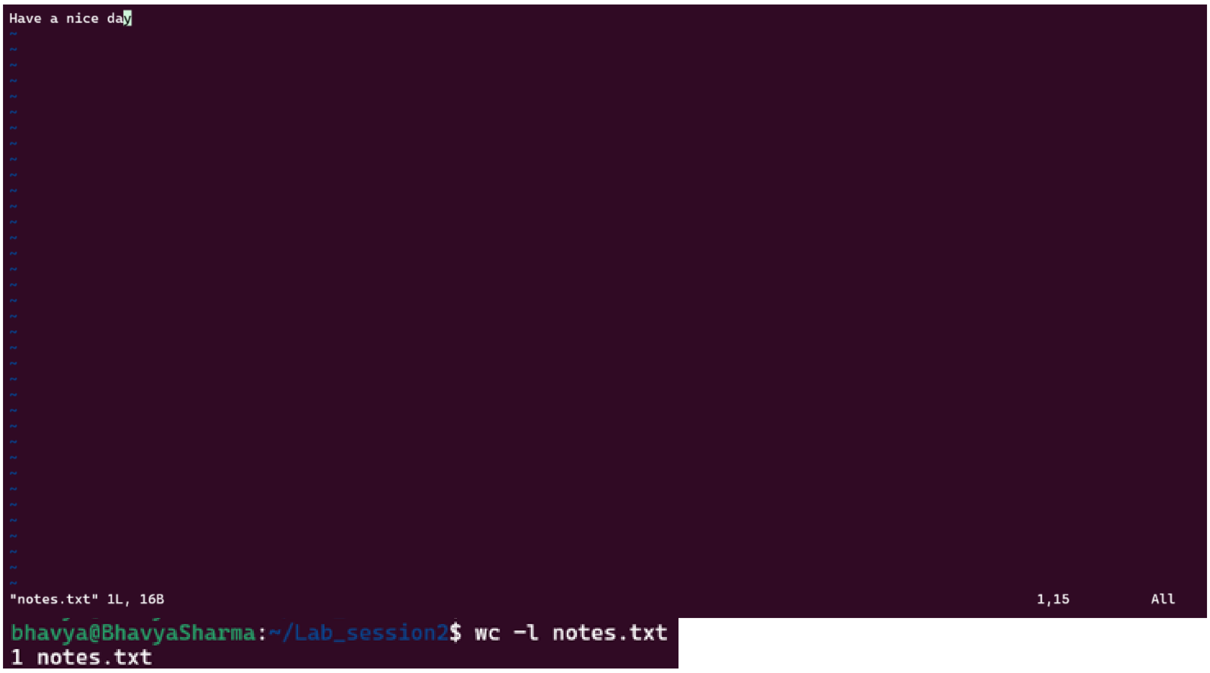
Linux & Shell Scripting with Biological Data Files

Name: Bhavya Sharma
Roll No.: 25210035
M.Tech Biological Engineering

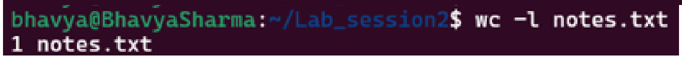
Q1)
Output-
Created notes.txt in vi



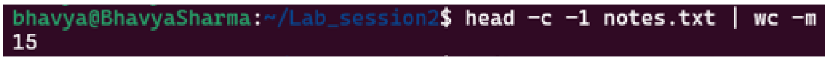
```
bhavya@BhavyaSharma:~$ cd Lab_session2
bhavya@BhavyaSharma:~/Lab_session2$ vi notes.txt
```



```
Have a nice day
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
"notes.txt" 1L, 16B                                              1,15          All
```



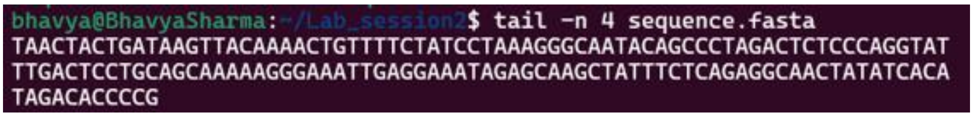
```
bhavya@BhavyaSharma:~/Lab_session2$ wc -l notes.txt
1 notes.txt
```



```
bhavya@BhavyaSharma:~/Lab_session2$ head -c -1 notes.txt | wc -m
15
```

Verified that the file contains exactly one line and 15 characters.
(Reference:- In order to understand the functioning of head command, took help from google.)

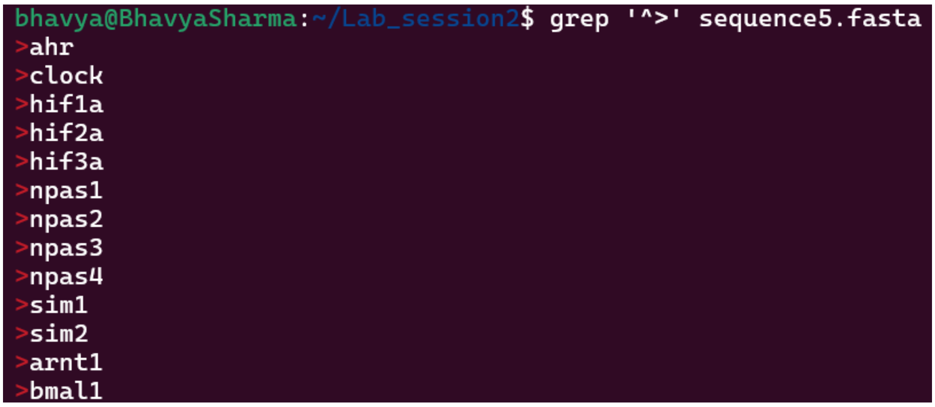
Q2)
Output-



```
bhavya@BhavyaSharma:~/Lab_session2$ tail -n 4 sequence.fasta
TAACTACTGATAAGTTACAAAACTGTTTTCTATCCTAAAGGGCAATACAGCCCTAGACTCTCCCAGGTAT
TTGACTCCTGCAGCAAAAAGGGAAATTGAGGAAATAGAGCAAGCTATTTCTCAGAGGCAACTATATCACA
TAGACACCCCG
```

Displayed the last four lines of sequence.fasta without opening the file in an editor.

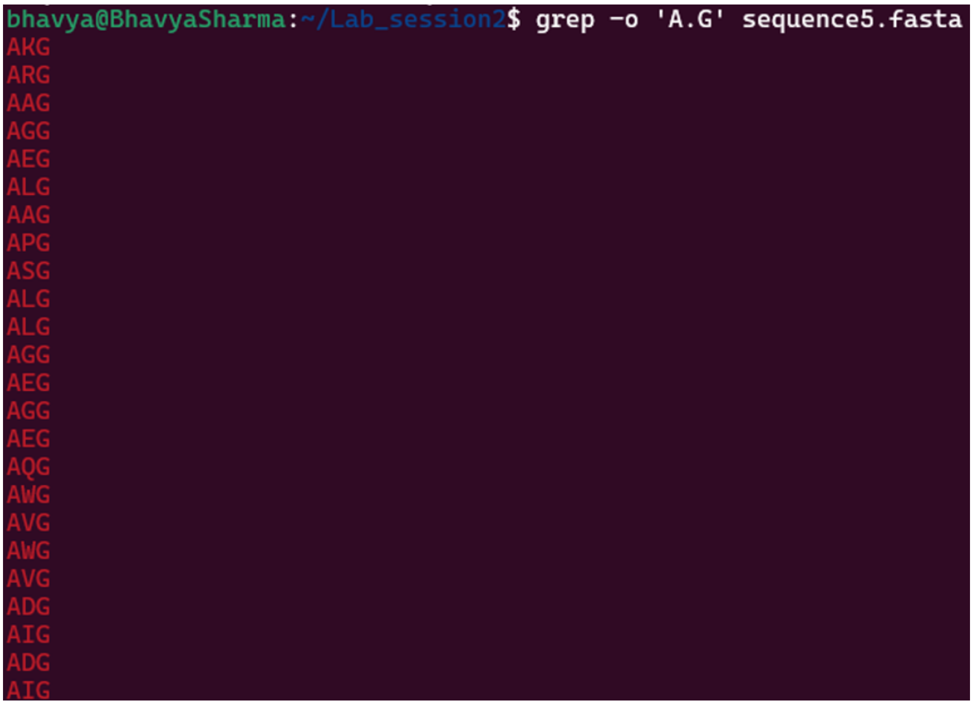
Q3)
Output-



```
bhavya@BhavyaSharma:~/Lab_session2$ grep '^>' sequence5.fasta
>ahr
>clock
>hif1a
>hif2a
>hif3a
>npas1
>npas2
>npas3
>npas4
>sim1
>sim2
>arnt1
>bmal1
```

In sequence5.fasta, printed all header lines (lines starting with >).

Q4)
Output-

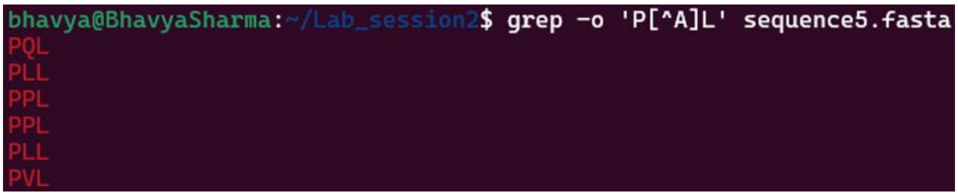


```
bhavya@BhavyaSharma:~/Lab_session2$ grep -o 'A.G' sequence5.fasta
AKG
ARG
AAG
AGG
AEG
ALG
AAG
APG
ASG
ALG
ALG
AGG
AEG
AGG
AEG
AQG
AWG
AVG
AWG
AVG
ADG
AIG
ADG
AIG
```

Found all matches in sequence5.fasta where A is followed by any single character and then G.

Q5)

Output-

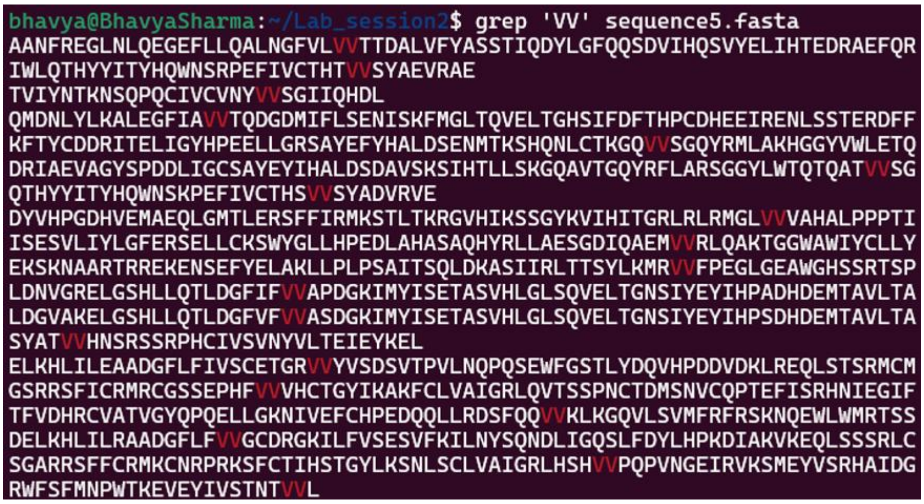


```
bhavya@BhavyaSharma:~/Lab_session2$ grep -o 'P[^A]L' sequence5.fasta
PQL
PLL
PPL
PPL
PLL
PVL
```

Found all matches in sequence5.fasta where P is followed by any character except A, then L.

Q6)

Output-

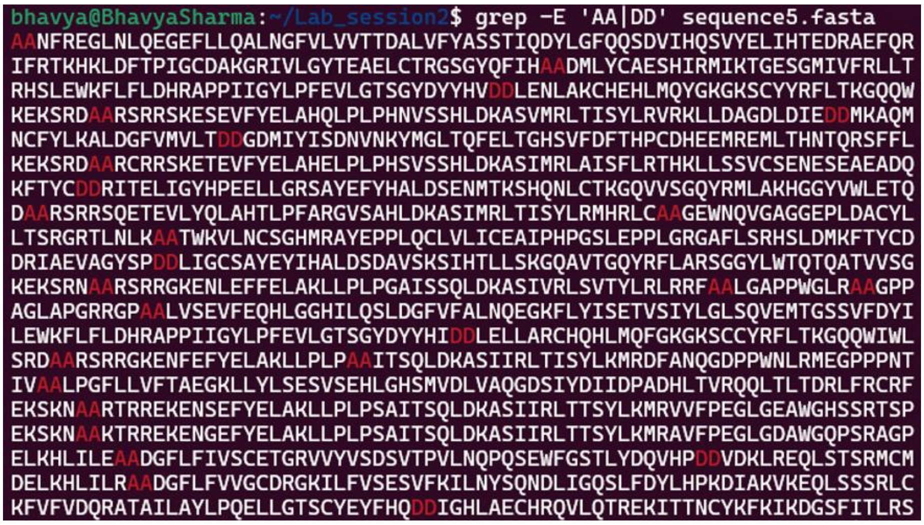


```
bhavya@BhavyaSharma:~/Lab_session2$ grep 'VV' sequence5.fasta
AANFREGLNLQEGEFLLQALNGFVLVVTTDALVFYASSTIQDYLGFQQSDVIHQSVYELIHTEDRAEFQR
IWLQTHYYITYHQWNSRPEFIVCTHTVVSYAEVRAE
TVIYNTKNSQPQCIVCVNYVVSGIIQHDL
QMDNLYLKALEGFIAVVTQDGDMIFLSENISKFMGLTQVELTGHSIFDFTHPCDHEEIRENLSSTERDFF
KFTYCDDRITELIGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSGQYRMLAKHGGYVWLETQ
DRIAEVAGYSPDDLIGCSAYEYIHALDSDAVSKSIHTLLSKGQAVTGQYRFLARSGGYLWTQTQATVVSG
QTHYYITYHQWNSKPEFIVCTHSVVSYADVRVE
DYVHPGDHVEMAEQLGMTLERSFFIRMKSTLTKRGVHIKSSGYKVIHITGRLRLRMGLVVVAHALPPPTI
ISESVLIYLGFERSELLCKSWYGLLHPEDLAHASAQHYRLLAESGDIQAEMVVRLQAKTGGWAWIYCLLY
EKSKNAARTRREKENSEFYELAKLLPLPSAITSQLDKASIIRLTTSYLKMRVVFPEGLGEAWGHSSRTSP
LDNVGRELGSHLLQTLDGFIFVVAPDGKIMYISETASVHLGLSQVELTGNSIYEYIHPADHDEMTAVLTA
LDGVAKELGSHLLQTLDGFVFVVASDGKIMYISETASVHLGLSQVELTGNSIYEYIHPSDHDEMTAVLTA
SYATVVHNSRSSRPHCIVSVNYVLTEIEYKEL
ELKHLILEAADGFLFIVSCETGRVVYVSDSVTPVLNQPQSEWFGSTLYDQVHPDDVDKLREQLSTSRMCM
GSRRSFICRMRCGSSEPHFVVVHCTGYIKAKFCLVAIGRLQVTSSPNCTDMSNVCQPTEFISRHNIEGIF
TFVDHRCVATVGYQPQELLGKNIVEFCHPEDQQLLRDSFQQVVKLKGQVLSVMFRFRSKNQEWLWMRTSS
DELKHLILRAADGFLFVVGCDRGKILFVSESVFKILNYSQNDLIGQSLFDYLHPKDIAKVKEQLSSSRLC
SGARRSFFCRMKCNRPRKSFCTIHSTGYLKSNLSCLVAIGRLHSHVVPQPVNGEIRVKSMEYVSRHAIDG
RWFSFMNPWTKEVEYIVSTNTVVL
```

Printed all lines in sequence5.fasta that have exactly 2 consecutive Vs anywhere in the line.

Q7)

Output-

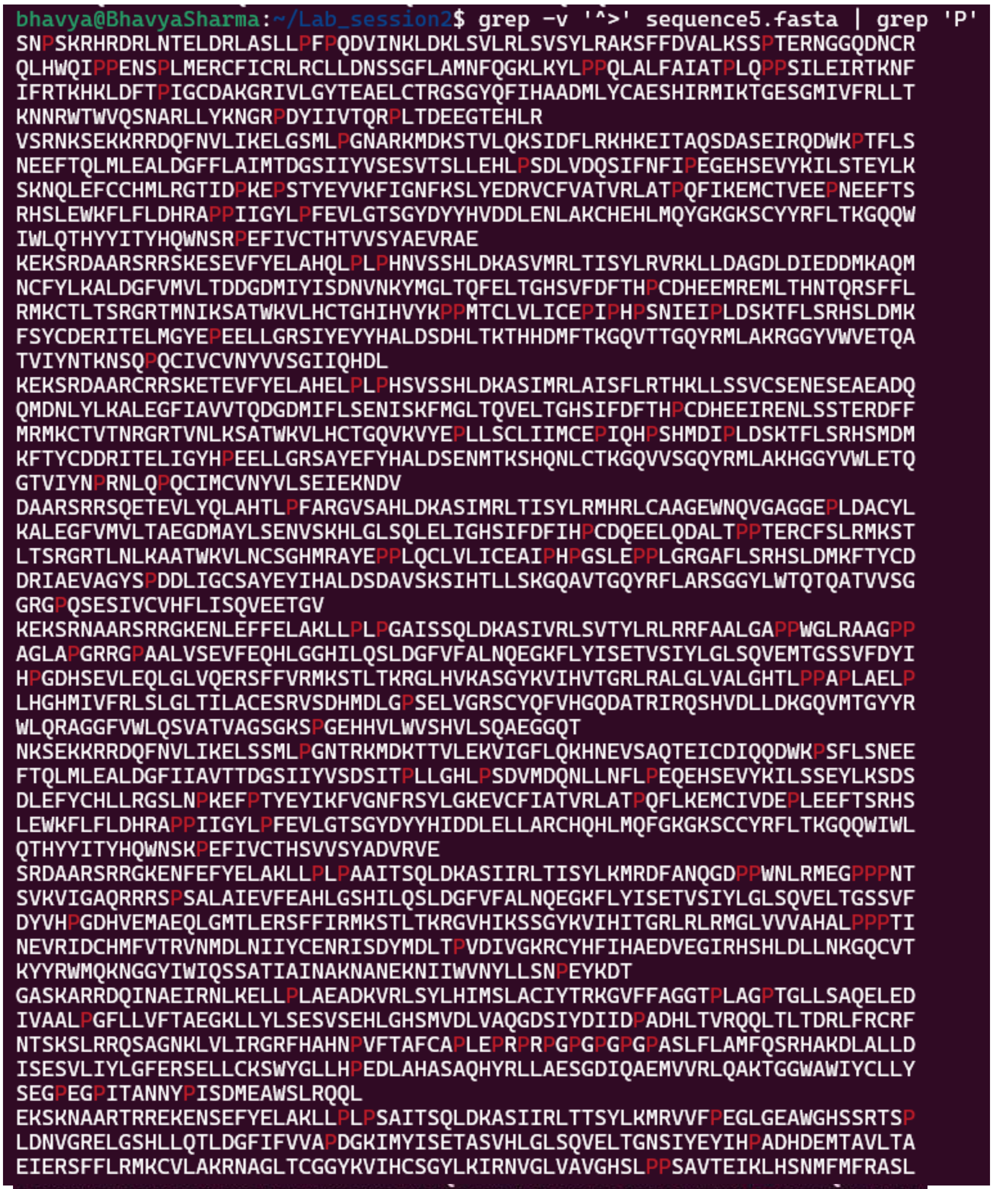


```
bhavya@BhavyaSharma:~/Lab_session2$ grep -E 'AA|DD' sequence5.fasta
AANFREGLNLQEGEFLLQALNGFVLVVTTDALVFYASSTIQDYLGFQQSDVIHQSVYELIHTEDRAEFQR
IFRTKHKLDFTPIGCDAKGRIVLGYTEAELCTRGSGYQFIHAADMLYCAESHIRMIKTGESGMIVFRLLT
RHSLEWKFLFLDHRAPPIIGYLPFEVLGTSGYDYYHVDDLENLAKCHEHLMQYGKGKSCYYRFLTKGQQW
KEKSRDAARSRRSKESEVFYELAHQLPLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQM
NCFYLKALDGFVMVLTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHNTQRSFFL
KEKSRDAARCRRSKETEVFYELAHELPLPHSVSSHLDKASIMRLAISFLRTHKLLSSVCSENESEAEADQ
KFTYCDDRITELIGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSGQYRMLAKHGGYVWLETQ
DAARSRRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHRLCAAGEWNQVGAGGEPLDACYL
LTSRGRTLNLKAATWKVLNCSGHMRAYEPPLQCLVLICEAIPHPGSLEPPLGRGAFLSRHSLDMKFTYCD
DRIAEVAGYSPDDLIGCSAYEYIHALDSDAVSKSIHTLLSKGQAVTGQYRFLARSGGYLWTQTQATVVSG
KEKSRNAARSRRGKENLEFFELAKLLPLPGAISSQLDKASIVRLSVTYLRLRRFAALGAPPWGLRAAGPP
AGLAPGRRGPAALVSEVFEQHLGGHILQSLDGFVFALNQEGKFLYISETVSIYLGLSQVEMTGSSVFDYI
LEWKFLFLDHRAPPIIGYLPFEVLGTSGYDYYHIDDLELLARCHQHLMQFGKGKSCCYRFLTKGQQWIWL
SRDAARSRRGKENFEFYELAKLLPLPAAITSQLDKASIIRLTISYLKMRDFANQGDPPWNLRMEGPPPNT
IVAALPGFLLVFTAEGKLLYLSESVSEHLGHSMVDLVAQGDSIYDIIDPADHLTVRQQLTLTDRLFRCRF
EKSKNAARTRREKENSEFYELAKLLPLPSAITSQLDKASIIRLTTSYLKMRVVFPEGLGEAWGHSSRTSP
EKSKNAAKTRREKENGEFYELAKLLPLPSAITSQLDKASIIRLTTSYLKMRAVFPEGLGDAWGQPSRAGP
ELKHLILEAADGFLFIVSCETGRVVYVSDSVTPVLNQPQSEWFGSTLYDQVHPDDVDKLREQLSTSRMCM
DELKHLILRAADGFLFVVGCDRGKILFVSESVFKILNYSQNDLIGQSLFDYLHPKDIAKVKEQLSSSRLC
KFVFVDQRATAILAYLPQELLGTSCYEYFHQDDIGHLAECHRQVLQTREKITTNCYKFKIKDGSFITLRS
```

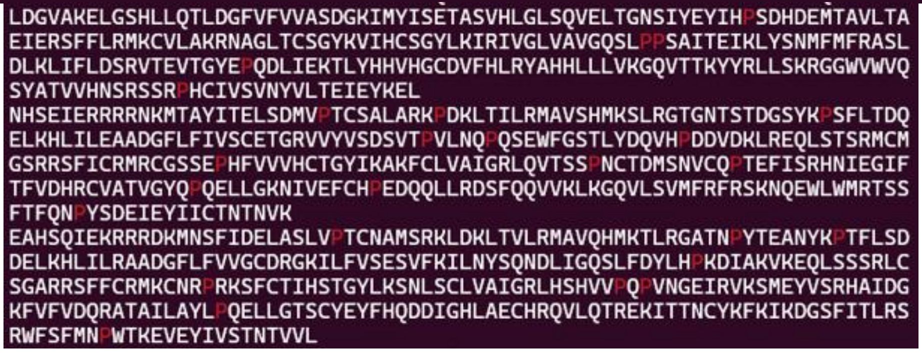
Printed all lines in sequence5.fasta that contain either AA or DD.

Q8)

Output-



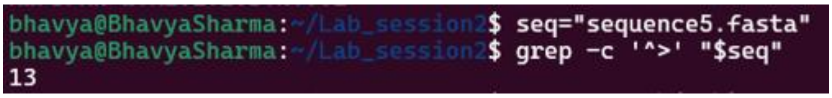
```
bhavya@BhavyaSharma:~/Lab_session2$ grep -v '^>' sequence5.fasta | grep 'P'
SNPSKRHRDRLNTELDRLASLLPFPQDVINKLDKLSVLRLSVSYLRAKSFFDVALKSSPTERNGGQDNCR
QLHWQIPPENSPLMERCFICRLRCLLDNSSGFLAMNFQGKLKYLPPQLALFAIATPLQPPSILEIRTKNF
IFRTKHKLDFTPIGCDAKGRIVLGYTEAELCTRGSGYQFIHAADMLYCAESHIRMIKTGESGMIVFRLLT
KNNRWTWVQSNARLLYKNGRPDYIIVTQRPLTDEEGTEHLR
VSRNKSEKKRRDQFNVLIKELGSMLPGNARKMDKSTVLQKSIDFLRKHKEITAQSDASEIRQDWKPTFLS
NEEFTQLMLEALDGFFLAIMTDGSIIYVSESVTSLLEHLPSDLVDQSIFNFIPEGEHSEVYKILSTEYLK
SKNQLEFCCHMLRGTIDPKEPSTYEYVKFIGNFKSLYEDRVCFVATVRLATPQFIKEMCTVEEPNEEFTS
RHSLEWKFLFLDHRAPPIIGYLPFEVLGTSGYDYYHVDDLENLAKCHEHLMQYGKGKSCYYRFLTKGQQW
IWLQTHYYITYHQWNSRPEFIVCTHTVVSYAEVRAE
KEKSRDAARSRRSKESEVFYELAHQLPLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLDIEDDMKAQM
NCFYLKALDGFVMVLTDDGDMIYISDNVNKYMGLTQFELTGHSVFDFTHPCDHEEMREMLTHNTQRSFFL
RMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYKPPMTCLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMK
FSYCDERITELMGYEPEELLGRSIYEYYHALDSDHLTKTHHDMFTKGQVTTGQYRMLAKRGGYVWVETQA
TVIYNTKNSQPQCIVCVNYVVSGIIQHDL
KEKSRDAARCRRSKETEVFYELAHELPLPHSVSSHLDKASIMRLAISFLRTHKLLSSVCSENESEAEADQ
QMDNLYLKALEGFIAVVTQDGDMIFLSENISKFMGLTQVELTGHSIFDFTHPCDHEEIRENLSSTERDFF
MRMKCTVTNRGRTVNLKSATWKVLHCTGQVKVYEPLLSCLIIMCEPIQHPSHMDIPLDSKTFLSRHSMDM
KFTYCDDRITELIGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSGQYRMLAKHGGYVWLETQ
GTVIYNPRNLQPQCIMCVNYVLSEIEKNDV
DAARSRRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHRLCAAGEWNQVGAGGEPLDACYL
KALEGFVMVLTAEGDMAYLSENVSKHLGLSQLELIGHSIFDFIHPCDQEELQDALTPPTERCFSLRMKST
LTSRGRTLNLKAATWKVLNCSGHMRAYEPPLQCLVLICEAIPHPGSLEPPLGRGAFLSRHSLDMKFTYCD
DRIAEVAGYSPDDLIGCSAYEYIHALDSDAVSKSIHTLLSKGQAVTGQYRFLARSGGYLWTQTQATVVSG
GRGPQSESIVCVHFLISQVEETGV
KEKSRNAARSRRGKENLEFFELAKLLPLPGAISSQLDKASIVRLSVTYLRLRRFAALGAPPWGLRAAGPP
AGLAPGRRGPAALVSEVFEQHLGGHILQSLDGFVFALNQEGKFLYISETVSIYLGLSQVEMTGSSVFDYI
HPGDHSEVLEQLGLVQERSFFVRMKSTLTKRGLHVKASGYKVIHVTGRLRALGLVALGHTLPPAPLAELP
LHGHMIVFRLSLGLTILACESRVSDHMDLGPSELVGRSCYQFVHGQDATRIRQSHVDLLDKGQVMTGYYR
WLQRAGGFVWLQSVATVAGSGKSPGEHHVLWVSHVLSQAEGGQT
NKSEKKRRDQFNVLIKELSSMLPGNTRKMDKTTVLEKVIGFLQKHNEVSAQTEICDIQQDWKPSFLSNEE
FTQLMLEALDGFIIAVTTDGSIIYVSDSITPLLGHLPSDVMDQNLLNFLPEQEHSEVYKILSSEYLKSDS
DLEFYCHLLRGSLNPKEFPTYEYIKFVGNFRSYLGKEVCFIATVRLATPQFLKEMCIVDEPLEEFTSRHS
LEWKFLFLDHRAPPIIGYLPFEVLGTSGYDYYHIDDLELLARCHQHLMQFGKGKSCCYRFLTKGQQWIWL
QTHYYITYHQWNSKPEFIVCTHSVVSYADVRVE
SRDAARSRRGKENFEFYELAKLLPLPAAITSQLDKASIIRLTISYLKMRDFANQGDPPWNLRMEGPPPNT
SVKVIGAQRRRSPSALAIEVFEAHLGSHILQSLDGFVFALNQEGKFLYISETVSIYLGLSQVELTGSSVF
DYVHPGDHVEMAEQLGMTLERSFFIRMKSTLTKRGVHIKSSGYKVIHITGRLRLRMGLVVVAHALPPPTI
NEVRIDCHMFVTRVNMDLNIIYCENRISDYMDLTPVDIVGKRCYHFIHAEDVEGIRHSHLDLLNKGQCVT
KYYRWMQKNGGYIWIQSSATIAINAKNANEKNIIWVNYLLSNPEYKDT
GASKARRDQINAEIRNLKELLPLAEADKVRLSYLHIMSLACIYTRKGVFFAGGTPLAGPTGLLSAQELED
IVAALPGFLLVFTAEGKLLYLSESVSEHLGHSMVDLVAQGDSIYDIIDPADHLTVRQQLTLTDRLFRCRF
NTSKSLRRQSAGNKLVLIRGRFHAHNPVFTAFCAPLEPRPRPGPGPGPGPASLFLAMFQSRHAKDLALLD
ISESVLIYLGFERSELLCKSWYGLLHPEDLAHASAQHYRLLAESGDIQAEMVVRLQAKTGGWAWIYCLLY
SEGPEGPITANNYPISDMEAWSLRQQL
EKSKNAARTRREKENSEFYELAKLLPLPSAITSQLDKASIIRLTTSYLKMRVVFPEGLGEAWGHSSRTSP
LDNVGRELGSHLLQTLDGFIFVVAPDGKIMYISETASVHLGLSQVELTGNSIYEYIHPADHDEMTAVLTA
EIERSFFLRMKCVLAKRNAGLTCGGYKVIHCSGYLKIRNVGLVAVGHSLPPSAVTEIKLHSNMFMFRASL
```



```
LDGVAKELGSHLLQTLDGFVFVVASDGKIMYISETASVHLGLSQVELTGNSIYEYIHPSDHDEMTAVLTA
EIERSFFLRMKCVLAKRNAGLTCSGYKVIHCSGYLKIRIVGLVAVGQSLPPSAITEIKLYSNMFMFRASL
DLKLIFLDSRVTEVTGYEPQDLIEKTLYHHVHGCDVFHLRYAHHLLLVKGQVTTKYYRLLSKRGGWVWVQ
SYATVVHNSRSSRPHCIVSVNYVLTEIEYKEL
NHSEIERRRRNKMTAYITELSDMVPTCSALARKPDKLTILRMAVSHMKSLRGTGNTSTDGSYKPSFLTDQ
ELKHLILEAADGFLFIVSCETGRVVYVSDSVTPVLNQPQSEWFGSTLYDQVHPDDVDKLREQLSTSRMCM
GSRRSFICRMRCGSSEPHFVVVHCTGYIKAKFCLVAIGRLQVTSSPNCTDMSNVCQPTEFISRHNIEGIF
TFVDHRCVATVGYQPQELLGKNIVEFCHPEDQQLLRDSFQQVVKLKGQVLSVMFRFRSKNQEWLWMRTSS
FTFQNPYSDEIEYIICTNTNVK
EAHSQIEKRRRDKMNSFIDELASLVPTCNAMSRKLDKLTVLRMAVQHMKTLRGATNPYTEANYKPTFLSD
DELKHLILRAADGFLFVVGCDRGKILFVSESVFKILNYSQNDLIGQSLFDYLHPKDIAKVKEQLSSSRLC
SGARRSFFCRMKCNRPRKSFCTIHSTGYLKSNLSCLVAIGRLHSHVVPQPVNGEIRVKSMEYVSRHAIDG
KFVFVDQRATAILAYLPQELLGTSCYEYFHQDDIGHLAECHRQVLQTREKITTNCYKFKIKDGSFITLRS
RWFSFMNPWTKEVEYIVSTNTVVL
```

Printed only the sequence lines (ignored headers) from sequence5.fasta that contain the letter P.

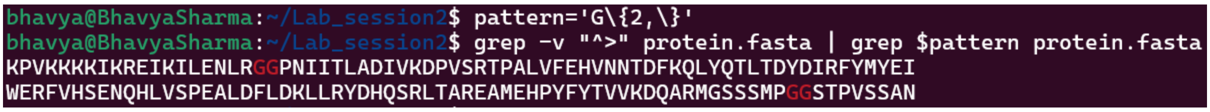
Q9)
Output-



```
bhavya@BhavyaSharma:~/Lab_session2$ seq="sequence5.fasta"
bhavya@BhavyaSharma:~/Lab_session2$ grep -c '^>' "$seq"
13
```

Stored the filename sequence5.fasta in a variable called seq and printed the number of sequences in it (headers count as sequences).

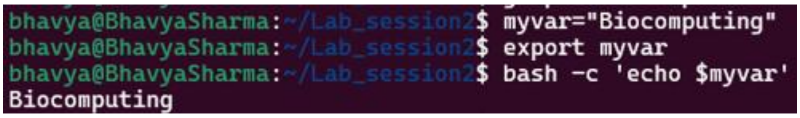
Q10)
Output-



```
bhavya@BhavyaSharma:~/Lab_session2$ pattern='G\{2,\}'
bhavya@BhavyaSharma:~/Lab_session2$ grep -v "^>" protein.fasta | grep $pattern protein.fasta
KPVKKKKIKREIKILENLRGGPNIITLADIVKDPVSRTPALVFEHVNNTDFKQLYQTLTDYDIRFYMYEI
WERFVHSENQHLVSPEALDFLDKLLRYDHQSRLTAREAMEHPYFYTVVKDQARMGSSSMPGGSTPVSSAN
```

Stored the pattern G\{2,\} in a variable and searched protein.fasta for sequence lines (ignored headers) with 2 or more consecutive Gs.

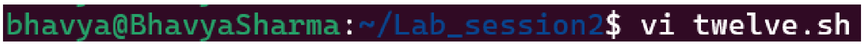
Q11)
Output-



```
bhavya@BhavyaSharma:~/Lab_session2$ myvar="Biocomputing"
bhavya@BhavyaSharma:~/Lab_session2$ export myvar
bhavya@BhavyaSharma:~/Lab_session2$ bash -c 'echo $myvar'
Biocomputing
```

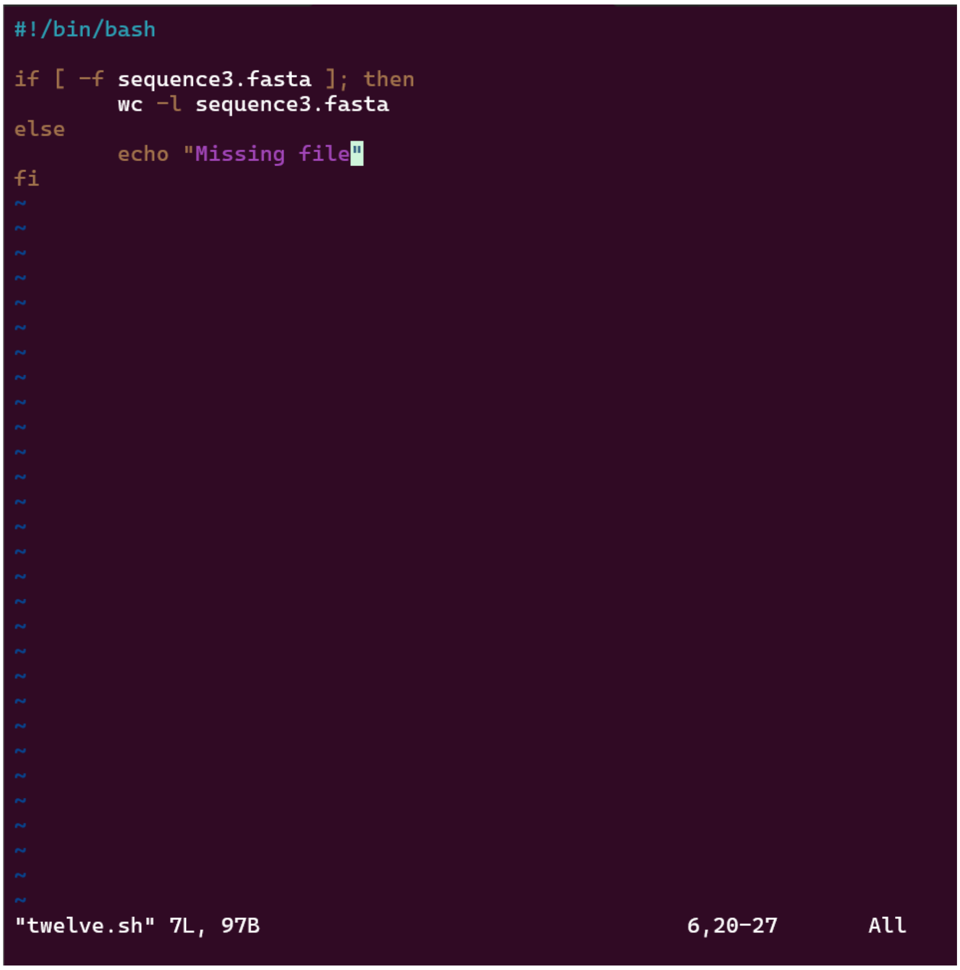
Stored "Biocomputing" in a variable, exported it, and verified that it is available inside a new shell started using: bash -c 'echo $VARIABLE_NAME'

Q12)
Output-

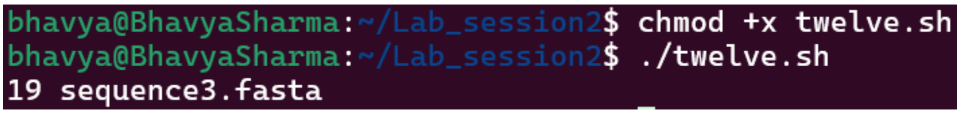


```
bhavya@BhavyaSharma:~/Lab_session2$ vi twelve.sh
```

Created a shell script that checks if sequence3.fasta exists in the current folder. If yes, it prints the number of lines. If no, it prints "Missing file".



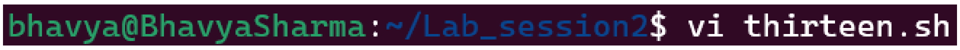
```
#!/bin/bash

if [ -f sequence3.fasta ]; then
        wc -l sequence3.fasta
else
        echo "Missing file"
fi
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
"twelve.sh" 7L, 97B                                      6,20-27       All
```



```
bhavya@BhavyaSharma:~/Lab_session2$ chmod +x twelve.sh
bhavya@BhavyaSharma:~/Lab_session2$ ./twelve.sh
19 sequence3.fasta
```

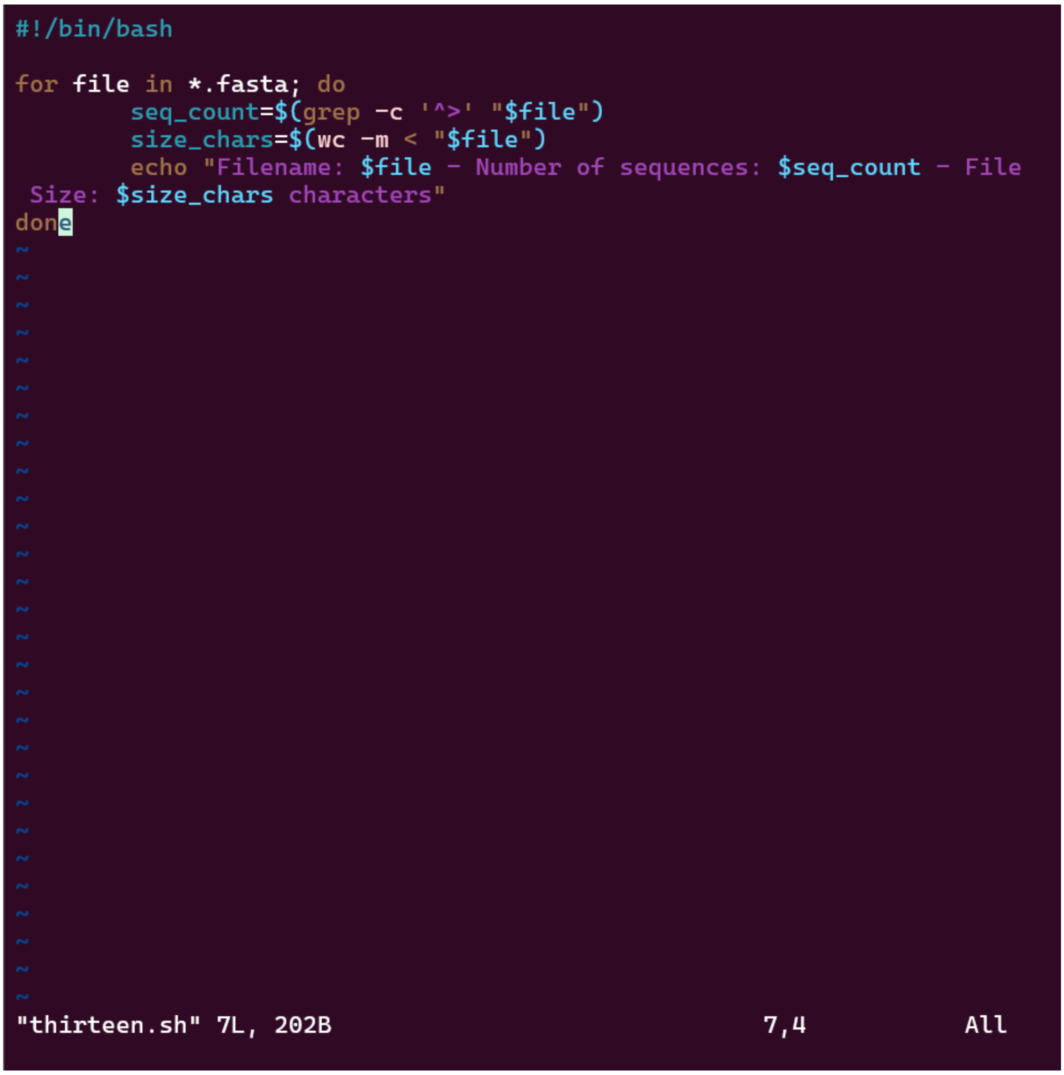
Using chmod command, the shell script is made executable and the output was displayed.

Q13)
Output-



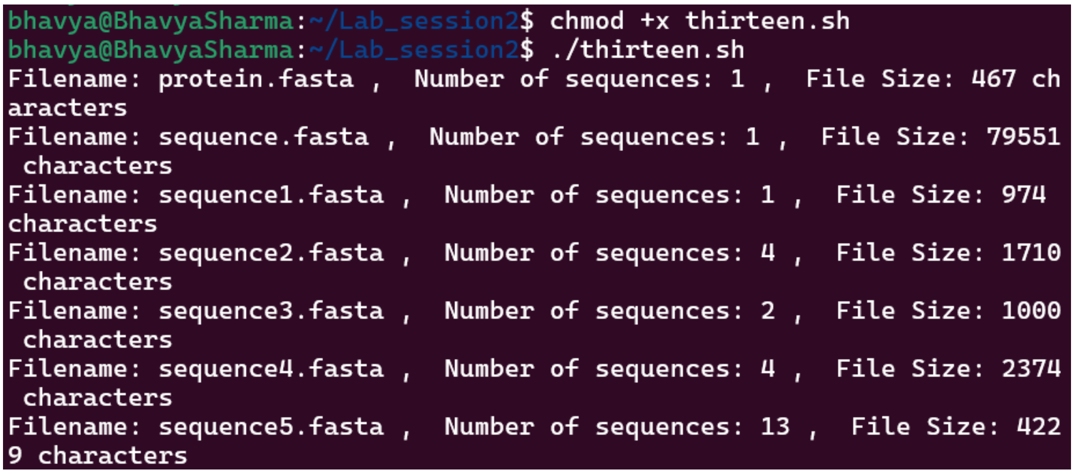
```
bhavya@BhavyaSharma:~/Lab_session2$ vi thirteen.sh
```

Created a shell script and using a for loop, went through all .fasta files in the current directory and printed: filename, number of sequences, and file size in characters.



```
#!/bin/bash

for file in *.fasta; do
        seq_count=$(grep -c '^>' "$file")
        size_chars=$(wc -m < "$file")
        echo "Filename: $file - Number of sequences: $seq_count - File
 Size: $size_chars characters"
done
```

```
"thirteen.sh" 7L, 202B                                        7,4           All
```

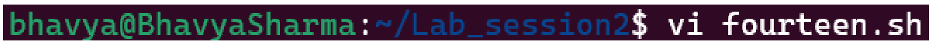


```
bhavya@BhavyaSharma:~/Lab_session2$ chmod +x thirteen.sh
bhavya@BhavyaSharma:~/Lab_session2$ ./thirteen.sh
Filename: protein.fasta ,  Number of sequences: 1 ,  File Size: 467 ch
aracters
Filename: sequence.fasta ,  Number of sequences: 1 ,  File Size: 79551
 characters
Filename: sequence1.fasta ,  Number of sequences: 1 ,  File Size: 974
characters
Filename: sequence2.fasta ,  Number of sequences: 4 ,  File Size: 1710
 characters
Filename: sequence3.fasta ,  Number of sequences: 2 ,  File Size: 1000
 characters
Filename: sequence4.fasta ,  Number of sequences: 4 ,  File Size: 2374
 characters
Filename: sequence5.fasta ,  Number of sequences: 13 ,  File Size: 422
9 characters
```

Using chmod command, the shell script is made executable and the output was displayed.

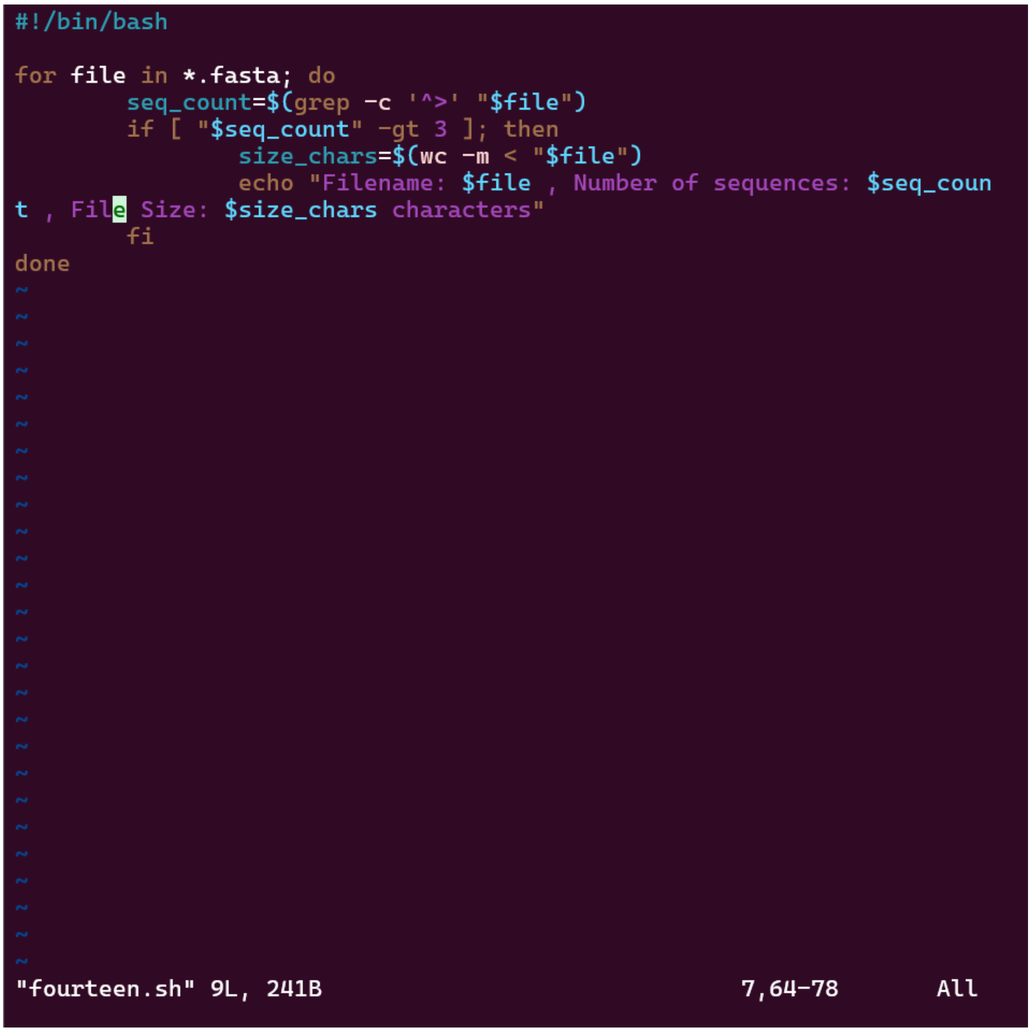
(Reference- Took help from google in order to understand the input redirection operator (<) that feeds the contents of the file to the wc command.)

Q14)
Output-

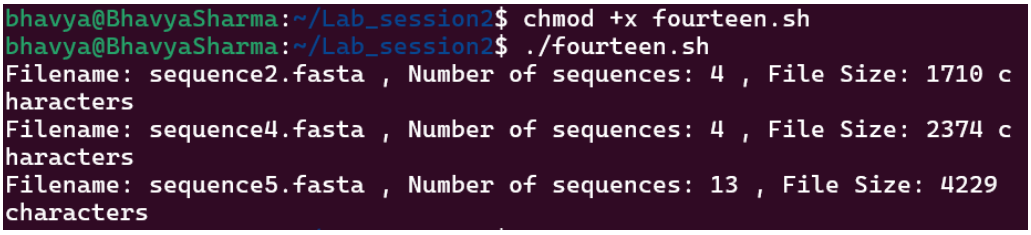


```
bhavya@BhavyaSharma:~/Lab_session2$ vi fourteen.sh
```

Modified the above loop in question number 13 so that it only prints files with more than 3 sequences. For this, a shell script is created.



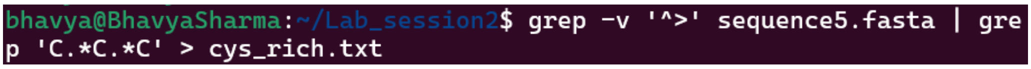
```
#!/bin/bash

for file in *.fasta; do
        seq_count=$(grep -c '^>' "$file")
        if [ "$seq_count" -gt 3 ]; then
                size_chars=$(wc -m < "$file")
                echo "Filename: $file , Number of sequences: $seq_coun
t , File Size: $size_chars characters"
        fi
done
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
"fourteen.sh" 9L, 241B                                 7,64-78       All
```



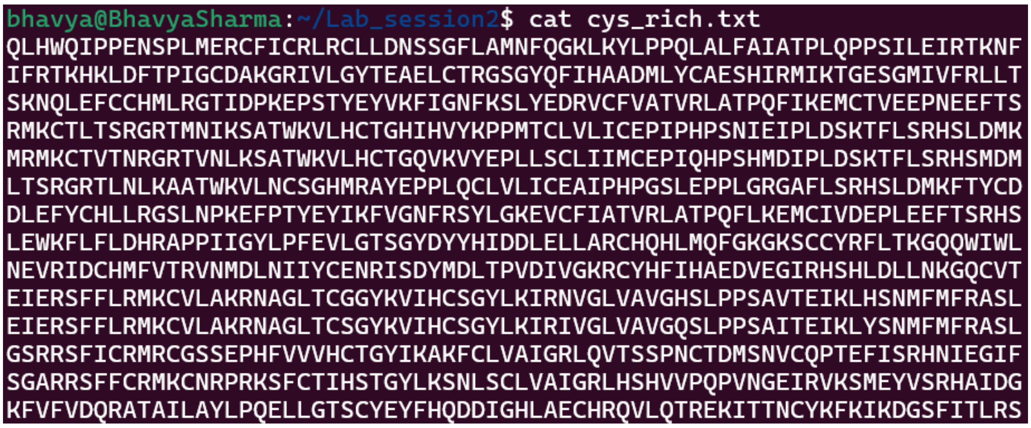
```
bhavya@BhavyaSharma:~/Lab_session2$ chmod +x fourteen.sh
bhavya@BhavyaSharma:~/Lab_session2$ ./fourteen.sh
Filename: sequence2.fasta , Number of sequences: 4 , File Size: 1710 c
haracters
Filename: sequence4.fasta , Number of sequences: 4 , File Size: 2374 c
haracters
Filename: sequence5.fasta , Number of sequences: 13 , File Size: 4229
characters
```

Using chmod command, the shell script is made executable and the output was displayed.

Q15)
Output-



```
bhavya@BhavyaSharma:~/Lab_session2$ grep -v '^>' sequence5.fasta | gre
p 'C.*C.*C' > cys_rich.txt
```



```
bhavya@BhavyaSharma:~/Lab_session2$ cat cys_rich.txt
QLHWQIPPENSPLMERCFICRLRCLLDNSSGFLAMNFQGKLKYLPPQLALFAIATPLQPPSILEIRTKNF
IFRTKHKLDFTPIGCDAKGRIVLGYTEAELCTRGSGYQFIHAADMLYCAESHIRMIKTGESGMIVFRLLT
SKNQLEFCCHMLRGTIDPKEPSTYEYVKFIGNFKSLYEDRVCFVATVRLATPQFIKEMCTVEEPNEEFTS
RMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYKPPMTCLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMK
MRMKCTVTNRGRTVNLKSATWKVLHCTGQVKVYEPLLSCLIIMCEPIQHPSHMDIPLDSKTFLSRHSMDM
LTSRGRTLNLKAATWKVLNCSGHMRAYEPPLQCLVLICEAIPHPGSLEPPLGRGAFLSRHSLDMKFTYCD
DLEFYCHLLRGSLNPKEFPTYEYIKFVGNFRSYLGKEVCFIATVRLATPQFLKEMCIVDEPLEEFTSRHS
LEWKFLFLDHRAPPIIGYLPFEVLGTSGYDYYHIDDLELLARCHQHLMQFGKGKSCCYRFLTKGQQWIWL
NEVRIDCHMFVTRVNMDLNIIYCENRISDYMDLTPVDIVGKRCYHFIHAEDVEGIRHSHLDLLNKGQCVT
EIERSFFLRMKCVLAKRNAGLTCGGYKVIHCSGYLKIRNVGLVAVGHSLPPSAVTEIKLHSNMFMFRASL
EIERSFFLRMKCVLAKRNAGLTCSGYKVIHCSGYLKIRIVGLVAVGQSLPPSAITEIKLYSNMFMFRASL
GSRRSFICRMRCGSSEPHFVVVHCTGYIKAKFCLVAIGRLQVTSSPNCTDMSNVCQPTEFISRHNIEGIF
SGARRSFFCRMKCNRPRKSFCTIHSTGYLKSNLSCLVAIGRLHSHVVPQPVNGEIRVKSMEYVSRHAIDG
KFVFVDQRATAILAYLPQELLGTSCYEYFHQDDIGHLAECHRQVLQTREKITTNCYKFKIKDGSFITLRS
```

From sequence5.fasta, extracted only the sequence lines (no headers) that contain 3 or more cysteines (C). Saved the output to a file named cys_rich.txt. Ensured the output file contains no empty lines. And using cat command, printed its entire content from beginning to end.