

Biocomputing

Lab Assignment –2

Part 1 - vi Basics & File Editing

1. Open a new file called notes.txt in vi.

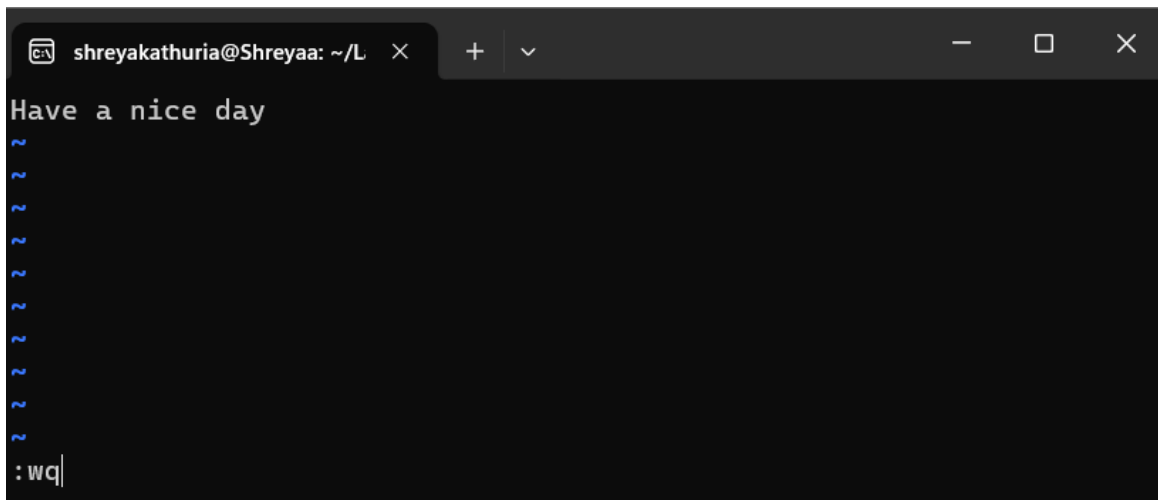
- Insert exactly one line of text:

Have a nice day

(Make sure there is no trailing space at the end.)

- Save and exit.

- Verify that the file contains exactly one line and 15 characters. v



```
shreyakathuria@Shreyaa: ~/L
Have a nice day
~
~
~
~
~
~
~
~
~
~
~
~
~
~
~
:wq
```

```
shreyakathuria@Shreyaa:~/Lab_session2$ wc -m notes.txt
15 notes.txt
shreyakathuria@Shreyaa:~/Lab_session2$ vi notes.txt
shreyakathuria@Shreyaa:~/Lab_session2$ wc -l notes.txt
1 notes.txt
```

Part 2 - Pattern Matching in FASTA Files

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

```
shreyakathuria@Shreyaa:~/Lab_session2$ tail -n 4 sequence.fasta
TAACTACTGATAAGTTACAAAAGTGTCTATCCTAAAGGGCAATACAGCCCTAGACTCTCCCAGGTAT
TTGACTCCTGCAGCAAAAAGGGAAATTGAGGAAATAGAGCAAGCTATTTCTCAGAGGCAACTATATCACA
TAGACACCCCG

shreyakathuria@Shreyaa:~/Lab_session2$ |
```

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

```
shreyakathuria@Shreyaa:~/Lab_session2$ grep "^>" sequence5.fasta
>ahr
>clock
>hif1a
>hif2a
>hif3a
>npas1
>npas2
>npas3
>npas4
>sim1
>sim2
>arnt1
>bmal1
shreyakathuria@Shreyaa:~/Lab_session2$ |
```

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

```
shreyakathuria@Shreyaa:~/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
shreyakathuria@Shreyaa:~/Lab_session2$ |
```

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

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

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

```
shreyakathuria@Shreyaa:~/Lab_session2$ grep 'VV' sequence5.fasta
AANFREGLNLQEGEFLLQALNGFVLVVTTDALVFYASSTIQDYLGFGQSDVIHQSVYELIHTEDRAEFQR
IWLQTHYYITYHQWNSRPEFIVCTHTVVSYAEVRAE
TVIYNTKNSQPQCIVCVNYVVSIGIIQHDL
QMDNLYLKALEGFIAVVTQDGMIFLSENISKFMGLTQVELTGHSIFDFTHPCDHEEIRENLSSTERDFF
KFTYCDDRITELIGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSQYRMLAKHGGYVWLETQ
DRIA EVAGYSPDDLIGCSAYEYIHALDSDAVSKSIHTLLSKGQAVTGQYRFLARSGGYLWTQTQATVVS
QTHYYITYHQWNSKPEFIVCTHSVVSADVRVE
DYVHPGDHVEAEQLGMTLERSFFIRMKSTLTKRGVHIKSSGYKVIHITGRLRLRMGLVVAHALPPPTI
ISESVLIYLGFERSELLCKSWYGLLHPEDLAHASAQHYRLLAESGDIQAEMVVRQLQAKTGGWAWIYCLLY
EKSKNAARTREKENSEFYELAKLLPLPSAITSQLDKASIRLTTSYLKMRVVFPEGLGEAWGHSSRTSP
LDNVGRELGSLLQTL DGFIFVVPDGGKIMYISETASVHLGLSQVELTGNSIYEYIHPADHDEMTAVLTA
LDGVAKELGSLLQTL DGFVFAVSDGKIMYISETASVHLGLSQVELTGNSIYEYIHPADHDEMTAVLTA
SYATVHNSRSSRPHCIVSVNYVLTEIEYKEL
ELKHLILEAADGFLFIVSCETGRVVVSDSVTPVLNQPQSEWFGSTLYDQVHPDDVDKLRQLSTSRMCM
GSRRSFCMRGCSSEPHFVVHCTGYIAKAFCLVAIGRLQVTSSPNDMSNVCQPTFISRHNIEGIF
TFVDHRCVATVGYQPQELLGKNIVEFCHPEDQQLRDSFQQVVKLGQVLSVMFRFRSKNQEWLWMRTSS
DELKHLILRAADGFLFVVGCDRGKILFVSESVFKILNYSQNDLIGQSLFDYLHPKDIKVKQELSSSRLC
SGARRSFFCRMKCNRPKSFCTIHSTGYLKSNSCLVAIGRLHSHVVPQPVNGEIRVKSMEYVSRHAIDG
RWFSSFMNPWTKEVEYIVSTNTVVL
shreyakathuria@Shreyaa:~/Lab_session2$
```

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

```
shreyakathuria@Shreyaa:~/Lab_session2$ egrep 'AA|DD' sequence5.fasta
AANFREGLNLQEGEFLLQALNGFVLVVTTDALVFYASSTIQDYLGFGQSDVIHQSVYELIHTEDRAEFQR
IFRTKHKLDFTPIGCDAGRIVLGYTEAELCTRGSGYQFIHAADMLYCAESHIRMIKTGESGMIVFRLLT
RHSLEWKFLFLDHRAPPIIGYLPFEVLGTSGYDYYHVDDLENLAKCHEHLMQYGGKSKCYRFLTKGQQW
KEKSRDAARSRRSKESEVFYELAHQLPLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLIEDDMKAQM
NCFYLKALDGFVMVLTDGDMIIYISDNVNKYMGLTQFELTGHSVDFTHPCDHEEMREMLTHNTQRSFFL
KEKSRDAARCRRSKETEVFYELAHPLPHSVSSHLDKASIMRLAISFLRTHKLLSSVCSENESEAEADQ
KFTYCDRITELIGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSQYRMLAKHGGYVWLETQ
DAARSRRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHRLCAAGEWNQVGAGGEPLDACYL
LTSRGRTLNLKAAATWKVLNCSGHMAYEPPLQCLVLICEAIPHPGSLEPPLGRGAFLSRHSLDMKFTYCD
DRIA EVAGYSPDDLIGCSAYEYIHALDSDAVSKSIHTLLSKGQAVTGQYRFLARSGGYLWTQTQATVVS
KEKSRNAARSRRGKENLEFFELAKLLPLPGAISSQLDKASIVRLSVTYLRLRRFAALGAPPWGLRAAGPP
AGLAPGRRGPAAALVSEVFEQHLGGHILQSLDGFVFALNQEGKFLYISETVSIYLGLSQVEMTGSSVFDYI
LEWKFLFLDHRAPPIIGYLPFEVLGTSGYDYYHIDDELLARCHQHLMQFGKGKSCCYRFLTKGQQWIWL
SRDAARSRRGKENFEFYELAKLLPLPAAITSQLDKASIRLTISYLMRDFANQGDPPWNLMEGPPPNT
IVAALPGFLLVFTAEGKLLYLSVSEHLGHSMDLVAQGDSIYDIIDPADHLTVRQQLTLDRLFRCRF
EKSKNAARTREKENSEFYELAKLLPLPSAITSQLDKASIRLTTSYLKMRVVFPEGLGEAWGHSSRTSP
EKSKNAAKTRREKENGEFYELAKLLPLPSAITSQLDKASIRLTTSYLKMRVVFPEGLGDAWQPSRAGP
ELKHLILEAADGFLFIVSCETGRVVVSDSVTPVLNQPQSEWFGSTLYDQVHPDDVDKLRQLSTSRMCM
DELKHLILRAADGFLFVVGCDRGKILFVSESVFKILNYSQNDLIGQSLFDYLHPKDIKVKQELSSSRLC
KFVFDQQRATAILAYLPQELLGTSCYEYFHQDDIGHLAECHRQVLQTREKITTNCYKFKIKDGSFITLRS
shreyakathuria@Shreyaa:~/Lab_session2$
```

8. Print only the sequence lines (ignore headers) from sequence5.fasta that contain the letter P.

```
shreyakathuria@Shreyaa:~/Lab_session2$ grep -v '^>' sequence5.fasta | grep 'P'
SNPSKRHRDRNLNTELDRLASLLPFQDVINKLDKLSVLRLSVSYLRAKSFFDVALKSSPTERNGGQDNCR
QLHWQIPPENSPLMERCFICRLRCLLDNSSGFLAMNFQGKLKYLPPQLALFAIATPLQPPSILEIRTKNF
IFRTKHKLDFTPIGCDAGRIVLGYTEAELCTRGSGYQFIHAADMLYCAESHIRMIKTGESGMIVFRLLT
KNNRWTWVQSNARLLYKNGRPPDYIIVTQRPLTDEEGTEHLR
VSRNKSEKKRRDQFNVLIKELGSMLPGNARKMDKSTVLQKSIDFLRKHKEITAQSDASEIRQDWKPTFLS
NEEFTQLMLEALDGGFLAIMTDGSIIVVSESVTSLLEHLPSDLVDQSIFNFIPEGEHSEVYKILSTEYLK
SKNQLEFCCHMLRGTIDPKEPSTYEYVKFIGNFKSLYEDRVCFVATVRLATPQFIKEMCTVEEPNEEFTS
RHSLEWKFLFLDHRAPPIIGYLFFEVLGTSGYDYHVDDLENLAKCHEHLMQYGKGKSCYYRFLTKGQQW
IWLQTHYYITYHQWNSRPEFIVCTHTVVSYAEEVRAE
KEKSRDAARSRRSKESEVFYELAHQLPLPHNVSSHLDKASVMRLTISYLRVRKLLDAGDLIEDDMKAQM
NCFYLKALDGFVMVLTDGDMIIYISDNVNKYMGLTQFELTGHSVDFTHPCDHEEMREMLTHNTQRSFFL
RMKCTLTSRGRTMNIKSATWKVLHCTGHIHVYKPPMTCLVLICEPIPHPSNIEIPLDSKTFLSRHSLDMK
FSYCDERITELMGYEPSELLGRSIYEYHALDSHDLTKTHDMFTKGQVTTGQYRMLAKRGGYVWVETQA
TVIYNTKNSQPQCIVCVNYVVSIGIIQHDL
KEKSRDAARCRRSKETEVFYELAHQLPLPHSVSSHLDKASIMRLAISFLRTHKLLSSVCSENESEAEADQ
QMDNLYLKALEGFIAVVTQDGMIFLSENISKFMGLTQVELTGHSIFDFTHPCDHEEIRENLSSTERDFF
MRMKCTVTNRGRTVNLKSATWKVLHCTGQVKVYEPLLSCLIIMCEPIQHPSHMDIPLDSKTFLSRHSDM
KFTYCDDRITELIGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSQYRMLAKHGGYVWLETQ
GTVIYNPRNLQPQCIMCVNYVLSEIEKNDV
DAARSRRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHRLCAAGEWNQVGAGGEPDACYL
KALEGFVMVLTAEQDMAYLSENVSKHLGLSLELIGHSIFDFIHPCDQEELQDALTPTERCFSLRMKST
LTSRGRTLNLKAATWKVLNCSGHMRAYEPPLQCLVLICEAIPHGSLEPPLGRGAFLSRHSLDMKFTYCD
DRIA EVAGYSPDDLIGCSAYEYIHALDSDAVSKSIHTLLSKGQAVTGQYRFLARSGGYLWTQTQATVVS
GRGPQSESIVCVHFLISQVEETGV
KEKSRNAARSRRGKENLEFFELAKLLPLPGAISSQLDKASIVRLSVTYLRRLRFAALGAPPWGLRAAGPP
AGLAPGRRGPAALVSEVFQHLGGHILQSLDGFVFALNQEGKFLYISETVSIYGLSQVEMTGSSVFDYI
HPGDHSEVLEQLGLVQERSFFVRMKSTLTRGLHVKASGYKVIHVTGRLRALGLVALGHTLPPAPLAELP
LHGHMIVFRLSLGLTILACESRVSDHMDLGPSELVGRSCYQFVHGQDATRIRQSHVDLLDKGQVMTGYR
WLQRAGGFVWLQSVATVAGSGKSPGEHHVLWVSHVLSQAEGGQT
NKSEKKRRDQFNVLIKELSSMLPGNTRKMDKTTVLEKVIQFGLQKHNEVSAQTEICDIQQDWKPSFSLNNEE
```

```
shreyakathuria@Shreyaa: ~/Lab_session2$ |
NKSEKKRRDQFNVLIKELSSMLPGNTRKMDKTTVLEKVIQFLQKHNEVSAQTEICDIQQDWKPSFLSNEE
FTQLMLEALDGFIIAVTTDGSIIYVSDSITPLLGHLPSDVMDQNLNLFPEQEHSEVYKILSSEYLKSDS
DLEFYCHLLRGSLLNPKFPTVEYIKFVGNFRSYLGKEVCFIATVRLATPQFLKEMCIVDEPLEEFTSRHS
LEWKFLFLDHRAPPIIGYLPEFVLGTSGYDYHIDDELLARCHQHLMQFGKGKSCCYRFLTkgqQWiwL
QTHYYITYHQWNSKPEFIVCTHSVVSYADVRVE
SRDAARSRRGKENFEFELAKLLPLPAAITSQLDKASIIRLTISYLKMRDFANQGDPPWNLRMEGPPPPNT
SVKVIGAQRRRSPSALAIEVFEHLGSHILQSLDGFVFNQEGKFLYISETVSIYLGLSQVELTGSSVF
DYVHPGDHVEMAEQLGMLTSLERSFFIRMKSTLTKRGVHIKSSGYKVIHITGRRLRMGLVVVAHALPPPTI
NEVRIDCHMFVTRVNMOLNIIYCENRISDYMDLTPVDIVGKRCYHFIHAEDVEGIRHSHDLLNKGQCVT
KYYRWQMKGNGGYIWIQSSATIAINAKNANEKNIWVNYLLSNPEYKDT
GASKARRDQINAEIRNLKELLPLAEADKVRLSYLHIMSLACIYTRKGVFFAGGTPLAGPTGLLSAQELED
IVAALPGFLLVFTAEGKLLYLSSEVSEHLGHSMDLVAQGDSDIYDIIDPADHLTVRQQLTLTDRLFRCRF
NTSKSLRRQSAGNKLVLIRGRFAHNPFVTAFCAPLEPRPRPGPGPGPASLFLAMFQSRHAKDLALLD
ISESVLIYLGFERSELLCKSWYGLLHPEDLAHASAQHYRLLAESGDIQAEMVVRLQAKTGGWAWIYCLLY
SEGPEGPITANNYPISDMEAWSLRQQL
EKSKNAARTRREKENSEFELAKLLPLPSAITSQLDKASIIRLTISYLKMRVVFPEGLGEAWGHSSRTSP
LDNVGRELGSLLQTLDFGFIVVAPDGMIMYISETASVHLGLSQVELTGNSIYEYIHPADHDEMTAVLTA
EIERSSFFLRMKCVLAKRNAGLTCGGYKVIHCSGYLKIRNVGLVAVGHSLLPPSAVTEIKLHNSMFMFRASL
DMKLIIFLDSRVAELTGYPEQDLIEKTLYHHVHGCDTFHLRCAHHLLLKVGQVTTKYRFLAKHGGWVWVQ
SYATIVHNSRSSRPHCIVSVNYVLTDEYKGL
EKSKNAARTRREKENGEFELAKLLPLPSAITSQLDKASIIRLTISYLKMRVVFPEGLGDWAGQPSRAGP
LDGVAKELGSHLLQTLDFGFVFNASDGMIMYISETASVHLGLSQVELTGNSIYEYIHPADHDEMTAVLTA
EIERSSFFLRMKCVLAKRNAGLTCGGYKVIHCSGYLKIRNVGLVAVGHSLLPPSAITEIKLYSNMFMFRASL
DLKLIIFLDSRVTEVTGYEPQDLIEKTLYHHVHGCDVFHLRYAHHLLLKVGQVTTKYRLLSKRGGWVWVQ
SYATVVHNSRSSRPHCIVSVNYVLTDEYKEL
NHSEIERRRRNKMTAYITELSDMVP TCSALARKPDKLTILRMAVSHMKSRLGTGNTSTDGSYKPSFLTQDQ
ELKHLILEAADGFLFIWSCETGRVYVSDSVTPVLNQPSSEWFGSTLYDQVHPDDVDKLRQLSTSRMCM
GSRSSFICRMRCGSSEPHFVVVHCTGYIKAKFCLVAIGRLQVTSSPNCTDMSNVCQPTFISRHNIIEGIF
TFVDHRCVATVGYPQELLGKNIVEFCHPEDQQLRDSFQQVVKLGQVLSVMFRFRSKNQEWLWMRTSS
FTFQNPYSDEIEYIICNTNTVK
EAHSQIEKRRRDKMNSFIDELASLVPTCNAMSRKLDKLTVLRMAVQHMKTLRGATNPYTEANYKPTFLSD
DELKHLILRAADGFLFVVGCDRGKILFVSESVFKILNYSQNDLIGQSLFDYLHPKDIKAVKEQLSSSRLC
SGARRSFFCRMKNRPRKSFCTIHSTGYLKNLSCLVAIGRLHSHVVPQPVNGEIRVKSMEYVSRHAIDG
KFVFVDQRATAILAYLPQELLGTSCYEFHQDDIGHLAECHRQVLQTREKITTNCYKFKIKDGSFITLRS
RWFSSFMNPWTKVEYIVSTNTVVL
shreyakathuria@Shreyaa:~/Lab_session2$ |
```

Part 3 - Using Variables

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

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

10. Store the pattern `G{2,}` in a variable and search protein.fasta for sequence lines (ignore headers) with 2 or more consecutive Gs.

```

shreyakathuria@Shreyaa:~/Lab_session2$ pattern1="G\{2,\}"
shreyakathuria@Shreyaa:~/Lab_session2$ grep -v '^>' sequence5.fasta
a | grep "$pattern1"
SNPSKRHRDRLNTELDRLASLLPFPQDVINKLDKLSVLRLSVSYLRAKSFFDVALKSSPTERNGGQ
DNCR
FSYCDERITELMGYEPEELLGRSIYEYHALDSDHLTKTHDMFTKGQVTTGQYRMLAKRGGYVWV
ETQA
KFTYCDDRITELIGYHPEELLGRSAYEFYHALDSENMTKSHQNLCTKGQVVSGQYRMLAKHGGYVW
LETQ
DAARSRRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHRLCAAGEWNQVGAGGEPLD
ACYL
DRIAEVAGYSPDDLIGCSAYEYIHALDSDAVSKSIHTLLSKGQAVTGQYRFLARSGGYLWTQTQAT
VVSG
AGLAPGRRGPAALVSEVFEQHLGGHILQSLDGFVFALNQEGKFLYISETVSIYLGLSQVEMTGSSV
FDYI
WLQRAAGGFVWLQSVATVAGSGKSPGEHHVLWVSHVLSQAEGGQT
KYYRWMQKNGGYIWIQSSATIAINAKNANEKNIIWVNYLLSNPEYKDT
GASKARRDQINAEIRNLKELLPLAEADKVRLSYLHIMSLACIYTRKGVFFAGGTPLAGPTGLLSAQ
ELED
ISESVLIYLGFERSELLCKSWYGLLHPEDLAHASAQHYRLLAESGDIQAEMVVRLQAKTGGWAWIY
CLLY
EIERSFFLRMKCVLAKRNAGLTCGGYKVIHCSGYLKIRNVGLVAVGHSLPPSAVTEIKLHSNMFMF
RASL
DMKLIFLDSRVAELTGYPQDLIEKTLYHHVHGCDTFHLRCAHHLLL VKGQVTTKYRFLAKHGGW
VWVQ
DLKLIFLDSRVTEVTGYEPQDLIEKTLYHHVHGCDVFHLRYAHHLLL VKGQVTTKYRLLSKRGGW
VWVQ
shreyakathuria@Shreyaa:~/Lab_session2$ |

```

11. Store "Biocomputing" in a variable, export it, and verify that it is available inside a new shell started using:

```

bash -c 'echo $VARIABLE NAME'
shreyakathuria@Shreyaa:~/Lab_session2$ variable1="Biocomputing"
shreyakathuria@Shreyaa:~/Lab_session2$ export variable1
shreyakathuria@Shreyaa:~/Lab_session2$ bash -c 'echo $variable1'
Biocomputing
shreyakathuria@Shreyaa:~/Lab_session2$

```

Part 4 - File Existence & Loops

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

```
shreyakathuria@Shreyaa:~/Lab_session2$ if [ -f sequence3.fasta ]; then  
  wc -l sequence3.fasta ; else echo "Missing file"; fi  
19 sequence3.fasta  
shreyakathuria@Shreyaa:~/Lab_session2$ |
```

13. Using a for loop, go through all .fasta files in the current directory and print: filename, number of sequences, and file size in characters.

```
shreyakathuria@Shreyaa: ~/L × + ▾  
shreyakathuria@Shreyaa:~/Lab_session2$ for file in *.fasta ; do echo "Filename: $file"; echo "Number of Sequences: $(grep -c '^>' "$file")"; echo "file size: $(wc -m < "$file")"  
> done  
Filename: protein.fasta  
Number of Sequences: 1  
file size: 467  
Filename: sequence.fasta  
Number of Sequences: 1  
file size: 79551  
Filename: sequence1.fasta  
Number of Sequences: 1  
file size: 974  
Filename: sequence2.fasta  
Number of Sequences: 4  
file size: 1710  
Filename: sequence3.fasta  
Number of Sequences: 2  
file size: 1000  
Filename: sequence4.fasta  
Number of Sequences: 4  
file size: 2374  
Filename: sequence5.fasta  
Number of Sequences: 13  
file size: 4229  
shreyakathuria@Shreyaa:~/Lab_session2$ |
```

14. Modify the above loop so that it only prints files with more than 3 sequences.


```
shreyakathuria@Shreyaa:~/Lab_session2$ for file in *.fasta; do seqs=$(
grep -c '^>' "$file"); if (( seqs > 3 )); then echo "Filename: $file";
echo "Number of Sequences: $seqs"; echo "File size: $(wc -m < "$file"
)"; fi; done
Filename: sequence2.fasta
Number of Sequences: 4
File size: 1710
Filename: sequence4.fasta
Number of Sequences: 4
File size: 2374
Filename: sequence5.fasta
Number of Sequences: 13
File size: 4229
shreyakathuria@Shreyaa:~/Lab_session2$
```

Part 5 - Applied Data Extraction

15. From sequence5.fasta, extract only the sequence lines (no headers) that contain 3 or more cysteines (C). Save the output to a file named cys_rich.txt. Ensure the output file contains no empty lines.

```
shreyakathuria@Shreyaa:~/Lab_session2$ grep -v '^>' sequence5.fasta |
egrep 'C.*C.*C' | grep . > cys_rich.txt
shreyakathuria@Shreyaa:~/Lab_session2$ grep -n '^$'
^C
shreyakathuria@Shreyaa:~/Lab_session2$ grep -n '^$' cys_rich.txt
shreyakathuria@Shreyaa:~/Lab_session2$
```

Extra Challenge (Optional)

Write a single shell command that finds the file in the current directory with the largest number of sequences (by header count) and prints:

<filename> has <count> sequences

Hint: You will likely need wc, grep, sort, and head

```
shreyakathuria@Shreyaa:~/Lab_session2$ grep -c '^>' *.fasta | sort -t: -k2 -nr
| head -n1
sequence5.fasta:13
shreyakathuria@Shreyaa:~/Lab_session2$
```

USED GEMINI FOR SORT IN OPTIONAL ONE
AND GEEKSFORGEEKS FOR THE COMMAND USAGE MANUAL