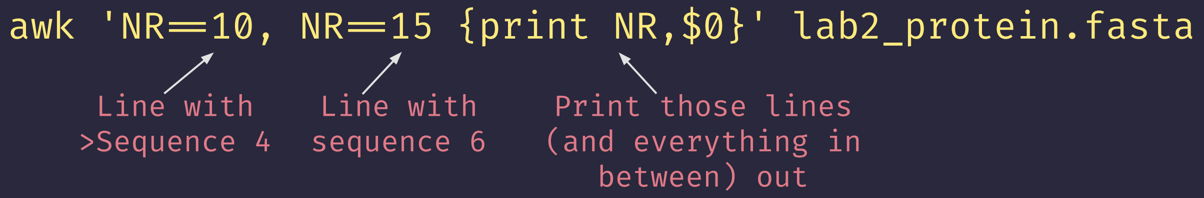
Lab #3

Commands to Edit Files

BINF 2111, Fall 2023

Bonus #1 Review

- BONUS 1: Give an awk command to display Sequence 4 through 6 (including headers) in lab2_protein.fasta

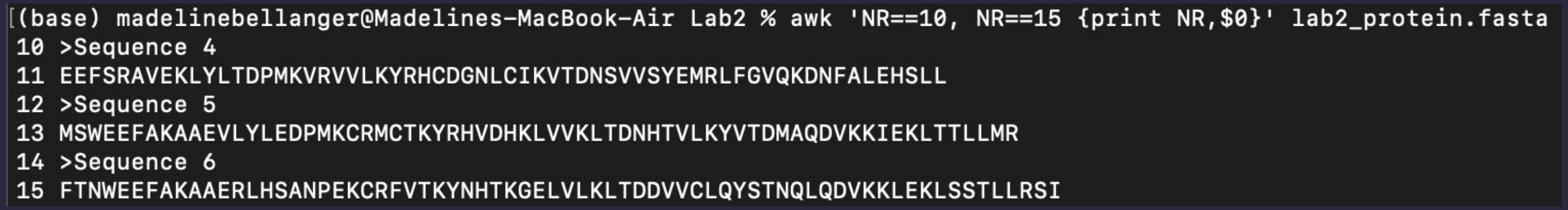


```
awk 'NR==10, NR==15 {print NR,$0}' lab2_protein.fasta
```

Line with >Sequence 4

Line with sequence 6

Print those lines (and everything in between) out



```
(base) madelinebellanger@Madelines-MacBook-Air Lab2 % awk 'NR==10, NR==15 {print NR,$0}' lab2_protein.fasta
10 >Sequence 4
11 EEFSRAVEKLYLTDPMKVRVVLKYRHCDGNLCIKVTDNSVVSYEMRLFGVQKDNFALEHSLL
12 >Sequence 5
13 MSWEEFAKAAEVLYLEDPMKCRMCTKYRHVDHKLVVKLTDNHTVLKYVTDMAQDVKKIEKLTTLLMR
14 >Sequence 6
15 FTNWEEFAKAAERLHSANPEKCRFVTKYNHTKGELVLKLTDDVVCLQYSTNQLQDVKKLEKLSSTLLRSI
```

Bonus #2 Review

- BONUS 2: Give an awk command to find the longest line in lab2_protein.fasta. What is the length?

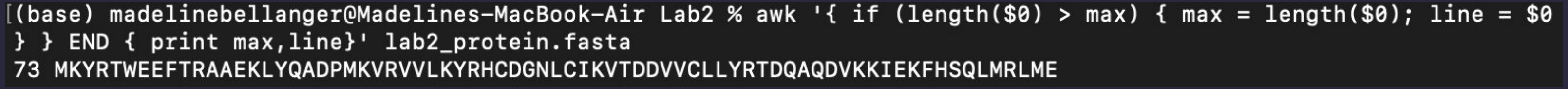
If the length of the current line is longer than the max

Make the length of the current line the max

Make the current line the line

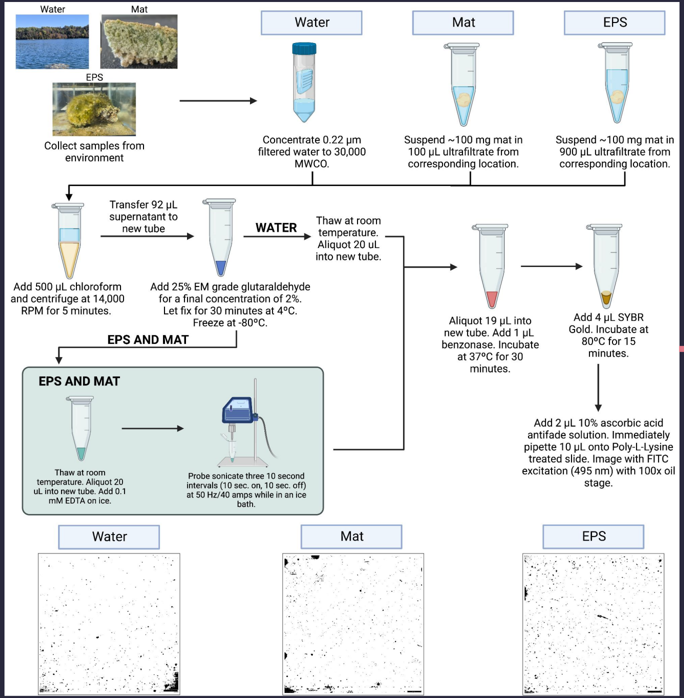
```
awk '{ if (length($0) > max) { max = length($0); line = $0 } } END
{ print max,line }' lab2_protein.fasta
```

The longest line is 73 characters long



```
[(base) madelinebellanger@Madelines-MacBook-Air Lab2 % awk '{ if (length($0) > max) { max = length($0); line = $0
} } END { print max,line}' lab2_protein.fasta
73 MKYRTWEEFTRAAEKLYQADPMKVRVVLKYRHCDGNLCIKVTDDVVCLLYRTDQAQDVKKIEKFHSQLMRLME
```

Lab 3 Data - EFM Counts



Collect samples from environment

Concentrate 0.22 µm filtered water to 30,000 MWCO.

Suspend ~100 mg mat in 100 µL ultrafiltrate from corresponding location.

Suspend ~100 mg mat in 900 µL ultrafiltrate from corresponding location.

Add 500 µL chloroform and centrifuge at 14,000 RPM for 5 minutes.

Transfer 92 µL supernatant to new tube

Add 25% EM grade glutaraldehyde for a final concentration of 2%. Let fix for 30 minutes at 4ºC. Freeze at -80ºC.

Thaw at room temperature. Aliquot 20 uL into new tube.

EPS AND MAT

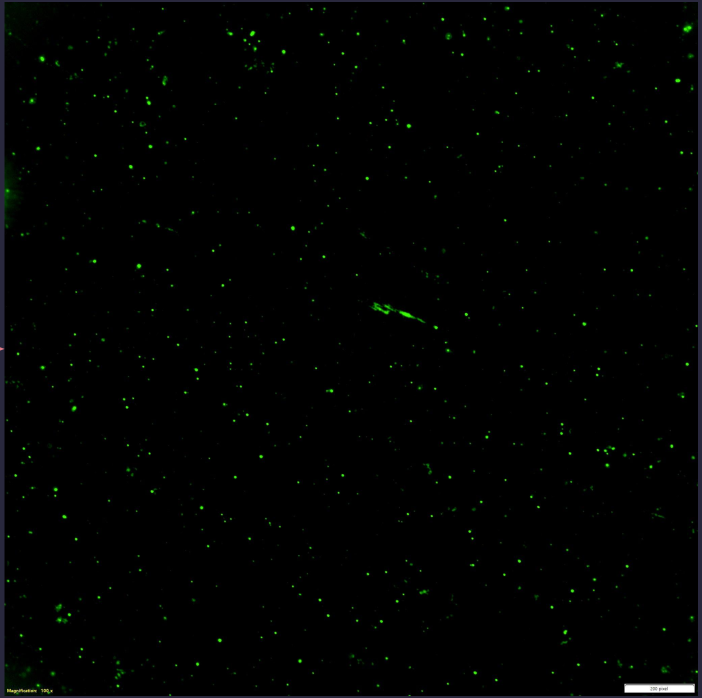
Thaw at room temperature. Aliquot 20 uL into new tube. Add 0.1 mM EDTA on ice.

Probe sonicate three 10 second intervals (10 sec. on, 10 sec. off) at 50 Hz/40 amps while in an ice bath.

Aliquot 19 µL into new tube. Add 1 µL benzonase. Incubate at 37ºC for 30 minutes.

Add 4 µL SYBR Gold. Incubate at 80ºC for 15 minutes.

Add 2 µL 10% ascorbic acid antifade solution. Immediately pipette 10 µL onto Poly-L-Lysine treated slide. Image with FITC excitation (495 nm) with 100x oil stage.



Bellanger et al, 2023

Terminology

TSV

Tab Separated Value - file where each column is separated by tabs

Text editor

A system or program that allows a user to edit text

CSV

Comma Separated Value - file where each column is separated by commas

EFM

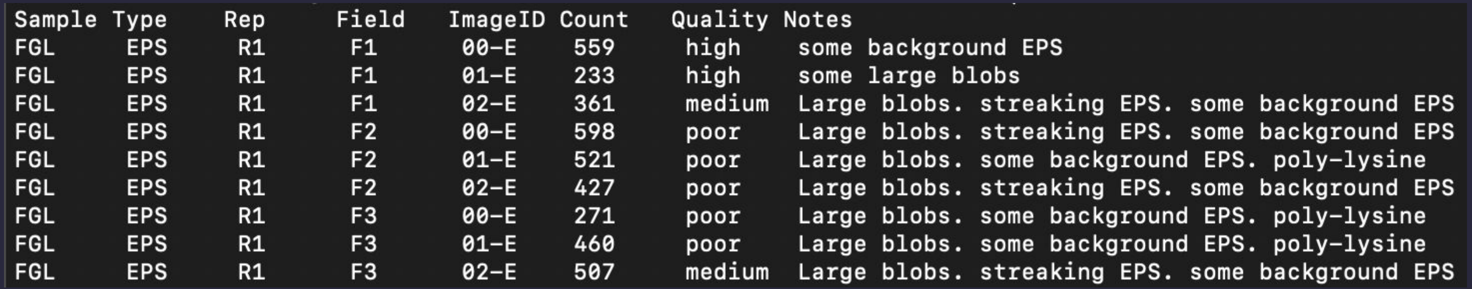
Epifluorescence microscopy. Microscopy technique that uses fluorescence to see particles. (Not necessary to know.)

Script

A list of programmatically-written instructions (commands) that can be carried out when ran

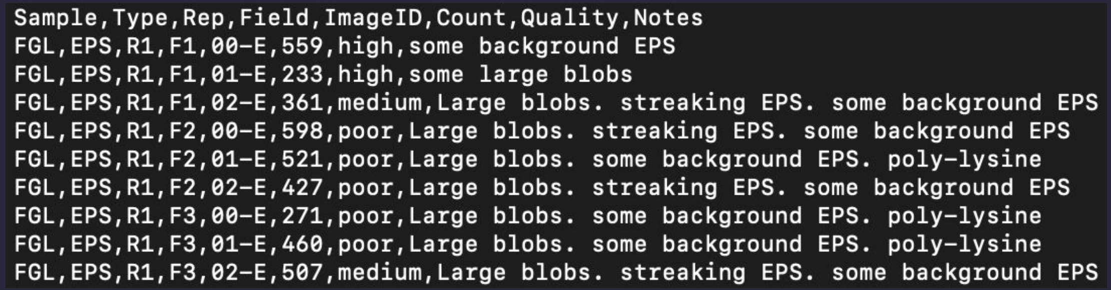
TSVs and CSVs

- TSV: Tab Separated Value



```
Sample Type    Rep     Field   ImageID Count    Quality Notes
FGL     EPS     R1      F1      00-E     559      high    some background EPS
FGL     EPS     R1      F1      01-E     233      high    some large blobs
FGL     EPS     R1      F1      02-E     361      medium  Large blobs. streaking EPS. some background EPS
FGL     EPS     R1      F2      00-E     598      poor    Large blobs. streaking EPS. some background EPS
FGL     EPS     R1      F2      01-E     521      poor    Large blobs. some background EPS. poly-lysine
FGL     EPS     R1      F2      02-E     427      poor    Large blobs. streaking EPS. some background EPS
FGL     EPS     R1      F3      00-E     271      poor    Large blobs. some background EPS. poly-lysine
FGL     EPS     R1      F3      01-E     460      poor    Large blobs. some background EPS. poly-lysine
FGL     EPS     R1      F3      02-E     507      medium  Large blobs. streaking EPS. some background EPS
```

- CSV: Comma Separated Value



```
Sample,Type,Rep,Field,ImageID,Count,Quality,Notes
FGL,EPS,R1,F1,00-E,559,high,some background EPS
FGL,EPS,R1,F1,01-E,233,high,some large blobs
FGL,EPS,R1,F1,02-E,361,medium,Large blobs. streaking EPS. some background EPS
FGL,EPS,R1,F2,00-E,598,poor,Large blobs. streaking EPS. some background EPS
FGL,EPS,R1,F2,01-E,521,poor,Large blobs. some background EPS. poly-lysine
FGL,EPS,R1,F2,02-E,427,poor,Large blobs. streaking EPS. some background EPS
FGL,EPS,R1,F3,00-E,271,poor,Large blobs. some background EPS. poly-lysine
FGL,EPS,R1,F3,01-E,460,poor,Large blobs. some background EPS. poly-lysine
FGL,EPS,R1,F3,02-E,507,medium,Large blobs. streaking EPS. some background EPS
```

Commands To Know

Commands are case sensitive !!

● Command

● Options

● Input (like a file or folder)

Command	Meaning	Usage
cut	Cut out sections of files	cut [options] [file]
sort	Sort a file line by line	sort [options] [file]
uniq	Prints or deletes the repeated lines in a file	uniq [options] [file]
wget	Web get. Download online files to your computer	wget [link]
tar	Tape archive, used to create Archive and extract the Archive files	tar [options] [file]
gzip	Compress a file to be gzipped	gzip [file]
gunzip	Uncompress a file that was gzipped	gunzip [file.gz]
printf	Format and print text	printf [options] [input]
head	Print the first 10 lines, unless specified	head [number] [file]

Command Breakdown - cut

Mac users:
brew install coreutils

Use gcut instead of cut!

- **cut**: Cut out sections of files
  - Useful Options
    - -c [#] Character, cut by character [#]
    - -f [#] Field, cut by column [#]
    - -d "[delim]" Delimiter, comma (,) or tab (\t)
    - --complement Get the opposite/complement of what is requested. Used with -f or -c.
  - Usage
    - `cut -c 2,3 --complement file.txt`
    - `cut -d "," -f 1 file.txt`

Command Breakdown - sort

- **sort**: Sort a file line by line
  - Useful Options
    - -t "[delim]" Type, delimiter used in file
    - -k [#] Sort column [#]
    - -n Sort numerically
    - -r Sort in reverse order
    - -u Sort and remove duplicates
  - Usage
    - `sort -k 2n file.txt`
    - `sort -ur file.txt`
    - `sort -t "," -nr file.txt`

Command Breakdown - uniq

- **uniq**: Prints or deletes the repeated lines in a file
  - Duplicate lines must be adjacent to each other! Sort before using uniq!
  - Useful Options
    - -c Count repeats
    - -d Only print repeated lines
    - -u Only print unique lines
  - Usage
    - `uniq -c file.txt`
    - `uniq -d file.txt`
    - `uniq -cu file.txt`

Commands to Extract and Compress Files

- **tar**: Tape archive, used to create Archive and extract the Archive files
    - Useful Options
        - -x Extracts files and directories from an existing archive
        - -v Displays verbose information
        - -f Specifies the filename of the archive to be created or extracted
        - -z Uses gzip compression when creating a tar file (gives .tar.gz)
        - -c Creates an archive by bundling files and directories together
    - Usage
        - `tar -xzvf file.txt.tar.gz`
        - `tar -czvf file.tar.gz file.txt`
- **gzip**: Compress a file to be gzipped
    - Usage
        - `gzip file.txt`
- **gunzip**: Uncompress a file that was gzipped
    - Usage
        - `gunzip file.txt.gz`

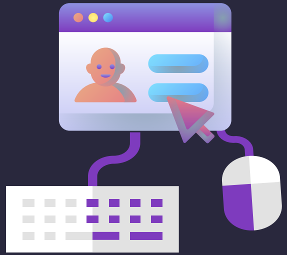
Command Breakdown - printf

- **printf**: Format and print text
  - Useful Formats
    - %d Signed decimal number
    - %s String
    - \n New line (like pressing Enter)
    - \t Tab (like pressing Tab)

  - Usage
    - `printf 'This is a line. \nThis is a new line'`
    - `printf "This is a number: %d\nThis is a string: %s" 72 "hello" > file.txt`
    - `printf "%s\n" "#!/bin/bash" "#This is a script" "echo "Hello World" > hello_world.sh`

Text Editors

- How to open/write in a new file with a text editor:
  - **nano**
    - nano file.txt
    - Begin typing in file
    - (My favorite)
  - **vim**
    - vim file.txt
    - Press I
    - Begin typing in file
  - **gedit**
    - gedit file.txt
    - Begin typing in file
  - **emacs**
    - emacs file.txt
    - Begin typing in file



Mac users need to install:

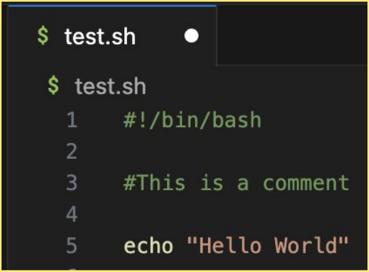
```
brew install gedit
brew install emacs
```

Text Editor Resources

- How to close your file with a text editor:
  - **nano**
    - Cheatsheet for GNU nano
    - Ctrl + X to exit, will be asked to save
    - (My favorite)
  - **vim**
    - Vim Cheat Sheet
    - Press Esc to stop writing
    - :wqa to save and exit
  - **gedit**
    - 37 Keyboard Shortcuts for Gedit
    - Ctrl + W to exit, will be asked to save
  - **emacs**
    - GNU Emacs Reference Card
    - Ctrl + X, Ctrl + C to save and exit

Quick Intro to Scripts

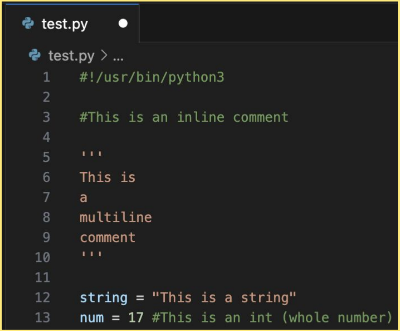
- Scripts - runnable file containing code
  - Bash scripts
    - End in .sh
    - Contain Bash commands
    - Comments start with a #
    - No multiline comments
  - Python scripts
    - End in .py
    - Contain python commands
    - Inline comments start with a #
    - Multiline comments start and end with '''



```bash
#!/bin/bash

#This is a comment

echo "Hello World"
```



```python
#!/usr/bin/python3

#This is an inline comment

'''
This is
a
multiline
comment
'''

string = "This is a string"
num = 17 #This is an int (whole number)
```