

Introduction to command line bioinformatics





Websites

NEOF: <https://neof.org.uk/>

NERC: <https://nerc.ukri.org/>

CGR:

<https://www.liverpool.ac.uk/genomic-research/>



Twitter

NEOF: @NERC_EOF

NERC: @NERCscience

CGR: @CGR_UoL

Upcoming workshops

<https://neof.org.uk/training/>

- Introduction to sequencing data and quality control
 - 1st & 3 November 2022
 - Register before 4pm today!
- R Primer for omics
 - 22nd & 23rd November 2022
- Python for bioinformatics
 - 13th & 15th December 2022
- Population genomics
 - 17th & 19th January 2023
- More!

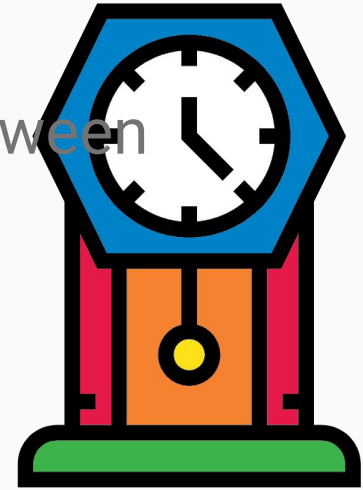


Format & Schedule

This intro
Bookdown
Theory
Practice
Exercises
Optional materials

Work at your own pace
We are here to help
Time with breaks in between

- 10:00-11:20
- 11:30-12:30
- 13:30-14:40
- 15:00-16:00





Plan

Linux, command line, and VNC

Files and Directories

Tips and tricks

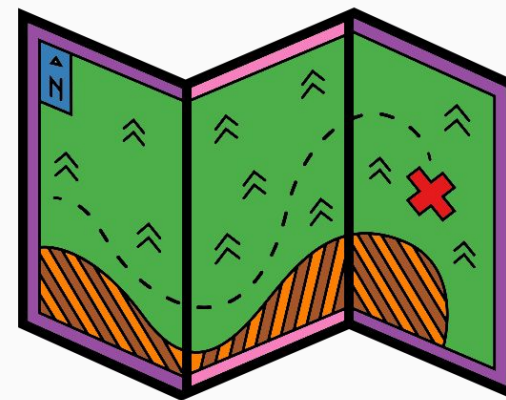
Directory manipulation

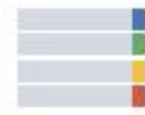
File manipulation

fastq file format

Advanced Linux

Other bioinformatic languages





Linux

Operating system

Free

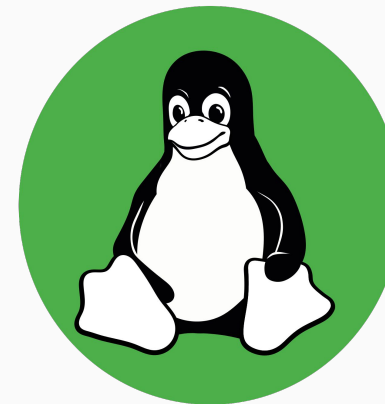
Open source

Multitasking

Multiuser

Bioinformatic programmes

Relatively easy to run and develop software



Applications

Our Future Workshops

Bioinformatics requires:

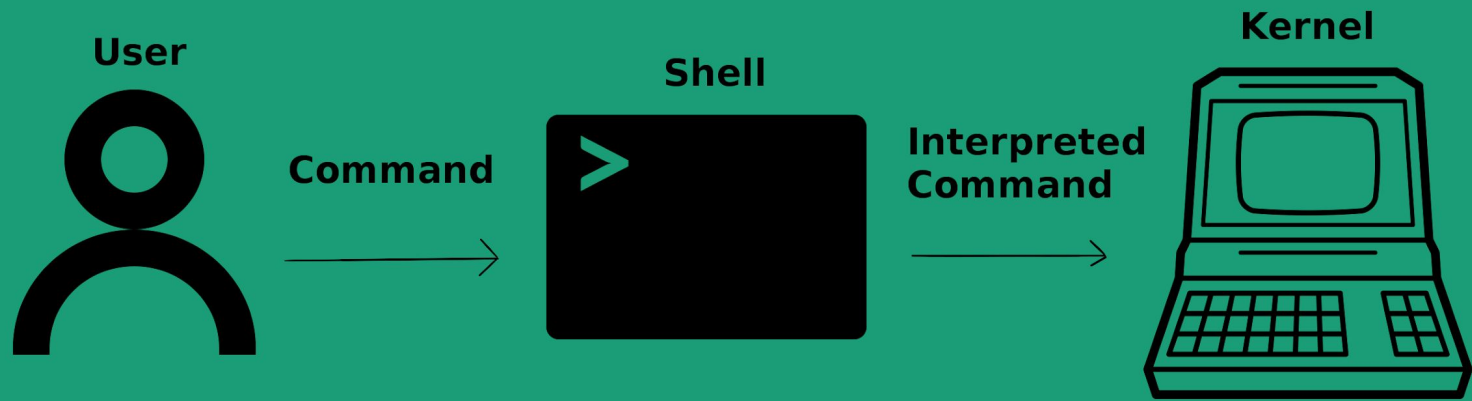
- Storage
- RAM
- Cores

Not available on laptops/desktops

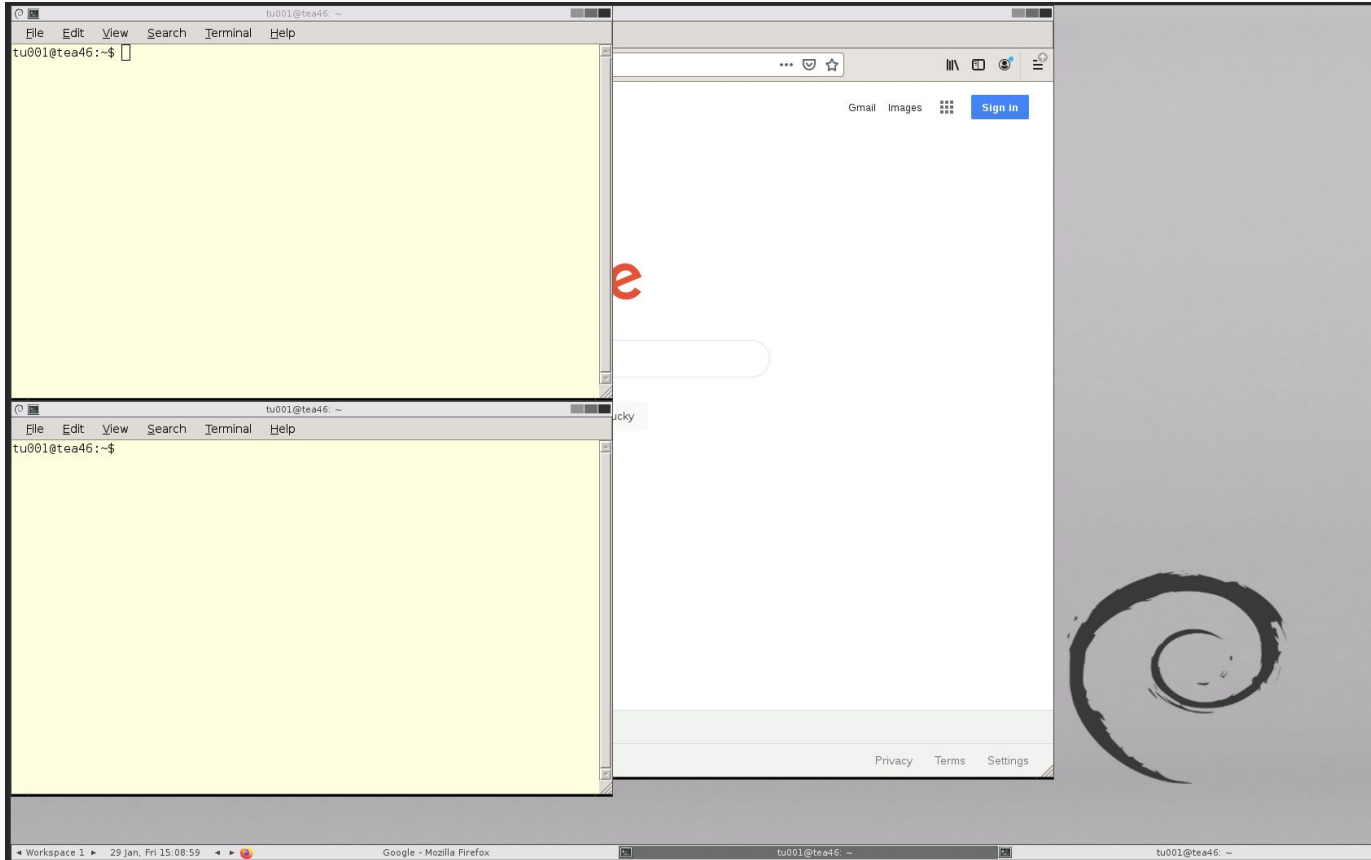
HPCs which utilise unix

>99% of bfx programs work in unix





VNC: Virtual Network Computing





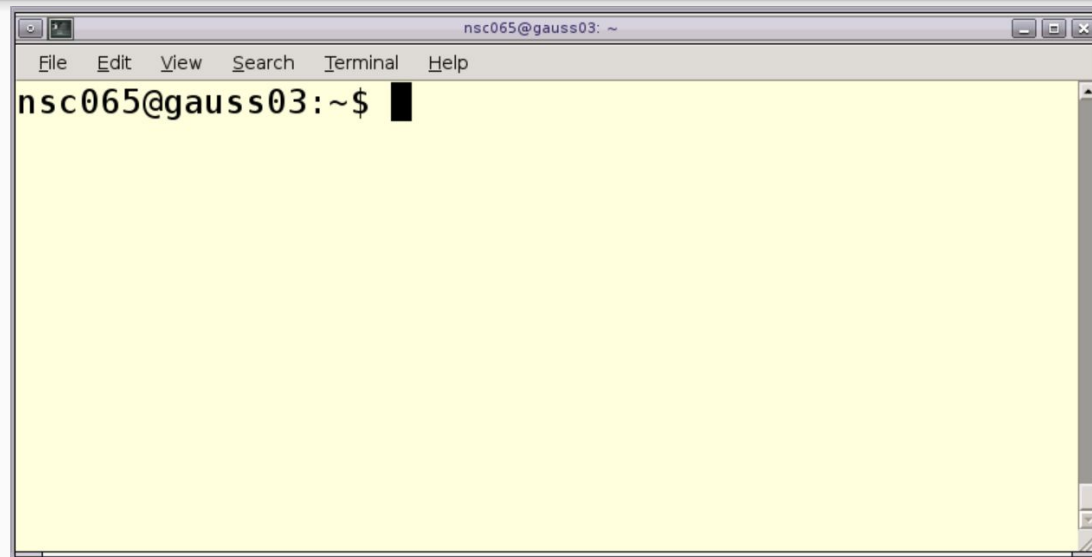
Terminal window

Login name

Machine name

Current directory path

Command line





Directories and files

cd

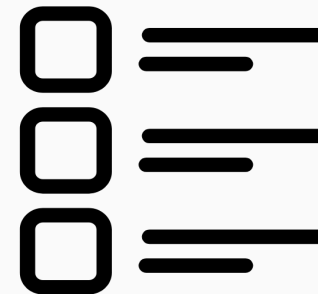
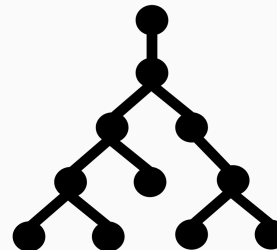
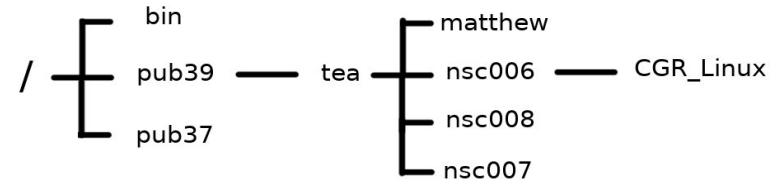
pwd

cp

Directory structure

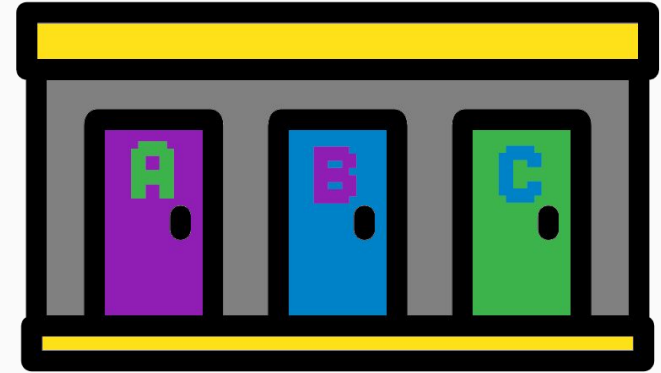
Paths

ls



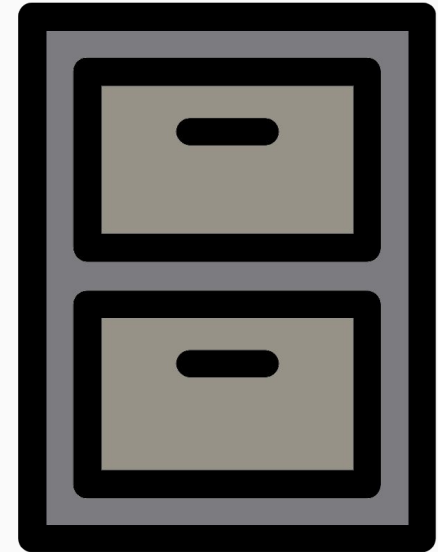
Directories/folders and files

- File storage Building
 - / (aka root)
 - Note: Every other / is a divider
- Rooms
 - **/room_a**
 - **/room_b**
 - **/room_c**, etc.



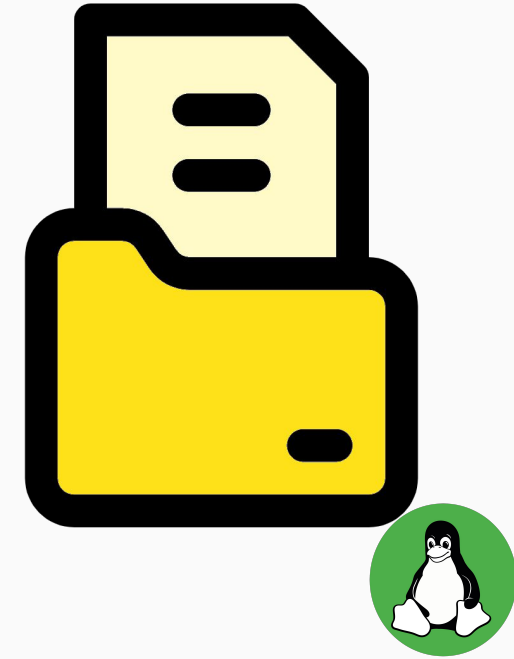
Directories/folders and files

- Filing cabinets
 - `/room_a/cabinet_a`
 - `/room_b/cabinet_b`
 - `/room_c/cabinet_g`
- Filing cabinets contain folders
 - `/room_a/cabinet_a/folder_a`
 - `/room_b/cabinet_b/folder_s`
 - `/room_c/cabinet_g/folder_d`



Directories/folders and files

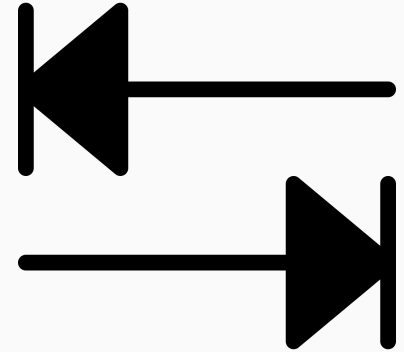
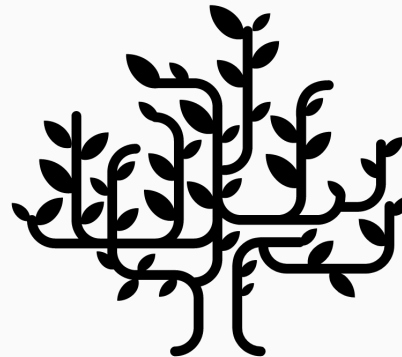
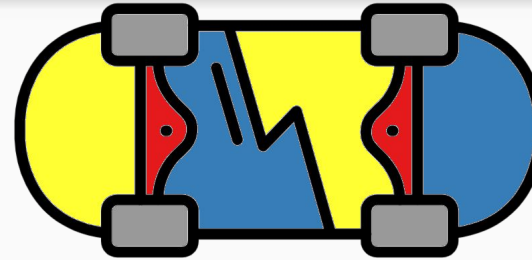
- Folders contain documents
 - `/room_a/cabinet_a/folder_a/file_a`
 - `/room_b/cabinet_b/folder_s/file_e`
 - `/room_c/cabinet_g/folder_d/file_z`
- Building (/) with folders of folders and files
 - Multiple sub folders/directories in a folder/directory
 - Multiple files in these folders/directories





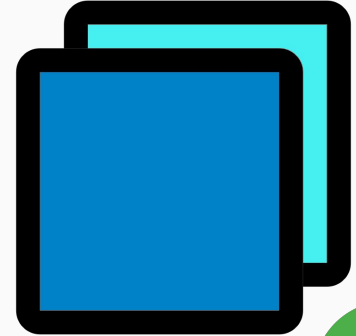
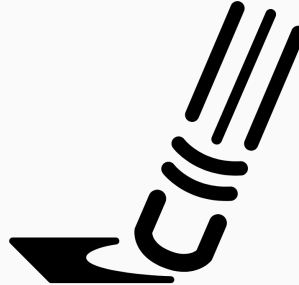
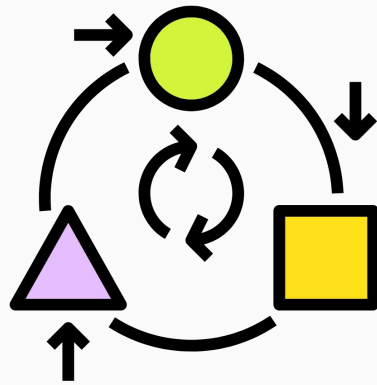
Tips and tricks

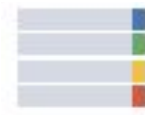
- Tab complete
- Ending a command
- History
- Clear
- Bash escape
- Navigating directories
- Annotations



Manipulating directories

mkdir
mv
cp -r
rm





File manipulation

cat

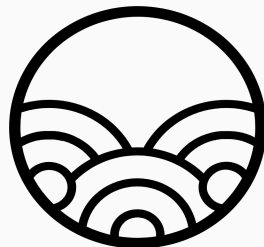
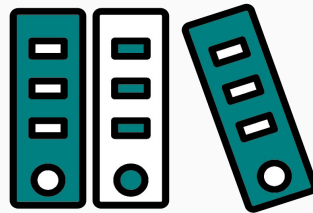
head & tail

less

wc

grep

nano & vim



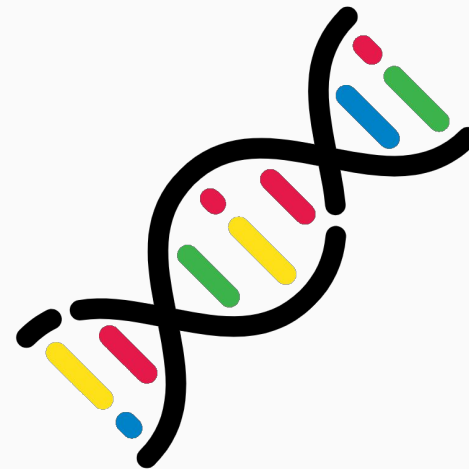
```
    :::  
iLE88dj.  :jd88888dj:  
.LG1tE888D.F8GjjjL8888E;  
ie  :8888Et.      .G8888,  
;i   E888,        :8888,  
      D888,       :8888:  
      D888,       :8888:  
      D888,       :8888:  
      D888,       :8888:  
      888W,       :8888:  
      W88W,       :8888:  
      W88W,       :8888:  
      DGGD:       :8888:  
                   :8888:  
                   :8888:  
                   E888I  
                   tW88D
```



Fastq file format

4 lines per sequence

1. Fastq header
2. Sequence
3. Quality header
4. Quality



@Sequence 1

CTGTTAAATACCGACTTGCGTCAGGTGCGTGAACAACCTGGGCCGCTTT

+

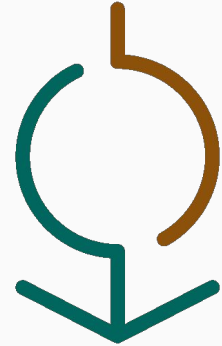
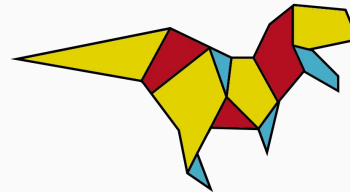
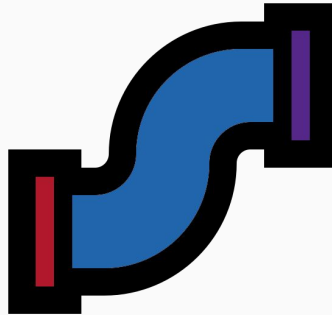
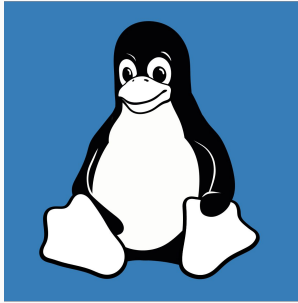
=<<<=>@@@ACDCBCDAC@BAA@BA@BBCBBDA@BB@>CD@A@B?B@@



Advanced Linux

Optional
Wildcards
Redirection
Pipes

Regular expressions
sed
Permissions



Bioinformatic languages

awk

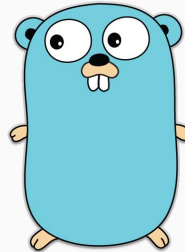
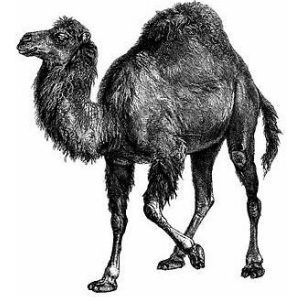
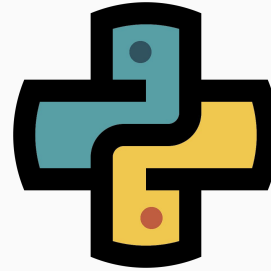
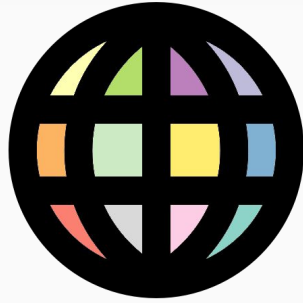
Python

Perl

Ruby

Golang

R





Reminders and Tips

Work at your own pace

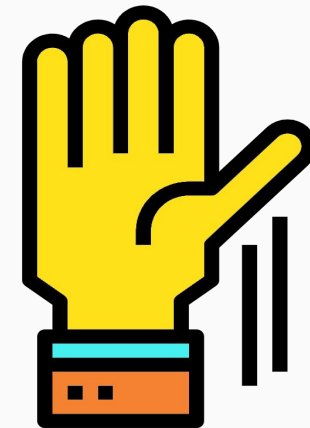
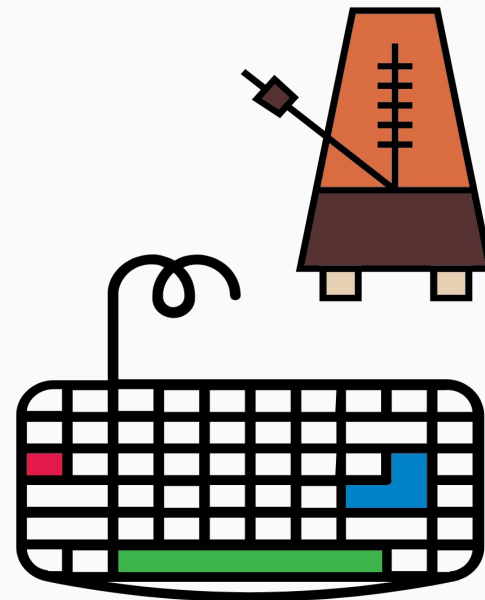
Typos

Ask questions

Breaks are important

Tab, space, and enter

Answers in back of bookdown





During sessions

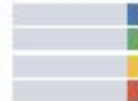
Zoom - Ask via microphone if no question currently being asked/answered

Slack - Ask questions via the channel or ask to go into a zoom breakout room with one of us

WebVNC - We can connect to your webVNC to see and help with issues.

Breakout rooms upon request





Thank you!

Questions?

