

TE&E Time: Bioinformatics I

frederick.jaya@anu.edu.au

caitlin.cherryh@anu.edu.au

Introductions

- Research background
- What bioinformatics are you (planning) on doing?
- One challenge

slido



What does E&E think of "bioinformatics"?

① Start presenting to display the poll results on this slide.



Analysis



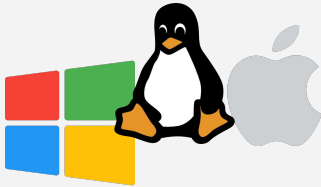
Programming



Software (installation)



Interface



Operating systems



Hardware

1. Best practices
2. Troubleshooting
3. Community



Analysis



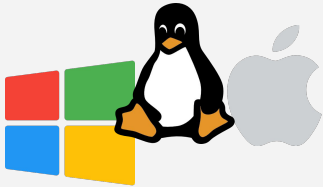
Programming



Software (installation)



Interface



Operating systems



Hardware

1. Best practices
2. Troubleshooting
3. Community

Operating systems (OS)

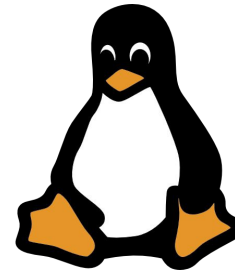
- Different OS → hardware and software managed differently
- **Software compatibility**
- Bioinformatics is increasingly OS-agnostic



Windows



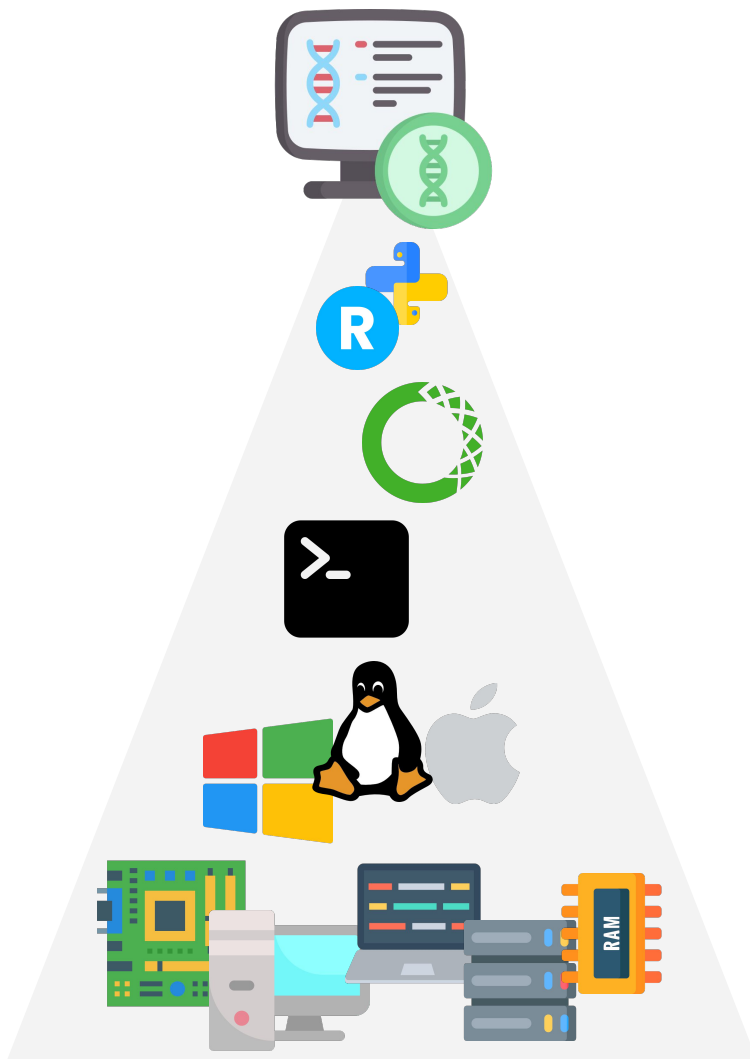
macOS



'Unix-based'

Linux

- Windows Subsystem for Linux (WSL)
- Most tools developed with Unix
- Generally more compatible for bioinformatics
- Computing servers likely use a linux OS



Analysis

Programming

Software (installation)

1. Best practices
2. Troubleshooting
3. Community

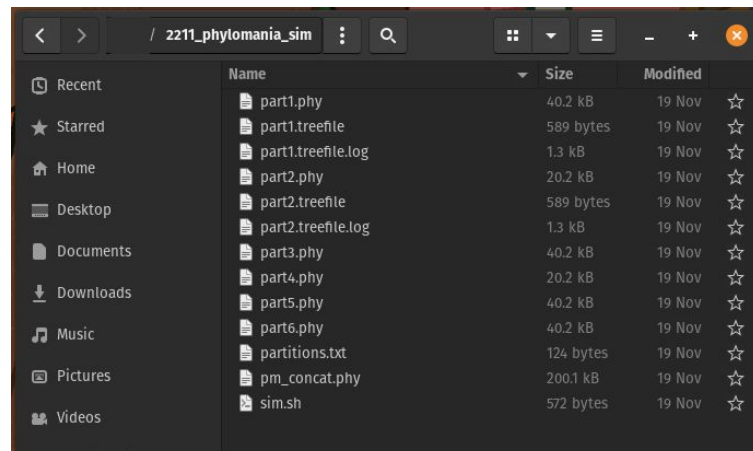
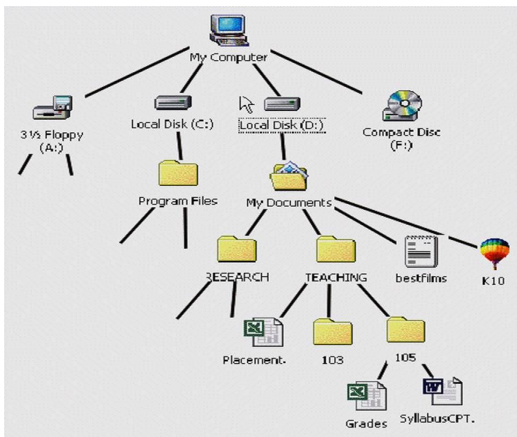
Interface

Operating systems

Hardware

Interfaces and files

- Graphical user interface (GUI) vs. command-line interface (CLI)
- Two ways of:
 1. Navigating the computer to store and access files
 2. Running programs
- Many tools and servers only have a CLI
- But it is also improving! (i.e. Galaxy)



```
(base) frederickjaya@pop-os:~/Dropbox/treemix_rc/02_working/2211_phylomania_sim$ ls -lh
total 420K
-rw-rw-r-- 1 frederickjaya frederickjaya 40K Nov 19 14:14 part1.phy
-rw-rw-r-- 1 frederickjaya frederickjaya 589 Nov 19 14:14 part1.treefile
-rw-rw-r-- 1 frederickjaya frederickjaya 1.3K Nov 19 14:14 part1.treefile.log
-rw-rw-r-- 1 frederickjaya frederickjaya 20K Nov 19 14:14 part2.phy
-rw-rw-r-- 1 frederickjaya frederickjaya 589 Nov 19 14:14 part2.treefile
-rw-rw-r-- 1 frederickjaya frederickjaya 1.3K Nov 19 14:14 part2.treefile.log
-rw-rw-r-- 1 frederickjaya frederickjaya 40K Nov 19 14:14 part3.phy
-rw-rw-r-- 1 frederickjaya frederickjaya 20K Nov 19 14:14 part4.phy
-rw-rw-r-- 1 frederickjaya frederickjaya 40K Nov 19 14:14 part5.phy
-rw-rw-r-- 1 frederickjaya frederickjaya 40K Nov 19 14:14 part6.phy
-rw-rw-r-- 1 frederickjaya frederickjaya 124 Nov 19 14:14 partitions.txt
-rw-rw-r-- 1 frederickjaya frederickjaya 196K Nov 19 14:14 pm_concat.phy
-rwxrwxr-x 1 frederickjaya frederickjaya 572 Nov 19 14:14 sim.sh
```




Analysis



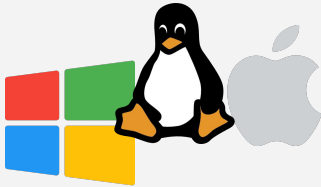
Programming



Software (installation)

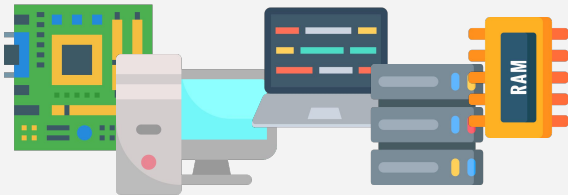


1. Best practices
2. Troubleshooting
3. Community



Interface

Operating systems



Hardware

Software and Installation

- On shared systems, check with your sysadmin



Binary

- + Software, tools, programs
- + Executables stored in `/bin`
- Machine readable

```
-----  
MAFFT: a multiple sequence alignment program  
version 7.471, 2020/Jul/3  
  
http://mafft.cbrc.jp/alignment/software/  
kato@ifrec.osaka-u.ac.jp  
-----  
  
1. COMPILE  
   To use Visual C++ compiler (windows only):  
   % set CC=cl  
  
   % cd core  
   % make clean  
   % make  
   % cd ..
```

Compilation

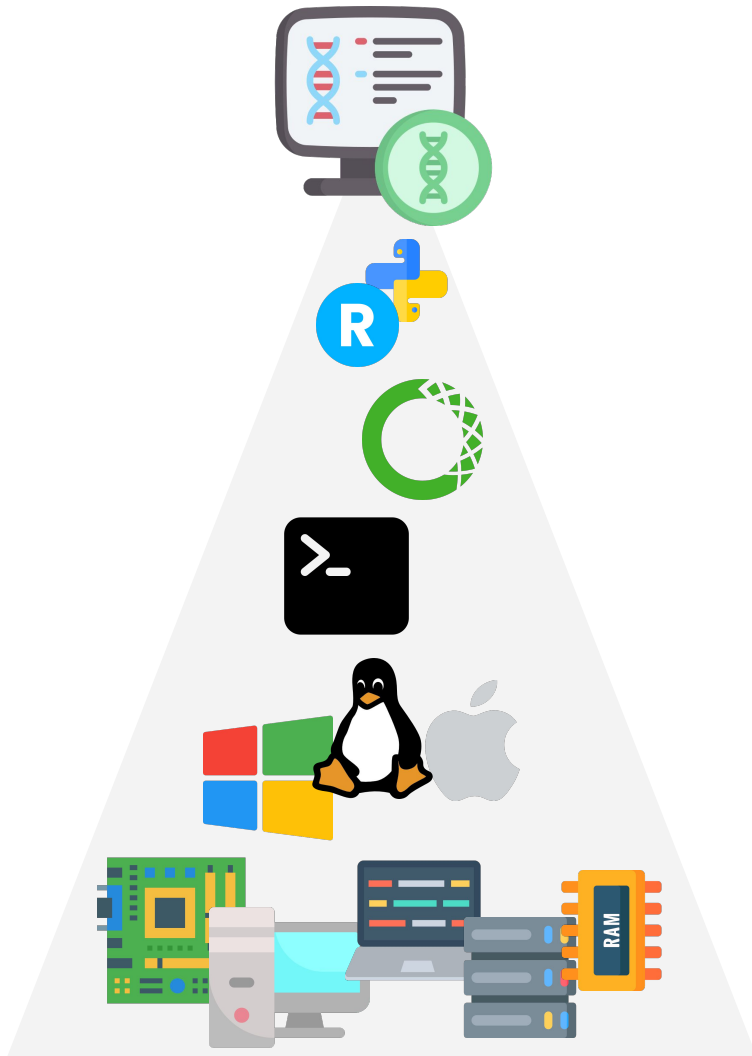
- If binaries are unavailable, you might have to compile it from the source code
- Avoid?



`conda install -c bioconda mafft`

conda

- + A few commands for installation
- + Version control and package management
- + Reproducible
- + Checks for conflicts
- Resolving conflicts can be tricky



Analysis

Programming

Software (installation)

1. Best practices
2. Troubleshooting
3. Community

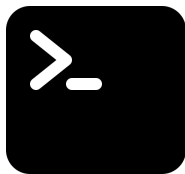
Interface

Operating systems

Hardware

Programming languages

- No superior language - pros and cons for all
- Choose and focus on the one your peers use
 - Package availability



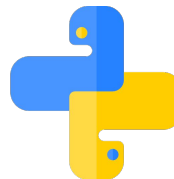
bash/zsh

- The 'shell'
- CLI navigation, moving and editing files
- Calling programs



R/RStudio

- Data wrangling, analysis and visualisation
- Statistical tests and modelling



python



Analysis



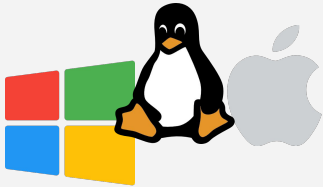
Programming



Software (installation)



Interface



Operating systems



Hardware

1. Best practices
2. **Troubleshooting**
3. **Community**

Troubleshooting

- Normal, necessary, inescapable
- Debugging code - reading the error carefully
- Read the manual

Manual

You



It STILL doesn't work - what now??

- Search online!
- Providing a reproducible example:
 1. Data and code sharing
 2. What have you done? (commands)
 3. What is the error? (logs)



Google it



[StackOverflow](#)

- + General coding issues
- + Issues already solved
- Can be hostile



[BioStars](#)

- + Bioinformatics-specific
- Can get spammy



[ChatGPT!!](#)

- + Fast
- + Detailed answers
- **BE CAREFUL**

Communities (so important)

- RSB HDR and computing teams?
- BDSI drop-in sessions
- [COMBINE/ABACBS](#)
 - [COMBINE Slack](#)

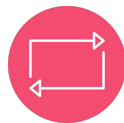
Resources and further reading

- Cheatsheets
- Cookbooks and tutorials (e.g. [R4DS](#))
- [Ten simple rules for getting started with command-line bioinformatics](#)
 - Hardware jargon, HPC, best practices
- [Biostars handbooks](#)
 - Step-by-step analysis-specific tutorials
- [A Quick Guide for Developing Effective Bioinformatics Programming Skills](#)
 - If you're interested in computing and programming
- This [twitter thread](#)
- Dive into your data!

Data (analysis)



Data storage +
management



Reproducible
research (.md)



Common tools and
file formats

Programming

Software



Containerisation

Interface



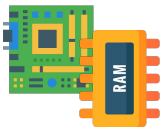
CLI tips + tricks



String processing

Operating systems

Hardware



Hardware



HPC +
(RSB) infrastructure

TE&E Time: Bioinformatics II?

- Hands-on deep dive: IQ-TREE installation + usage
- Specific topics?