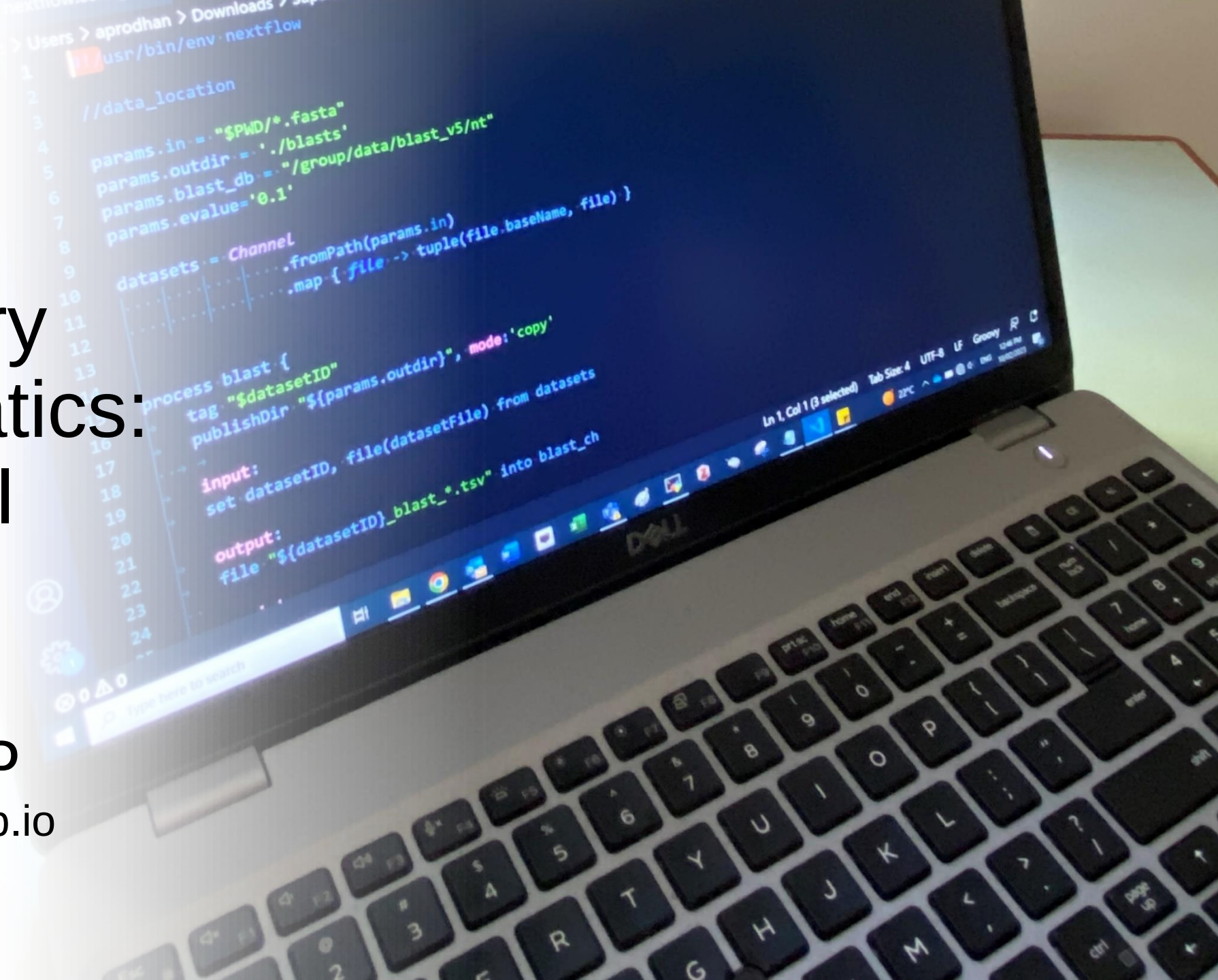
Introductory Bioinformatics: Where do I start?

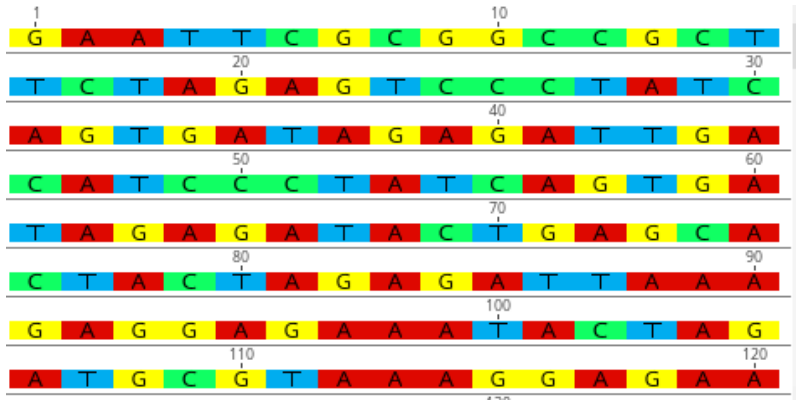


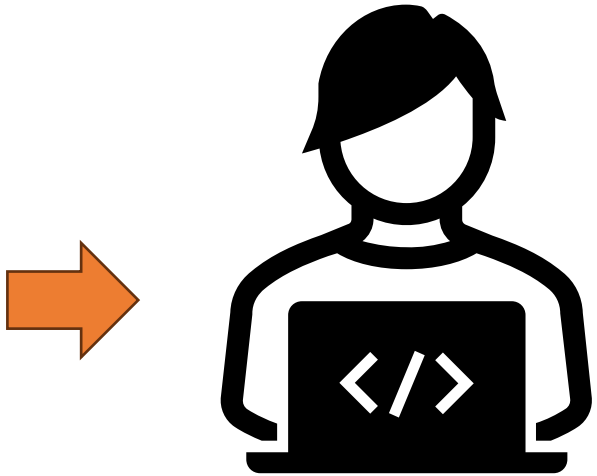
Asad Prodhan PhD

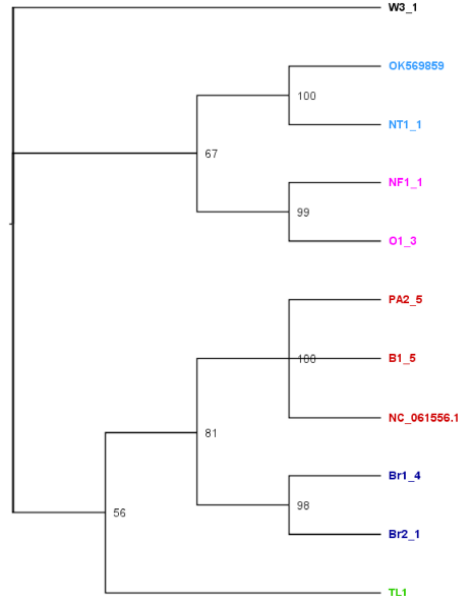
asadprodhan.github.io

What is bioinformatics?

Bioinformatics = Biological Information Processing





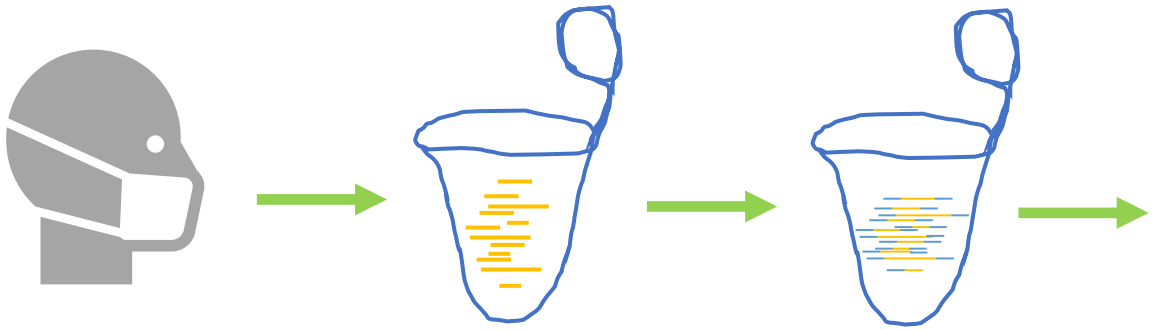


Biological Information
(DNA sequence data)

Processing
(Computer + Computing)

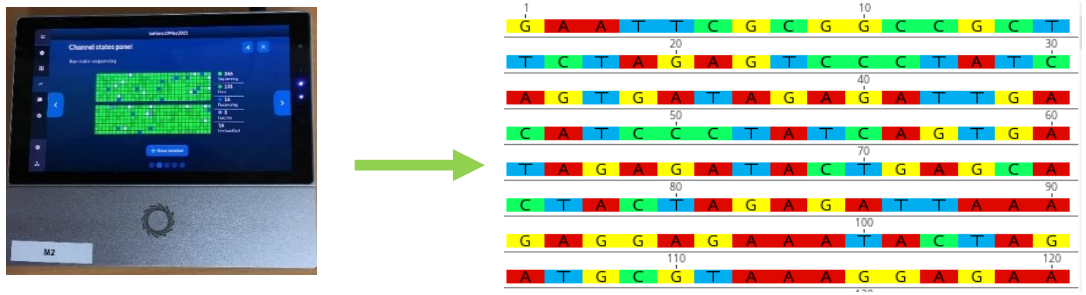
Biological Insight
(Phylogenetic tree)
nature.com/articles/s41598-023-37425-1

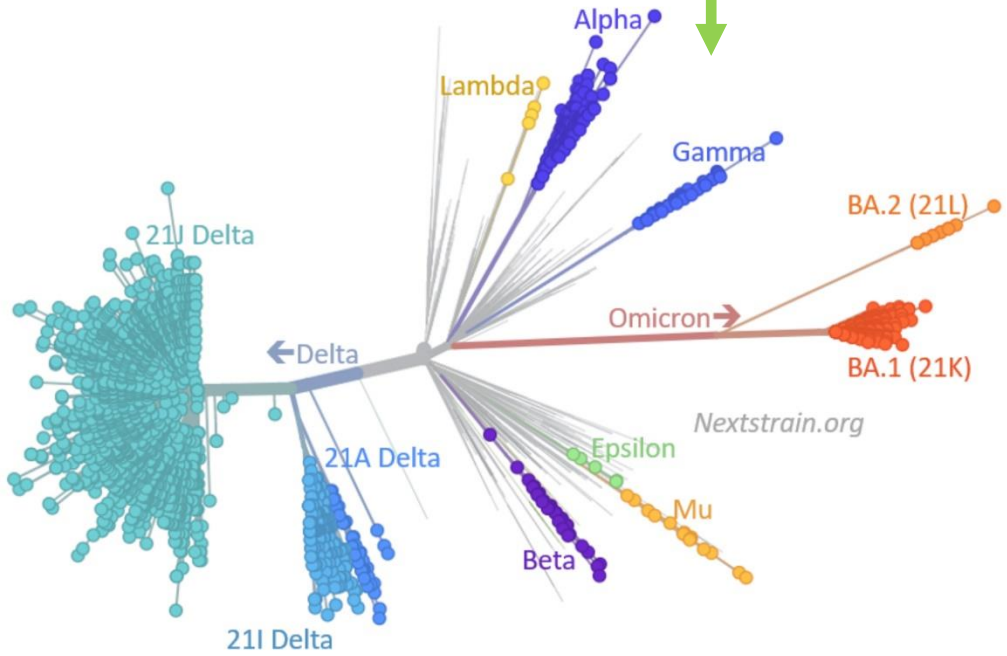
Bioinformatics applications. An example



SARS-CoV-2 | DNA extraction | Library prep | DNA sequencing

Tested positive
but what variant?





Bioinformatics analysis of SARS-CoV-2 evolution, adopted from Dr Emma Hodcroft, Nextstrain

Where do I start?

- A biological question
- Data (new/secondary)
- A computer (GUI/CLI)
- If CLI, then programing

Where do I start?

- A biological question
- Data (new/secondary)
- A computer (GUI/CLI)
- If CLI, then programing

❑ PhD research proposal

❑ Bioinformatics project ideas,

https://www.youtube.com/watch?v=VoqbmLuw5Tw

Where do I start?

- A biological question
- Data (new/secondary)
- A computer (GUI/CLI)
- If CLI, then programing

❑ What data do you need?

❑ Read publications (M & M)

❑ New data?

❑ How to generate? An example

https://github.com/asadprodhan/Invited_talks/blob/main/Nanopore_Workshop_AsadProdhan_DPIRD.pdf

❑ Secondary data?

❑ How to download? Tutorials:

https://www.linkedin.com/feed/update/urn:li:activity:7161193395183325184/

Where do I start?

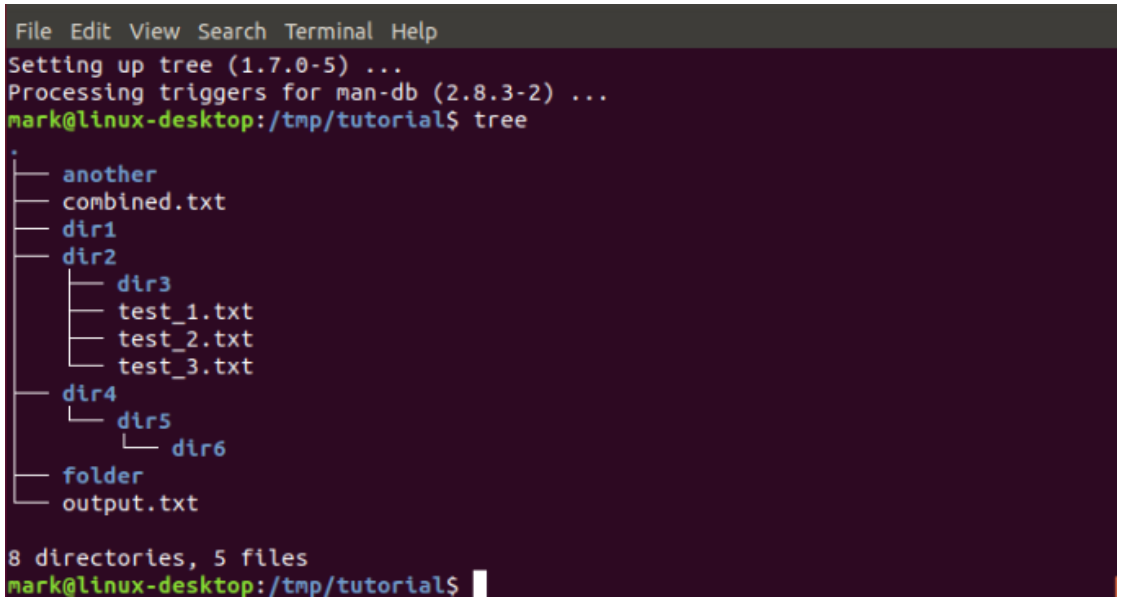
- A biological question
- Data (new/secondary)
- A computer (GUI/CLI)
- If CLI, then programing

- ❑ GUI: any operating system
- ❑ CLI: Mac/Linux/WSL/VirtualBox

GUI

https://galaxy-au-training.github.io/tutorials/modules/galaxy/

CLI



https://ubuntu.com/tutorials/command-line-for-beginners#1-overview

Where do I start?

- A biological question
- Data (new/secondary)
- A computer (GUI/CLI)
- If CLI, then programing

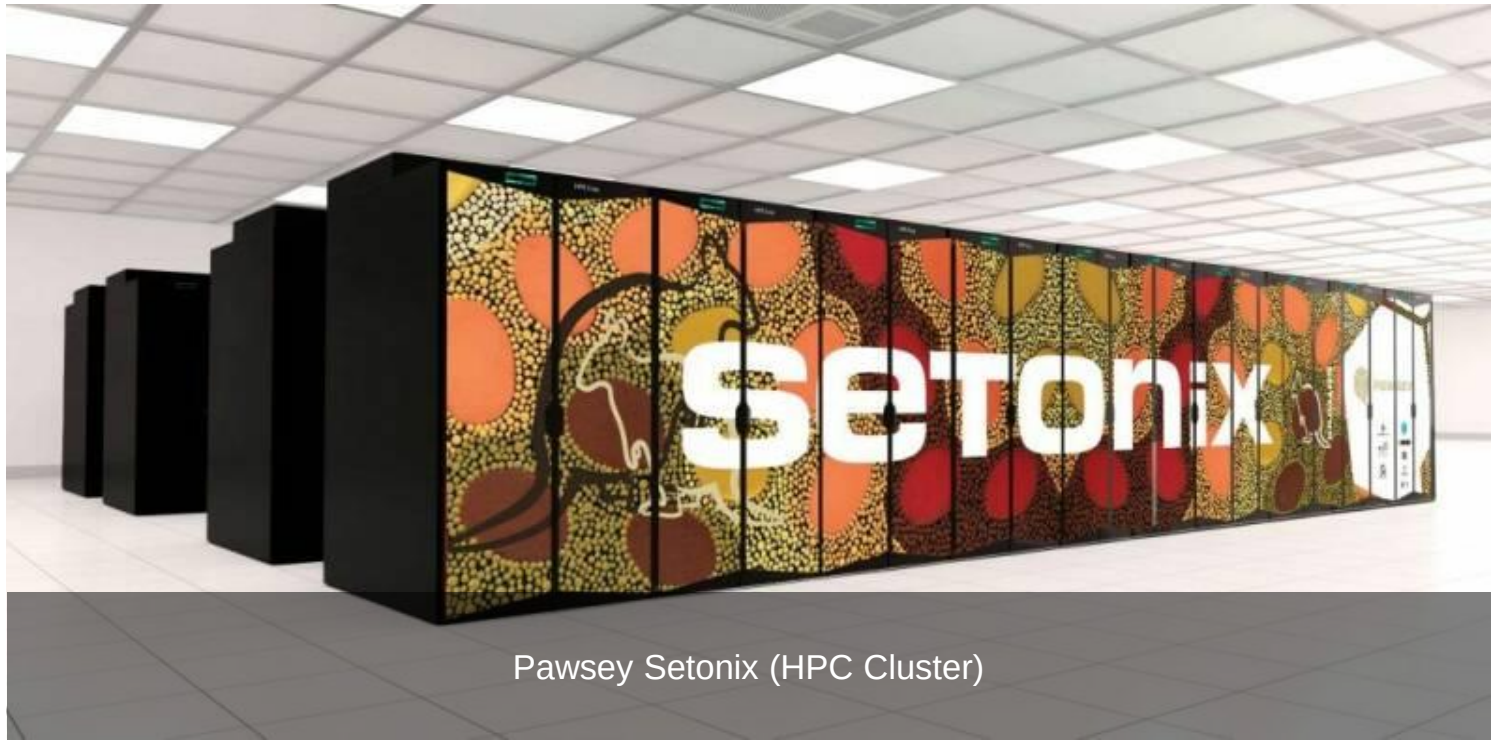
❑ CLI computing facilities

❑ HPC Cluster

❑ Commercial Cloud services

Google Cloud, Microsoft Azure, and Amazon Web Services (AWS)

❑ Pawsey Nimbus



Pawsey Setonix (HPC Cluster)

Pawsey Nimbus (Cloud Service)

Where do I start?

- A biological question
- Data (new/secondary)
- A computer (GUI/CLI)
- If CLI, then programing

❑ CLI computing facilities

❑ HPC Cluster, how to connect?

❑ Commercial Cloud services

Google Cloud, Microsoft Azure, and Amazon Web Services (AWS)

❑ Pawsey Nimbus, how to connect?

❑ **Apply for Pawsey HPC or Nimbus**

Where do I start?

- A biological question
- Data (new/secondary)
- A computer (GUI/CLI)
- If CLI, then programing

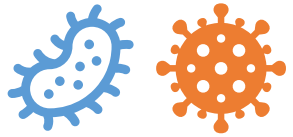
Where do I start?

- A biological question
- Data (new/secondary)
- A computer (GUI/CLI)
- If CLI, then programing

❑ Question > Data > Access to the CLI-based Computer > Now what?

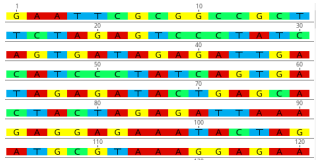
❑ Coding!

Where do I start?

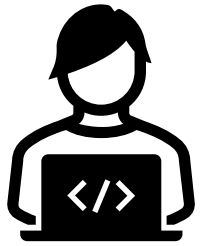


Question

Data generation

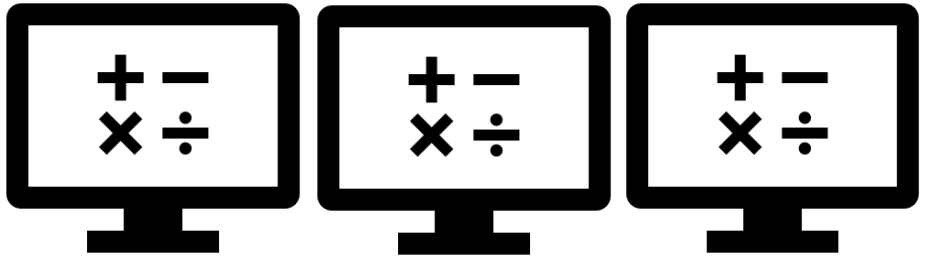


Raw Data



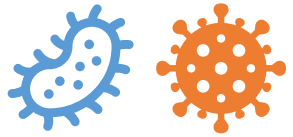
Your laptop

Data processing



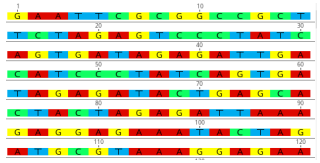
CLI-based computers

Where do I start?

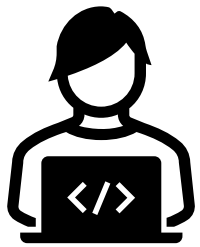


Question

Data generation

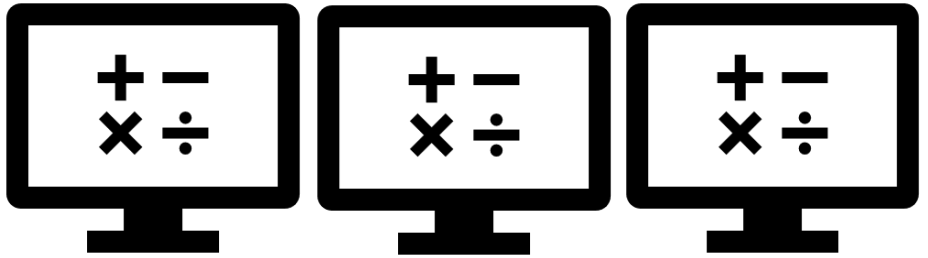


Raw Data



Your laptop

Data processing



CLI-based computers

What to learn?

How to set up your laptop for coding?

How to connect to remote computers?

https://github.com/asadprodhan/How-to-connect-to-a-remote-Linux_computer

https://github.com/asadprodhan/How_to_connect_to_Pawsey_nimbus_instance

How to navigate files using CLI?

https://github.com/asadprodhan/Invited_talks/blob/main/Command_Line_Basics_AsadProdhan.pdf

How to transfer data?

How to install bioinformatics tools?

https://github.com/asadprodhan/A-beginner-s-guide-to-Bioinformatics

How to carry out your analyses?

Hands-on workshop, YouTube

Steep learning curve!

Steep learning curve but REWARDING