Environmental metagenomics

Working with the command line

Environmental metagenomics
April 2022

Igor S. Pessi & Antti Karkman, University of Helsinki

What is Unix?

A family of computer operating systems (OSs)
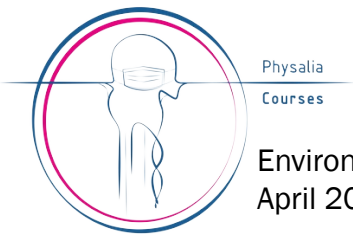
- Linux, MacOS, Solaris, OpenBSD

Key characteristics

Multitasking: multiple software processes can run at the same time

Multiuser: several users can use the same computer at the same time

Multiprocessing: can use more than one computer processor

**Portable:** can be used in various hardware architectures

The Unix philosophy

"The idea that the power of a system comes more from the relationships among programs than from the programs themselves"

- Use of a large number of simple programs performing a limited, well-defined function

- Use of a command-line interpreter ("shell") to combine these programs to perform complex tasks

- Use of plain text for storing data

- Use of a hierarchical filesystem

Physalia
Courses

Environmental metagenomics
April 2022

Igor S. Pessi & Antti Karkman, University of Helsinki
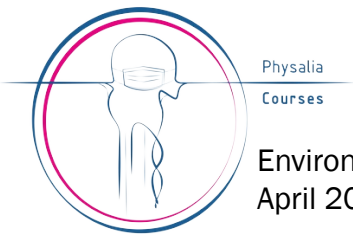
The Unix shell (command-line interpreter)

Interprets sequences of text
- Entered by a user
- From a file
- From a data stream

Primary interface before
graphical user interfaces (GUIs) appeared

Still widely used today
- Efficient
- Low memory footprint
- Advanced scripting

Igor S. Pessi & Antti Karkman, University of Helsinki

Physalia
Courses

Some basic Unix commands

pwd: <u>p</u>rint <u>w</u>orking <u>d</u>irectory ("where am I?")

ls: <u>lis</u>t ("show folder contents")

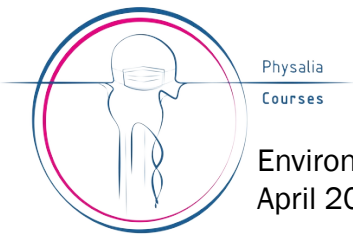mkdir: <u>m</u>a<u>k</u>e <u>dir</u>ectory (a.k.a. folder)

cd: <u>c</u>hange <u>d</u>irectory ("go to folder")

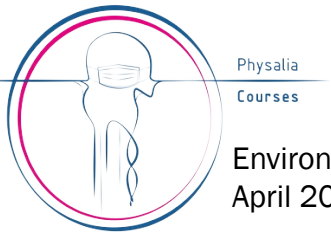cp: <u>c</u>o<u>p</u>y

mv: <u>m</u>o<u>v</u>e

rm: <u>r</u>e<u>m</u>ove

Some additional notes

Case-sensitive
- photo.jpg ≠ PHOTO.jpg

Does not like spaces and special characters in file/folder names
- genome report.txt ❌
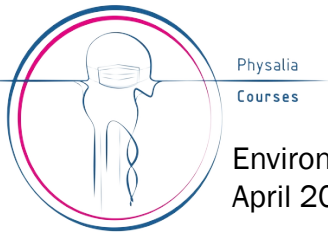- genome_report.txt ✅
- Tromsø.txt ❌
- Tromso.txt ✅

Some additional notes

Running commands
- Space after each "word" in the command
- Typed in a single line, one at a time
- After each command, hit "Enter" to execute it
- Lines starting with "#" are comments

Directory navigation:
- One dot (.) means "here"
- To go up one folder: ../
- To go up two folders: ../../

More advanced usage

Piping (|)

Stream redirection (>)

- Stdout ("output"): >

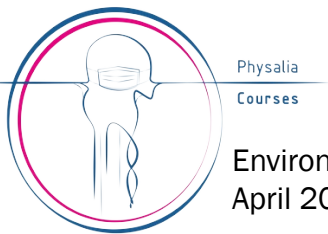- Stderr ("error messages"): 2>

- Stdout + stderr: &>

Some tricks

- Tabulator

- History

- Wildcards

Variable assignment

```
> NUMBER_OF_CPUS=40
> echo $NUMBER_OF_CPUS
40
```

For/while loops

```
for FILE in file01 file02 file03; do
    mv $FILE.txt $FILE.old.txt
done

while read FILE; do
    rm $FILE.txt
done < files_to_remove.txt
```

Physalia
Courses

How to learn UNIX?

By using it!

- Trial and error
- Don't copy and paste it, type it

Ask the internet

- http://stackoverflow.com/
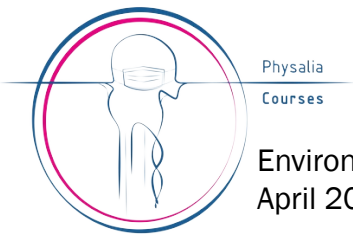- http://stackexchange.com/
- http://askubuntu.com/
- Google!

Cheat sheets

- https://www.guru99.com/linux-commands-cheat-sheet.html

Manual ("man") pages

- `man ls`

Online courses/tutorials

- http://codecademy.com

Igor S. Pessi & Antti Karkman, University of Helsinki

Physalia
Courses

Let's practice the command line a bit

https://github.com/karkman/Physalia_EnvMetagenomics_2022

Physalia
Courses