

# Environmental metagenomics

Working with the command line



Physalia  
Courses

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

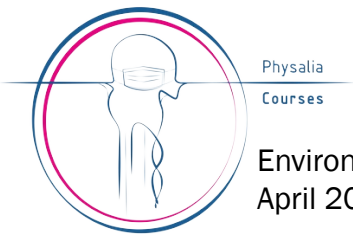
## Key characteristics

**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



Physalia  
Courses

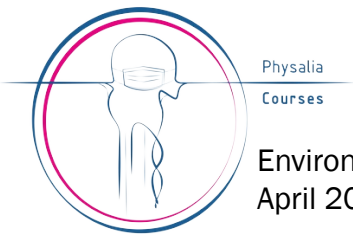
Environmental metagenomics  
April 2022

Igor S. Pessi & Antti Karkman, University of Helsinki

# 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



# 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

```
ipessi — stelmach@taito-login3:~ — ssh stelmach@taito.csc.fi — 90x50
Last login: Wed Feb 27 11:18:56 2019 from dsl-hkibng42-5673dc-24.dhcp.inet.fi
Welcome
CSC - Tieteen tietotekniikan keskus - IT Center for Science
HP Cluster Platform SL230s Gen8 / XL230a Gen9 TAITO
Contact
  Servicedesk : 09-457 2821, servicedesk@csc.fi   Switchboard : 09-457 2801
User Guide
  https://research.csc.fi/taito-user-guide
Software
Available modules can be listed with command: module avail | module spider
Partitions
parallel : 16-672 cores / 5mins/3days      def/max runtime
serial   : 1-24 cores / 5mins/3days        def/max runtime
longrun  : 1-24 cores / 5mins/14days       def/max runtime
test     : 1-48 cores / 5mins/30mins       def/max runtime
hugemem  : 1-40 cores / 5mins/7days        def/max runtime
News
2014-04-25: Quota cutter for batch jobs is now in use on Taito. It is not possible to run batch jobs if quota is exceeded. One can apply for more resources in the usual way, see:
https://www.csc.fi/laskentaresurssien-haku (suomeksi)
https://www.csc.fi/applying-for-computing-resources (in English)
2014-10-02: If you wish to change the project you're using for billing, please use "newgrp groupname". You can find more information at:
http://tinyurl.com/kozfa6t
2014-11-27: For jobs requiring more than 16 GB memory per core, please use the 'hugemem' queue consisting of six 1.5 TB memory nodes with 32/40 cores each.
2017-01-13: Monthly $WRKDIR clean-up will start on Feb. 1, 2017.
Data storage info: http://bit.ly/2jDk7SP and http://bit.ly/2jIjydh
2017-01-19: Remember that memory is also billed. That is, the basis of billing is either the number of cores or number (rounded up to the nearest integer) of 4 GB blocks of memory, depending which is larger.
For example,
- an 8 core job with 16 GB of memory will be billed with 2 x 8 = 16 bu/h
- an 8 core job with 64 GB of memory will be billed with 2 x 16 = 32 bu/h
CRITICAL: You belong to the project 10004688, which has used all its CPU time quota. Project 10004688 is not allowed to submit any new jobs.
Please apply for more CPU time, the instructions are at: https://research.csc.fi/applying-for-computing-resources
stelmach@taito-login3 [ ~ ] >
```

# Some basic Unix commands

**pwd:** print working directory (“where am I?”)

**ls:** list (“show folder contents”)

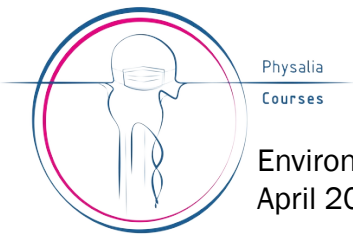
**mkdir:** make directory (a.k.a. folder)

**cd:** change directory (“go to folder”)

**cp:** copy

**mv:** move

**rm:** remove



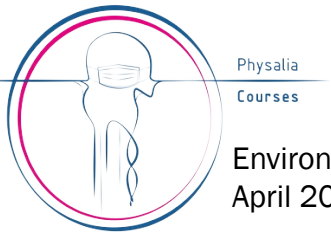
# Some additional notes

## 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 ✅



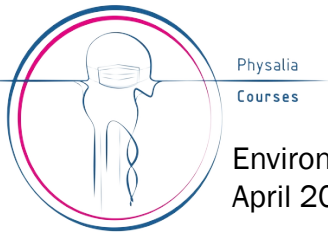
# Some additional notes

## 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

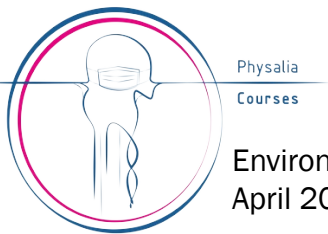
- Tabulator
- History
- Wildcards

## 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
```





# 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

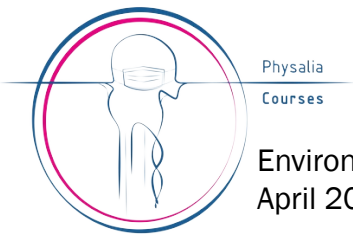
- <https://www.guru99.com/linux-commands-cheat-sheet.html>

## Manual (“man”) pages

- `man ls`

## Online courses/tutorials

- <http://codecademy.com>



Physalia  
Courses

Environmental metagenomics  
April 2022

Igor S. Pessi & Antti Karkman, University of Helsinki

# Let's practice the command line a bit

[https://github.com/karkman/Physalia\\_EnvMetagenomics\\_2022](https://github.com/karkman/Physalia_EnvMetagenomics_2022)



Physalia  
Courses

Environmental metagenomics  
April 2022

Igor S. Pessi & Antti Karkman, University of Helsinki