

# Algorithms for Bioinformatics

Exercises

2019-2020

# Contact for any question

Toma Tebaldi

t.tebaldi@unitn.it

*You can also use the course forum on doodle*

<https://didatticaonline.unitn.it/dol/mod/forum/view.php?id=403040>

# Calendar

Lesson 1 – April 20

Lesson 2 – April 22

Lesson 3 – April 27

**Lesson 4 – April 29**

Lesson 5 – May 4

**Lesson 6 – May 6**

Lesson 7 – May 11

**Lesson 8 – May 13**

Grey: asynchronous lesson  
(material available on Moodle)

**Red: online synchronous  
lesson (Zoom)**

# Topics

- Introduction to Bash
- Introduction to Python
- Implementation of the Smith and Waterman algorithm
- Exercises for retrieving biological sequences
- Exercises with BLAST
- Exercises with Multiple sequence alignments
- Exercises with phylogenetic analysis

# Exam and evaluation

For the exam, each candidate should present a personal implementation of the Smith and Waterman algorithm in Python.

During the exam, the candidate will be asked to modify the program, by adding functionalities or the way results are displayed.

The original and the modified version of the program will be used for the evaluation of the laboratory part of the course

# Lesson 1

20 April 2020

## Bash

the Bourne-Again SHell

A small guide to the Bash terminal

# Bash

- Command-line interpreter
- Allows to perform actions by inserting text commands
  - List directories and files
  - Move, rename, delete, edit and view files
  - Execute any type of software (even the ones with a graphical interface)
  - View information on the status of the system
  - Run a set of commands
  - Programming

*Please try to use the commands presented in the following slides if you are not confident with them*

```
tomateba@DESKTOP-MEC5C18: ~/input
To run a command as administrator (user "root"), use "sudo <command>".
See "man sudo_root" for details.

tomateba@DESKTOP-MEC5C18:~$ cd input
tomateba@DESKTOP-MEC5C18:~/input$ ls -l
total 0
-rwxrwxrwx 1 tomateba tomateba 88 Mar 26 2019 0.tsv
-rwxrwxrwx 1 tomateba tomateba 102 Mar 26 2019 1.tsv
-rwxrwxrwx 1 tomateba tomateba 98 Mar 26 2019 2.tsv
-rwxrwxrwx 1 tomateba tomateba 112 Mar 26 2019 3.tsv
-rwxrwxrwx 1 tomateba tomateba 162 Mar 26 2019 4.tsv
-rwxrwxrwx 1 tomateba tomateba 136 Mar 26 2019 5.tsv
-rwxrwxrwx 1 tomateba tomateba 104 Mar 26 2019 6.tsv
-rwxrwxrwx 1 tomateba tomateba 138 Mar 26 2019 7.tsv
-rwxrwxrwx 1 tomateba tomateba 144 Mar 26 2019 8.tsv
-rwxrwxrwx 1 tomateba tomateba 92 Mar 26 2019 9.tsv
tomateba@DESKTOP-MEC5C18:~/input$
```

# Print & List

- Print something to the shell

```
echo "Hello World"
```

- Print current time and date

```
echo $(date)
```

- Write some text to a file

```
echo "Ode to the west wind" > shelley.txt
```

- List current directory: *list storage* (-l for long format)

```
ls
```

- List with a long format only files in the current directory that ends with ".txt"

```
ls -lh *.txt
```



# Create directories and files

- Create a new directory/folder

`mkdir poetry`

- Create an empty file

`touch empty.txt`

- Print the current working directory (full path)

`pwd`

# Find and identify files

- Find a file by name  
locate empty
- Recognize the type of data contained in a file  
file empty.txt

# Copy & Moving

- Copying a file or a directory (-r)

```
cp <source> <destination>
```

- Copy a file to a directory

```
cp shelley.txt ./poetry/shelley.txt
```

- Moving a file

```
mv <source> <destination>
```

- Rename a file (when source = destination)

```
mv shelley.txt percy.txt
```

- Move and rename a file to a different folder

```
mv percy.txt ./poetry/bysshe.txt
```

# Change path & remove files

- Go to your home directory  
`cd`
- Go up (or back) to the parent folder  
`cd ..`
- Go to the “home” folder present in “/”  
`cd /home/`
- Remove a file  
`rm <filename>`
- Remove a directory  
`rm -r <directory>`

# Viewing the content of files

- Visualize (without modifying) the content of a file (q for quitting)  
`less <filename>`
- Print the first 10 (default) lines of a file on the console (-n to control lines)  
`head <filename>`
- Print the last 10 (default) lines of a file on the console  
`tail <filename>`
- Print all the content of a file on the console  
`cat <filename>`

# Concatenating commands

- Execute a series of commands for which the output of the previous command becomes the input of the following command

`cmd1 | cmd2 | ... | cmdN`

- In the bash, the character “|” is used to chain commands
- The output of the command on the left becomes the input of the command on the right of the “|”
- Examples
  - `man less | head`
  - `cat shelly.txt | tail | head -n 3`

# Redirection

- The output of a command may be redirected using a special notation, >, interpreted by the shell

- Write on a file the result of the `ls -l` command

```
ls -l > list_files.txt  
less list_files.txt
```

- To append the output to an existing file, the >> notation can be used

```
ls >> list_files.txt  
less list_files.txt
```

# Text manipulation

- Search for occurrences of <what> in <where> (lines as output)  
`grep <what> <where>`
- Add context (lines) after (–A), before (–B), or both (–C)  
`grep –A3 <what> <where>`  
`grep –B3 <what> <where>`  
`grep –C3 <what> <where>`
- Case-sensitive or not?  
`grep –i <what> <where>`
- Invert the match (i.e., everything that does not match <what>)  
`grep –v <what> <where>`



# Text manipulation

- Count lines that match

```
grep -c <what> <where>
```

- Match whole word

```
grep -w <what> <where>
```

- Display the line number of the matches

```
grep -n <what> <where>
```

# What do to if I need help with a command?

- Find what a command does

`whatis <command>`  
`whatis ls`

- Use the man!

`man <command>`  
`man ls`

- `man` provides you the manual for that command that explains what it does, what are the parameters, and other useful stuff
- Be aware that, in general, the `man` on a Linux system is simply displaying a text file using the `less` command. So, to exit just press `q`!

# Bash exercise

1. Lets go to your home folder
2. Create a folder named "tmp"
3. Create a file that contains the manual of the ls command
4. Put this file into the "tmp" folder
5. Count how many occurrences of "ls" are present (case sensitive, whole word)
6. Print the first 7 lines that matched (case sensitive, not whole word)
7. Save the output to a file named "10.txt"
8. Print the first 3 lines of the last 5 lines that matched
9. Concatenate this output to the file "10.txt"
10. Now remove the "tmp" folder