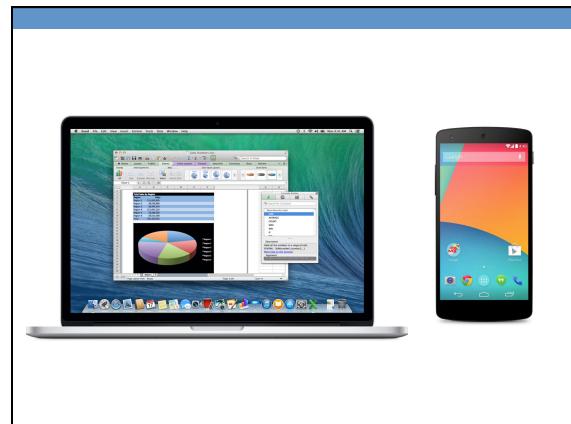


UNIX INTRODUCTION

Janos Binder - janos.binder@embl.de



Biology in 2014

- Omics Era
- Several Go of data / experiment
- Personalized Medicine

Why using Unix in Biomedical Research?

Free Fast
Ethics Open

Easy & documented!

What do you do with Unix in Biology?

- File manipulation
 - Sequences
 - Interaction networks
 - Research articles
 - Spectrum
 - ...
- Analysis
 - Calling programs (R, Perl, Python, Java, etc...)
 - Generating knowledge from raw information
 - Statistics
 - Filtering
 - ...

OSX terminal

Name	Date Modified	Size	Type
Applications	14 Apr 2014 03:53	24.4 KB	Application
Dock	23 Feb 2013 03:38	12.4 KB	Application
All My Files	10 Jan 2014 08:11	50.3 MB	Application
Finder	13 Mar 2014 03:53	12.4 KB	Application
Utilities	29 Apr 2014 13:06	11.4 MB	Application
Security Monitor	20 Aug 2013 09:49	10.6 MB	Application
Adobe AIR Application Installer	10 Apr 2010 05:36	896 KB	Application
Adobe Flash Player	03 Dec 2013 09:49	90.9 KB	Application
Adobe Flash Player Manager	29 Apr 2014 13:06	580 KB	Application
Adobe Flash Player Update	29 Apr 2014 13:06	580 KB	Application
Adobe Utilities - CSS	10 Jan 2014 17:57	10.3 KB	Application
Apache Utilities	21 May 2013 22:07	10.2 MB	Application
Apple Software Update	20 Aug 2013 07:21	10.9 MB	Application
Austin M2D Serial	23 Aug 2013 07:12	10.2 MB	Application
Barcode Reader	29 Apr 2014 13:06	32.4 KB	Application
Boot Camp Assistant	28 Feb 2014 02:26	5.2 MB	Application
CloudKit	23 Aug 2013 07:21	15.2 MB	Application
Console	9 Mar 2013 03:49	10.2 MB	Application
Disk Utility	29 Apr 2014 13:06	22 MB	Application
Disk Speed Test	29 Apr 2014 13:06	2.7 MB	Application
GIMP	7 Mar 2013 13:31	35.6 MB	Application
Karabiner Access	20 Sep 2013 02:20	12.6 MB	Application
Migrate Assistant	29 Aug 2013 07:05	3.7 MB	Application
System Information	29 Feb 2013 07:08	8.5 MB	Application
Terminal	29 Feb 2013 07:08	27.5 MB	Application
VoiceOver Utility	28 Feb 2014 03:16	10.9 KB	Application
N Squeezed	19 Oct 2013 03:53	7.8 MB	Application
Others	22 Oct 2008 08:16	82.6 KB	Application
VM	29 Sep 2013 03:46	105.8 MB	Application
Honey Fenn	29 Oct 2013 03:07	925.1 KB	Application

```
james@volcano:~$ login as: James  
James@volcano's password:  
Linux volcano 2.6.22-14-server #1 SMP Sun Oct 14 23:34:23 GMT 2007 1686  
  
The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*copyright.  
  
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.  
  
james@volcano:~-> [REDACTED]
```

What do you do with Unix in Biology?

- File manipulation
 - Sequences
 - Interaction networks
 - Research articles
 - Spectrum
 - ...
- Analysis
 - Calling programs (R, Perl, Python, Java, etc...)
 - Generating knowledge from raw information
 - Statistics
 - Filtering
 - ...

Part 1

Part 2

What do you do with Unix in Biology?

- File manipulation
 - Protein Sequences
 - Name of the file: proteins.txt

The Shell – What the (s)hell is this?

- *Piece of software that provides an interface for users of an operating system*



A screenshot of a terminal window titled "johnec246:~". The window shows the command "python test.py" being run, followed by its output: "foobar" repeated eight times, and then the command "stty sane" being run. The terminal window has a dark background and light-colored text.

```
johnec246:~$ python test.py
foobar
foobar
foobar
foobar
foobar
foobar
foobar
foobar
stty sane
johnec246:~$
```

The Shell – Tips

- Previous command:

- Auto-Complete:

- Emergency Stop:


- Going forward/back:



STEP 0

Getting help

info [command]

```
Terminal - info - 80x24
File: info.info, Node: Top, Next: Getting Started, Up: (dir)
Info: An Introduction
.....
The GNU Project distributes most of its on-line manuals in the "Info
format", which you read using an "Info reader". You are probably using
an Info reader to read this now.

There are two primary Info readers: 'info', a stand-alone program
designed just for reading Info files (note that is info:(info-stnd)top), and
the 'info' package in GIMP Emacs mode, a screen editor. At
present, only the Emacs reader supports using a mouse.

If you are new to the Info reader and want to learn how to use it,
type the command 'h' now. It brings you to a programmed instruction
sequence.

To read about advanced Info commands, type 'n' twice. This brings
you to 'Advanced Info Commands', skipping over the 'Getting Started'
chapter.

-----Info: (info.info)Top 46 lines --Top-----
Welcome to Info version 5.2. Type h for help, m for menu item.
```

The info command provides help

man [command]

```
Terminal - less - 80x24
man(1)                                     Terminal - less - 80x24
man(1)

NAME
    man - format and display the on-line manual pages

SYNOPSIS
    man [-odffHkKtw] [-path] [-m system] [-p string] [-c config_file]
    [-M catalog] [-P pager] [-B browser] [-H helpviewer] [-S section_list]
    [-s section_name ...]

DESCRIPTION
    man formats and displays the on-line manual pages. If you specify a section,
    man only looks in that section of the manual. man is normally the name of the manual page, which is typically the name of a command,
    function, or file. However, if man contains a slash (/), then man
    interprets it as a file specification, so that you can do man ./foo.5
    or even man /cd/for/bar-3.gp.

    See below for a description of where man looks for the manual page
    files.

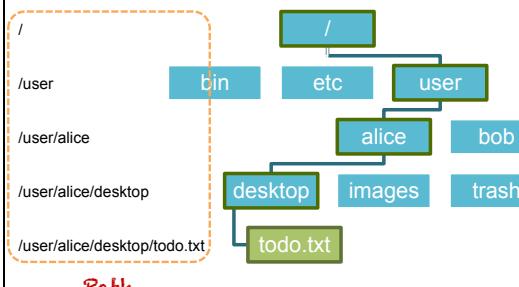
OPTIONS
    -c config_file
    :|||
```

The man command is a more traditional way to get help

STEP 1

Browsing around

Everything's a file



ls [directory] (list)

```
Terminal - bash - 130x30
Last login: Mon May 11 16:09:08 on ttys001
...
Desktop/Transcribers_Documents
Desktop
Documents
Downloads
Music
Pictures
Public
Videos
...

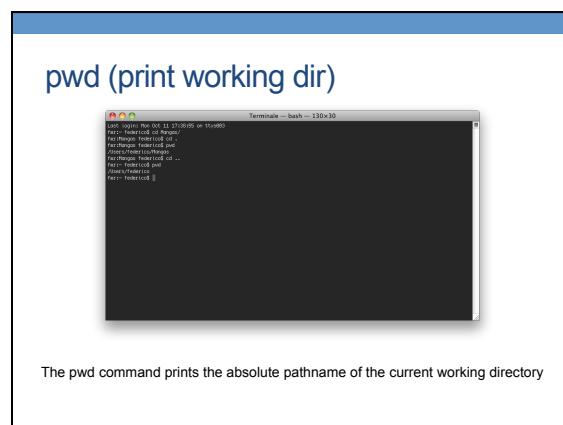
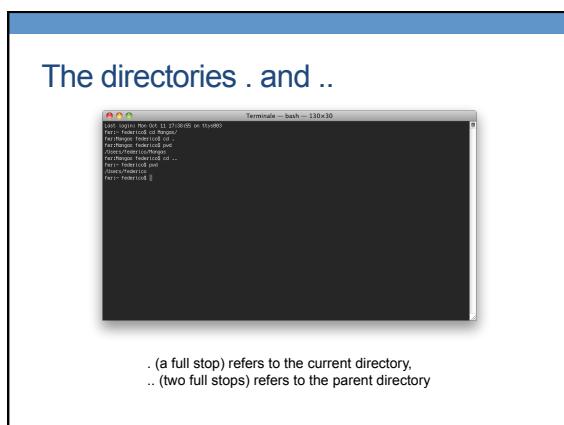
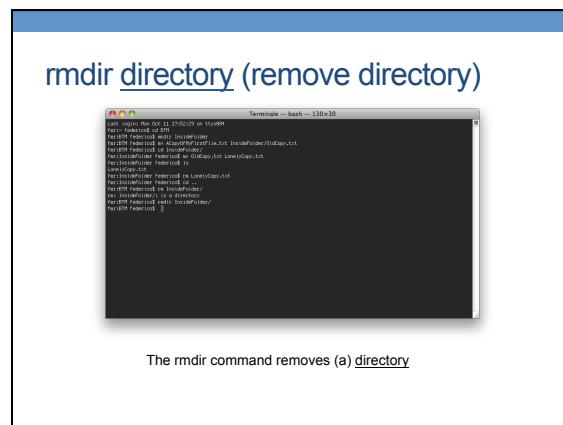
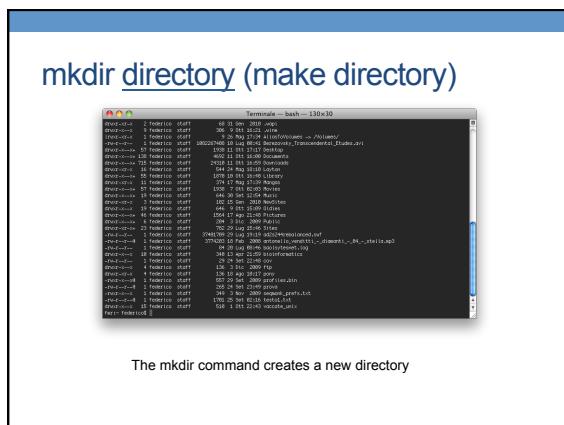
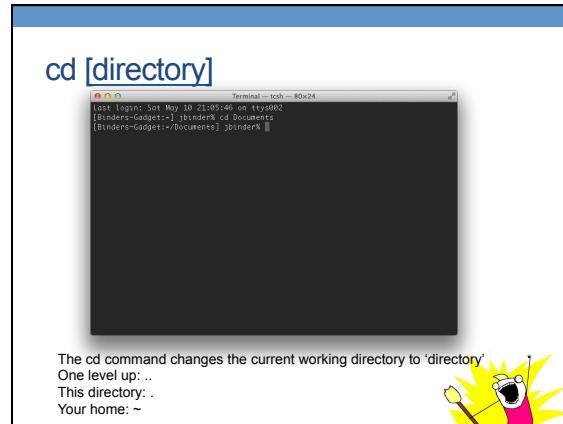
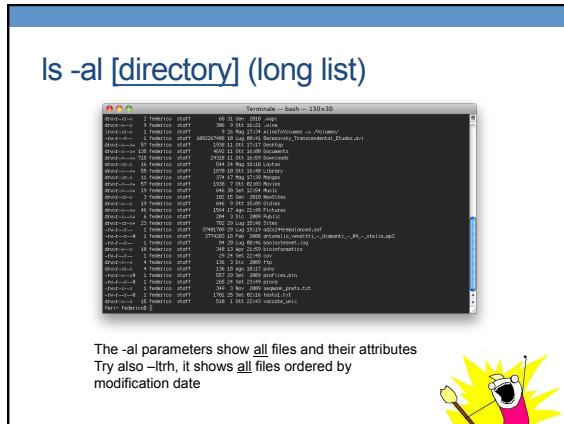
```

The ls command lists the contents of your current working directory

clear (clear screen)

```
Terminal - bash - 130x30
[clear]
```

The clear command clears the screen



STEP 2

Handling files

cp file1 file2 (copy)

```
laptop:~ jblinder$ cp file1 file2
laptop:~ jblinder$ ls
file1  file2  file3  file4  file5
```

The cp command makes a copy of file1, naming it file2

mv file1 file2 (move)

```
laptop:~ jblinder$ mv file1 file2
laptop:~ jblinder$ ls
file2  file3  file4  file5
```

The mv command moves file1 into file2
(if file2 is in the same directory, file1 is renamed)

rm file (remove)

```
laptop:~ jblinder$ rm file
laptop:~ jblinder$ ls
file1  file2  file3  file4  file5
```

The rm command removes (a) file

scp server:file1 file2 (secure copy)

```
[dhcp-1blinder-13132:~] jblinder% scp orginise.cbs.dtu.dk:/~Pf1_analysis.tar.gz .
[dhcp-1blinder-13132:~] jblinder%
```

scp copies a file from a server to a local place
scp file1 server:file2 also works

wget url

```
[Binders-Gadgets:~/Downloads] jblinder% wget ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/completedb/uniprot_sprot.dat.gz
--2014-05-31 10:18:14--  ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/completedb/uniprot_sprot.dat.gz
                   => 'uniprot_sprot.dat.gz'
Resolving ftp.uniprot.org (ftp.uniprot.org)... 141.161.100.107
Connecting to ftp.uniprot.org (ftp.uniprot.org)|141.161.100.107|:21...
Logging in as anonymous ... Logged in.
=> PORT command successful.
=> TYPE I ...
done. => CMD (1) /pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.dat.gz ...
done.
=> PASV command successful.
=> RETR uniprot_sprot.dat.gz ...
Length: 491964874 (469M) (unauthoritative)

  1% [          ] 6,748,344  849KB/s  eta 9m 32s
```

wget can download files

curl url > filename

```
[Binders-Godot:-/Downloads] jbinder% curl ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot_dot.gz > uniprot_dot.gz
  % Total    % Received % Xferd  Average Speed   Time   Time  Current
  0  469M     0  4413k   0      539k  0  24:50  0:00:08  0:14:42  813k
```

curl retrieves data from various internet services

STEP 3

Files content

less file1

```
Terminal — less — 130x30
[Binders-Godot:-/Downloads] jbinder% less file1
L.  1 Nel mezzo del cammin di nostra vita
L.  2 mi ritrovai per una selva oscura
L.  3 ché la diritta via era smarrita
L.  4 Ahor quanto è dir dico! era e cosa dura
L.  5 nello smarrirsi e vagare e farla
L.  6 cercare.
L.  7 Tant'è vero che poco a più volte;
L.  8 mi per la selva del son che mi costava
L.  9 mi per la selva del son che mi costava.
L. 10
L. 11 Io non so che cosa l'è 'ntreto,
L. 12 se la verità sia abbastanza.
L. 13 Nel mezzo del cammin di nostra vita,
L. 14 mi ritrovai per una selva oscura
L. 15 ché la diritta via era smarrita
L. 16 Ahor quanto è dir dico! era e cosa dura
L. 17 nello smarrirsi e vagare e farla
L. 18 cercare.
L. 19 Fu in pechi un poco fata
L. 20 che nel son che mi costava
L. 21 mi per la selva del son che mi costava.
Internetto.txt
```

The less command shows the contents of a file on the screen, one page at a time

head [-n] file

```
Terminal — bash — 130x30
[Binders-Godot:-/Downloads] jbinder% head Internetto.txt
L.  1 Nel mezzo del cammin di nostra vita
L.  2 mi ritrovai per una selva oscura
L.  3 ché la diritta via era smarrita
L.  4 Ahor quanto è dir dico! era e cosa dura
L.  5 nello smarrirsi e vagare e farla
L.  6 cercare.
Internetto.txt
```

head [-n] file

The head command shows the first 10 (or n) lines of a file

tail [-n] file

```
Terminal — bash — 130x30
[Binders-Godot:-/Downloads] jbinder% tail Internetto.txt
L. 129 Nel mezzo del cammin di nostra vita
L. 130 mi ritrovai per una selva oscura
L. 131 ché la diritta via era smarrita
L. 132 Ahor quanto è dir dico! era e cosa dura
L. 133 nello smarrirsi e vagare e farla
L. 134 cercare.
L. 135 come ho ai pari di non far mestier.
L. 136 al oh! io voglio la porta di non festier
L. 137 al oh! io voglio la porta di non festier
L. 138 Ahor quanto è dir dico! era e cosa dura
Internetto.txt
```

The tail command shows the last 10 (or n) lines of a file

wc [-lwm] file

```
Terminal — bash — 130x30
[Binders-Godot:-/Downloads] jbinder% wc Internetto.txt
129 130 130 Internetto.txt
Internetto.txt w.l Internetto.txt
Internetto.txt w Internetto.txt
```

The wc command returns lines/words/characters number of a file

Take home message

- Unix/Linux is the most widely used operating system in bioinformatics
- Unix/Linux is used to handle files and to run analysis on very large experiments, by calling external programs
 - You can manipulate your files through command lines
 - You can launch programs by the command line (next talk)
- Google is the best way to find information about particular command lines

STEP 4

Setting up programs

A screenshot of a Mac OS X terminal window. The title bar says "Terminal - ssh - 107x43". The window shows a command-line interface with the following text:

```
Last login: Wed May 20 13:07:22 on ttys0001  
Last login: Wed May 20 13:07:22 on ttys0001  
Last login: Wed May 20 13:07:22 2014 from 140.254.88.96  
root@elb-140-254-88-96: ~ % cat /etc/issue  
Ubuntu/12.04 LTS \n \l  
root@elb-140-254-88-96: ~ %
```

The user has just typed "ssh" at the prompt.

chmod a+x file

chmod changes permission of a file. Here it becomes executable

configure

```
Terminal — clang — 107x43
$ ./configure
...
configure adopt a program package for the local environment
```

make

make compiles and set ups a program package

STEP 4

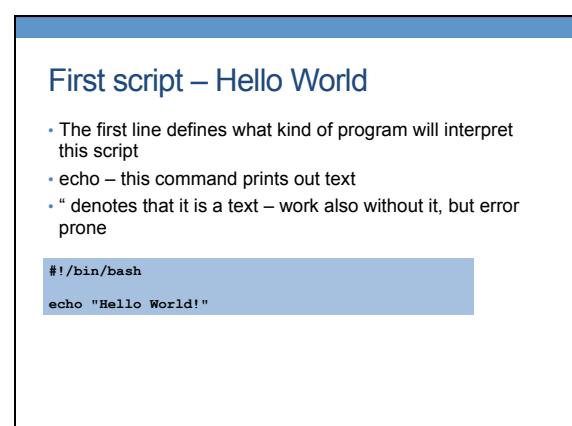
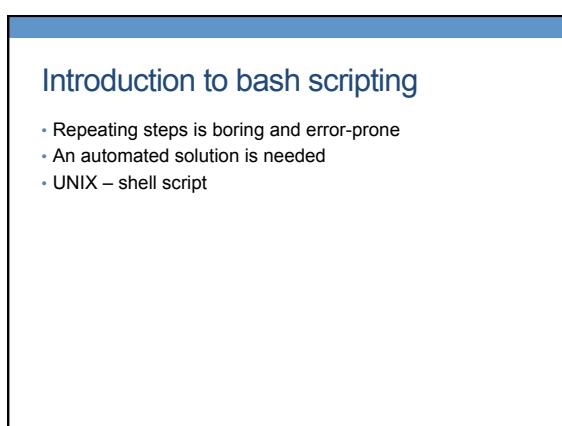
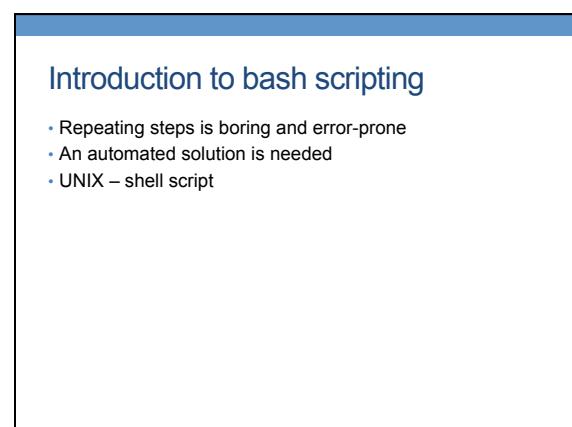
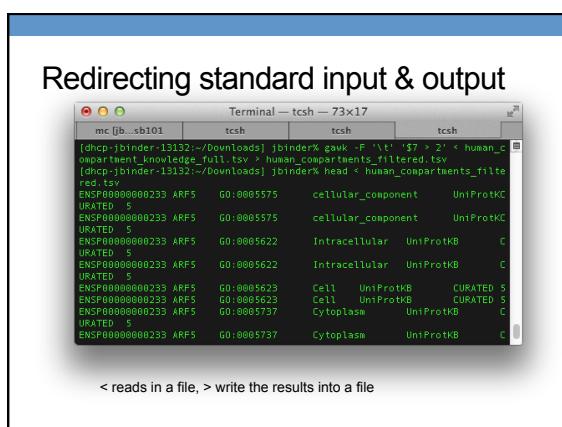
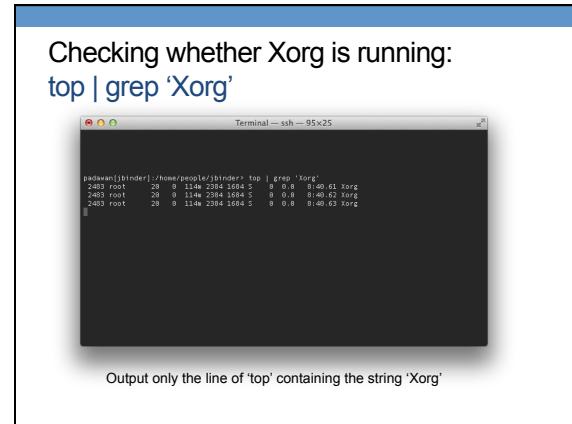
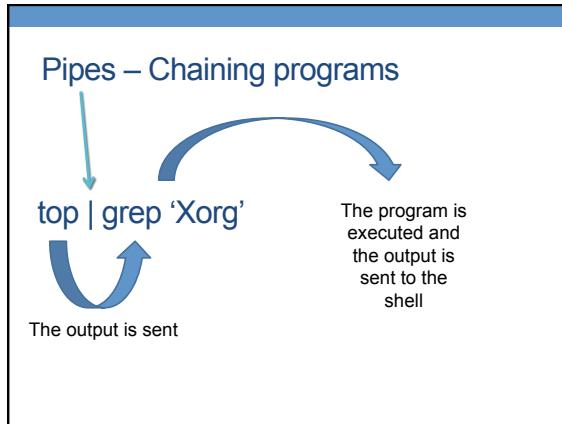
Pipelines and streams

CHAINING PROGRAMS

What's going on: top

```
Terminal — top — 80x24
Processes: 237 total, 2 running, 5 stuck, 238 sleeping, 1487 threads
Load average: 0.00 0.00 0.00  CPU usage: 0.0% user, 0.0% sys, 69.6% idle
SharedLibs: 1564K resident, 98 data, 98 linked.
MemRegions: 63639K total, 2430K resident, 979 private, 1014M shared.
PhysMem: 6142M used (1304M wired), 6739M free.
VM: 13825580K total, 8133M used, 1309220K free, 0(0) swapouts.
Networks: packets: 13825580/8133M in, 2779720/346K out.
Disks: 2204757/534G read 129139/376 written.
```

Display the list of running processes
To kill a program type 'k' on the keyboard followed by the PID



Variables

- Storing often used texts, numbers
- Storing results of a calculation
- Remember to the \$

```
#!/bin/bash
HW="Hello World!"
echo $HW
echo "It is boring to print" $HW
```

Conditionals

- Let you decide whether to perform an action or not, which based on the evaluation of the expression

```
#!/bin/bash
FILE="3_basic_conditionals.sh"
if [ -f $FILE ]; then
    echo "FILE:" $FILE "exists!"
fi
```

- An if-then-else construct:

```
if [ -f $FILE ]; then
    echo "FILE:" $FILE "exists!"
else
    echo "FILE:" $FILE "does not exists!"
fi
```

Streams and pipes - revisited

- Storing the names of the files in a directory

```
ls -1 > files_in_the_directory.txt
less files_in_the_directory.txt
```

- But it shows also lists files_in_the_directory.txt

- Piping the other way

```
grep -v files_in_the_directory.txt <
files_in_the_directory.txt | less
```

- A more clever solution:

```
ls -1 | grep -v files_in_the_directory.txt >
files_in_the_directory.txt
less files_in_the_directory.txt
```

Loops

- To do stuff the "same" stuff more than once

```
#!/bin/sh
echo "We count to ten!"
for (( i = 1 ; i <= 10 ; i++ )) ; do
    echo $i
done
```

```
#!/bin/sh
(( i = 1 ))
while (( i <= 10 )) ; do
    echo $i
    (( i++ ))
done
```

How do I execute a series of commands?

Bash scripting

- You can create scripts (e.g, "my_series_of_command.sh") executing a series of Unix commands

```
#!/bin/bash
for i in $(seq 1 5)
do
    echo "i is equal to: $i"
    ls -l
done
```

You can execute the scripts via: sh my_series_of_command.sh
or chmod a+x my_series_of_command.sh; ./my_series_of_command.sh

"Hey, Unix looks cool, but I have a Mac!"





Good news:

- The recent Mac OS are Unix based.
- So you were using it without realizing it
- Mac OS has a shell, most of the commands (used in biology) are working directly!




“I’m using Windows, how can I try or install a Unix-like system?”



Try Wubi!

- It will install Ubuntu on your Windows machine
- You install it like a normal program inside Windows
- When you will boot your computer, you will have the choice to run Ubuntu

Take home message

- Command lines are helpful to run programs or handle computers with no visual interface
- You have much more commands in the real life, helping you for a variety of tasks → Google is your friend
- Install Linux on your PC: Ubuntu




Acknowledgements

- Samuel Croset and others at EMBL for the Unix introduction slides.

Important parts

- Understand the file/directory structure of Unix (+ special characters ... ~ /)
- Knowledge of basic commands
- Pipes, redirection (| < >)
- Basic manipulation and filtering of text files
- Writing simple scripts (conditionals, loops)