Jon K. Lærdahl,
Structural Bioinformatics

Basic UNIX

Oslo
universitetssykehus

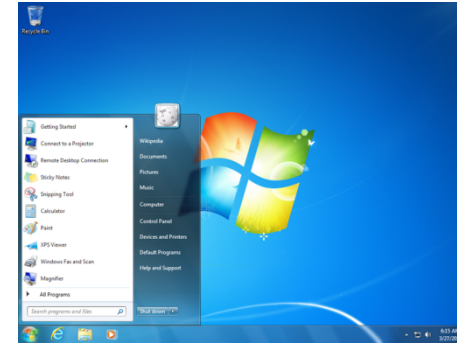UiO : **Department of Informatics**
University of Oslo

Today's Programme

- Biological databases

- **Brief introduction**
  - What is UNIX?
  - Why should you learn UNIX?

- **Setting up your laptops**

- **Very briefly on the Unix shell, file system and some commands**

- **UNIX basics exercise**

- **Tomorrow, continue on databases & working with biological sequences**

What about those of you that know Unix and Python very well?
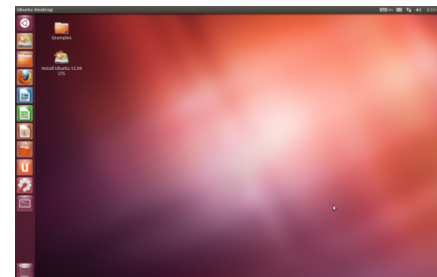
Operating systems

- Software that manages computer hardware
  - Reading input from keyboard and pointing devices
  - Sending output to screen
  - Keeping track of, reading and writing files
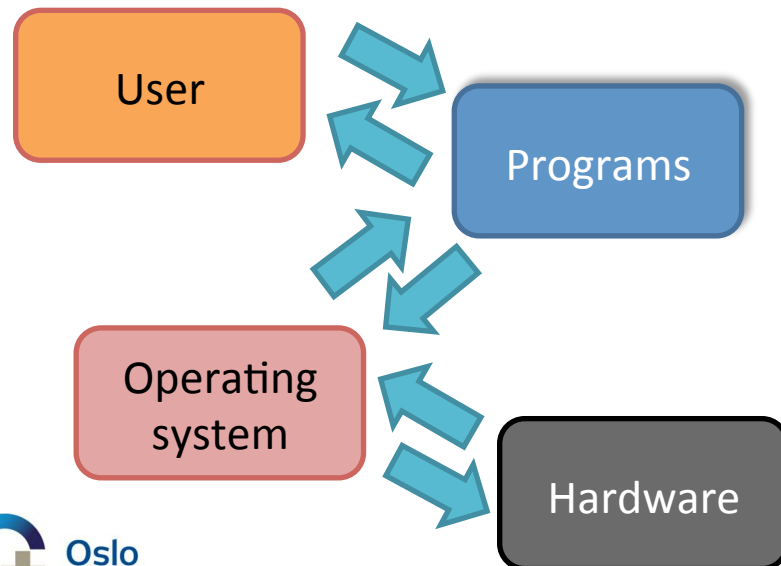- Provides common services for computer programs

Microsoft Windows

(Mac) OS X

UNIX (Linux, Ubuntu)

User

Programs

Operating system

Hardware

Oslo universitetssykehus

UiO : **Department of Informatics**
University of Oslo

UNIX

Jon K. Lærdahl,
Structural Bioinformatics

- UNIX was developed in the early 1970s
- Is extremely well-tested and powerful
  - multiuser
  - multi-tasking
  - available on many architectures
  - extremely rich is commands, possibilities, flexibility
- Extensively used in academic institutions/universities (and industry) for decades, *e.g.* in physics, computational chemistry, and meteorology
- The operating system of the WWW – most web servers run on UNIX machines
- Most software developed at academic institutions for scientific data analysis and simulations

New supercomputer Abel (and old Titan) at UiO

Weather forecast or climate modelling – developed and run under UNIX

Average of all IPCC Models: Temperature Change in 2070
IPCC SRES Scenarios a2 (left) und b2 (right)

-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6   [Deg. C]

Oslo universitetssykehus

UiO : **Department of Informatics**
University of Oslo

Why should (some) life scientists learn UNIX?

Jon K. Lærdahl,
Structural Bioinformatics

- More or less all development of bioinformatics tools and algorithms is done on UNIX computers
  - If you want to run the programs with all options and all flexibilities in the same environment as the program developer, you must use UNIX
- Extremely well tested, flexible and constantly evolving
  - A major fraction of the world's information technology scientists and program developers, including "all" bioinformaticians, are using UNIX in their daily work
- One can easily set up pipelines and automatically run scripts that make it simple to set up work flows and reduce the need for manually manipulating data
- Sharing and limiting access to data is trivial, secure and extremely well tested
  - Data can be protected by using established technology based on encryption and password protection and users can be grouped in user groups that can access each others data while other data is kept private
- *Researchers using bioinformatics tools beyond the most elementary level should definitely invest a week or two in order to learn UNIX*

Oslo
universitetssykehus

UiO **: Department of Informatics**
University of Oslo

More on UNIX

- Security
  - multiple users with multiple environments/data
  - extremely well tested systems for keeping information private
  - (almost) free from viruses and other forms of malware
- University of Oslo
  - user database, systems for e-mail and backup, web servers and much more is running under UNIX
  - everyone with a UiO user account also have a personal UNIX user account (you can log onto a UNIX machine)
- Many variants
  - open source
  - closed source

http://www.levenez.com/unix

Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo
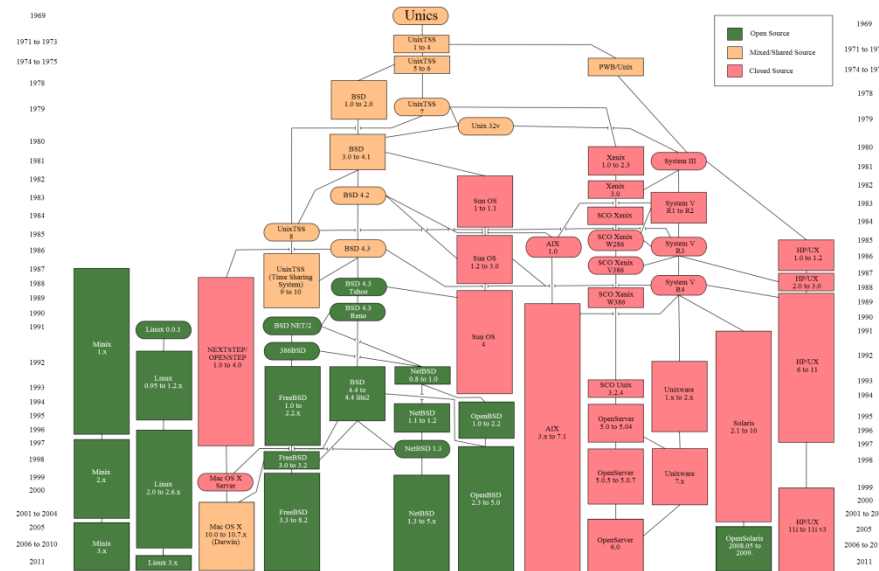
UNIX variants

- Solaris (Sun Microsystems, now Oracle)
- HP-UX (Hewlett-Packard UniX)
- AIX (IBM)
- OS X (Apple Inc., preloaded on all Macs since 2002)
- Linux

http://www.levenez.com/unix

Oslo universitetssykehus

UiO **: Department of Informatics**
University of Oslo
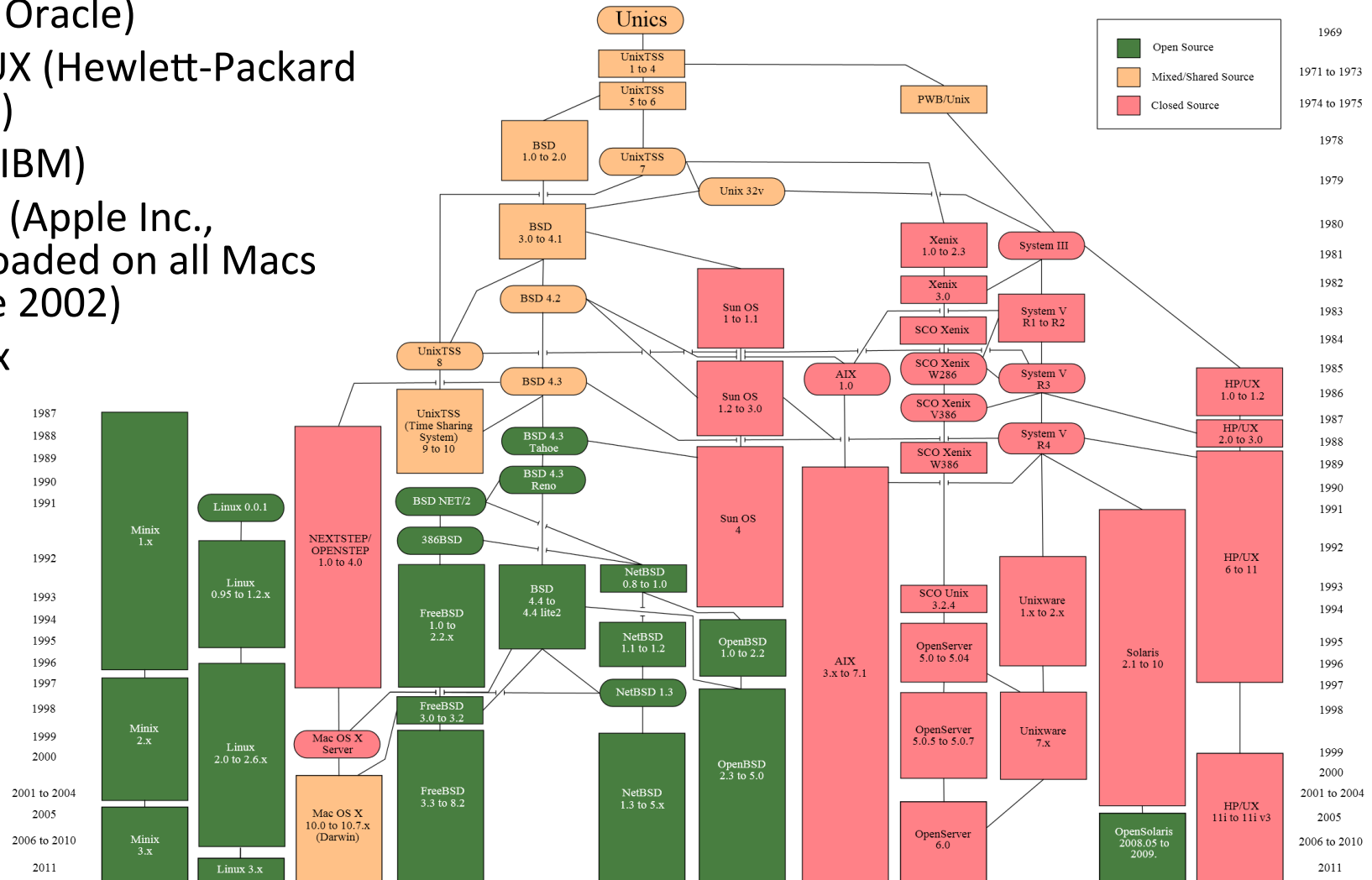
Linux

- UNIX-like OS
  - free and open source software
  - Linux kernel first developed by Linus Torvalds in the early 1990's
- Variants
  - Red Hat Enterprise Linux
  - CentOS
  - Ubuntu
  - Debian
- Can be installed on nearly all standard Windows laptops/ desktops
- Is the OS of >92% of the worlds Top500 supercomputers (including Abel at UiO)

User

Programs

Kernel

Hardware
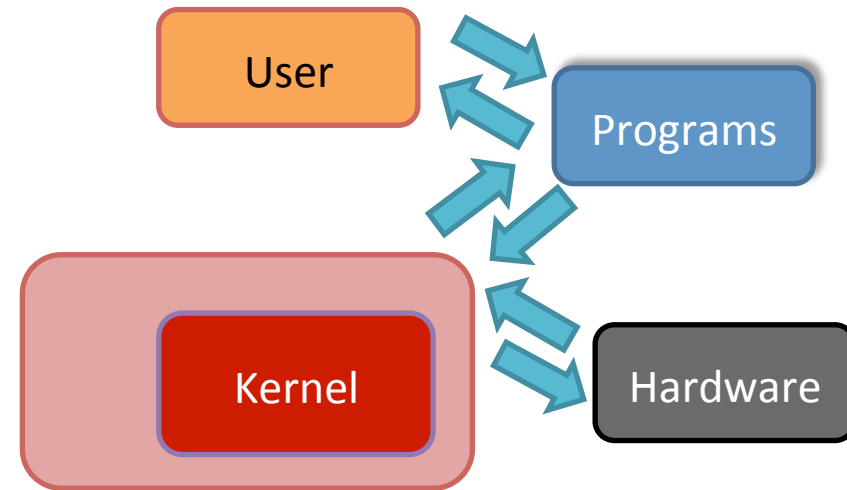
TOP500 Statistics

| TOP500 Release: | June 2012 |
| Category: | Operating system Family |

Submit
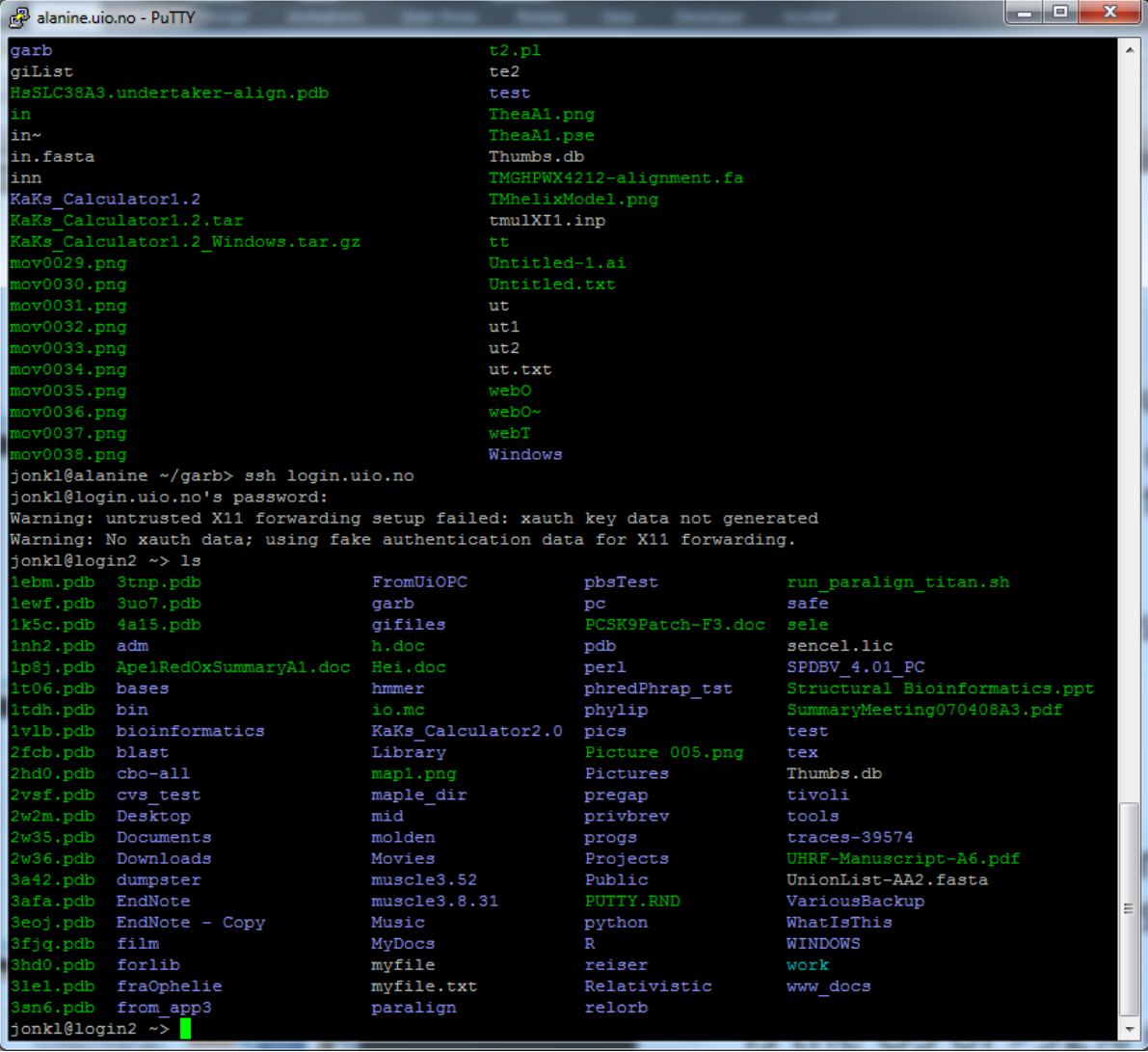
R_{max} and R_{peak} values are in GFlops. For more details about other fields, check the TOP500 description.

Operating system Family System Share

- Linux
- Unix
- Mixed
- Windows
- BSD Based

92.4%

http://www.top500.org

Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo
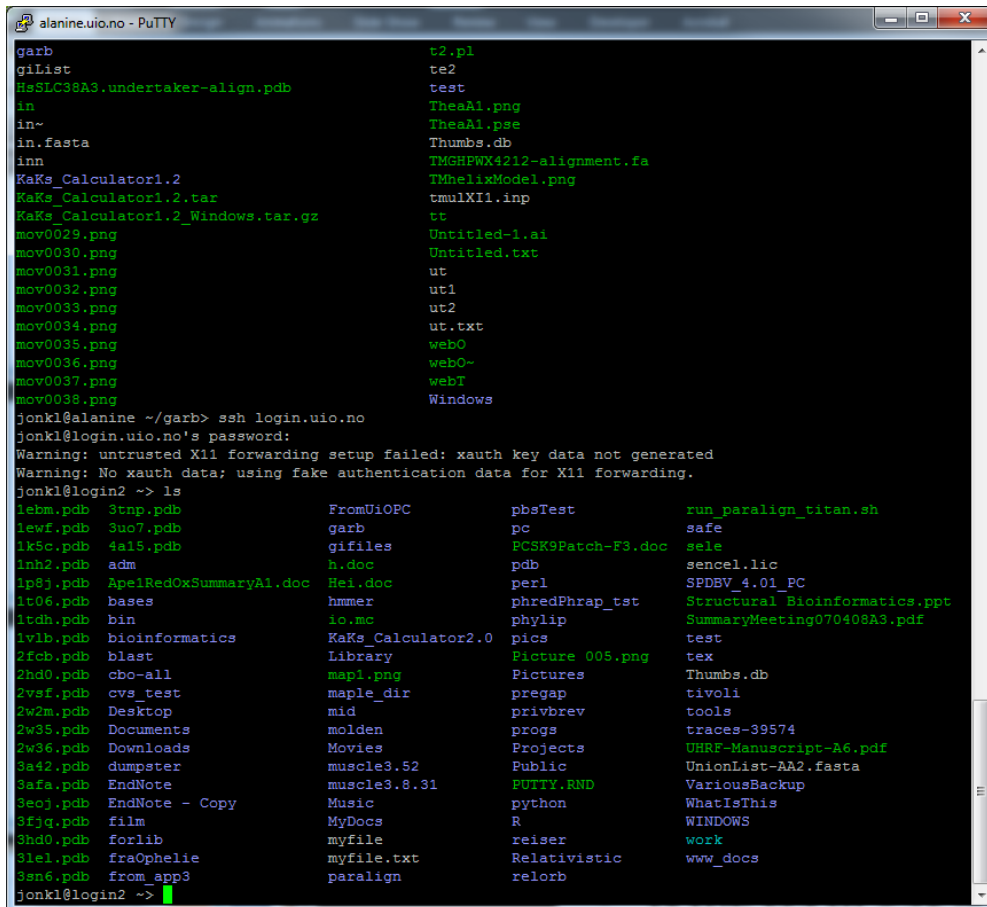
Jon K. Lærdahl,
Structural Bioinformatics

UNIX is text-based OS

The user writes commands on the **command line**, usually in a **terminal window**

Command-line interface - interaction with a computer program where the user gives commands to the program in the form of successive lines of text

As opposed to **graphical user interfaces (GUI)**

Oslo universitetssykehus

UiO **: Department of Informatics**
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

UNIX is text-based OS

Primitive and old-fashioned? *No!*

- Line-by-line
- Step-by-step
- Logical order of things
- Logical workflow
- Same way of thinking as in programming

Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

X Window System (X11) with X display manager – basis for GUI
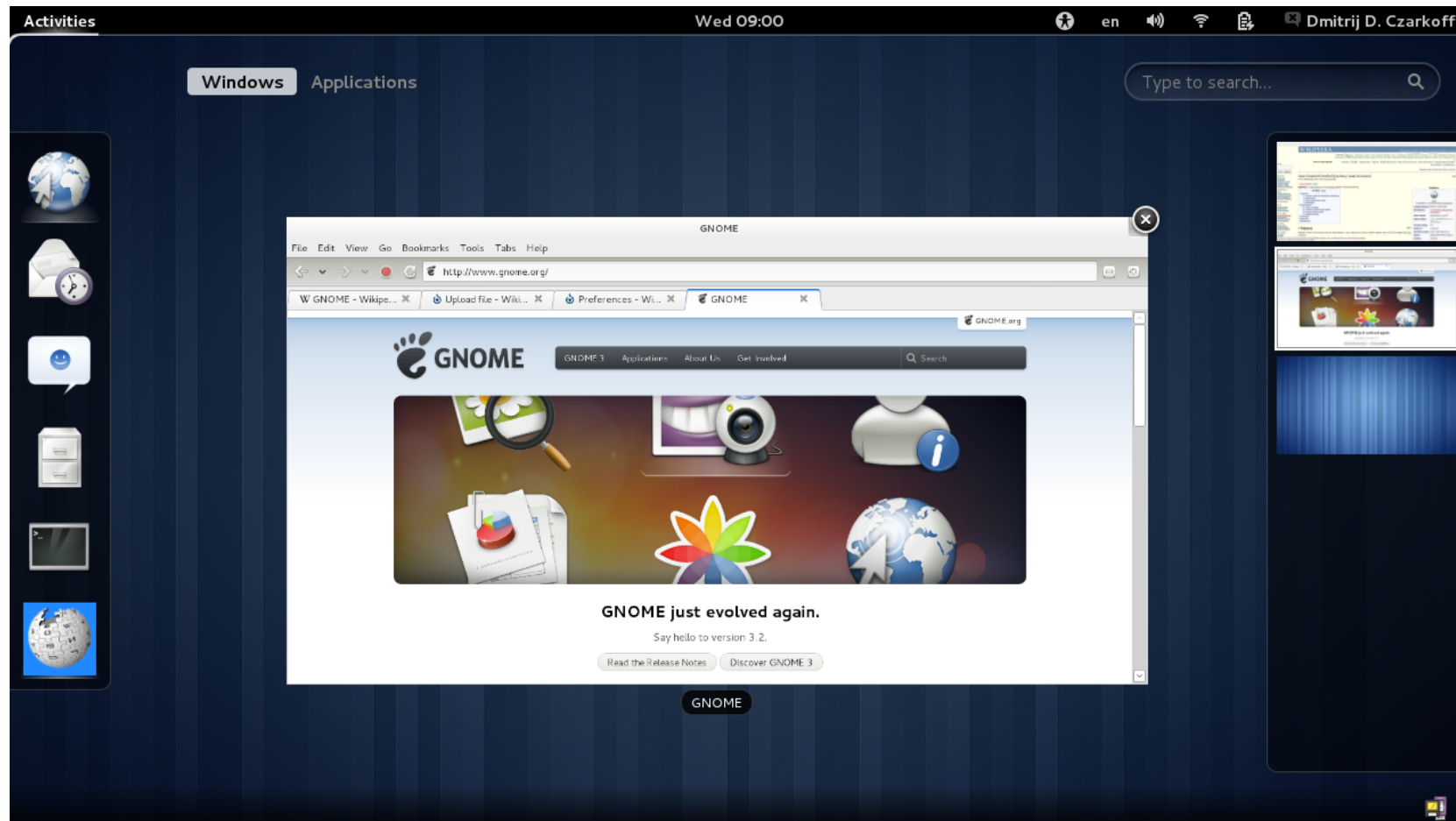
Setup – how to get access to a UNIX computer

Jon K. Lærdahl,
Structural Bioinformatics

- Linux (ok!)
- Mac (ok!)
- Windows
  - Use PuTTY, an open source and free UNIX terminal emulator
  - Installed already on all Windows laptops with a UiO image (UiO setup)
  - Can be downloaded and installed from here:

    http://www.chiark.greenend.org.uk/~sgtatham/putty
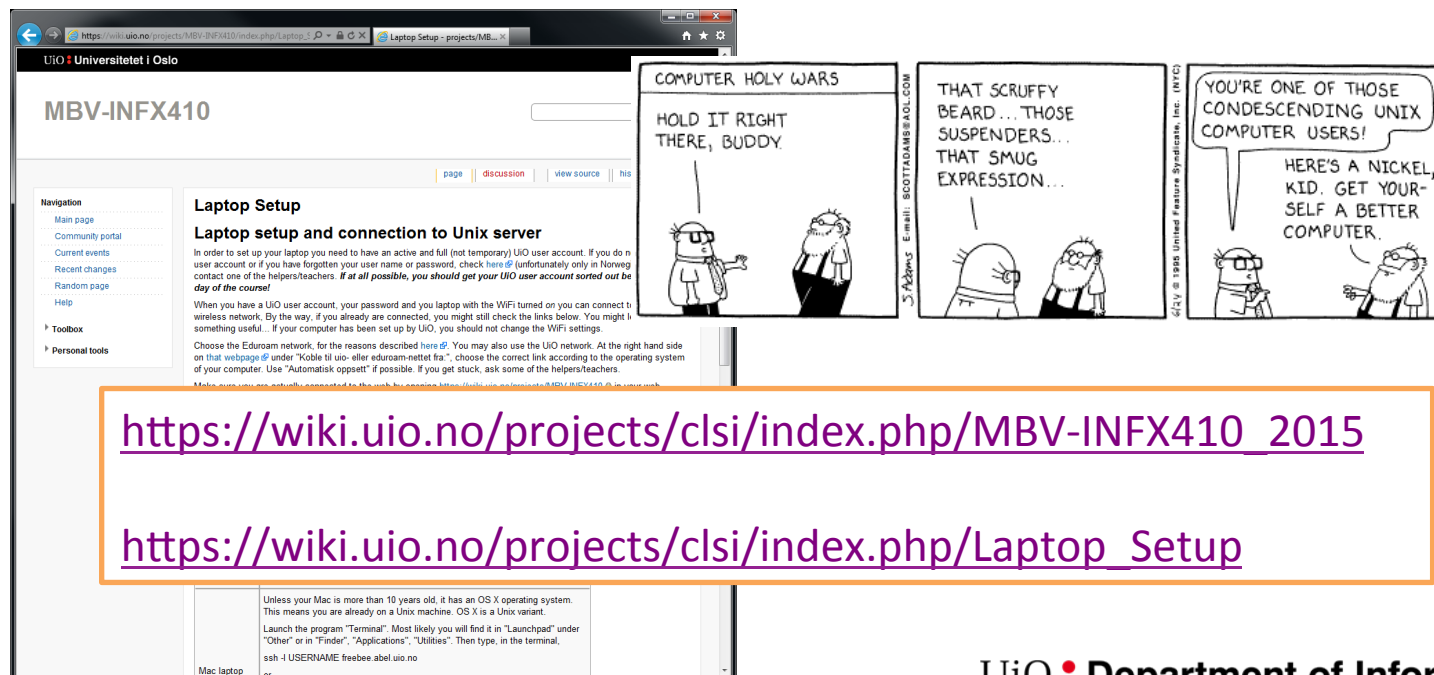
Simon Tatham,
developer of PuTTY

Oslo
universitetssykehus

UiO : **Department of Informatics**
University of Oslo
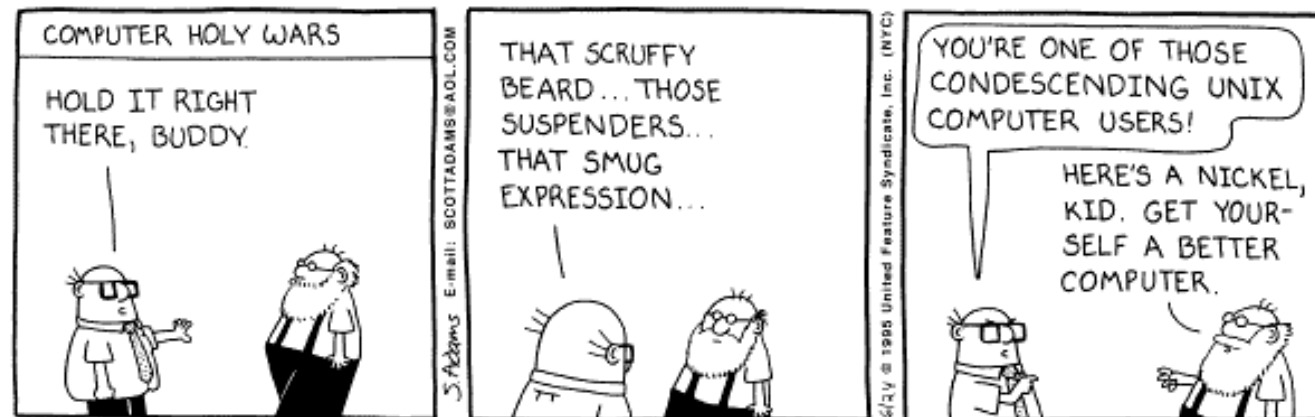
Time to set up the laptops and get connected!

Jon K. Lærdahl,
Structural Bioinformatics

- Set up all laptops correctly
  - *Do it yourself so that you learn something!*
- Today: Log in on the freebee.abel.uio.no UNIX server
- Another possibility: Log in on the login.uio.no UNIX server

https://wiki.uio.no/projects/clsi/index.php/MBV-INFX410_2015

https://wiki.uio.no/projects/clsi/index.php/Laptop_Setup

Today's Programme

- Biological databases
- Brief introduction
  - What is UNIX?
  - Why should you learn UNIX?

- Setting up your laptops

> What about those of you that
> know Unix and Python very well?

They have left the room...

- Very briefly on the Unix shell, file system and some commands
- UNIX basics exercise
- Tomorrow, continue on databases & working with biological sequences

Oslo
universitetssykehus

UiO : Department of Informatics
University of Oslo

UNIX file system hierarchy

Jon K. Lærdahl,
Structural Bioinformatics

File system hierarchy

Jon K. Lærdahl,
Structural Bioinformatics

~/ $HOME

At UiO ~/ or $HOME is exactly the same as M: on a Windows machine – your own home area

~/dir1

~/dir2

~/projects

~/pictures

~/bin

~/projects/p1

~/projects/p2

~/projects/p3

~/projects/p1/subdir

~/projects/p3/prt1

~/projects/p3/prt2

Oslo universitetssykehus

UiO : Department of Informatics
University of Oslo
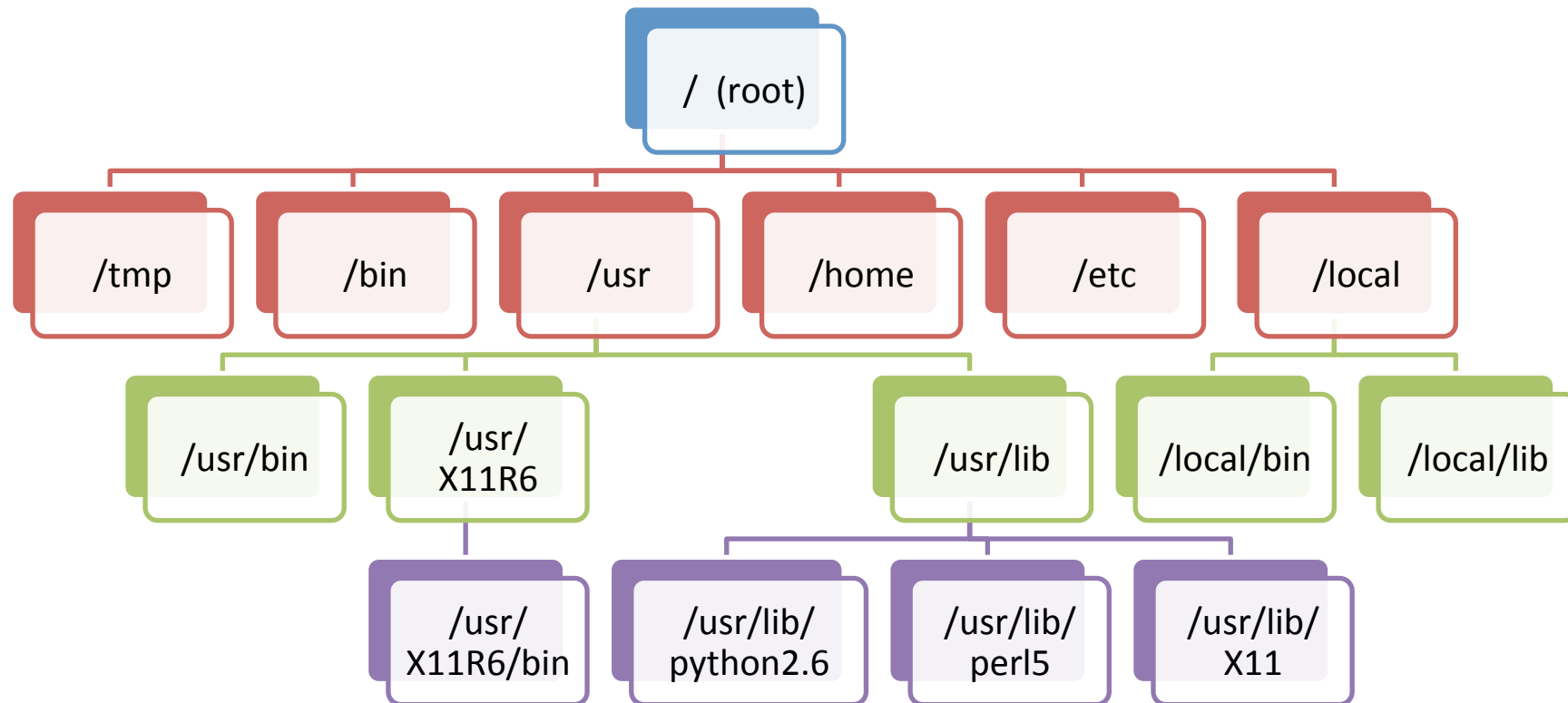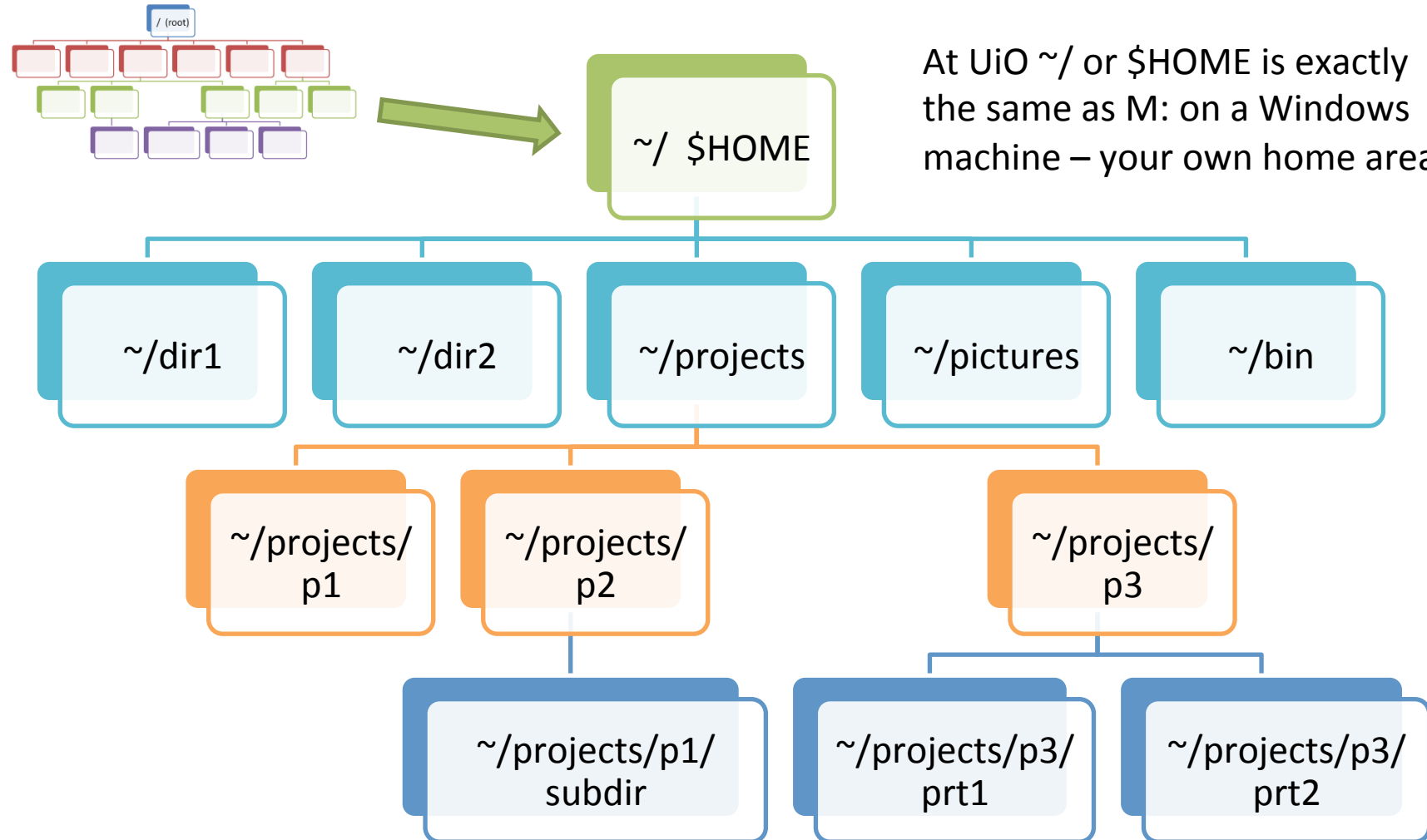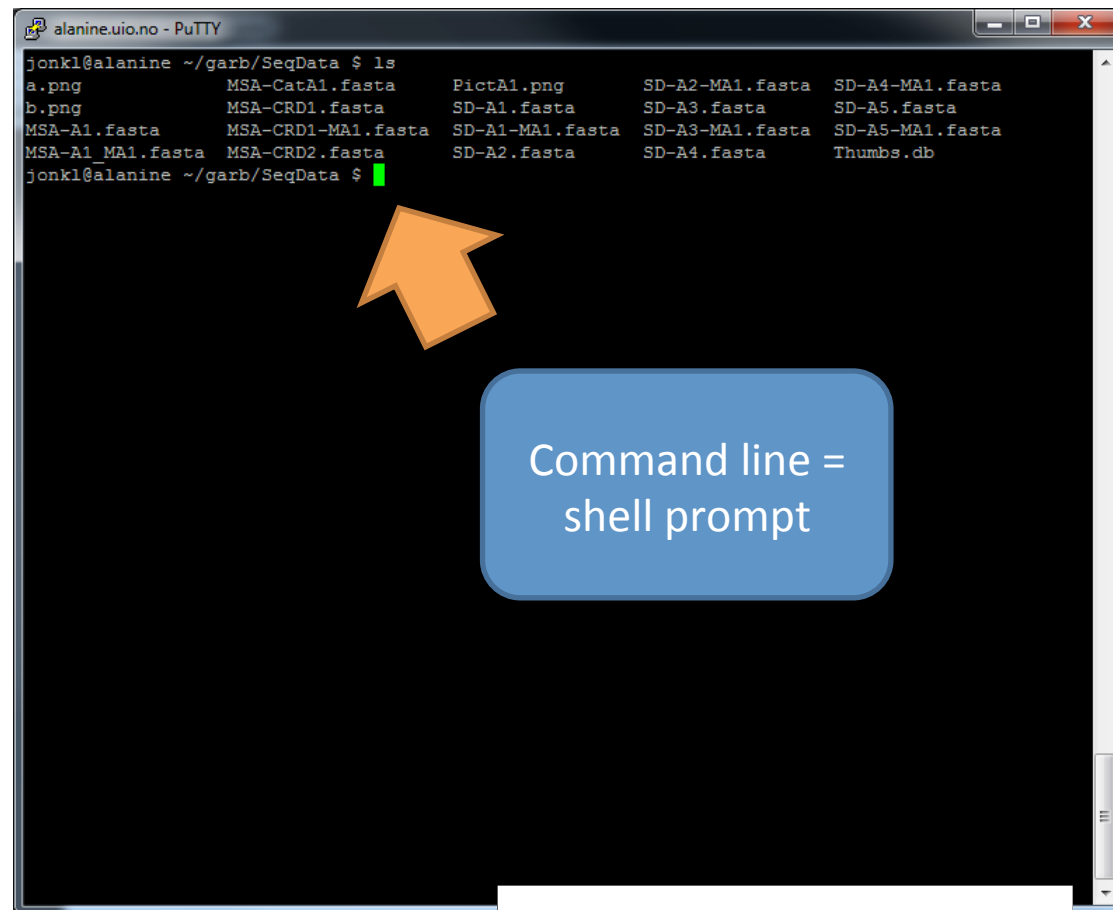
UNIX and file system basics

- ~/ or $HOME is your home area directory
- . is your current directory
- .. is the directory above the one your are in
- File names and commands are **case sensitive**
- MyFile.txt and myfile.txt are not the same
- ~/Pictures/Family/anna1.jpg and ~/Pictures/family/anna1.jpg are not the same (they are in different directories)
- Avoid spaces, special characters and Norwegian letters in file and directory names as this sometimes causes trouble
  - ~/Pictures/Ølfest/anna1.jpg
  - ~/Pics/Summer 2012/anna1.jpg

~/Pictures/Ol-fest/anna1.jpg
~/Pics/Summer_2012/anna1.jpg

Oslo
universitetssykehus

UiO : **Department of Informatics**
University of Oslo

UNIX command line

```
alanine.uio.no - PuTTY
jonkl@alanine ~/garb/SeqData $ ls
a.png            MSA-CatA1.fasta     PictA1.png        SD-A2-MA1.fasta  SD-A4-MA1.fasta
b.png            MSA-CRD1.fasta      SD-A1.fasta       SD-A3.fasta      SD-A5.fasta
MSA-A1.fasta     MSA-CRD1-MA1.fasta  SD-A1-MA1.fasta   SD-A3-MA1.fasta  SD-A5-MA1.fasta
MSA-A1_MA1.fasta MSA-CRD2.fasta      SD-A2.fasta       SD-A4.fasta      Thumbs.db
jonkl@alanine ~/garb/SeqData $
```

Command line =
shell prompt

The UNIX shell is a command-line interpreter
• a program that waits for your commands and executes them
• it is a shell around all programs being run

Various versions
• C shell (csh)
• tcsh
• Bourne shell (sh)
• Bourne-again shell (bash)
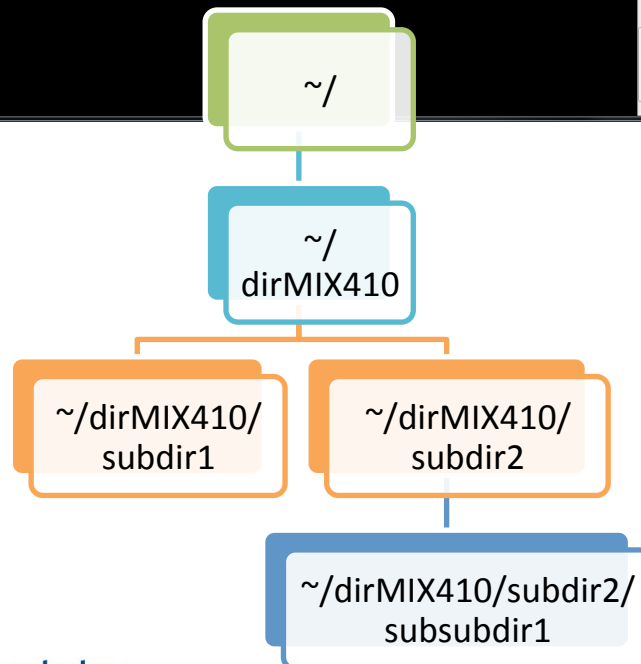   • Default on most Linux systems and Mac OS X

We will use bash

Nerd humor

Oslo
universitetssykehus

UiO **:** **Department of Informatics**
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics


```
jonkl@alanine ~ $ mkdir dirMIX410
jonkl@alanine ~ $ mkdir dirMIX410/subdir1
jonkl@alanine ~ $ mkdir dirMIX410/subdir2
jonkl@alanine ~ $ mkdir dirMIX410/subdir2/subsubdir1
jonkl@alanine ~ $ cd dirMIX410
./dirMIX410
jonkl@alanine ~/dirMIX410 $ cd ..
jonkl@alanine ~ $ cd .
jonkl@alanine ~ $ cd dirMIX410/subdir2/subsubdir1
./dirMIX410/subdir2/subsubdir1
jonkl@alanine ~/dirMIX410/subdir2/subsubdir1 $ cd
jonkl@alanine ~ $ cd dirMIX410/subdir2
./dirMIX410/subdir2
jonkl@alanine ~/dirMIX410/subdir2 $ pwd
/ifi/bifrost/a04/jonkl/dirMIX410/subdir2
jonkl@alanine ~/dirMIX410/subdir2 $ ls
subsubdir1
jonkl@alanine ~/dirMIX410/subdir2 $ █
```

UNIX commands

• To make a new directory use mkdir (make directory)
  • mkdir dirMIX410
• To make subdirectories
  • mkdir dirMIX410/subdir1
  • mkdir dirMIX410/subdir2
  • mkdir dirMIX410/subdir2/subsubdir1
• To navigate use cd (change directory)
  • cd dirMIX410
  • cd .. (takes you one level up)
  • cd . (nothing happens, you stay were you are)
  • cd (takes you to your home directory)
• To find out where you are use pwd (print working directory)
• To see the files and directories in your current directory use ls

Oslo
universitetssykehus

UiO : Department of Informatics
University of Oslo

Jon K. Lærdahl,
Structural Bioinformatics

Time to try yourself!

Unix basics exercises