Bash for Bioinformatics

(MBIO 4030 T10)

Abdullah Zubaer; Hausner Lab, University of Manitoba, zubaera@myumanitoba.ca

What you will learn

- Bash basics (e.g., directories and navigation)
- File manipulation (e.g., text editing)
- Shell scripting basics (e.g., running programs)

Bioinformatics and Computer

- Computer is the main tool for bioinformatics

- A bioinformatician must know his/her computer very well

What is a Computer?

- Computer is a machine/device
- Can do arithmetic or logical operations
- Takes input/instructions and gives output
- Components - hardware, software (operating system)

Operating System (OS)

- An interactive interface to communicate with the machine
- Two types:
  - Command-line interface
  - Graphical interface

Operating System (OS)

- Early computers were like calculators – run one program only
- Later “operating system” was developed – allow to run multiple programs and applications
- Unix – developed by AT&T Bell Laboratories in 1960s - Written in C language
- Unix was originally developed for reserach

Unix OS

- Developed by the experts for the experts
- ‘No nonsense’ design
- Multi-user facility
- Powerful command-line
- Core of a number of modern operating systems

Contemporary OS

Operating system	Kernel type	Command line	Comment
Windows	Windows NT	cmd	Commercial
MacOS	Darwin (Unix-based)	bash	Commercial
Ubuntu (Debian)	Linux (Unix-like)	bash	Open source

Command-line Interface (CLI)

- Also know as Command-line User Interface (CUI) or Console User Interface (CUI)
- More powerful than Graphical User Interface (GUI) in certain ways
- e.g., sh, bash etc.

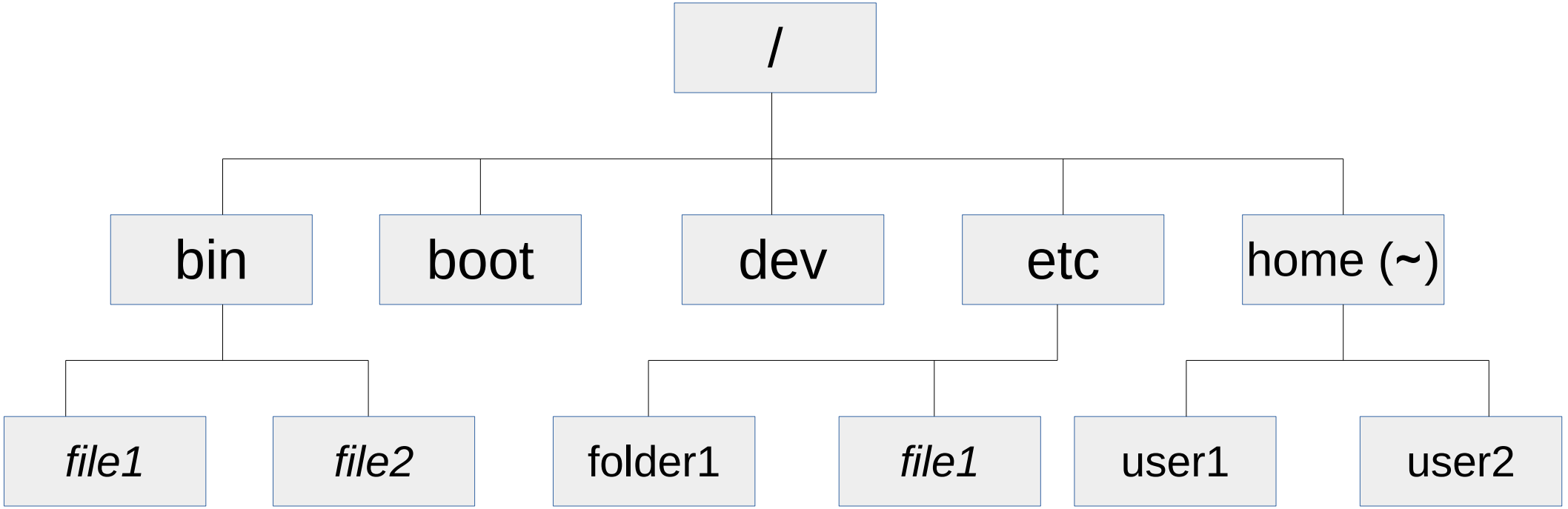
Command-line Interface (CLI)

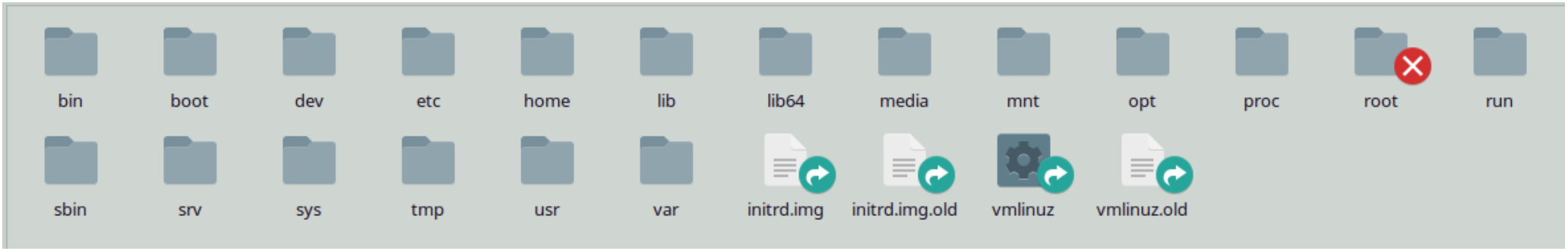
- shell = an outer layer, an interpreter
- CLI shell types:
  - sh (Bourne shell, 1976, Unix)
  - csh (and also tcsh, C shell, 1978, BSD)
  - ksh (KornShell, 1983, Unix)
  - zsh (improved sh, 1990, Unix)
  - bash (Bourne-again shell, 1989, GNU/Linux)

Bash

- A modern command language and shell created by GNU
- Simple syntax
- More utilities (super-set of the previous shells)
- Auto completion with ‘tab’ key

Navigating the Unix/Linux System





Directory path

If we have a file named **`class.txt`** in a user's home folder, then -

- The **absolute path** is **`/home/username/class.txt`**
- The **relative path** is **`~/class.txt`**
- Notations:

  `~` means home folder

  `.` means current folder

  `..` means previous folder

First step to bash

- `pwd`
- `whoami`
- `who`
- `cd directory`
- `cd`
- `cd ..`
- `ls`

Finger exercise 1

- Navigate your file system. List the files in your home directory, then move to different directories such as **root**, **bin**, **etc** and **lib**.
- Use **pwd** and `ls` every time you change directory
- Find out the users logged into the system

Known technical issues:

- **For MacOS, the home folder is `/Users/username/`. The `/home/username/` directory does not exist.**
- **For Windows 10, you can access your Windows filesystem by navigating to `/mnt/c/` from your Linux subsystem.**

Manipulating files and directories

- Copy – cp

  cp original_directory/file_name destination_directory/

- Cut/Move/Rename – mv

  mv original_directory/file_name destination_directory/file_name

- Delete/Remove – rm

  `rm directory/file_name`

Note: For current directory, there is no need to mention the directory details.

Finger Exercise 2

- Download a nucleotide sequence file from NCBI (Hint: you can use an accession - MK026450.1)
- Move it to your home directory
- Copy it to the Desktop folder and the Document folder
- Remove all the files

Note: Your home directory is `/home/your_username/`