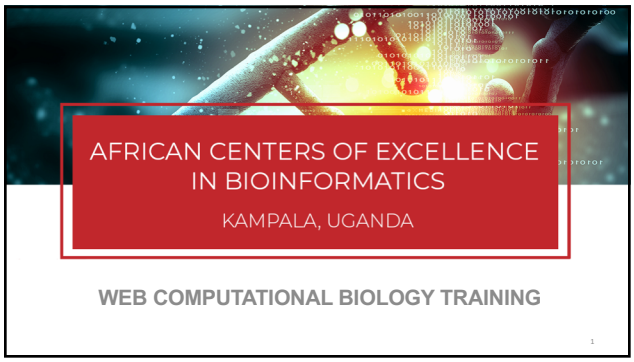


Today's Instructor

Dr. Amitava Roy,
Ph.D. in Physics

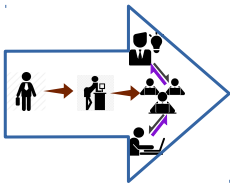
Ongoing Computational Biology projects:
- Vaccine development
- Structure determination of prion

- Bioinformatics and Computational Biosciences Branch (BCBB), Rocky Mountain Laboratories (RML), NIAID,
- NIH, Hamilton, MT USA.
- Contact our team via email:
  - Email: ace@icermali.org
  - Listserv: ACE-MALI-L@LIST.NIH.GOV
  - Instructor: amitava.roy@nih.gov

2

Unix for Computational Biology

Computational biology requires putting together many several tools together to create a pipeline.



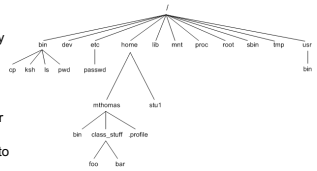
Fundamental design concept in Unix is like a 'workbench' of small tools, each of which only does one thing but does it very well. Unix allows easy integration of different tools in to a pipeline. Programs can run on background – do not need their own pop up window

3

Unix – file structure

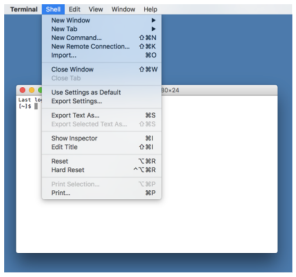
Unix uses a hierarchical file system. Easy to to make programs and databases available to all users.

A HOME directory is a default location for all of a given user's files. It gives the system a standardized location in which to find configuration and preference files.



4

Unix – command line



There are many kinds of tasks in bioinformatics that are much more difficult, or perhaps impossible to do by pointing and clicking, as compared to using a command line.

Unix has thousands of commands and makes it easy to integrate graphic interface based programs with text based command line programs.

5

Unix – example of few tools

First hands on exercise - handson_unix_I_1.sh

ls
cd
man
mkdir
touch
mv
cp
rm
less

6

Unix – aliases and profile

From output of **"ls"**, it is not often clear what is a directory and a file.
"ls –l" put a "d" at the start of each line.
"ls –p" put a trailing slash at the end of a directory.

"rm –i" makes **"rm"** a little bit safer by asking confirmation before deleting.

Luckily, there is a way of doing this by using Unix aliases.
alias ls='ls -p'

To Unalias temporarily - **"\ls"**, **"command ls"**

Permanently - **"unalias ls"**

7

Unix – aliases and profile

Aliases only exist in the current terminal session. If you log out, or use new terminal window, then you will have to retype the alias.

We can create a file storing all the aliases and load them automatically every time.We are going to use a text editor (gedit, TextEdit…) to create a file "profile". Write the aliases you want

```
# some useful command line short-cuts
alias ls='ls -p'
alias rm='rm -i'
```

To use the aliases type source **"source profile"**

You can also write the aliases directly to .bashrc file in your home directory, or sourcing the profile file containing the aliases from .bashrc file.

8

Unix – aliases and profile

If you name your file **.profile** in your home directory, then it will be loaded automatically every time you log in.

If a filename starts with a dot, Unix will treat it as a hidden file. To see it, you can use **"ls –a"** which lists all hidden files.

All Unix commands are just files that live in directories somewhere on your computer. Unix uses something called $PATH (which is an environment variable) to store a list of places to look for programs to run. If we develop some scripts, and want them to be able to run from any directory, we can update the variable $PATH In our **.profile** file by adding a line

```
PATH=$PATH":$HOME/script_directory"
```

9

Unix – Match making

You will often want to search files to find lines that match a certain pattern. The Unix command grep does this (and much more).

You can also replace a pattern by another pattern by the command sed.

Second hands on exercise - handson_unix_I_2.sh

10