# Bioinformatic Summer School - Comparative Genomics, Nanopore Metagenomics and Metataxonomic Analysis

# Homework

## Linux install and BASH

Installation of a Linux environment. Most bioinformatics is done with the **Linux** operating system, rather than for example windows. Ubuntu is a modern version of Linux which is free and widely used. If you run windows, you can install an Ubuntu terminal directly with the Ubuntu subsystem.

1. Open the Windows Store on your machine and install **Ubuntu** (just the first result – **not** version 20.04 and not 24.04!).
2. An Ubuntu terminal will open and show you an error message, but just close it. It is complaining about finding the ‘Windows Subsystem for Linux’ (WSL), which we will install next
3. Press the windows key and type ‘cmd’, then enter. This will open the windows command line.
4. Type this and then enter:
   1. wsl --unregister ubuntu
5. Type this and then enter:
   1. wsl –install
6. Enter a username and password you can remember
   1. I use the same as my university account to make it easier to remember
7. Now open your Ubuntu terminal!
   1. One way is to press the windows key and type ‘Ubuntu’

When you have successfully installed ubuntu, make sure you update it. The next two commands will do exactly this, although they are a bit unintuitive: first we **update** the list of things that can be updated and next we **upgrade** them. The anatomy of these commands are as follows:

1. **sudo** is for ‘superuser’ access, because you are messing with the essential files of the operating system. Linux will not allow you to do so unless you specifically ask.
2. **apt** is a program to handle packages. A constant issue with software installation is to make sure that software packages have the appropriate versions, which you will be spending a lot of time on the next few days
3. **update/upgrade** updates the index of which versions are needed and then upgrades them, respectively.

**sudo apt update**

**sudo apt upgrade**

**\*\*IMPORTANT\*\***

**THE COMMANDS OF THE COURSE ARE SPECIFICALLY MADE FOR A WINDOWS SYSTEM RUNNING THE WINDOWS SUBSYSTEM FOR LINUX (WSL) AS INSTALLED ABOVE – I STRONGLY SUGGEST THAT YOU ONLY USE THAT**

**IF YOU HAVE AN OLD VERSION OF UBUNTU YOUR WINDOWS, GET RID OF IT AND INSTALL AS ABOVE**

**IF YOU HAVE UBUNTU THROUGH VIRTUAL BOX, GET RID OF IT AND INSTALL AS ABOVE**

**IF YOU HAVE ANYTHING OTHER THAN AS DESCRIBED ABOVE, GET RID OF IT AND REINSTALL AS ABOVE**

**WE ALWAYS SPEND WAY TOO MUCH TIME TRYING TO MAKE ALTERNATIVES WORK, SO PLEASE JUST DO AS INSTRUCTED**

**MAC USERS**

For **MAC OS X** users, you are already running Linux underneath the OS X interface. You have access to the command line with a standard program called ‘Terminal’. Things should work in the Mac terminal as in the Ubuntu terminal, but be aware of two key issues:

1. Mac uses the **zshell** instead of the Ubuntu **bash** as the language in the terminal. Often exactly the same, sometimes not.
2. Newer Macs have a highly optimized and super fast CPU, which works great for native Mac-software. Not so much for anything else, though, which means that software made for Intel CPUs might not run at all. There is a workaround for this, which we will get to shortly

**VIRTUAL MACHINE USERS**

If you have a virtual machine installation of Ubuntu already, you will have to ignore it and install a real version as described above. **It will not work.**

**Email me if you have any issues.**

**Regardless of your operating system, make sure you have a working terminal and then do a Linux tutorial here:**

[https://app.datacamp.com/learn/courses/introduction-to-bash-scripting](https://app.datacamp.com/learn/courses/introduction-to-bash-scripting?fbclid=IwAR0efqQIXcikI_RaKUHWFK87wBJKvHxkDkuwp1IUgiceqNkA1CQmV0zshmo).

The basic programming language of the Linux command line is called bash, and we will be using this extensively**. If you show up with no knowledge of this, you will probably not learn a whole lot**, so do the tutorial and play around as much as you can.

## Conda installation

One of the most tricky parts of bioinformatics is the installation of packages. Package A needs version X of package B, but package C need version Y of Package B, which might not be possible. Luckily, the conda package manager takes care of this for us by working out the details of these dependencies and allows us to make individual ‘environments’ for each set of packages for each analysis.

Follow the instructions here (go all the way to the bottom and remember you are now running **Linux** and not windows). Copy and paste (and run!) each command individually:

<https://www.anaconda.com/docs/getting-started/miniconda/install#quickstart-install-instructions>

For MAC users, you should obviously use the **macOS** instructions.

Remember to do steps 2 and 3!

**\*\*\*Close your terminal and open it again to finalize the installation.\*\*\*\***

## Installation of R and Rstudio

We need the programming language R for the metataxonomy. R is a statistically minded programming language, which is great for statistics and for plotting – all my statistics and plots are made in R.

We will use a great integrated development environment (IDE) for R called Rstudio. Rstudio makes writing and running code exceptionally easy. **Rstudio runs directly in windows (or macOS) so don’t mix it up with Linux!**

First you install R:

<https://www.r-project.org/>

Then you install Rstudio

<https://posit.co/download/rstudio-desktop/>

## Consider freeing up some space on your machine

We will need several Gb of harddrive space for the exercises, so this might be a good time to clean out your PC.