# Things to prepare for Command Line Bioinformatics:

We'll spend some time Friday setting everything up, but you can ask James and Ifeanyi for help in the meantime, so I'm sending this through – also your colleagues in the class (I know many of you already have your set-up in place, or managed to set it up during the first workshop!). I'm also very happy to hear many of you are now on Linux machines!

## 1. Windows subsystem for Linux (WSL)

a. If you're on a Mac, or Linux machine, you can skip this!
b. For the command line assembly, participants who are working on a Windows machine will have to install a Linux Subsystem for Windows. The subsystem basically allows us to run tools built for the Linux operating system on a Windows system – very important for bioinformatics and most tools are not built for Windows! Instructions on how to do install WSL and a good explanation of the system is here: https://www.howtogeek.com/249966/how-to-install-and-use-the-linux-bash-shell-on-windows-10/

## 2. A package manager like Anaconda or Miniconda or Mamba

a. Anaconda is just a package manager – it is basically just a big software grocery store (that's for free) that we can use to download all the software we're going to use from. Instructions on how to install Anaconda after you've downloaded the Windows subsystem is here: https://www.how2shout.com/how-to/install-anaconda-wsl-windows-10-ubuntu-linux-app.html
b. If you're on a Mac, you can just download Anaconda! Instructions here: https://tecadmin.net/how-to-install-anaconda-on-macos/
c. If Anaconda is foiling you, there's a good alternative in Miniconda, which is actually a bit faster too:
    i. You can grab the right version for your system here:
        1. https://docs.conda.io/en/latest/miniconda.html
    ii. If you're on the Linux subsystem, remember you should use the Linux version:
        1. E.g Miniconda3-py38_4.9.2-Linux-x86_64.sh
    iii. You can then run this installer by just typing "sh Miniconda3-py38_4.9.2-Linux-x86_64.sh" into the terminal – you need to be in the right directory, where your installer file is.

iv. This is probably in your downloads folder – I'll add some instructions below to find your home directory and user name. Change to your Downloads folder and then run "sh Miniconda3*.sh" – whatever the name of your installer file is! And without the quotations…
v. If you on Mac and Linux, I trust you can find where you saved the installer (look in ~/Downloads/), so just run the installer haha

# 3. The Mpox read files we'll work with in the command line tutorial:

a. Available from: https://www.dropbox.com/sh/qmwz86ii8hld1ta/AABsx53lOlAzTGHpH4SRuWLNa?dl=0
b. These files are slightly large, so download them over the weekend on the hotel's wifi haha
c. Additionally, we may attempt to deplete human reads from our sample. For this we'll need the human genome indices. We can download this from:
d. https://genome-idx.s3.amazonaws.com/bt/GRCh38_noalt_as.zip
e. However, this is 3.75 GB, so if you can't download it, don't worry too much for now!

# 4. Pipeline tools

a. You can install most of the pipeline tools we need with Anaconda, so run the following command in your terminal (anything in a gray box is a separate command here) – don't worry, we'll talk about what this all means next week:

```
conda install -c bioconda fastqc
```

```
conda install -c bioconda trimmomatic
```

```
conda install -c bioconda bwa
```

```
conda install -c bioconda samtools
```

```
conda install -c bioconda ivar
```

```
conda install -c bioconda bowtie2
```

*(Bowtie might have conflicts with your version of python – don't worry about this for now, as the human read depletion step is a bonus for now, and we can resolve this down the line)*

We can also try to install the assembler Spades with Anaconda:

```
conda install -c bioconda spades
```

But for me, this throws up some conflicts – so I've installed from source (basically, just from the original source), that I downloaded here: https://github.com/ablab/spades#sec2.1

What I did (I'm on a Mac)

```
curl http://cab.spbu.ru/files/release3.15.5/SPAdes-3.15.5-Darwin.tar.gz -o SPAdes-3.15.5-Darwin.tar.gz
```

Which should create the folder spades-spades_3.15.4 in the directory (I was in my Downloads folder ~/Downloads) – we'll figure out how to use that in class.

If you're on Linux, try (maybe in your Downloads folder – instruction on how to find that below for WSL):

```
wget http://cab.spbu.ru/files/release3.15.5/SPAdes-3.15.5-Linux.tar.gz
```

then:

```
tar -xzf SPAdes-3.15.5-Linux.tar.gz
```

Which should create a folder called SPAdes-3.15.5-Linux/

# That's all for now! Instructions for WSL home directory below…

# How to find your home directory on a linux subsystem:

In Windows subsystems, this should be somewhere like

**/mnt/c/Users/<username>/Downloads**

Where **<username>** is the name of your account on your computer. If you're unsure what that is try typing the following into the terminal:

```
cd  /mnt/c/Users/
```

This will change directory (the command cd) to the folder **/mnt/c/Users/,** which is the home directory of Windows Linux subsystem. Now type:

```
ls
```

Which is the list command. Do you see your username in the folders and files listed? Note that down that!

So, to get to the downloads folder we type:

```
cd  /mnt/c/Users/YOURNAME/Downloads
```

Where **YOURNAME** is your username you just noted down. For example, I'd type

```
cd  /mnt/c/Users/edythp/Downloads
```

Now type the list command again and check if our read files are there!

```
ls
```