Installing TrimGalore for use on Quest genomics

For the class demo on raw read processing, I chose to use the program TrimGalore for trimming adapter sequences. Quest genomics does not include TrimGalore as a standard module for easy loading. Quest seems to want you to use a different application: Trimmomatic. If you have experience with Trimmomatic, or are agnostic about what trimmer you use, then good: use Trimmomatic or learn to use it.

If instead you think you would like to use TrimGalore in your own research, it may be useful for you to know how I installed it on the class Quest allocation so that you can install it on your home lab's allocation in the future.

TrimGalore operates best within a virtual environment. We need to install that. TrimGalore is really just a wrapper for two other programs, cutadapt and fastQC. We need to install cutadapt (fastQC is available through Quest though, and we do not need to install it redundantly).

Procedure:

1) log in to quest through ssh

```
ssh -X <netID>@quest.northwestern.edu
```

2) navigate to the appropriate directory in your allocation. As part of this installation, you will create a new directory, `TrimGaloreEnv` which will contain all of the software that we download. You will want to create this directory in a location that you will be able to find easily, and that won't get lost in a mess of other directories and files.

3) Load an appropriate python3 module.

```
module load python/3.8.4
```

4) Install virtualenv

```
pip3 install virtualenv
```

5) Now we will create the Trim Galore virtual environment, which is contained in the directory `TrimGaloreEnv`, and can be activated by running a command from an executable within that directory. (Navigate to the parental directory where you want to create the virtual environment…)

```
virtualenv -p python3 TrimGaloreEnv
```

This uses virtualenv to make a virtual environment named TrimGaloreEnv that is pre-loaded with python3 capabilities.

6) Navigate into TrimGaloreEnv

```
cd TrimGaloreEnv
```

7) Activate the virtual environment

```
source bin/activate
```

This "sources" the executable `activate` from within the `bin` directory in the folder you just created. (Note, to turn off the virtual environment, just enter the command `deactivate` from the command line…)

8) Now that it is activated, when we install python stuff, it specifically is installed within the search path for the virtual environment and is guaranteed to be seen when you call it from the command line. Nice. Let's first install TrimGalore itself using git.

```
git clone https://github.com/FelixKrueger/TrimGalore.git
```

9) change into the TrimGalore directory.

```
cd TrimGalore
```

10) note that there is an executable in this directory named `trim_galore`. We need to copy this into a directory within the search path of this virtual environment. One way to do this is:

```
cp trim_galore ../bin
```

I've used relative notation for the path assignment because I know that `bin` is one directory *up* from my current position in the filesystem. You could give and absolute path here if you would be more comfortable with that.

11) There is one Python dependency that we need to install. Do that with pip.

```
pip3 install cython
```

12) Now we need to install cutadapt, which we will also do with git. Note, for some reason this is installing a beta version. It works, so I am not concerned but this might be nice to fix at some point.

```
git clone https://github.com/marcelm/cutadapt.git
```

13) Again, navigate into the `cutadapt` directory.

```
cd cutadapt
```

14) Cutadapt is distributed as source code that we need to compile, or 'make'. Let's "make" cutadapt.

```
python3 setup.py install
```

15) If we want to make full use of the parallel processing on Quest, we will need one additional utility called Pigz that allows for parallel gzipping. We also install this using git.

```
git clone https://github.com/madler/pigz.git
```

16) We also have to switch to the pigz directory, 'make' it, and copy the executable into the parental `bin` directory.

```
cd pigz
make
cp pigz ../../bin
```

Note, my installation and directory system is a little sloppy here. Pigz is nested within cutadapt it looks like. This works, but you might want to have it be a bit neater.

17) Installation is done.

###
In any case, after these steps, it should be possible to test that trim galore and cutadapt are installed properly.

```
trim_galore --version
cutadapt --version
```

If you get an error, please check that the virtual environment is activated.

When you are done, exit the virtual environment by entering:

`deactivate`