PRAGUI Practical

I) Logging onto the cluster

I-a) Via terminal on Linux or Mac

Generic usage:

command: ssh -X <your-lmb-username>@hal

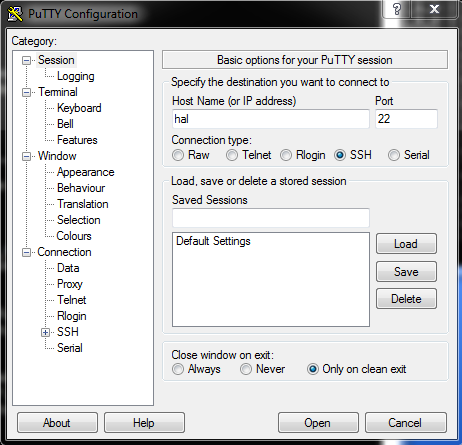
If logged in to your account on an LMB computer:

command: ssh -X hal

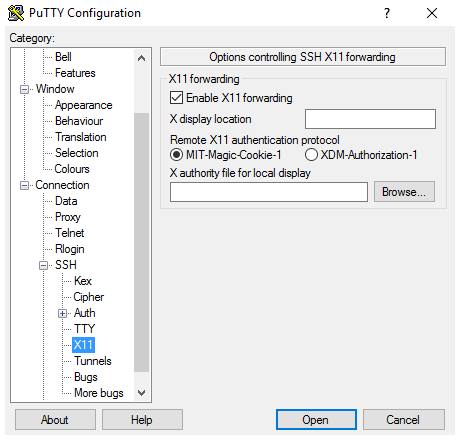
(XQuartz is required)

I-b) Via PuTTY on Windows

Use PuTTY’s graphical interface



X11 is required



II) Setting up PRAGUI ready to run

Commands to run:

module load python3/3.7.1

python3 /public/genomics/PRAGUI/add\_soft\_paths\_cshrc.py

source ~/.cshrc

This last command is going to ask you if you want to modify your ~/.cshrc file. Please type “y” to agree.

Check that it worked by typing the following command:

echo $PRAGUI

You should get back the following:

/net/nfs1/public/genomics/PRAGUI/rnaseq\_pip\_gui.py

You’re now ready to run PRAGUI!

III) Running PRAGUI

III-a) File location

By default, PRAGUI will save your results in the same folder where your fastq files are stored. Therefore, you should be careful to store them in a separate folder.

In this practical, we will create a folder where we will store all our data.

Please type the following commands on the terminal:

cd /beegfs2/

mkdir -p <your-lmb-username>/PRAGUI\_course/single\_end\_data

cd <your-lmb-username>/PRAGUI\_course/single\_end\_data

foreach f (/beegfs2/paulafp/live\_demo/chr3/\*fq.gz)

ln -s $f

end

ls -ltr

By running these commands we have created symbolic links to our data. This avoids having to make multiple copies of the data on the cluster. If you would like to copy the files into your new folder you would need to use the “cp” command:

cp /beegfs2/paulafp/live\_demo/chr3/\*fq.gz .

(please do not do this with the toy data set)

III-b) Opening the GUI

The command to launch PRAGUI is the following:

python3 $PRAGUI

Now you’re ready to start uploading your data for processing!

III-c) Filling in the GUI

Example 1

In this example we will be comparing data from two different C elegans mutant strains. We have single-end data for two replicates in each mutant (i.e four different fastq files). In order to speed up the process, no other replicates were included and we will only be using reads from chr3.

We will start by creating our samples file. To do so, we need the following information about our samples:

- AX6130 (IL-17 overexpression):

/beegfs2/paulafp/live\_demo/chr3/SLX-12097.D706\_D503.HGJCFBBXX.s\_6.r\_1\_chr3.fq.gz

/beegfs2/paulafp/live\_demo/chr3/SLX-12097.D706\_D504.HGJCFBBXX.s\_6.r\_1\_chr3.fq.gz

- AX6186 (NFKI-1 overexpression):

/beegfs2/paulafp/live\_demo/chr3/SLX-12097.D706\_D508.HGJCFBBXX.s\_6.r\_1\_chr3.fq.gz

/beegfs2/paulafp/live\_demo/chr3/SLX-12097.D707\_D501.HGJCFBBXX.s\_6.r\_1\_chr3.fq.gz

Other information required:

- The library used is directional in the reverse strand.

- Location of genome sequence and annotation files

/beegfs2/paulafp/live\_demo/WS255

- Filter reads with mapq below 20.

- Tick cuff\_gtf option so that cufflinks can use your annotation file.

In this example we will be running Cufflinks rather than DESeq2. Since our library is directional, we need to add that information as an option to Cufflinks (use cuff\_opt box). That is done in the following manner for reversed directional libraries:

--library-type fr-firststrand

(if it were a standard directional library the term would be “--library-type fr-secondstrand”)

Example 2

In this example, we will be comparing how two different environmental conditions affect worm neurons. WT animals have been exposed to 21% O2 and 7% O2. We have paired-end data for two replicates in each condition and we will only be using data from chr3.

Firstly, we will need to create a new working directory where our data will be saved (as we have done for the first example)

cd /beegfs2/<your-lmb-username>/PRAGUI\_course/

mkdir -p paired\_end\_data

cd paired\_end\_data

foreach f (/beegfs2/paulafp/live\_demo/paired\_end/\*fq.gz)

ln -s $f

end

We will also copy and edit a samples file designed for this example by running the command:

/beegfs2/paulafp/live\_demo/data/paired\_end/edit\_samples.csv.sh

ls -ltr

If you are curious to see what your samples\_file.csv file looks like, you can print it onto the terminal by typing the following command **(optional)**:

more samples\_file.csv

Launch new session of PRAGUI:

python3 $PRAGUI

Since we already have a samples file in our folder, in this example we will not need to create a new one. We can just use the “Upload” option on PRAGUI instead.

Other information required to run PRAGUI:

- The library used is non-directional.

- Location of genome sequence and annotation files

/beegfs2/paulafp/live\_demo/WS255

- Filter reads with mapq below 20.

- paired-end tags: r1 r2

In this example we will be running DESeq2.

Once PRAGUI has finished running, let’s have a look at the data that was generated:

ls -ltr

Let’s see our multiqc report and exploratory analysis plots:

firefox multiqc\_report.html

xdg-open samples\_file.csv\_sclust.pdf

IV) Download data from the cluster

IV-a) Cyberduck for Windows or Mac

https://cyberduck.io/

IV-b) FileZilla for Linux

https://filezilla-project.org/