JCU Zodiac HPC Intro

Martha Cooper

2020/06/09. Last updated on 2020-10-01

About

A collection of information, links and resources for (human) RNA-seq analysis on the Zodiac HPC by researchers in AITHM.

Dislaimer: This documentation is not endorsed by the JCU HPC team/eResearch. I created this resource when I was learning and am sharing it the AITHM researchers in Cairns to (hopefully) make the HPC learning curve a little less steep. Any mistakes are my own.

---

1. Before you start

The JCU HPC is a linux HPC, and you'll need to be comfortable in the unix/bash command line. Things like navigating the file system as well making/removing files and directories are essential, and understanding bash scripting will be very helpful. There are lots of online resources for this purpose. I like LearnLinuxTV, a youtube channel with intro series on bash commands and scripting to get you started.

Please also make sure you've read the official documentation for the JCU HPC. I'll point back to this documentation a lot throughout.

2. Getting a HPC account

The JCU HPC (zodiac.hpc.jcu.edu.au) is managed by the eResearch Team in Townsville. Important people to ask for help include Wayne Mallet and Wayne Spagnol in Townsville, and Andrew Grey Spence in Cairns.

To get your own account, **email ITHelpDesk to raise a Service Now request**.

3. Logging on to the HPC

When you get your account set up, you can log in using `ssh` from the terminal application on your computer. On a **mac**, open the terminal and log in using your user name is (your jc number) and password (the normal one for your jc account) by the following command (remembering to replace the x's with your jc number):

When on the JCU network, use the following;

```
ssh jcxxxxxx@zodiac.hpc.jcu.edu.au
```

When outside the JCU network, add the port flag (-p) 8822;

```
ssh jcxxxxxx@zodiac.hpc.jcu.edu.au -p 8822
```

And enter your password when prompted.

If you'll be using the HPC a lot, you might want to set up an alias (a command line shortcut) on your computer to make logging on quicker. You can read more about setting up aliases for commonly used commands here. I set up an alias called `connect_zodiac` that `ssh`'s me on to the zodiac HPC.

Unless using **windows 10** (can use computer terminal), instructions for **windows** computers are a bit different and requires ssh client software like PuTTY. Instructions can be found in the official JCU HPC documentation and in the Software Carpentries Introduction to High Performance Computing.

4. Where am I?

You are in your home directory on the HPC `log in node`. This node is used for uploading and downloading files from/to your computer, running quick tests and submitting jobs to the `compute nodes` on the cluster, which are nodes where the actual work is performed. A short introductory lecture to HPC by Andrew Turner at the University of Edinburgh can be found here.

5. Software/Package availability

Use the following command to see what software packages are available for use:

```
module avail
```

Some packages are available in conda environments rather than as environment modules. As example is MultiQC, which is available in Anaconda3.

If you need other softwares, raise another Service Now request by emailing ITHelpDesk.

6. Transfering files to my HPC home directory

`scp` can be used to sending one file from your computer to your account on the log in node.

FileZilla can be used to interactively send many files from your computer to your account on the log in node. connect to HPC in FileZilla. Drag and drop files.

You can keep up to 5TB of data in your HPC home directory. *This is not a permenant data storage solution, only keep working data files on in your HPC home directory.* I'm still not sure how to permenantly store large datasets like sequencing data at JCU and will update this document when/if I find out! *Note: if you run out of space in your home directory, processed data will not be able to be sent back from the compute nodes when your jobs are processing. They will just stop halfway through and you will get an email saying there was a "post processing job error". So, don't store excess data here, and delete already processed files.*

7. Submitting jobs

The JCU HPC wuses a PBSPro Job Submission system. This requires you to write the job you require doing in a special bash script called a PBS script. The PBS script is then submitted to a queue using the command `qsub`, and will be executed when there are enough compute resources available on the cluster to perform your job.

8. PBS Scripts

You can see an example PBS script is below.

The first line `#!/bin/bash` is called the 'shebang' and is required as the first line in every bash script, not just PBS scripts.

The following lines that start with `#PBS` are information for the PBS job submission system (there are loads more options, these are some essential ones and the ones I use).

- -N defines the script name
- -l defines the compute resources the job needs

- -j defined the output made. oe means output and errors will be in the same output file
- -m defined when the user will be emailed (abort, begin, end)
- -M defines the email address to send the above information to

In each PBS script, we then need to: + load the modules (software) that are required to perform that job + move to the correct directory to perfome the job, in the above case that is the directory that the PBS script was submitted from `$PBS_O_WORKDIR` + set any necessary environment variables (What are environment variables?) + make any necessary files for job output + execute the job.

```
#!/bin/bash
#PBS -P ChmiInVivoRnaseq
#PBS -N TestFastqc
#PBS -l select=1:ncpus=1:mem=1gb
#PBS -l walltime=10:00
#PBS -j oe
#PBS -m abe
#PBS -M martha.cooper@jcu.edu.au

##### Load modules #####
module load fastqc/0.11.7

##### Change to current working directory #####
cd $PBS_O_WORKDIR

##### Set environment vars #####
INPUTDIR="$PBS_O_WORKDIR/data"
NCPU=1
OUTDIR="$PBS_O_WORKDIR/outputs"

##### Make output dirs #####
mkdir -p $OUTDIR

##### Execute Program #####
fastqc -t $NCPU -o $OUTDIR $INPUTDIR/*.fastq.gz
```

Make your own PBS script in your favourite text editor, such as *nano* or *Notepad++*. (NB. If using text editors outside the the HPC's linux system

you will need to convert the EOL for your scripts from your OS to linux before submitting to the HPC. In *Notepad++* this can be done under the edit tab, other text editors may require use of *dos2unix*)

```
nano example.pbs
```

Once created and saved, submit the job in pbs script form to the queue using `qsub`

```
qsub fastq.pbs
```

View the job queue with qstat

```
qstat
```

To view only your jobs use qstat with the -u flag and your HPC user name (your JC number)

```
qstat –u jcXXXXXX
```

To delete a job, use qdel followed by the job number

```
qdel jobnumber
```

Once your job is complete, you can view the output of your job using

```
less TestFastqc.o1572010
```

Once complete, you may want to send the output to your own laptop e.g. FastQC or MultiQC html files.

If you're on a JCU computer, you'll need to make yourself an admin by clicking the small JCU flag in the top right bar and clicking "Make me an admin." You'll also need to be on the JCU network, so VPN in with Forticlient if you're not on campus wifi/internet.

On a mac, you can the address of your computer in System Preferences -> Sharing -> Remote login. It should be your jc number followed by @ and then an IP address.

Then, use `scp` to transfer the files

```
scp D0_1505_R106_HHWK3DSXX_ATGAGGCC-GTTAATTG_L001_R1_fastqc.htı
```

Enter your password when prompted.