Pronghorn Homework Module – **Kelli McKeegan**

All personal directories are located: /data/gpfs/assoc/biol\_bids-0/

Necessary files and directories are located:

1. /data/gpfs/assoc/biol\_bids-0/**[NETID]**
2. /data/gpfs/assoc/biol\_bids-0/atram\_files

In your personal directory you will need to create two directories:

mkdir temp

mkdir atram\_out

And create two files:

1. run\_atram.sh
2. class.slurm.sh

The ‘run\_atram.sh’ script is sent into the container to execute aTRAM and the path needs to be altered. Below is an example of the script and the bold sections within square brackets should be changed accordingly.

*run\_atram.sh*

#!/usr/bin/bash

python3 /aTRAM/atram.py -b /biol\_bids-0/atram\_files/atram\_lib/AlspHabad/AlspHabad -t /biol\_bids-0/**[NETID]**/temp -Q /biol\_bids-0/atram\_files/query\_files/**[Query\_File]** -a trinity -o /biol\_bids-0/**[NETID]**/atram\_out/ --cpus 1

The ‘class.slurm.sh’ script needs to be altered as well. Add the name of the job (this can be anything you want), how many CPUs per task (you decide what you think is appropriate), and your email to get updates on the status of your job.

*class.slurm.sh*

#!/usr/bin/bash -l

#SBATCH --job-name=**[Job\_Name]**

#SBATCH --account=cpu-s5-biol\_bids-0

#SBATCH --partition=cpu-core-0

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=**[n]**

#SBATCH --hint=compute\_bound

#SBATCH --time=14-00:00

#SBATCH --mem-per-cpu=3500M

#SBATCH --output=hostname\_%j.out

#SBATCH --error=hostname\_%j.err

#SBATCH --mail-type=ALL

#SBATCH --mail-user=**[Email]**

srun singularity exec -B /data/gpfs/assoc/biol\_bids-0:/biol\_bids-0 /app/atram/aTRAM.sif **[path\_to\_your\_script]**/run\_atram.sh

Submit the job using:

sbatch class.slurm.sh

Answer the following questions:

1. What is the purpose of High Performance Computers (HPCs) and when might you use them?

**To perform complex projects that require higher computing power, with the ability to more sustainably schedule/queue projects with added security to avoid crashing lesser computer systems. These “supercomputers” such as Pronghorn might be utilized when dealing with big data or analyses that run through many steps (e.g. assembling a genome!).**

1. What is a slurm script? Paste the path your slurm script here.

**Slurm is the workload manager used by Pronghorn, and the slurm script is the file of commands and parameters that will be executed. My path: /data/gpfs/assoc/biol\_bids-0/kmckeegan/atram\_files/example\_atram.sl**

1. Paste the path to your gene assemblies on pronghorn.

**My path: /data/gpfs/assoc/biol\_bids-0/kmckeegan/atram\_files/out**

1. How many CPUs did you ask for?

**8**

1. How many genes assembled?

**79 (out of 100)**

1. Did all genes assemble? Why or why not?

**No.**

**Missing or repeat sequences, or unclean/inconsistent data files, may have prevented some genes from assembling correctly or at all.**

1. How long did it take for them to assemble? What processes might have slowed this down?

**03:34:54 (just under 4 hrs)**

**Computer processing power, number of bases or size of datasets, and other users being scheduled with higher priorities in the queue could all slow down assembly.**