Pronghorn Homework Module

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?**

The purpose of High-Performance Computers (HPCs) is to use aggregating computing power (from multiple computers) to complete computational tasks that are often not possible to complete on a typical desktop computer. We use the HPCs when we work with large datasets and complex computational processes related to scientific research. The HPCs delivers high performance with more capacity (CPUs) and less time.

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

A slurm script is basically a workload manager, where we can assign how and when the a job will be done. As shown above, the slurm script allows a user to specify how many CPUs per task can be used, or whether the user wants to be notified when the job is complete (or when errors occur).

Here’s a path to my slurm script

/data/gpfs/assoc/biol\_bids-0/csudta/example\_atram.sl

Note: I first created a new slurm script called *class.slurm.sh* , but it ran into an error. Instead, I copied an example script and modified the slurm script. I was able to submit the job (without errors) using the script from the class example).

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

/data/gpfs/assoc/biol\_bids-0/csudta/atram\_files/out

The output files are stored in this directory.

1. **How many CPUs did you ask for?** 8
2. **How many genes assembled?**

157 output files

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

No. With 157 output files generated, it indicated that not all genes assemble as the expected numbers of output files when all genes (100 proteins) assemble should be 200 files.

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

It took 4 hours 38 minutes for them assemble A low number of CPUs used and a large number of users utilizing the HPCs might slow this job process down.