Pronghorn Homework Module

All personal directories are located: /data/gpfs/assoc/biol\_bids-2/

Setting up

Submission script is located:

/data/gpfs/assoc/biol\_bids-2/atram.slurm.sh

Copy this file to your personal directory.

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

mkdir temp

mkdir atram\_out

Below is an example of the submission script and the bold sections within square brackets should be changed accordingly.

*atram.slurm.sh*

#!/usr/bin/bash -l

#SBATCH --job-name=**homework\_processing\_genes**

#SBATCH --account=cpu-s5-biol\_bids-2

#SBATCH --partition=cpu-core-0

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=1

#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=**aripastore@nevada.unr.edu**

singularity exec /apps/atram/aTRAM.sif python3 /aTRAM/atram.py -b /data/gpfs/assoc/biol\_bids-2/Course\_content/atram\_files/AlspHabad/AlspHabad -t /data/gpfs/assoc/biol\_bids-2/**arielle**/temp -Q /data/gpfs/assoc/biol\_bids-2/Course\_content/atram\_files/query\_files/**20**\_proteins.fasta -a trinity -o /data/gpfs/assoc/biol\_bids-2/**arielle**/atram\_out/

Target files

All target files are found here: /data/gpfs/assoc/biol\_bids-2/Course\_content/atram\_files/query\_files/

Each file begins with the number of genes it contains. In class you will assemble 2 genes. For your homework you will assemble 20 genes.

/data/gpfs/assoc/biol\_bids-2/arielle

Processors

You will need to select the number of processors you want to use per job. You can change the number of tasks and CPUs per task in the SBATCH options in the submission script. Pronghorn has 32 CPUs per node. Pronghorn will allocate the resources you request to aTRAM as it runs. You can start with a single task and CPU per task, measure the gene assembly rate, and change the computational resources from there.

Submitting

Submit the job using:

sbatch atram.slurm.sh

Queue

To check on the status of your job, use squeue in the command line. This will pull up a list of all of the jobs currently running or waiting in the queue. You will see your job by the job name and your user name. If your job is waiting in the queue, the time will say 0:00. Once it starts running the timer will start. You can check on your jobs only with squeue -u <USERNAME>

**For your homework you will assemble 20 genes.**

**Answer the following questions:**

1. What is the purpose of High Performance Computing (HPC) and when might you use an HPC cluster?

**HPC allows you to use the power of super computers on your personal device. HPC is useful when you need to simultaneously compute many problems at once in a cost and time effective manner. It is also useful when you need to scale up (vertical scaling) and increase the computer resources available to a problem. Finally, there is scaling out (horizontal scaling) that also increases the computer resources available to the problem.**

2. What is a slurm script? Print the path of your slurm script here (including file name).

**A slurm script is almost like a shell script, but you have to set parameters on resource allocations and requirements. The path to my slurm script is /data/gpfs/assoc/biol\_bids-2/arielle/atram.slurm.sh.**

3. What is the path to your gene assemblies on pronghorn (i.e., output files).

**/data/gpfs/assoc/biol\_bids-2/arielle/atram\_out**

4. How many CPUs per task did you ask for?

**1 CPU per task.**

5. How many genes assembled?

**19**

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

**No. I am not sure why, but it is only showing genes 1-19.**

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

**It took a total time of 2:16:40.** **It could have been slowed down by other students in the queue.**