**BIOL 792 (Biological Datasets)**

**Dr Julie Allen**

**Homework 8– High Performance Clusters**

**Pankaj Bhatta**

Pronghorn Homework Module

\*\*\* Next time remind them how to log in and where to go in this sheet then– then get into the homework.

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?

Ans: High performance computers (HPCs) are basically an aggregation of computing power to solve problems which are either too larger for standard computers or would too long. It uses supercomputing system which enables the analysis of huge volumes of data.

We use them when we have large number of files or data to process.

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

Ans: Slurm script is the scripting that includes a list of SLURM directives or commands to tell the job scheduler what to do.

The path of my slurm script:

[pbhatta@login-0 ~]$ cp -r \

> /data/gpfs/assoc/biol\_bids-0/Course\_content/hpc/submission\_scripts/ \

> /data/gpfs/assoc/biol\_bids-0/pbhatta

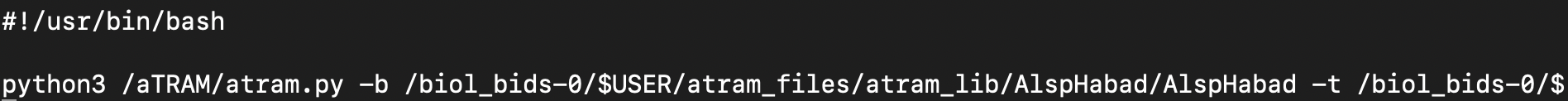
[pbhatta@login-0 ~]$ cp -r \

> /data/gpfs/assoc/biol\_bids 0/Course\_content/atram\_files \

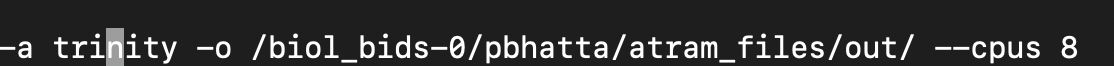
> /data/gpfs/assoc/biol\_bids 0/pbhatta

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

Ans: run\_atram.sh







class.slurm.shText

Description automatically generated

sbatch class.slurm.sh

1. How many CPUs did you ask for?

Ans: I asked for 8 CPUs (4 CPUs takes twice as long to run than 8 CPUs).

1. How many genes assembled?

Ans: Only 16 genes got assembled out of 100. Text

Description automatically generated

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

Ans: No, it did assemble all the genes, because I think there is some issue with the container where it can’t find the python pathway, or it stops suddenly during running.

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

Ans: For me, when I submitted, it immediately failed after completing 16 gene assemblies. This might be because of inefficient resource allocation, other people running the jobs and taking up the queue.

**Thank You! ☺**