Ellen Wilson

BIOL792

4/27/2021

wilson\_run\_atram.sh

#!/usr/bin/bash

python3 /aTRAM/atram.py -b /biol\_bids-0/ellenw/atram\_files/atram\_lib/AlspHabad/AlspHabad -t /biol\_bids-0/ellenw/atram\_files/tmp -Q /biol\_bids-0/ellenw/atram\_files/query\_files/100\_proteins.fasta -a trinity -o /biol\_bids-0/ellenw/atram\_files/out --cpus 8

~

class.slurm.sh

#!/usr/bin/bash -l

#SBATCH --job-name=Wilson\_Job

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

#SBATCH --partition=cpu-core-0

#SBATCH --reservation=cpu-s5-biol\_bids-0\_41

#SBATCH --nodes=1

#SBATCH --ntasks=1

#SBATCH --cpus-per-task=8

#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=ellenw@nevada.unr.edu

#^^^^^^^ Resource Request Section

#--------------------------------

#vvvvvvv Execution Section

srun singularity exec -B /data/gpfs/assoc/biol\_bids-0:/biol\_bids-0 /apps/atram/aTRAM.sif ./wilson\_run\_atram.sh

Answer the following questions:

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

HPCs are very powerful computers that use cumulative computer power to complete highly complex and big jobs. What would take a regular laptop something of 31 years to complete a single job, a HPC can complete that job in mere hours or less. They are essentially clusters of computers that are all interconnected with many nodes and cores. We would use HPCs when analyzing large datasets like entire genomes.

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

A slurm script is similar to a bash or Unix script. They give directions and code to the computing system to complete a job. They must stay within slurm perameters.

/data/gpfs/assoc/biol\_bids-0/ellenw is the path to my class.slurm.sh script. I believe it should have been in /data/gpfs/assoc/biol\_bids-0/ellenw/atram\_files, but I was having difficulties with my script so I was testing it in different locations.

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

/data/gpfs/assoc/biol\_bids-0/ellenw/atram\_files is where my assemblies are. I think they should have been in my out/ directory, but I may have missed something in my script.

1. How many CPUs did you ask for?

8

1. How many genes assembled?

76 genes completely assembled. 3 were assembled differently. \_11 was PROT, \_40 was AUGMASK and \_83 only had one file and it was also AUGMASK. 21 genes did not assemble at all.

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

No, not all genes assembled. This could be because of inconsistencies in the data, missing sequences or nucleotides, and computing processing.

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

My slurm job took about 5 hours (05:01:51 exactly). Even though I gave it plenty of time to run, this felt like a longer task for it to complete. This could because of the queue I was in, and other people had priority on pronghorn over my run.