Busco on HPC cluster:

What is Busco?

BUSCO provides measures for quantitative assessment of genome assembly, gene set, and transcriptome completeness based on evolutionarily informed expectations of gene content from near-universal single-copy orthologs.

The full documentation for Busco is found in the following links:

User Guide

GitLab

Versions Available:

- Busco –v4.0.2

How to load a version of Busco?

To load a version of Busco on the HPC, use the following command:

```
module avail bio/busco
```

The version will be listed. To use a version of software, use following command:

```
module load bio/busco/4.0.2
```

Verify by using this command:

```
module list
```

The loaded software and dependencies, a lot of secondary softwares in this case, will be shown.

How to use Busco on the cluster?

There are two methods to run Busco on the cluster.

The Interactive Way:

To run the program interactively, follow the steps:

```
#Open a bash session on compute node
srun -p main --qos main -n 1 -c 12 --mem 10G --pty bash

#Load the module
module load bio/busco

# Start your commands here
busco --help

#Follow with commands to execute
```

This method is ideal for a short job run which produces runtime output and to debug the codes.

The Script (Preferred):

To run a slurm job, the user must prepare input files. For this example, get input files with,

```
# Copy test files from busco installation folder
cp -r /share/apps/busco/test-day/eukaryota/ .

cd eukaryota
#Copy Augustus config file to the same directory since most don't have
```

```
#write permissions to AUGUSTUS cofig file
cp -r /share/apps/augustus/augustus-3.3.2/config .
#Now export the AUGUSTUS_CONFIG_PATH to the copied directory
#Note: You must load Busco first
export AUGUSTUS_CONFIG_PATH="<path_to_copied_directory>"

#Now we are ready to run busco throuch a script
touch script.sbatch
```

Use the following template for the script,

```
#!/bin/bash
#SBATCH -J Busco-test
#SBATCH -n 1
#SBATCH -mem=16G
#SBATCH -p main
#SBATCH --qos main
#SBATCH -o Busco_sim_out-%J.txt
#SBATCH -e Busco_sim_out-%J.txt
#SBATCH -t 30

module load bio/busco        # load the module

#Copy and export augustus config file
export AUGUSTUS_CONFIG_PATH="<path_to_copied_directory>"

cd $SLURM_SUBMIT_DIR       # Moving to the test directory

#Run Busco
busco -i genome.fna --augustus_species saccharomyces_cerevisiae_S288C
-c 8 -m geno -f --out test_eukaryota
```

Schedule the job with the following sbatch command.

sbatch script.sbatch

All the processed files will be generated in the test_eukaryota directory.

Where to find help?

If you are stuck on some part or need help at any point, please contact OIT at the following address.

https://ua-app01.ua.edu/researchComputingPortal/public/oitHelp