MAKER on HPC cluster:

What is MAKER?

MAKER is a portable and easily configurable genome annotation pipeline. Its purpose is to allow smaller eukaryotic and prokaryotic genome projects to independently annotate their genomes and to create genome databases. MAKER identifies repeats, aligns ESTs and proteins to a genome, produces ab-initio gene predictions and automatically synthesizes these data into gene annotations having evidence-based quality values. MAKER is also easily trainable: outputs of preliminary runs can be used to automatically retrain its gene prediction algorithm, producing higher quality gene-models on seusequent runs. MAKER's inputs are minimal and its ouputs can be directly loaded into a GMOD database. They can also be viewed in the Apollo genome browser; this feature of MAKER provides an easy means to annotate, view and edit individual contigs and BACs without the overhead of a database. MAKER should prove especially useful for emerging model organism projects with minimal bioinformatics expertise and computer resources.

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

[Official Website](#)

[GitHub](#)

Versions Available:

- Maker –v 2.31.10
- Maker –v 3.1.3

How to load a version of MAKER?

To see all versions of MAKER on the HPC, use the following command:

```
module avail bio/maker/
```

To load a desired version of maker (maker- v 3.1.3 in this case), use the following command

```
module load bio/maker/3.1.3
```

Verify dependencies using this command:

```
module list
```

This program has a lot of runtime dependencies, mostly biology software, but includes Perl and openMPI as well.

Important binaries loaded with MAKER module:

- maker
- fasta_tool
- evaluator

Moreover, binaries in the dependency's module loaded are also available for use if necessary.

How to use MAKER on the cluster?

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

```
#Copy the test files from the maker module example files into test
dir
```

```
cp /share/apps/maker_3_1_3/maker/data/* .


#Now to generate configuration files use following command on working
#directory
maker -CTL
```

This will generate three extra files along with input data which is needed to run MAKER. To see more information about all the configuration files, read the docs linked at the top.

Now, edit one of the configuration file, namely maker_opts.ctl, to point to input files. Use any text editor you want and make the following changes.

```
genome=dpp_contig.fasta
est=dpp_est.fasta
protein=dpp_protein.fasta
est2genome=1
```

Put in above values in the field highlighted by yellow in the file. Do not edit other fields.

Use the following script as a template,

```
#!/bin/bash
#SBATCH -J Test
#SBATCH -n 1
#SBATCH -c 8
#SBATCH --mem=10g
#SBATCH -p main
#SBATCH --qos main
#SBATCH -e err.txt
#SBATCH -o out.txt

# Load the MAKER module
module load bio/maker/3.1.3

#Move to test directory
cd $SLURM_SUBMIT_DIR

# Run maker in directory containing input and configuration files.
maker -R
```

Schedule the job with the following sbatch command.

```
sbatch script.sbatch
```

All the processed files will be generated in the same directory as the script.

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