

Braker on HPC cluster:

What is Braker?

Braker is a combination of GeneMark-ET and AUGUSTUS, designed to automatically generate full gene structure annotation in novel genome using genomic and RNA-Seq data.

BRAKER2 is an update to BRAKER which enables fully automated training of the gene prediction tools GeneMark-ET and AUGUSTUS from RNA-Seq and/or protein homology information and that integrates the extrinsic evidence from RNA-Seq and protein homology information into the prediction.

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

[GitHub](#)

Versions Available:

- BRAKER v. 36x1

How to load a version of Braker?

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

```
module load bio/braker/v54322a4
```

Verify by using this command:

```
module list
```

Above command should list 16 dependencies, including itself, since this is software uses functionality of other software in its pipeline.

Note:

To execute the braker, use `braker.pl` command after loading the braker module.

See all options using the following command,

```
braker.pl --help
```

Moreover, there is a user guide in the following location `/share/apps/braker/BRAKER/docs`.

How to use Braker on the cluster?

Pre-Setup:

Before running the braker, the user should set up an environment for Augustus and GeneMark in their home.

By default, the Augustus home directory is not writable for HPC-users to avoid changing global settings for software. Therefore, users should copy Augustus config and download Augustus scripts folder from GitHub to any writable directory, in this case user home directory.

```
#Copy Config Folder  
cp -r /share/apps/augustus/augustus-3.3.2/config $HOME/
```

Go to the \$HOME folder and Download scripts,

```
#Download Scripts from Augustus GitHub  
svn export https://github.com/Gaius-Augustus/Augustus/trunk/script
```

Now, user should export two environment variables for Augustus to know which config and scripts to use before running,

```
#Export Config Path
export AUGUSTUS_CONFIG_PATH="$HOME/config"

#Export Scripts Path
export AUGUSTUS_SCRIPTS_PATH="$HOME/scripts"
```

Note: It is recommended to create a text file, write this command, and source the file in each run because these environment variables will change in each new login. Users can edit their `.bashrc` file as well.

This should be set up Augustus. However, for GeneMark, users need to download the license key from GeneMark Website. Here is the link to the website,

[Genemark Liscence](#)

Fill in the credentials and download GeneMark-ES/ET/EP ver 4.69_lic 64-bit license.

Unzip the license and copy it to \$HOME directory.

Rename the file to `.gm_key`, which will be hidden in the home directory.

Create a script to run Braker on the cluster.

The Script:

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

```
#Copy the test input file from Braker installation directory
cp /share/apps/braker/BRAKER/example ~/test/

# Make a script to run Braker
touch script.sbatch
```

Note: ~/test refers to a test directory in the home folder. Feel free to use any other directory. Script should be in the same directory as in the input files.

Use the following template for the script,

```
#!/bin/bash
#SBATCH -J test
#SBATCH -p main
#SBATCH --qos main
#SBATCH -N 1
#SBATCH -c 16
#SBATCH --mem=16G
#SBATCH -e err.log
#SBATCH -o out.log

# Go to the script's directory
cd $SLURM_SUBMIT_DIR

#Load the module
module load bio/braker

# Source the environment variable file | See pre-setup
source source.me

#Run Braker

braker.pl --genome=$(pwd)/genome.fa --prot_seq=$(pwd)/proteins.fa --
softmasking
```

Schedule the job with the following sbatch command.

```
sbatch script.sbatch
```

All the processed files will be generated in the same directory as the sbatch 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>