GeneMark on HPC

What is GeneMark?

GeneMark developed in 1993 was the first gene finding method recognized as an efficient and accurate tool for genome projects. GeneMark was used for annotation of the first completely sequenced bacteria, *Haemophilus influenzae*, and the first completely sequenced archaea, *Methanococcus jannaschii*. The GeneMark algorithm uses species specific inhomogeneous Markov chain models of protein-coding DNA sequence as well as homogeneous Markov chain models of non- coding DNA. Parameters of the models are estimated from training sets of sequences of known type. The major step of the algorithm computes a posteriory probability of a sequence fragment to carry on a genetic code in one of six possible frames (including three frames in complementary DNA strand) or to be "non-coding". [Source]

Links:

Official Website

Documentation

Versions Available:

The following versions are available on the cluster:

- GeneMark-v4.38
- GeneMark-v4.6.8

How to load GeneMark?

To load GeneMark, use the following commands:

```
#Load the genemark module
module load bio/genemark/4_6_8
```

To verify if the module and dependencies are loaded correctly, use the following command.

```
#Show all the modules loaded
module list
```

This should list all the GeneMark and dependencies that are loaded. In this case, it should be miniconda and genemark itself.

How to get License?

Users must have license key to access this software, which is free. Follow these steps to get a license key file,

- Get to the following website: [Genemark Liscence](#)
- Set up credential and download GeneMark-ES/ET/EP ver 4.69_lic 64-bit license.
- Unzip the liscense file and copy it to user's home directory
- Rename file to .gm_key, with "." in front so that it will be hidden, and software will recognize it.

How to use GeneMark?

For this tutorial, copy the input data from the following directory,

```
# Copy the data in user's working directory
cp -r /share/apps/GeneMark_ESETEP_468/gmes_linux_64/GeneMark-E-
tests/GeneMark.hmm ~/test
# Go the the copied directory
cd ~/test/GeneMark.hmm/
```

The input files should be in the input directory. Use the following command to use genemark,

```
# Run genemark in a compute node
srun -p main --qos main -n 1 -c 16  --mem 16G gmhmme3  -o
genemark.gff3  -m ./input/athaliana.mod  -f gff3
./input/sequence.fasta
```

The processed file should be in the same directory- genemark.gff3. To see additional tutorial, please refer to the documentation.

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