EVidenceModeler (EVM) on HPC

What is EvidenceModeler (EVM)?

The EVidenceModeler (aka EVM) software combines ab intio gene predictions and protein and transcript alignments into weighted consensus gene structures. EVM provides a flexible and intuitive framework for combining diverse evidence types into a single automated gene structure annotation system.

Inputs to EVM include the genome sequence, gene predictions and alignment data in GFF3 format, and a list of numeric weight values to be applied to each type of evidence. The weights can be configured manually.[Source]

Links:

Official Website

Documentation

Versions Available:

The following versions are available on the cluster:

- Evidence Modeler v1.1.1

How to load EVM?

To load EVM, use the following commands:

```
#Load the  EVidenceModeler module
module load bio/evm/1.1.1:
```

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 software and dependencies that are loaded. In this case, 8 dependencies should be loaded.

How to use EVM?

For this tutorial, copy all the input file from following directory,

```
# Copy essential input file from installation directory to a test
#directory. 8 files should be copied over
cp /share/apps/evm/EVidenceModeler-1.1.1/simple_example/* ~/test/
```

To run EVM, paste the following new content in runMe.sh file which was copied over.

New Content of file should be :

```
#!/bin/sh


## Partitioning the Inputs

echo -e "Partitioning the Inputs files\n"

partition_EVM_inputs.pl --genome genome.fasta \
     --gene_predictions gene_predictions.gff3 --protein_alignments
protein_alignments.gff3 \
     --transcript_alignments transcript_alignments.gff3 \
     --segmentSize 100000 --overlapSize 10000 --partition_listing
partitions_list.out
```

```
##Generating the EVM Command Set

echo -e "Generating the EVM Command Set\n"

write_EVM_commands.pl --genome genome.fasta --weights
$(pwd)/weights.txt \
      --gene_predictions gene_predictions.gff3 --protein_alignments
protein_alignments.gff3 \
      --transcript_alignments transcript_alignments.gff3 \
      --output_file_name evm.out  --partitions partitions_list.out
>  commands.list

## Executing EVM Command

echo -e "Executing EVM Command\n"

execute_EVM_commands.pl commands.list | tee run.log



##Combining the Partitions

echo -e "Combining the Partitions\n"
recombine_EVM_partial_outputs.pl --partitions partitions_list.out --
output_file_name evm.out

##Convert to GFF3 Format
echo -e "Convert to GFF3 Format\n"

convert_EVM_outputs_to_GFF3.pl  --partitions partitions_list.out --
output evm.out  --genome genome.fasta

## Combining into single output file

echo -e "Combining into single output file\n"

find . -regex ".*evm.out.gff3" -exec cat {} \; > EVM.all.gff3


echo -e "Script Completed\n"
```

To understand the function of all the commands, please refer to the documentation of the software.

To run the script, simple use the following command,

```
bash runMe.sh
```

All the output files should be in the script's directory including the final output file – EVM.all.gff3.

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