MIRA on HPC cluster:

What is MIRA?

MIRA is a multi-pass DNA sequence data assembler/mapper for whole genome and EST/RNASeq projects. MIRA assembles/maps reads gained by

- electrophoresis sequencing (aka Sanger sequencing)
- Illumina (Solexa) sequencing
- Less so: 454 pyro-sequencing (GS20, FLX or Titanium)
- Less so: Ion Torrent

into contiguous sequences (called *contigs*). One can use the sequences of different sequencing technologies eithear in a single assembly run (a *true hybrid assembly*) or by mapping one type of data to an assembly of other sequencing type (a *semi-hybrid assembly (or mapping)*) or by mapping a data against consensus sequences of other assemblies (a *simple mapping*).

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

Documentation

GitHub

Versions Available:

- MIRA v4.0.2

How to load a version of MIRA?

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

```
module avail bio/mira/4.0.2
```

Verify by using this command:

```
module list     #Should only list bio/mira module
```

Since this is a standalone software, it does not have dependencies on other software.

How to use MIRA on the cluster?

There are two methods to run MIRA 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/mira

# Start your commands here
mira --help

#Follow with commands to execute here
```

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

The Script:

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

```
#Copy all the required files to the test folder
cp /share/apps/salmon/salmon-1.3.0_linux_x86_64/sample_data/reads_* .
```

```
# Make a MIRA config file to run mira
touch manifest.conf && nano manifest.conf
```

Template for config file can be found in MIRA docs. For this example, use the following template:

```
#Example for a manifest describing a genome de-novo assembly with
#Unpaired Illumina data

#First part: defining some basic things
#In this example, we just give a name to the assembly
# and tell MIRA it should map a genome in accurate mode

project = MyFirstAssembly
job = genome,denovo,accurate

# The second part defines the sequencing data MIRA should load and
assemble
# The data is logically divided into "readgroups"

# here comes the unpaired Illumina data

readgroup = reads_1.fastq
data =reads_2.fastq
technology = solexa
#SUPPRESS NETWORK STORAGE FILESYSTEM ERROR OR USE TEMP DIR IN CURRENT
#DIR
parameters= -NW:cnfs=warn
```

Now make a batch script with the following template,

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


module load bio/mira        # load the module


cd $SLURM_SUBMIT_DIR        # Moving to the test directory
```

```
#RUN MIRA
mira manifest.conf >&log_assembly.txt
```

Schedule the job with the following sbatch command.

sbatch script.sbatch

All the processed files, including log 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