Miniasm on HPC cluster:

What is Miniasm?

Miniasm is a fast OLC-based *de novo* assembler for noisy long reads. It takes all-vs-all read self-mappings (typically by minimap) as input and outputs an assembly graph in the GFA format. Different from mainstream assemblers, miniasm does not have a consensus step. It simply concatenates pieces of read sequences to generate the final unitig sequences. Thus, the per-base error rate is like the raw input reads.

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

GitHub

Versions Available:

- Miniasm –v0.3

How to load a version of Miniasm?

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

```
module avail bio/miniasm
```

The version will be listed. To use a version of software, use following command:

```
module load bio/miniasm/0.3
```

Verify by using this command:

```
module list
```

The loaded software and dependencies will be shown.

How to use Miniasm on the cluster?

There are two methods to run Miniasm 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/miniasm
# Start your commands here
Miniasm--help

#Follow with commands to execute
```

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,

```
# Download sample PacBio from the PBcR website
wget -O-
```

```
http://www.cbcb.umd.edu/software/PBcR/data/selfSampleData.tar.gz | tar
zxf -
ln -s selfSampleData/pacbio_filtered.fastq reads.fq

# Download E. coli K-12 sequence
wget -O NC_000913.fa
'http://www.ncbi.nlm.nih.gov/sviewer/viewer.cgi?sendto=on&db=nuccore&d
opt=fasta&val=556503834'

#Load minimap2
module load bio/minimap2

# Overlap
minimap2 -Sw5 -L100 -m0 -t8 reads.fq reads.fq | gzip -1 > reads.paf.gz
```

Use the following template for the script,

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


module load bio/miniasm         # load the module



cd $SLURM_SUBMIT_DIR         # Moving to the test directory



#Run miniasm on the cluster
miniasm -f reads.fq reads.paf.gz > utg.gfa
#Convert in to FASTA
awk '/^S/{print ">"$2"\n"$3}' utg.gfa > utg.fa
#Map assembly to ref
minimap NC_000913.fa utg.fa | minidot - > utg.eps
```

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