MaSuRCA on HPC:

What is MaSuRCA?

MaSuRCA stands for Maryland Super Read Cabog Assembler, which is a Genome Assembly and Analysis Toolkit. It has tools loke MaSuRCA genome assembler, QuQRUM error corrector for Illumine data, POLCA genome polishing software, chromosome scaffolder, and so on. It also merges the benefits of Debruijn graph and Overlap-Layout-Consensus assembly approach.

The documentation for the MaSuRCA is found on:

https://github.com/alekseyzimin/masurca

https://academic.oup.com/bioinformatics/article/29/21/2669/195975

Versions Available:

MaSuRCA – 3.3.4

MaSuRCA – 3.3.5

MaSuRCA – 4.0.3

MaSuRCA – 4.0.5

How to load a version of MaSuRCA?

To view readily available builds of Samtools on the HPC, use the following command:

module avail bio/masurca

You will see the versions listed. Select one which fits your project. For this example, I am loading:

module load bio/masurca /4.0.5

Verify by using this command:

module list

This will show all modules loaded and all the dependencies required to run.

How to use MaSuRCA on the cluster?

Here is one example of how to use the MaSuRCA on the cluster.

```
cd ~ && mkdir masurca_test # Make a test directory on home folder


cd masurca


wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR848/SRR848976/SRR848976_1.fastq.gz     #download the data
wget ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR848/SRR848976/SRR848976_2.fastq.gz


gunzip SRR848976_1.fastq.gz SRR848976_2.fastq.gz
# Now after you load your module and download test files, make a script to run MaSuRC
A
```

The Script:

```
#!/bin/bash
$SBATCH -J Jobname # Jobname
#SBATCH -n 1       # Nodes per task
#SBATCH -p main    # Partition
#SBATCH --qos main  # Quality of service
```

```
#SBATCH -o MaSuRCA_sim_out-%J.txt # STDOUT Out file
#SBATCH -e MaSuRCA_sim_out-%J.txt # Error file



module load bio/masurca/4.0.5       # load the module

cd ~/masurca_test/                # Moving to the test directory

masurca -t 24 -i SRR848976_1.fastq,SRR848976_2.fastq

sh assemble.sh

# This will run the masurca on both Illumina sequencing fastq files.
```

Now, you can schedule the job with sbatch command.

sbatch myscript.sh

All the processed file will be written in a directory named after the test-subject.

cd ~/masurca_test/

You can do advance jobs with configurations file: input more data, use new functions. Plese refer to the GitHub page for additional information about masurca.

Where to find help?

If you are stuck on some part or need help at any point, please contact at the following address.

https://ua-app01.ua.edu/researchComputingPortal/public/oitHelp

Resources:

https://github.com/alekseyzimin/masurca