Nanosim on HPC cluster:

What is Nanosim?

NanoSim is a fast and scalable read simulator that captures the technology-specific features of ONT data, and allows for adjustments upon improvement of nanopore sequencing technology.

NanoSim (v2.5) can simulate ONT transcriptome reads (cDNA / direct RNA) as well as genomic reads. It also models feature of the library preparation protocols used, including intron retention (IR) events in cDNA and directRNA reads. Multiprocessing option allows for faster runtime for large library simulation.

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

GitHub

Versions Available:

- Nanosim –v2.0.0

How to load a version of Nanosim?

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

```
module avail bio/nanosim
```

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

```
module load bio/nanosim/2.0.0
```

Verify by using this command:

```
module list
```

The loaded software and dependencies will be shown.

How to use Nanosim on the cluster?

There are two methods to run Nanosim on the cluster.

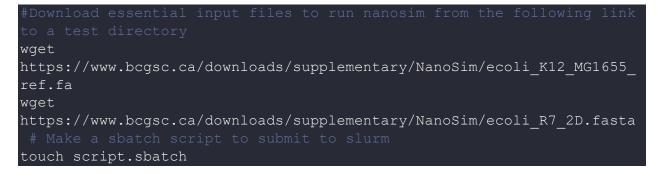### The Interactive Way:

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/nanosim
# Start your commands here
nanosim -h
#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 essential input files to run nanosim from the following link
to a test directory
wget
https://www.bcgsc.ca/downloads/supplementary/NanoSim/ecoli_K12_MG1655_
ref.fa
wget
https://www.bcgsc.ca/downloads/supplementary/NanoSim/ecoli_R7_2D.fasta
 # Make a sbatch script to submit to slurm
touch script.sbatch
```

Use the following template for the script,

```bash
#!/bin/bash
#SBATCH -p threaded
#SBATCH -q threaded
#SBATCH --mem-per-cpu=4G
#SBATCH -n 1
#SBATCH -c 16

#Load the module
module load bio/nanosim

#Go to the test directory
cd $SLURM_SUBIMT_DIR

# Run nanosim on the downloaded files
read_analysis.py -i ecoli_R73_2D.fasta  -r ecoli_K12_MG1655_ref.fa  -o
ecoli  -t 10
```

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