Nanopolish on HPC cluster:

What is Nanopolish?

Software package for signal-level analysis of Oxford Nanopore sequencing data. Nanopolish can calculate an improved consensus sequence for a draft genome assembly, detect base modifications, call SNPs and indels with respect to a reference genome and more (see Nanopolish modules, below).

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

GitHub

Full Documentation

Versions Available:

- Nanopolish –v 0.9.0

How to load a version of Nanopolish?

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

```
module avail bio/nanopolish
```

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

```
module load bio/nanopolish/0.9.0
```

Verify by using this command:

```
module list
```

The loaded software and dependencies will be shown.

How to use Nanopolish on the cluster?

There are two methods to run Nanopolish 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/nanopolish
# Start your commands here
nanopolish --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 essential input files to run Nanopolish
wget http://s3.climb.ac.uk/nanopolish_tutorial/ecoli_2kb_region.tar.gz
#Unzip the tarball
tar -xvf ecoli_2kb_region.tar.gz && cd ecoli_2kb_region

# Download ref genome
curl -o ref.fa
https://ftp.ncbi.nih.gov/genomes/archive/old_genbank/Bacteria/Escherichia_coli_K_12_substr__MG1655_uid225/U00096.ffn
```

```
# Make a sbatch script to submit to slurm
touch script.sbatch
```

Use the following template for the script,

```
#!/bin/bash
#SABTCH -J test
#SBATCH -n 1
#SBATCH -c 12
#SBATCH -mem=20G
#SBATCH -p main
#SBATCH --qos main


#Load the module

module load bio/nanopolish

#Go to the test directory
cd $SLURM_SUBIMT_DIR

# Run Nanopolish on the downloaded files
#Example taken from Documentation (example 3.1)

#Index the data
nanopolish index -d fast5_files/ reads.fasta
# Align to reference
minimap2 -ax map-ont -t 8 ref.fa reads.fasta | samtools sort -o reads-
ref.sorted.bam -T reads.tmp

samtools index reads-ref.sorted.bam

# Align the nanopore event to ref
nanopolish eventalign --reads reads.fasta --bam reads-ref.sorted.bam -
-genome ref.fa --scale-events > reads-ref.eventalign.txt
```

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