

NextPolish on HPC

What is NextPolish?

NextPolish is a software tool for correcting errors in genome sequences generated by long-read sequencing technologies, such as third-generation sequencing (TGS). It is specifically designed for correcting base errors (SNV/Indel) in the genome, and uses a stepwise approach that can be applied to long-read data alone or in combination with short-read data.

The software contains two core modules: a base-calling module and an error-correction module. The base-calling module uses machine-learning algorithms to assign the most likely nucleotide at each position of the raw long reads. The error-correction module then utilizes the base-calling results and the information from both long reads and short reads to correct errors in the genome.

Links:

[Official Website](#)

Versions Available:

The following versions are available on the cluster:

- nextPolish v1.3.1

How to load NextPolish?

To load NEXTPOLISH, use the following commands:

```
#Load the NEXTPOLISH module
module load bio/nextpolish
```

To verify if the module and dependencies are loaded correctly, use the following command.

```
#Show all the modules loaded
module list
```

This should list all the NextPolish dependencies that are loaded – miniconda in this case.

How to use NextPolish?

See the command line arguments for the software using following syntax:

```
# Command Line Args
nextPolish -h
```

To run nextPolish, users must create a configuration file with all the parameters mentioned. See the [tutorial](#) for more info about configuration file.

For this demonstration, copy all the test files from following directory,

```
# Copy the files to test directory
cp -r /share/apps/NextPolish/NextPolish/test_data ~/test/
```

Go to the directory and edit following parameters in run.cfg,

```
parallel_jobs = 4
multithread_jobs = 8
```

Use the following slurm script to submit job to the scheduler,

```
#!/bin/bash
#SBATCH --job-name=nextpolish_run
#SBATCH --output=nextpolish_run.out
#SBATCH --error=nextpolish_run.err
#SBATCH --nodes=1
```

```
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=32
#SBATCH --mem=64GB
#SBATCH -p main
#SBATCH --qos main

# load NextPolish module
module load bio/nextpolish

nextPolish run.cfg
```

All the output file should be in same directory as the script and test files.

Where to find help?

If you are confused or need help at any point, please contact OIT at the following address.

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

