

Minimap2 on HPC cluster:

What is Minimap2?

Minimap2 is a versatile sequence alignment program that aligns DNA or mRNA sequences against a large reference database. Typical use cases include:

- Mapping PacBio or Oxford Nanopore genomic reads to the human genome;
- Finding overlaps between long reads with error rate up to 15%;
- Splice-aware alignment of PacBio Iso-Seq or Nanopore cDNA or Direct RNA reads against a reference genome;
- Aligning Illumina single- or paired-end reads;
- Assembly-to-assembly alignment;
- Full-genome alignment between two closely related species with divergence below 15%.

The full documentation for the Minimap2 is found in the following links:

[GitHub](#)

Versions Available:

- Minimap2- v2.10

How to load a version of Minimap2?

To load a version of Minimap2s on the HPC, use the following command:

```
module avail bio/minimap2
```

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

```
module load bio/minimap2/2.10
```

Verify by using this command:

```
module list
```

The loaded software and dependencies will be shown.

How to use Minimap2s on the cluster?

There are two methods to run Minimap2 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/minimap2/2.10

# start your commands here
minimap2 --help

#Follow with commands to execute
```

This method is ideal for 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 the input files required to run minimap
mkdir test && cd test

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

#Now all the test files are downloaded
# Make a sbatch script
touch script.sbatch
```

Use the following template for the script,

```
#!/bin/bash
#SBATCH -J Minimap2s-test
#SBATCH -n 1
#SBATCH -mem=16G
#SBATCH -p main
#SBATCH --qos main
#SBATCH -o Minimap2s_sim_out-%J.txt
#SBATCH -e Minimap2s_sim_out-%J.txt
#SBATCH -t 30

module load bio/Minimap2          # load the module

cd $SLURM_SUBMIT_DIR              # Moving to the test directory

#Run Minimap2
minimap2 -Sw5 -L100 -m0 -t8 reads.fq reads.fq | gzip -1 > reads.paf.gz
```

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>