

## ***PBSuite on HPC***

### **What is PBSuite?**

PBSuite is a software package for analyzing PacBio long-read DNA sequencing data. It includes tools for error correction, assembly, and annotation of genomes and transcriptomes, as well as tools for analyzing long-read sequencing data from other technologies such as Oxford Nanopore. It is designed to improve the accuracy and completeness of genome assemblies and to enable the discovery of previously inaccessible genomic regions.

Links:

[Official Website](#)

### **Versions Available:**

The following versions are available on the cluster:

- bio/pbsuite/15.8.4

### **How to load PBSuite?**

To load PBSuite, use the following commands:

```
#Load the PBSuite module  
module load bio/pbsuite/15.8.4
```

To verify if the module is loaded correctly, use the following command,

```
# List all the module loaded in the environment  
module list
```

In a fresh environment, this should show anaconda, pbioconda and PBSuite module loaded.

## How to use PBSuite?

Many tools are included in this software collection. Refer to the documentation for more information about all the command line tools.

Here are general outline of available command line tools,

- **pbmm2**: aligns PacBio reads to a reference genome
- **pbsv**: structural variant (SV) detection using PacBio long reads
- **pbalign**: align PacBio reads to a reference genome using the BLASR aligner
- **pbindex**: creates an index file for PacBio BAM files
- **pbmm2-make-library-metadata**: creates a library metadata file for **pbmm2**
- **pbmm2-build**: builds a reference genome index for **pbmm2**
- **pbh5tools**: tools for working with PacBio HDF5 files
- **pbvalidate**: validation of PacBio BAM files

To run a program in the cluster, use the following slurm script,

```
#!/bin/bash
#SBATCH --job-name=pbmm2_align
#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=16
#SBATCH --mem=64GB
#SBATCH --time=12:00:00
#SBATCH --output=pbmm2_align.log
#SBATCH --p main
#SBATCH --qos main

# Load the module
module load bio/pbsuite/15.8.4

pbmm2 align \
```

```
--sort \  
--preset CCS \  
--threads $SLURM_CPUS_PER_TASK \  
--fofn input_files.fofn \  
-j $SLURM_CPUS_PER_TASK \  
-o output.bam
```

This script requests 1 node with 16 CPU cores and 64GB of memory and runs for 12 hours. The PacBio software is loaded using the **module load bio/pbsuite/15.8.4** command, and the **pbmm2 align** command is run with the specified options. The input files are specified in the fofn file, and the output is a bam file.

Users may need to adjust the resources requested depending on your specific needs and the size of your dataset.

### ***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>



