

## ***Meraculous on HPC***

### **What is Meraculous?**

Meraculous is a software tool that is used to assemble DNA sequences from Next Generation Sequencing (NGS) data. It is specifically designed to work with large genomes, such as those of plants and animals. Meraculous takes raw NGS data as input and uses algorithms to piece together the complete DNA sequence of an organism's genome. This can be useful for a variety of applications in biology and medicine, including identifying genetic mutations, studying evolution, and developing new treatments for genetic diseases.

**IMPORTANT:** Meraculous-2D has been superseded by the HipMer assembler, available here: <https://sourceforge.net/projects/hipmer/>

Links:

[Official Website](#)

[Manual](#)

### **Versions Available:**

The following versions are available on the cluster:

- Meraculous v2.2.2.4

### **How to load MERACULOUS?**

To load MERACULOUS, use the following commands:

```
#Load the MERACULOUS module
module load bio/meraculous/2.2.2.4
```

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 MERACULOUS dependencies that are loaded – perl in this case.

## How to use MERACULOUS?

Here are all the command line arguments to,

```
# See all command line args
meraculous.pl --help
```

To use meraculous, follow the following steps:

- Prepare your input data. This will typically involve generating NGS data using a sequencer and then processing it to create a set of reads that can be input into Meraculous.
- Configure the parameters for the genome assembly. This may include specifying the type of input data (e.g., Illumina, PacBio), the size of the genome, and any additional options that are relevant to your specific assembly.
- Submit the genome assembly job to the HPC cluster. This will typically involve creating a script that specifies the input data, the desired output location, and the necessary resources (e.g., number of cores, amount of memory).
- Once the assembly is complete, you can use tools provided by Meraculous or other software to analyze the resulting genome assembly and extract meaningful insights from the data.

A sample config is located at  
/share/apps/meraculous/etc/meraculous/pipeline/meraculous.config

Here is a sample job submission script:

```
#!/bin/bash
#SBATCH --job-name=my_assembly
#SBATCH --output=my_assembly.out
#SBATCH --error=my_assembly.err
#SBATCH --nodes=1
#SBATCH --ntasks=8
#SBATCH --mem=32GB
#SBATCH --time=24:00:00
#SBATCH --p main
#SBACTH --qos main
# Load the required module
module load bio/meraculous/2.2.2.4

# Run the assembly
meraculous -c config.file
```

### ***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>

