scMSI Installation Guide

Version: 1.0
Date Compiled: September 2024

Document Overview

This installation guide outlines an protocol for installing scMSI, a high-performance bioinformatics tool used to estimate microsatellite instability in complex genomic data sets. This guide is designed for bioinformatics professionals managing large HPC clusters with high I/O demands.

Installation Prerequisites

- Operating System: CentOS 8.4, Ubuntu 20.04, or compatible HPC system
- Minimum 512 GB RAM
- Python 3.7 environment with Intel MKL optimizations
- Required Python Libraries: numpy, scipy, pandas, scikit-learn, matplotlib
- BWA 0.7.17, Samtools 1.15.1, Bedtools 2.30.0
- Gurobi Optimizer 9.1.2 (commercial license required)

Software Acquisition

Obtain the scMSI codebase from the GitHub repository and clone it to your HPC system:

```
git clone https://github.com/SeqAnalysis/scMSI.git
cd scMSI
```

Installation Procedure

1. **System Configuration**
   Ensure the HPC system has Intel MKL installed for matrix optimizations. Install essential development tools:
   ```
   sudo yum groupinstall "Development Tools"
   sudo apt-get install libopenblas-dev liblapack-dev -y
   ```

2. **Set Up Python Environment**

Create a Conda environment with the necessary libraries:
```
conda create -n scMSI python=3.7 --strict-channel-priority
conda activate scMSI
```

3. **Install Python Libraries**
Install required libraries with HPC optimizations:
```
conda install numpy==1.21.0 scipy==1.7.3 pandas==1.3.2 scikit-learn==1.0.2 matplotlib==3.4.2 -c intel
```

4. **Install Additional Tools**
Install bwa, samtools, bedtools with multi-threading:
```
module load BWA/0.7.17
module load SAMtools/1.15.1
```

5. **Configure Gurobi**
Install and configure Gurobi with your HPC license:
```
conda install gurobi
gurobi_cl --license
```

Installation Verification

To verify installation, execute the following command and ensure the version information is returned correctly:

```
python main.py --version
```

Additionally, you can verify Gurobi installation by solving a test problem using the Gurobi optimizer:

```
gurobi_cl ResultFile=test_result.sol scMSI_test_problem.lp
```