ChloroMiner: Advancing Chloroplast Genome Research through Computational Extraction and Assembly from Whole Genome Data

1. Introduction

"ChloroMiner" the chloroplast genome assembly Pipeline is a user-friendly tool designed to facilitate the assembly of chloroplast genomes from the raw sequencing data. This manual provides a step-by-step guide on how to use the pipeline effectively. Please ensure that you have the following dependencies installed before proceeding:

- Python (version 3 or higher)

  sudo apt-get update

  sudo apt-get install python3.6

  For more details: https://docs.python-guide.org/starting/install3/linux/

  After installing Python install the required Pandas library by using following commands:

  sudo apt-get install python3-pip

  sudo -H pip3 install panda

- Trimmomatic

  Download the Trimmomatic tool using

  git clone --recursive https://git.launchpad.net/ubuntu/+source/trimmomatic/log/?h=ubuntu/focal

  Then install the trimmomatic using these commands

  sudo apt-get update

  sudo apt-get install trimmomatic

2. Installation

To install the ChloroMiner, follow these steps:

Step 1: Download the pipeline package from the github using https://github.com/ICAR-BIOINFORMATICS/ChloroMiner.git

Step 2: Downloaded python scripts to a desired location on your computer.

The ChloroMiner GitHub directory contains two files (Figure 1), named automation_script.py and c_miner.py. Download both the files and paste into the folder where you want to perform the task.

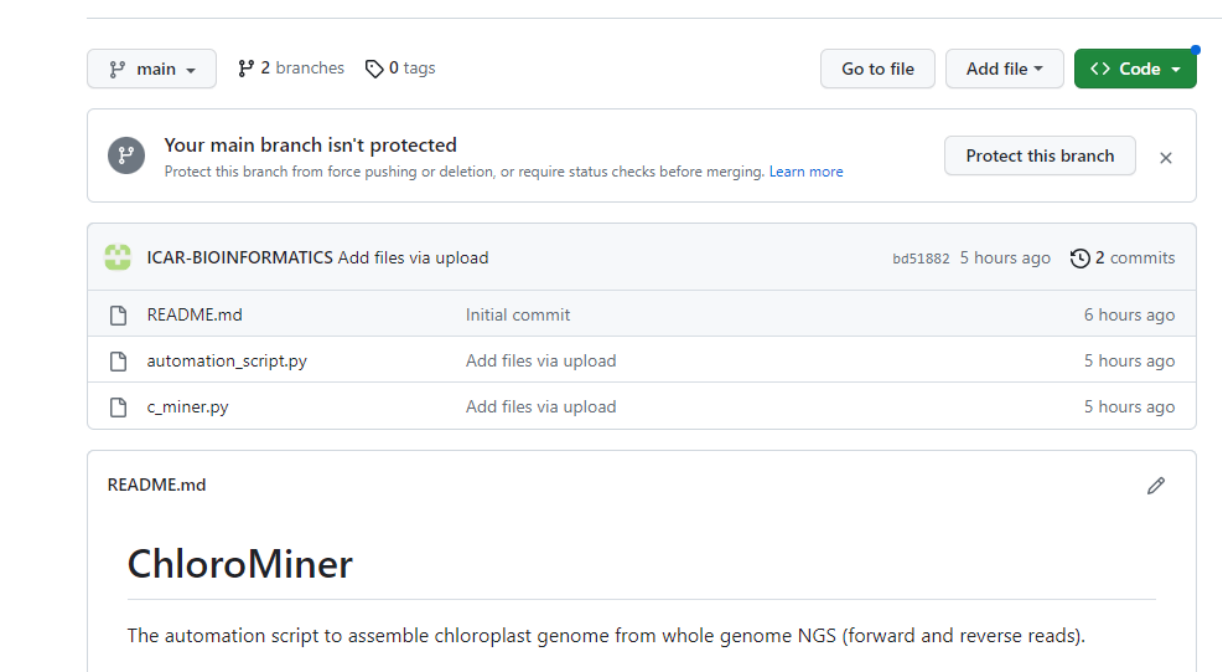

Figure1: GitHub page of ChloroMiner

3. Preparing the Input Data Before running the pipeline, make sure you have the following input data ready:

- Raw sequencing reads in farstq.gz format.

Note: Suppose you want to assemble the chloroplast genome from two files (Figure 2). Keep both the whole genome raw data file in the folder with both the python scripts

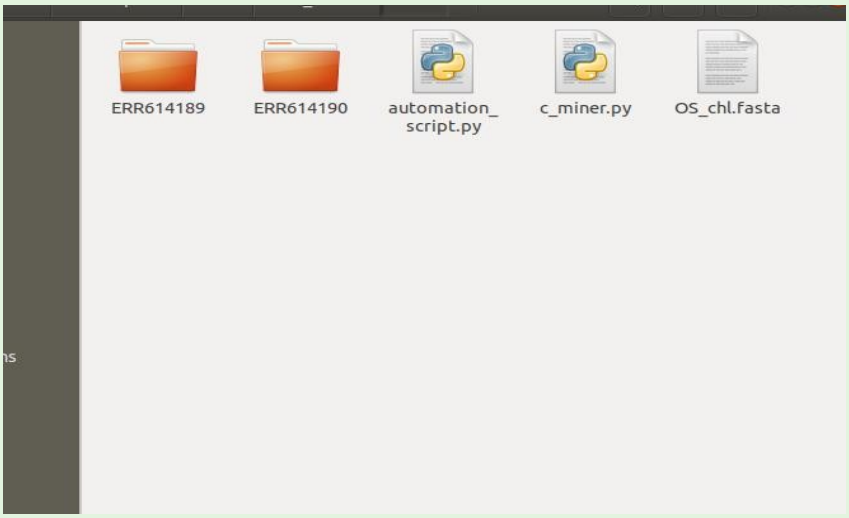

Figure 2: Working folder with raw data and python scripts

4. Running the Pipeline Follow these steps to run the ChloroMiner:

Step 1: Open a terminal or command prompt on your computer (Ctrl+Alt+T).Step 2: Navigate to the directory where the pipeline is installed.

Step 3: Execute the pipeline script using the following command:

python3 automatioon_script.py

python3 automatioon_script.py -adap <Trimomatic_adapter_path>
-ref <reference_chloroplast_genome_file>
-threads <no of threads >

Here

- -adap is the path of adapters which downloaded under the Trimmomatric folder

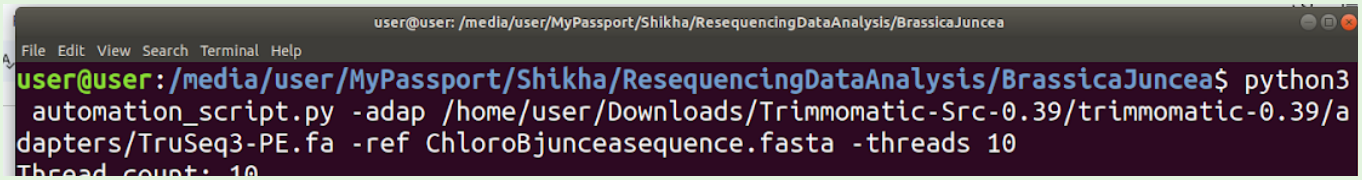

Figure 3: ChloroMiner Commands

- ref is the reference chloroplast genome (fasta file) which you would like to take as a reference
- -threads represent the no of threads can be adjusted as per the computer system's specification

Note : python script, adapter path and reference genome are space separated

5. Monitoring the Progress During the execution of the pipeline, you will see the progress displayed on the terminal or command prompt. This will include information about the trimming and assembly steps being performed.
6. Obtaining the Output Once the pipeline completes the assembly process, you can find the output files in the working directory. The output typically includes the assembled chloroplast genome in FASTA format, along with cleaned/trimmed read files.
7. Troubleshooting If you encounter any issues while running the pipeline, consider the following troubleshooting steps:
    - Ensure that you have the correct versions of Python and Trimmomatic installed.
    - Verify that the input reads file is in the correct format (fq.gz).
    - Double-check the command used to run the pipeline, ensuring that the input file and reference file paths are accurate and in the correct sequence (adapter path followed by reference genome file).

8. Conclusion

The ChloroMiner pipeline, with Python and Trimmomatic as its dependencies, provides an efficient and user-friendly solution for assembling chloroplast genomes from sequencing data. By following this user manual, you can successfully run the pipeline and obtain the assembled chloroplast genome for further analysis.

For any further assistance or inquiries, please refer to the official documentation or contact our support team.

Developed by:

Dr. Samarth Godara, Scientist, Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute (IASRI), Library Avenue, Pusa, New Delhi-110012 For any query, contact: samarth.godara@gmail.com.

Dr. Shbana Begam, Scientist, ICAR-National Institute for Plant Biotechnology, New Delhi-110012