

MosMAP is a fully integrated, user-friendly pipeline designed for the analysis of mosquito metagenomes, enabling high-throughput metagenomic analysis without the need for advanced bioinformatics expertise. The pipeline automates key steps in the metagenomic workflow, including quality control, taxonomic classification, species abundance estimation, and data visualization.

MosMAP automates essential tasks including:

- Quality control of sequencing data.
- Taxonomic classification of sequences.
- Viral species abundance estimation.
- Interactive data visualization.

MosMAP integrates several widely-used bioinformatics tools such as:

- **Trimgalore** for quality filtering.
- **Kraken2** for taxonomic classification.
- **Bracken** for abundance estimation.
- **Krona** for interactive visualization.

MosMAP is designed to be simple to use and is suitable for non-expert users in various fields of research, particularly in entomological studies and mosquito-borne viral surveillance.

Key Features

- Fully automated workflow for mosquito metagenome analysis.
- Seamless integration of multiple bioinformatics tools.
- Minimal system prerequisites and dependencies.
- High-quality, reproducible results.
- Supports interactive taxonomic data visualization using Krona.
- Simple to install and use with just a list of input FASTQ files.

System Requirements

- Unix-based operating system (Linux/macOS).
- Anaconda or Miniconda (for managing dependencies).
- Approximately 8GB of RAM for processing.

Installation

Step 1: Clone the Repository

Download the *MosMAP* repository from GitHub to your local system:

Step 2: Install Dependencies

MosMAP includes an installation script (`install.sh`) that automatically installs all the required dependencies. You need to have Anaconda installed beforehand.

If Anaconda is not installed, install it using:

```
wget https://repo.anaconda.com/archive/Anaconda3-2022.05-Linux-x86_64.sh
```

And after download is finished do:

```
bash Anaconda3-2022.05-Linux-x86_64.sh
```

Run the following command to install dependencies:

```
bash install.sh
```

This will set up all required tools, including **Trimgalore**, **Kraken2**, **Bracken**, and **Krona**.

Step 3: Prepare Input Files

To use *MosMAP*, you need a list of input FASTQ files (name it as `list.txt`). Place this file in the same directory as the *MosMAP* folder, and ensure it contains names of your sequencing data files.

You should also specify the path to your Anaconda installation when asked while running the pipeline, which will be used by *MosMAP* to configure the environment. The path is usually `/home/username/anaconda3/bin/anaconda` but you can get the path by the command:

```
which anaconda
```

Usage

Running the Pipeline

Once the dependencies are installed and the input files are prepared, you can run the entire metagenomic analysis workflow by executing the `pipeline.sh` script:

```
bash pipeline_unzipped.sh (if your FASTQ files are not compressed)
bash pipeline_zipped.sh   (if your FASTQ files are compressed)
```

This will process your FASTQ files, applying quality control, taxonomic classification, abundance estimation, and generating visualization files.

Output Files

- **QC:** QC output stored in QC folder
- **Taxonomic Classification:** Kraken2 output stored in `Kraken_result` folder.
- **Abundance Estimation:** Bracken output stored in `Bracken_result` folder.
- **Visualization:** Interactive Krona visualization saved as `krona_output.html`.

- **Combined bracken output:** Stored in combined_bracken file