BIOINFORMATION
Discovery at the interface of physical and biological sciences

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 21(2)

Research Article

Received February 1, 2025; Revised February 28, 2025; Accepted February 28, 2025, Published February 28, 2025

DOI: 10.6026/973206300210110

SJIF 2025 (Scientific Journal Impact Factor for 2025) = 8.478
2022 Impact Factor (2023 Clarivate Inc. release) is 1.9

Declaration on Publication Ethics:
The author's state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

Declaration on official E-mail:
The corresponding author declares that lifetime official e-mail from their institution is not available for all authors

License statement:
This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Comments from readers:
Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article without open access charges. Comments should be concise, coherent and critical in less than 1000 words.

Disclaimer:
Bioinformation provides a platform for scholarly communication of data and information to create knowledge in the Biological/Biomedical domain after adequate peer/editorial reviews and editing entertaining revisions where required. The views and opinions expressed are those of the author(s) and do not reflect the views or opinions of Bioinformation and (or) its publisher Biomedical Informatics. Biomedical Informatics remains neutral and allows authors to specify their address and affiliation details including territory where required.

Edited by P Kangueane
Citation: Kulsum *et al*. Bioinformation 21(2): 110-112 (2025)

MOSMAP: Mosquito metagenome analysis pipeline

Umay Kulsum#, Chitra Patankar# & Debasis Biswas*

Department of Microbiology, All India Institute of Medical Sciences, Bhopal, India; #Equal Contribution for first author; *Corresponding author

Affiliation URL:
https://aiimsbhopal.edu.in/

Author contacts:
Umay Kulsum - E - mail: umay.project@aiimsbhopal.edu.in
Chitra Patankar - E - mail: chitra.phd2019@aiimsbhopal.edu.in
Debasis Biswas - E - mail: debasis.microbiology@aiimsbhopal.edu.in; Phone: +91 7747010312

Abstract:

MosMAP is a bioinformatics pipeline designed for mosquito metagenome analysis. MosMAP automates essential processes like quality control, taxonomic classification, species abundance estimation and visualization by integrating tools such as Trimgalore, Kraken2, Bracken and Krona into a user-friendly workflow. Each of these tools is integrated to ensure a smooth and efficient workflow from raw data to interpretable results. The pipeline simplifies complex bioinformatics tasks, making them accessible to researchers with limited computational expertise. MosMAP demonstrated high concordance with standard bioinformatics workflows such as Kraken and Bracken in terms of read retention, taxonomic accuracy and abundance estimation when applied to metagenomes of mosquito collected in Bhopal, India. This accessible pipeline promotes the simplification of meta-genomics, supporting research in microbiology, ecology and vector-borne diseases.

Keywords: Mosquito metagenome analysis (MosMAP), bioinformatics pipeline, taxonomic classification, species abundance estimation

Background:

Metagenomic analysis has become an indispensable tool for exploring the diversity of microbial and viral communities in environmental, clinical and biological samples **[1]**. The advent of Next-Generation Sequencing (NGS) technologies has significantly enhanced our knowledge of genomic data, providing deeper insights into the structure and function of microbial and viral populations **[2]**. However, the process of analysing metagenomic data is often complicated by the large volume of sequencing reads, the diversity of organisms present and the necessity for multiple bioinformatics tools to process, classify and visualize the data **[3]**.

Tools like *SqueezeMeta* and *MOCAT2* have attempted to ease Metagenomic workflows that often require manual configuration of multiple bioinformatics tools, making the process time-consuming and technically demanding **[4, 5]**. Therefore, it is of interest to present MosMAP, an integrated and automated pipeline designed to streamline mosquito metagenome analysis. MosMAP combines several established bioinformatics tools into a single, user-friendly shell script that performs essential tasks such as sequence quality filtering, taxonomic classification, viral species abundance estimation and interactive data visualization without the need for extensive computational expertise.

Input:

MosMAP accepts raw sequencing data in FASTQ format, either compressed or uncompressed. The pipeline requires a simple text file (list.txt) containing the names of the input FASTQ files. Users must install all the prerequisites including the MiniKraken database for taxonomic classification by running the *install.sh* script and then execute *pipeline.sh* to process and analyse the data. Additionally, the user must specify the path to their Anaconda installation, as several of the tools utilized in MosMAP are distributed via Anaconda. MosMAP is hosted on Git Hub, with detailed instructions available in the README file. The minimal hardware requirement for running MosMAP is 8 GB of RAM, making it suitable for most standard computational setups.

Output:

MosMAP generates a series of outputs, including:

[1] Quality control reports: Trimgalore and FastQC outputs that summarize read quality and trimming statistics.
[2] Taxonomic classification reports: Kraken2 outputs detailing taxonomic assignments of sequencing reads.
[3] Abundance estimation: Bracken-generated tables that provide taxonomic composition.
[4] Interactive visualization: Krona plots in HTML format that allow users to explore taxonomic data interactively.
[5] Combined report: A combined report summarizing the abundance estimations across all samples is produced.

Performance highlights:

Tools like Trimgalore ensure high-quality sequence data by trimming low-quality bases and adapter sequences **[6],** while FastQC provides comprehensive quality control reports **[7]**. The effectiveness of these pre-processing steps in MosMAP was evaluated by comparing read retention rates after quality control. The results showed a high level of concordance between MosMAP and the standard tool, indicating that MosMAP's pre-processing methods effectively retained high-quality reads while removing low-quality sequences and adapters. Kraken, integrated within MosMAP, was used to perform taxonomic classification of the metagenome reads **[8]**. The classification results were compared with those obtained from Kraken run individually, revealing a strong agreement in the identification of taxa. This consistency underscores the reliability of MosMAP for taxonomic identification in mosquito metagenomes.

The relative abundance of taxa estimated by MosMAP was compared to estimates from standard pipeline Bracken, which complements Kraken by refining abundance estimates and providing detailed insights into the microbial community structure **[9]**. This comparison showed that MosMAP's abundance estimation was highly consistent with those from Bracken run individually, providing accurate and reliable abundance profiles for the mosquito metagenome samples. The integration of Krona for visualization facilitates the exploration of metagenomic data through interactive charts **[10]**. By developing a bash script that integrates all these standard tools, we have demonstrated that MosMAP can generate results that are highly consistent with those obtained from running the tools individually. MosMAP, thus, advances scientific research on mosquito Metagenomic analysis by consolidating multiple bioinformatics tools into a single, stream-lined, automated workflow that ensures high concordance with standard methodologies and significantly reduces analysis time and technical barriers. Furthermore, its potential for integration with expanded taxonomic databases and functional annotation tools makes it a scalable resource for facilitating large microbial and viral surveillance studies in mosquito populations and thus contributing to research on vector-borne diseases.

Caveats:
Despite being impactful in improving accessibility to bioinformatics analysis, we understand the current version of MosMAP has sufficient scope for further improvement. Running the analysis on the standard Kraken database instead of MiniKraken, will enable the discovery of novel genetic variants and emerging viral strains. Secondly, the functional annotation of the identified genes, if enabled in this pipeline will further enrich the knowledge output generated from the Metagenomic analysis, providing deeper insights and advancing our understanding in this field.

Future development:
The future development of MosMAP will focus on incorporating the standard Kraken database to improve taxonomic classification accuracy and the identification of novel species. Additionally, efforts will be made toward developing a user-friendly graphical user interface (GUI) to make the pipeline even more accessible to researchers without command-line experience. These developments will ensure that MosMAP remains a robust and versatile tool for the metagenomics community.

Conclusion:
MosMAP is a user-friendly bioinformatics pipeline that streamlines mosquito metagenome analysis by integrating essential tools for quality control, taxonomic classification, abundance estimation and visualization. By enhancing accessibility and efficiency, MosMAP has the potential to advance research in microbial ecology, vector-borne diseases and environmental microbiology.

Availability and requirements:

Project home page:
https://github.com/Biomedinformatics/MosMAP

Operating system(s): Linux

Programming language: Bash script

Other requirements: Anaconda

Ethics approval and consent to participate: Not Applicable

Consent for publication: Not Applicable

Availability of data and materials: Not Applicable

Competing interests: The author(s) declare no competing interests.

Funding: None

Authors' contributions:
The conceptualization of the study was carried out by UK, CP and DB. UK developed the methodology, while data curation was handled by CP and DB. The formal analysis was conducted by UK, CP and DB. The original draft of the manuscript was written by UK and DB and CP and DB reviewed and edited the draft. The project was administered and supervised by DB.

Acknowledgements:
Not Applicable

References:

[1] Riesenfeld CS *et al. Annu Rev Genet*. 2004 **38**:525. [PMID: 15568985]
[2] Mardis ER. *Annu Rev Genomics Hum Genet*. 2008 **9**:387. [PMID: 18576944].
[3] Thomas T *et al. Microb Inform Exp*. 2012 **2**:3. [PMID: 22587947].
[4] Tamames J & Puente-Sánchez F. *Front Microbiol*. 2018 **9**:3349. [PMID: 30733714].
[5] Kultima JR *et al. PLoS One*. 2012 **7**:e47656. [PMID: 23082188].
[6] Krueger F. Babraham Bioinformatics, *Babraham Institute*; 2015. [https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/]
[7] Andrews S. Babraham Bioinformatics, *Babraham Institute*; 2010. [https://www.bioinformatics.babraham.ac.uk/projects/fastqc/]
[8] Wood DE & Salzberg SL. *Genome Biol*. 2014 **15**:R46. [PMID: 24580807].
[9] Lu J *et al. PeerJ Comput Sci*. 2017 **3**:e104. [DOI: 10.7717/peerj-cs.104]
[10] Ondov BD *et al. BMC Bioinformatics*. 2011 **12**:385. [PMID: 21961884].