BMC Genomic Data

DATA NOTE

Open Access

Shotgun metagenomic dataset of leaf endophytic microbiome of the garden sage (*Salvia officinalis* L.)

Mouliraj Palanisamy[1], Olubukola Oluranti Babalola[2] and Sathishkumar Ramalingam[1*]

Abstract

Objectives Garden sage (*Salvia officinalis* L.) is a traditional medicinal plant known for its rich bioactive secondary metabolites. However, there is limited information about the diversity of endophytic microbial communities, including bacteria, fungi, archaea, and viruses. Therefore, the study employs shotgun metagenomics to generate and make publicly available a dataset representing the leaf endophytic microbiome of *Salvia officinalis*.

Data description Metagenomic DNA was extracted from leaves of *S. officinalis* collected as three biological replicates and sequenced using the Illumina NovaSeq X platform. Host-derived and contaminant sequences were removed by mapping reads to the *S. officinalis* reference genome using BWA-MEM. The resulting high-quality FASTQ files were analyzed to characterize the taxonomic composition of the endophytic microbiome using Kraken2-based classification.

Keywords *Salvia officinalis*, Endophytic microbiome, Shotgun metagenome sequencing, Taxonomic profiling, Plant-microbe interactions

Objective

Garden sage (*Salvia officinalis* L.) belongs to the family Lamiaceae. It is a traditional medicinal plant known for its rich bioactive secondary metabolites. It is native to the Mediterranean region (the Dalmatian region of Yugoslavia) and naturalized in various parts of the world [1]. In India, it is not widespread but is sparingly cultivated at high altitudes in the Himalayas (Jammu, Himachal Pradesh, Uttarakhand) and the Western Ghats (Tamil Nadu, Karnataka, Kerala) [2, 3]. As it is a medicinal plant, the endophytic microbiome plays a significant role in plant growth and secondary metabolism [4, 5]. However, information on the microbial diversity and functional insights into the endophytic microbes of garden sage is limited. To our knowledge, the number of studies on endophytic microbes on *S. officinalis* is very few (in single digits) [6, 7]. So far, most studies focus on its allied species, such as *Salvia miltiorrhiza* and *S. abrotanoides* [8]. To address the lack of information on microbial diversity and the functional genes involved in metabolite pathways, culture-dependent methods are often not sufficient. Therefore, this study presents the shotgun metagenome datasets of the endophytic microbiome inhabiting the leaves of garden sage. The data from this study provide insights into the diversity and abundance of bacteria, fungi, viruses, and archaea, as well as their functions. Such information enhances our understanding

*Correspondence:
Sathishkumar Ramalingam
rsathish@buc.edu.in
[1]Plant Genetic Engineering Laboratory, Department of Biotechnology, Bharathiar University, Coimbatore, Tamil Nadu 641046, India
[2]Food Security and Safety Focus Area, Faculty of Natural and Agricultural Sciences, North-West University, Mmabatho 2735, South Africa

of the diversity of endophytic microorganisms and their potential roles in plant growth promotion and secondary (specialized) metabolism in *S. officinalis*.

Data description

Sampling

The dataset comprises raw next-generation sequencing reads generated through whole-metagenome (shotgun) sequencing of DNA isolated from *Salvia officinalis* leaf tissues. Leaf samples were collected from 9-month-old healthy *S. officinalis* plants grown in the plant containment facility, Department of Biotechnology, Bharathiar University, Coimbatore, Tamil Nadu, India (11.042496° N latitude, 76.876982° E longitude). The plants were grown under consistent cultivation conditions throughout the experiment. Vermicompost was applied monthly, and plants were regularly irrigated to ensure optimal growth and minimize abiotic stress that could affect microbial community composition. For shotgun metagenomic analysis, three independent biological replicates, each from a different plant, were selected to strengthen the reliability of the study. From each plant, leaves representing three distinct developmental stages: young, intermediate, and fully matured leaves were collected. Leaves from the three stages of each plant were pooled to obtain a single composite sample per plant, yielding a total of 3 biological replicates ($n = 3$). Immediately after collection, samples were placed in sterile containers, transported to the laboratory on ice, and stored at 4 °C prior to metagenomic DNA extraction.

Metagenome DNA extraction, library construction, and sequencing

100 mg of plant leaf tissue was processed for metagenomic DNA extraction with the NucleoSpin Plant II Mini Kit (MACHEREY-NAGEL GmbH & Co. KG, Germany) according to the manufacturer's guidelines. Following isolation, DNA concentration was quantified using a Qubit 4 Fluorometer (Thermo Fisher Scientific, USA), and DNA quality was assessed by Agarose gel electrophoresis. Qualified DNA samples were used for Illumina library preparation with the TruSeq DNA PCR-Free Library Prep Kit (Illumina Inc., USA). Library construction involved DNA fragmentation, end repair, A-tailing, size selection, and ligation of Illumina-compatible indexed adapters. Concentration of the prepared library was determined using a QuantStudio Real-Time PCR System (Applied Biosystems, USA). After dilution to 1 ng/µl, library insert size was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., USA). Libraries meeting quality requirements, including an appropriate insert size of 300–350 bp and a concentration of more than 2 nM, were subjected to high-throughput paired-end sequencing on the Illumina NovaSeq X platform (Illumina Inc., USA) with a read length of 150 bp. High-throughput sequencing generated 48,406,856 reads for sample S1/R1, 64,038,698 reads for sample S2/R2, and 60,334,676 reads for sample S3/R3, with an average sequencing depth of 57,593,410 reads per sample across the three libraries, providing sufficient coverage for downstream metagenomic analyses.

Data processing

The reads obtained after sequencing were first QC checked with the tool FastQC v0.10.0 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The reads were then trimmed with Trimgalore v0.6.10 [9] using default parameters to remove adapters. The clean reads were mapped against the *S. officinalis* reference genome (GenBank Accession Number: GCA_047404455.1) retrieved from the NCBI database using BWA-MEM v0.7.19 [10] to remove host contamination. The tool Kraken2 [11] was used to assign taxonomy and calculate relative abundances for unmapped reads using a customized reference database constructed for metagenomic analysis. Approximately 10% of the reads remained unclassified after Kraken2 taxonomic assignment. The Kraken classification reports were then converted to BIOM format using the KrakenTools v1.2.1 [12]. The Krona charts of the microbiome distribution were generated using Krona Tools v2.8.1 [13] with the Kraken report as input. The Sankey plots were generated using the Pavian R package v1.2.1 [14] with the same Kraken2 reports as input.

The metagenomic dataset was deposited in the SRA database at NCBI under the BioProject accession PRJNA1400257. The submitted metadata can be retrieved and downloaded via https://identifiers.org/ncbi/bioproject:PRJNA1400257.

Limitations

The dataset was generated from leaf samples collected at a single geographical location and time point, which may limit its representativeness across different environments or seasons. Leaves from multiple developmental stages were pooled for each biological replicate, restricting stage-specific resolution of the endophytic microbiome.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (file extension)	Data repository and identifier (DOI or accession number)
Data file 1	General information on dataset	Microsoft Excel file (.xlsx)	Mendeley Data (https://doi.org/10.17632/2yc77x-m6mv.1) [15]
Data file 2	Raw Sequence Reads of the Endophytic Microbiome of *Salvia officinalis* (Garden Sage) – S1/R1	FASTQ file (fastq.gz)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRR36746394) [16]
Data file 3	Raw Sequence Reads of the Endophytic Microbiome of *Salvia officinalis* (Garden Sage) – S2/R2	FASTQ file (fastq.gz)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRR36746393) [17]
Data file 4	Raw Sequence Reads of the Endophytic Microbiome of *Salvia officinalis* (Garden Sage) – S3/R3	FASTQ file (fastq.gz)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRR36746392) [18]
Data file 5	Overview of microbial community structure inferred from the dataset using Krona chart visualizations (S1, S2 & S3 or R1, R2 & R3)	HyperText Markup Language file (.html)	Mendeley Data (https://doi.org/10.17632/2yc77x-m6mv.1) [15]
Data file 6	Core endomicrobiome of *Salvia officinalis* leaves illustrated through Sankey plot analysis (S1, S2 & S3 or R1, R2 & R3)	Joint Photographic Experts Group (.jpg)	Mendeley Data (https://doi.org/10.17632/2yc77xm6mv.1) [15]

Abbreviations

DNA	Deoxyribonucleic Acid
QC	Quality Check
PCR	Polymerase Chain Reaction
NCBI	National Center for Biotechnology Information
R1/S1, R2/S2, & R3/S3	Represent biological replicates/samples of leaf tissue collected from healthy *Salvia officinalis* plants

Acknowledgements
M.P. gratefully acknowledges the fellowship support provided by the Chief Minister`s Research Grant (CMRG), Government of Tamil Nadu, India and the American Society for Microbiology (ASM), United States, for the mentorship support through the ASM Future Leaders Mentorship Fellowship (ASM-FLMF-2024) (ASM Member ID: 200385040). The authors also thank Dr. Kesavan Markkandan of Oneomics Private Limited, Tamil Nadu, India, for valuable support with data analysis.

Author contributions
MP: Conceptualization, Investigation, Data Curation, Software Validation, Formal Analysis, Writing; Original Draft, Writing; Review, and Editing. OOB: Proofread Original Draft, Review, and Editing. SR: Project Administration, Supervision, Resource Acquisition, Proofread Original Draft, Editing, and Review.

Funding
This work was supported under the "Student-Led Project" scheme of the Chief Minister`s Research Grant (CMRG), Directorate of Technical Education (DoTE), Department of Higher Education, Government of Tamil Nadu, India (Sanction Order No. CMRG/3005/H3/2024/007; Project No. CMRG2400927), awarded to MP (Principal Investigator), SR (Co-Principal Investigator), and OOB (Collaborator).

Data availability
The data described in this Data note can be freely and openly accessed on the NCBI Sequence Read Archive under the BioProject accession number PRJNA1400257 (https://identifiers.org/ncbi/bioproject:PRJNA1400257). Supporting datasets and associated metadata are deposited in the Mendeley Data repository at https://doi.org/10.17632/2yc77xm6mv.1. Detailed information on sample identifiers, accession numbers, and corresponding datasets, along with direct access links, is provided in Table 1 and the cited references [15–18].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 30 January 2026 / Accepted: 24 April 2026
Published online: 06 May 2026

References

1. Ghorbani A, Esmaeilizadeh M. Pharmacological properties of *Salvia officinalis* and its components. J Traditional Complement Med. 2017;7:433–40. https://doi.org/10.1016/j.jtcme.2016.12.014.
2. Sankara Rao K, Raja KS, Deepak K, Arun Singh R, Gopalakrishna Bhat K. *Salvia officinalis* L. In: Flora of Peninsular India. https://indiaflora-ces.iisc.ac.in/FloraPeninsular/herbsheet.php?id=5134&cat=7. Accessed 10 Jan 2026.
3. Spices Board India. Sage (*Salvia officinalis*). Spices Board India, Ministry of Commerce & Industry, Government of India. https://www.indianspices.com/spice-catalog/sage.html. Accessed 10 Jan 2026.
4. Tsipinana S, Husseiny S, Alayande KA, Raslan M, Amoo S, Adeleke R. Contribution of endophytes towards improving plant bioactive metabolites: a rescue option against red-taping of medicinal plants. Front Plant Sci. 2023;14. https://doi.org/10.3389/fpls.2023.1248319.
5. Singh M, Naskar A, Rupashree A, Rajput M, Singh VK. Analysis of endophytic microbes harboring in medicinal plants of Himalayan region with their medicinal properties. Biocatal Agric Biotechnol. 2023;53:102857. https://doi.org/10.1016/j.bcab.2023.102857.
6. Debbab A, Hassan Aly A, Angelie Edrada-Ebel R, Müller EG, Mosaddak W, Hakiki M. A, Bioactive secondary metabolites from the endophytic fungus *Chaetomium* sp. isolated from *Salvia officinalis* growing in Morocco. Biotechnol Agron Soc Environ. 2009.
7. Schulz M, Knop M, Muellenborn C, Steiner U. Root-Associated Microorganisms Prevent Caffeine Accumulation in Shoots of *Salvia officinalis* L. Int J Agric Forestry. 2013;3:152–8.
8. Zimowska B, Bielecka M, Abramczyk B, Nicoletti R, Zimowska B, Bielecka M, et al. Bioactive Products from Endophytic Fungi of Sages (*Salvia* spp). Agriculture. 2020;10. https://doi.org/10.3390/agriculture10110543.
9. Krueger F, James F, Ewels P, Afyounian E, Weinstein M, Schuster-Boeckler B et al. FelixKrueger/TrimGalore: v0.6.10 - Add Default Decompression Path. 2023. https://doi.org/10.5281/zenodo.7598955
10. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26:589–95. https://doi.org/10.1093/bioinformatics/btp698.
11. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:257. https://doi.org/10.1186/s13059-019-1891-0.
12. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. Nat Protoc. 2022;17:2815–39. https://doi.org/10.1038/s41596-022-00738-y.
13. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. BMC Bioinformatics. 2011;12:385. https://doi.org/10.1186/1471-2105-12-385.

14. Breitwieser FP, Salzberg SL. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. Bioinformatics. 2020;36:1303–4. . https://doi.org/10.1093/bioinformatics/btz715
15. Palanisamy M, Babalola OO, Ramalingam S. General Information and Graphical Data Visualizations of the Leaf Endomicrobiome of Garden Sage (*Salvia officinalis* L.). 2026;1. https://doi.org/10.17632/2yc77xm6mv.1
16. NCBI Sequence Read Archive -. Endophytic Microbiome of *Salvia officinalis* (Garden Sage) - S1 (SRR36746394). https://identifiers.org/ncbi/insdc.sra:SRR36746394
17. NCBI Sequence Read Archive -. Endophytic Microbiome of *Salvia officinalis* (Garden Sage) - S2 (SRR36746393). https://identifiers.org/ncbi/insdc.sra:SRR36746393
18. NCBI Sequence Read Archive -. Endophytic Microbiome of *Salvia officinalis* (Garden Sage) - S3 (SRR36746392). https://identifiers.org/ncbi/insdc.sra:SRR36746392

Publisher's note