Data in Brief 60 (2025) 111544

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Endophytic bacteriome data of *Litchi chinensis* established by metagenomic 16S rRNA gene sequencing

Dinh Sy Nguyen[a,b], Dinh Minh Tran[a,*]

[a] *Institute of Biotechnology and Environment, Tay Nguyen University, Vietnam*
[b] *Faculty of Natural Science and Technology, Tay Nguyen University, Buon Ma Thuot, Dak Lak 630000, Vietnam*

ARTICLE INFO

Article history:
Received 26 February 2025
Revised 24 March 2025
Accepted 1 April 2025
Available online 10 April 2025

Dataset link: Data on endophytic microbiome of lychee (Litchi chinensis S.) (Original data)

Keywords:
Lychee
Endophytic bacteriome
Metagenomic 16S rRNA gene sequencing
Actinomycetota
Biosynthesis

ABSTRACT

This work reported the diversity profiling and predicted metabolic function of the endophytic bacteriome of lychee (*Litchi chinensis* S.) cultivated in Dak Lak Province of Vietnam for the first time. Roots of lychee were collected from three different fields in Krong Ana District in Dak Lak. 16S rRNA primers were used to sequence the metagenomic library. Kraken 2 was used to analyze the taxonomic distribution, while the MetaCyc database was used to predict the metabolic function. We identified 10 phyla, 14 classes, 27 orders, 30 families, and 27 genera of the endophytic bacteria from the sample. Actinomycetota was the most predominant phylum (84.49%), and biosynthesis was the bacteriome's primary function (75.42%). Data provided insight into the taxonomic distribution and metabolic function of lychee endophytic bacteria and might be helpful for the next steps concerning sustainable lychee cultivation using endophytic bacteria.

* Corresponding author.
E-mail address: tmdinh@ttn.edu.vn (D.M. Tran).

https://doi.org/10.1016/j.dib.2025.111544

Specifications Table

Subject	Biology
Specific subject area	Microbiology, Microbiome
Type of data	Raw (fastq.gz files), filtered, analyzed, and figures
Data collection	- Collection of the root sample of lychee (*Litchi chinensis* S.) cultivated in Dak Lak Province of Vietnam - Isolation of the metagenomic DNA from the sample - Preparation of the metagenomic library - Sequencing of the metagenomic library - Analysis of the metagenomic sequences
Data source location	Institution: Institute of Biotechnology and Environment, Tay Nguyen University District/Province/Country: Krong Ana/Dak Lak/ Vietnam Latitude and longitude coordinates for collected samples: 12°32′14′′N, 107°58′56′′E; 12°32′12′′N, 107°59′06′′E; 12°32′10′′N, 107°58′48′′E
Data accessibility	Raw sequences (fastq.gz files) Repository name: Mendeley Data Data identification number: doi: 10.17632/cfkpjd4322.2 Direct URL to data: https://data.mendeley.com/datasets/cfkpjd4322/2

1. Value of the Data

- The data provided insight into the taxonomic distribution and metabolic function of the endophytic bacteriome of lychee (*Litchi chinensis* S.) cultivated in Dak Lak, Vietnam.
- The data might be helpful in comparing the endophytic bacteriome of lychee and others.
- The data might be helpful for further work on applying endophytic bacteria for sustainable lychee cultivation.

2. Background

Lychee (*Litchi chinensis* S.) is one of the primary fruit crops of Vietnam and is cultivated throughout the country, including Dak Lak Province [1]. Chemical fertilizers were frequently used for lychee production in this province. However, it is clear that chemical fertilizers can reduce soil fertility and microorganisms [2]. Therefore, to produce sustainably lychee, endophytic microorganisms are thought to be the best solution [3–5]. Previously, we have established data on the endophytic microbiome of numerous crops, including coffee, black pepper, sugarcane, rice, banana, cashew, and citrus [6–13]; however, the endophytic microbial data of lychee are still unknown. Hence, the aim of this work was to establish data on the endophytic bacteriome of lychee (*Litchi chinensis* S.) grown in this province using the metagenomic 16S rRNA gene sequencing for subsequent experiments towards sustainable lychee cultivation by applying endophytic bacteria.

3. Data Description

This work obtained 271,080 reads filtered from 342,998 raw reads and used to analyze the taxonomic distribution and metabolic function. We found that 10 phyla, 14 classes, 27 orders, 30 families, and 27 genera of endophytic bacteria were identified from the sample. Actinomycetota was found to be the most abundant phylum among the phyla, with 84.49%. Actinobacteria was the most predominant class (82.49%), followed by Gammaproteobacteria (5.12%), Alphaproteobacteria (3.2%), Ktedonobacteria (2.08%), and Thermoleophilia (1.64%). Among bacteria belonging to identified orders, Corynebacteriales was abundant with 74.54%, followed by Frankiales (4.76), Burkholderiales (2.76%), Micrococcales (2.32%), Ktedonobacterales (2.08%), Rhizobiales (1.56%), and Solirubrobacterales (1.52%). Of the found families, *Mycobacteriaceae* was predominant (74.54%), followed by *Acidothermaceae* (4.76%), *Intrasporangiaceae* (2.24%), *Ktedonobacteraceae* (2.08%), *Burkholderiaceae* (2.48%), *Solirubrobacteraceae* (1.52%), and *Rhizobiaceae* (1.4%).

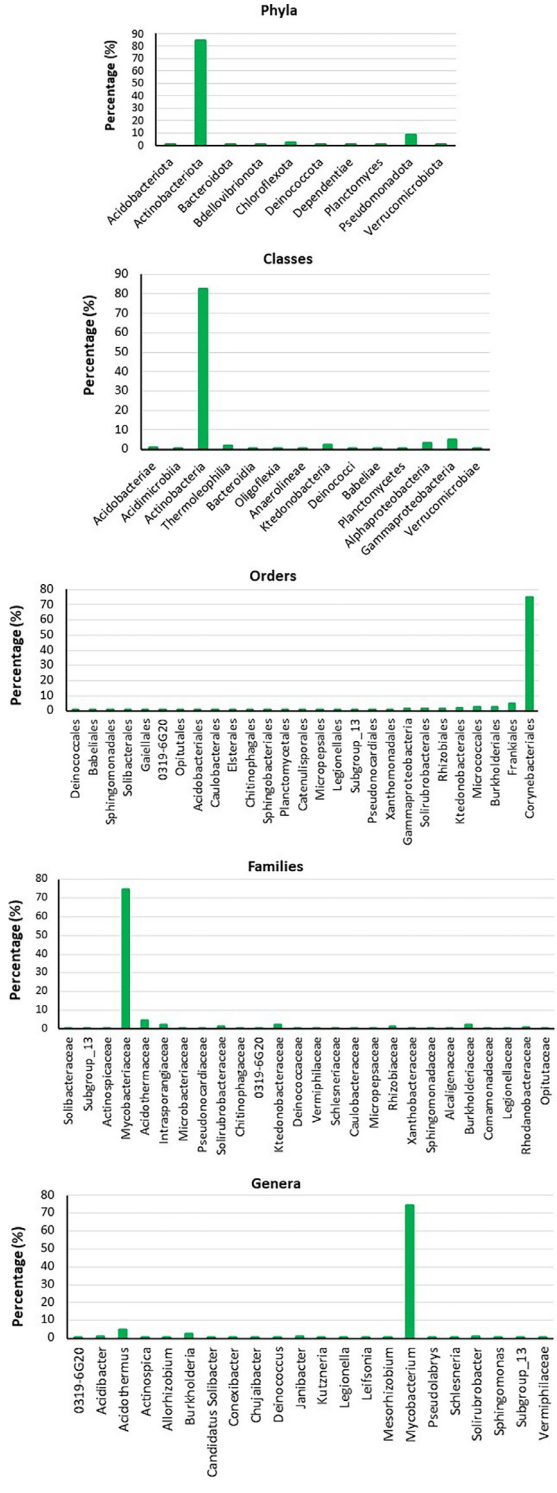

Fig. 1. The taxonomic distribution of the endophytic bacteriome of *Litchi chinensis.*

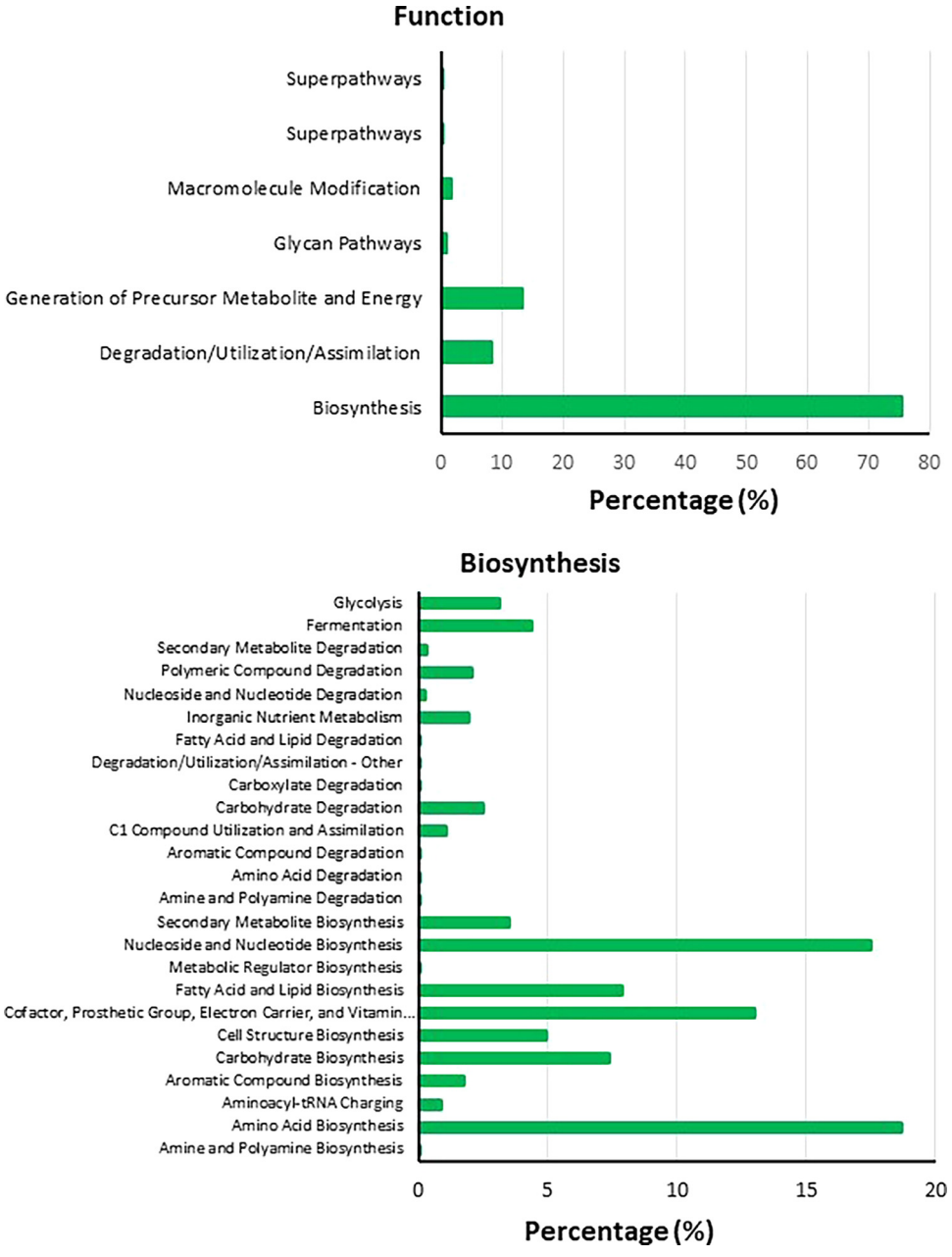

Fig. 2. The metabolic function of the endophytic bacteriome of *Litchi chinensis*.

Among the genera, *Mycobacterium* was the most abundant genus (74.54%), followed by *Acidothermus* (4.76%), *Burkholderia* (2.48%), *Acidibacter* (1.24%), *Janibacter* (0.92%), and *Solirubrobacter* (0.84%) (Fig. 1). The raw read sequences (fastq.gz files) were deposited in Mendeley Data (https://data.mendeley.com/datasets/cfkpjd4322/1).

Fig. 2 shows that biosynthesis (75.42%) was the primary function of the *Citrus nobilis* L. endophytic microbiome, followed by the generation of precursor metabolite and energy

(13.33%), and degradation/utilization/assimilation (8.32%). Of the functions concerning biosynthesis, amino acid biosynthesis (18.66%) was the primary function, followed by nucleoside and nucleotide biosynthesis (17.47%), cofactor, prosthetic group, electron carrier, and vitamin biosynthesis (12.96%), fatty acid and lipid biosynthesis (7.86%), carbohydrate biosynthesis (7.35%), cell structure biosynthesis (4.92%), and biosynthesis, secondary metabolite biosynthesis (3.52%). The raw read sequences (fastq.gz files) were deposited in Mendeley Data (https://data.mendeley.com/datasets/cfkpjd4322/1).

4. Experimental Design, Materials and Methods

Approximately 100 g of root samples were collected from each of the three gardens (12°32′14″N, 107°58′56″E; 12°32′12″N, 107°59′06″E; 12°32′10″N, 107°58′48″E) in Kong Ana District of Dak Lak, on 30 October 2021. Roots of five trees were collected from each garden. These samples were combined to create a representative sample. Thereafter, the sample was sterilized to remove the microbiome from the surface of the roots [8]. The DNeasy PowerSoil kit (Qiagen, USA) was used to isolate the metagenomic DNA from 300 mg of the sample, following the manufacturer's instructions. The 16S rRNA gene (the V1–V9 region) was amplified from the metagenomic DNA using primers F1: 5′–GAGTTTGATCMTGGCTCAG–3′, F2: 5′–CCTACGGGAGGCAGCAG–3′, F3: 5′–GCCAGCAGCCGCGGTAA–3′, F4: 5′–ATGGCTGTCGTCAGCT–3′, F5: 5′–GYAACGAGCGCAACCC–3′, R1: 5′–CTACCAGGGTATCTAATCC–3′, R2: 5′–CCGTCAATTCMTTTGAGTTT–3′, R3: 5′–GACGGGCGGTGTGTACAA–3′, and R4: 5′–TACCTTGTTACGACTT–3′ [14]. The Swift amplicon 16S plus internal transcribed spacer panel (Swift Biosciences, USA) was applied to prepare the 16S rRNA gene metagenomic library based on the supplier's instructions. The Illumina MiSeq platform (2 × 150 PE) was used to sequence the library. The Trimmomatic 0.39 [15] and Cutadapt 2.10 [16] were used to filter sequence data. Kraken 2 version 2.0.8 [17] was used to analyze taxonomic profiles, while the MetaCyc database [18] was used to predict metabolic functions.

Limitations

None.

Ethics Statement

The current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Credit Author Statement

Dinh Sy Nguyen: Sampling, Investigation, Formal analysis, Software, Data curation, Validation, Visualization. **Dinh Minh Tran:** Conceptualization, Methodology, Sampling, Investigation, Formal analysis, Software, Data curation, Validation, Visualization, Writing, Review and Editing.

Data Availability

Data on endophytic microbiome of lychee (Litchi chinensis S.) (Original data) (Mendeley Data).

Acknowledgement

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] General Statistics Office, Statistical Yearbook of Vietnam. Statistical Publishing House, Ha Noi. (2023) 570–637.
[2] Z. Wu, X. Chen, X. Lu, Y. Zhu, X. Han, J. Yan, L. Yan, W. Zou, Impact of combined organic amendments and chemical fertilizers on soil microbial limitations, soil quality, and soybean yield, Plant Soil. 507 (2025) 317–334, doi:10.1007/s11104-024-06733-4.
[3] S.G. Burragoni, J. Jeon, Applications of endophytic microbes in agriculture, biotechnology, medicine, and beyond, Microbiol. Res. 245 (2021) 126691, doi:10.1016/j.micres.2020.126691.
[4] J.F. White, K.L. Kingsley, Q. Zhang, R. Verma, N. Obi, S. Dvinskikh, M.T. Elmore, S.K. Verma, S.K. Gond, K.P. Kowalski, Review: Endophytic microbes and their potential applications in crop management, Pest. Manage. Sci. 75 (2019) 2558–2565, doi:10.1002/ps.5527.
[5] K.L. Rana, D. Kour, T. Kaur, R. Devi, A.N. Yadav, N. Yadav, H.S. Dhaliwal, A.K. Saxena, Endophytic microbes: biodiversity, plant growth-promoting mechanisms and potential applications for agricultural sustainability, Antonie Van Leeuwenhoek 113 (2020) 1075–1107, doi:10.1007/s10482-020-01429-y.
[6] D.M. Tran, Taxonomic and functional profiles of *Coffea canephora* endophytic microbiome in the Central Highlands region, Vietnam, revealed by analysis of 16S rRNA metagenomics sequence data, Data Brief 43 (2022) 108372, doi:10.1016/j.dib.2022.108372.
[7] D.M. Tran, T.H. Nguyen, T.O. Huynh, T.O. Do, Q.V. Nguyen, A.D. Nguyen, Analysis of endophytic microbiome dataset from roots of black pepper (*Piper nigrum* L.) cultivated in the Central Highlands region, Vietnam using 16S rRNA gene metagenomic next-generation sequencing, Data Brief. 42 (2022) 108108, doi:10.1016/j.dib.2022.108108.
[8] T.H. Nguyen, D.M. Tran, Root endophytic microbiome dataset of sugarcane (*Saccharum officinarum* L.) cultivated in the Central Highlands, Vietnam, established by the 16S rRNA metagenomics, Data Brief. 48 (2023) 109103, doi:10.1016/j.dib.2023.109103.
[9] D.M. Tran, T.H. Nguyen, Rice (*Oryza sativa* L.) cultivated in the Central Highlands of Vietnam: Dataset on the endophytic microbiome, Data Brief. 50 (2023) 109551, doi:10.1016/j.dib.2023.109551.
[10] D.M. Tran, T.H. Nguyen, Endophytic bacterial dataset of the Cavendish banana grown in Dak Lak Province of Vietnam using 16S rRNA gene metabarcoding, Data Brief 52 (2023) 109863, doi:10.1016/j.dib.2023.109863.
[11] D.M. Tran, T.H. Nguyen, 16S rRNA metagenomic dataset on endophytic bacterial community of the cashew plant (*Anacardium occidentale* L.) grown in Dak Lak Province of Vietnam, Data Brief. 52 (2024) 110039, doi:10.1016/j.dib.2024.110039.
[12] D.M. Tran, D.S. Nguyen, T.H. Nguyen, T.P.H. Tran, A.D. Nguyen, Shotgun metagenomic dataset of root endophytic microbiome of citrus (*Citrus nobilis* L.), Data Brief 56 (2024) 110777, doi:10.1016/j.dib.2024.110777.
[13] D.M. Tran, T.H. Nguyen, A.D. Nguyen, Shotgun metagenomics sequencing data of root microbial community of Huanglongbing-infected *Citrus nobilis*, Data Brief. 57 (2024) 111061, doi:10.1016/j.dib.2024.111061.
[14] D.M. Tran, T.O. Huynh, T.H. Nguyen, T.O. Do, H.T.P. Tran, Q.V. Nguyen, A.D. Nguyen, Soil microbiome dataset from Yok Don national park in the Central Highlands region of Vietnam, Data Brief 40 (2022) 107798, doi:10.1016/j.dib.2022.107798.
[15] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics. 30 (2014) 2114–2120, doi:10.1093/bioinformatics/btu170.
[16] M. Martin, CUTADAPT removes adapter sequences from high-throughput sequencing reads, EMBnet J. 17 (2011) 10–12, doi:10.14806/ej.17.1.200.
[17] D.E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2, Genome Biol. 20 (2019) 257, doi:10.1186/s13059-019-1891-0.
[18] R. Caspi, R. Billington, L. Ferrer, H. Foerster, C.A. Fulcher, I.M. Keseler, A. Kothari, M. Krummenacker, M. Latendresse, L.A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D.S. Weaver, P.D. Karp, The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases, Nucleic. Acids. Res. 44 (2016) D471–D480, doi:10.1093/nar/gkv1164.