Data Article

Shotgun metagenomic dataset of root endophytic microbiome of citrus (*Citrus nobilis* L.)

Dinh Minh Tran [a,1,*], Dinh Sy Nguyen [b,1], Thi Huyen Nguyen [a], Thi Phuong Hanh Tran [b], Anh Dzung Nguyen [a]

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

ARTICLE INFO

ABSTRACT

Citrus (*Citrus nobilis* L.) is one of the main fruit crops in Dak Lak Province of Vietnam; however, a dataset on the endophytic microbiome of this plant has yet to be discovered. This article presented the endophytic microbial dataset from roots of healthy *Citrus nobilis* L. collected in Dak Lak for the first time. We found that 4 kingdoms, 30 phyla, 58 classes, 125 orders, 242 families, 722 genera, and 1637 species of endophytic microorganisms were identified from the sample. *Actinomycetota* was shown to be the main phylum (64.36 %) and biosynthesis to be the most abundant function (55.64 %) of the endophytic microbial community. Data provided insights into the composition and functional diversity of the *Citrus nobilis* L. endophytic microbiome, especially novel microbial resources. They could be used for the next works towards applying the endophytic microbiome for sustainable citrus production.

---

* Corresponding author.
  E-mail address: tmdinh@ttn.edu.vn (D.M. Tran).
[1] These authors contributed equally to this work and share first authorship.

Specifications Table

Subject	Microbiology: Microbiome
Specific subject area	Molecular biology, Bioinformatics
Data format	Raw (fastq.gz files), Filtered, and Analyzed
Type of data	Figures, Table, and Fastq files
Data collection	Roots of *Citrus nobilis* L. grown in Dak Lak Province of Vietnam were collected, and the genomic DNA of the sample was isolated using the DNeasy PowerSoil Pro kit. The shotgun metagenomic library was prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina and sequenced using the Illumina platform (2 × 150 PE). The taxonomic distribution and functional categories of the endophytic microbiome were analyzed using the Kraken2 and the MetaCyc database, respectively.
Data source location	• Institution: Institute of Biotechnology and Environment, Tay Nguyen University • District/Province/Country: Buon Don/Dak Lak/Vietnam • Latitude and longitude coordinates for collected samples: 12°43′02′′N, 107°57′35′′E; 12°43′03′′N, 107°57′36′′E; 12°43′01′′N, 107°57′31′′E
Data accessibility	Raw sequences (fastq.gz files) Repository name: Mendeley Data Data identification number: doi: 10.17632/khn2zv8jf8.1 Direct URL to data: https://data.mendeley.com/datasets/khn2zv8jf8/1

1. Value of the Data

- The dataset provided insights into the composition and functional diversity of the endophytic microbiome of *Citrus nobilis* L. grown in Dak Lak, Vietnam
- The dataset could be useful for comparing the endophytic microbiome of *Citrus nobilis* L. and others.
- The dataset provided the prospect of finding new microbial resources from *Citrus nobilis* L.
- The dataset could be valuable for further experiments concerning the application of the endophytic microbiome for sustainable citrus production.

2. Background

Citrus is one of the main exporters of fruits in Vietnam. In 2022, Vietnam had 120,000 hectares and produced 3.14 million tons of citrus [1]. Citrus was grown throughout the country, including Dak Lak Province. Among them, *Citrus nobilis* L. was the primary cultivar grown in Dak Lak. Previously, we have reported data on the endophytic microbiome of black pepper, coffee, rice, sugarcane, cashew, and banana grown in Dak Lak [2–7]; however, no data on the endophytic microbiome of citrus cultivated in this province have been reported. Therefore, this work aimed to establish a dataset on the endophytic microbiome of healthy *Citrus nobilis* L. using shotgun metagenomics for further experiments concerning the application of endophytic microorganisms for sustainable citrus production.

3. Data Description

The current work obtained 37,453,139 raw reads with a total of 11,235,941,700 bp. After qualifying and filtering the raw reads, 4334,168 reads were used to analyze the taxonomic distribution and functional categories. Fig. 1 and Table S1-S7 show that 4 kingdoms, 30 phyla, 58 classes, 125 orders, 242 families, 722 genera, and 1637 species of endophytic microorganisms were identified from the sample. Among the kingdom, bacteria (98.859 %) were the most predominant, followed by eukaryota (1.12 %), archaea (0.018 %), and viruses (0.003 %). Of the bacteria, *Actinomycetota* (64.36 %), *Actinomycetes* (63.17 %), *Mycobacteriales* (35.71 %), *Mycobacteriaceae* (30.12 %),
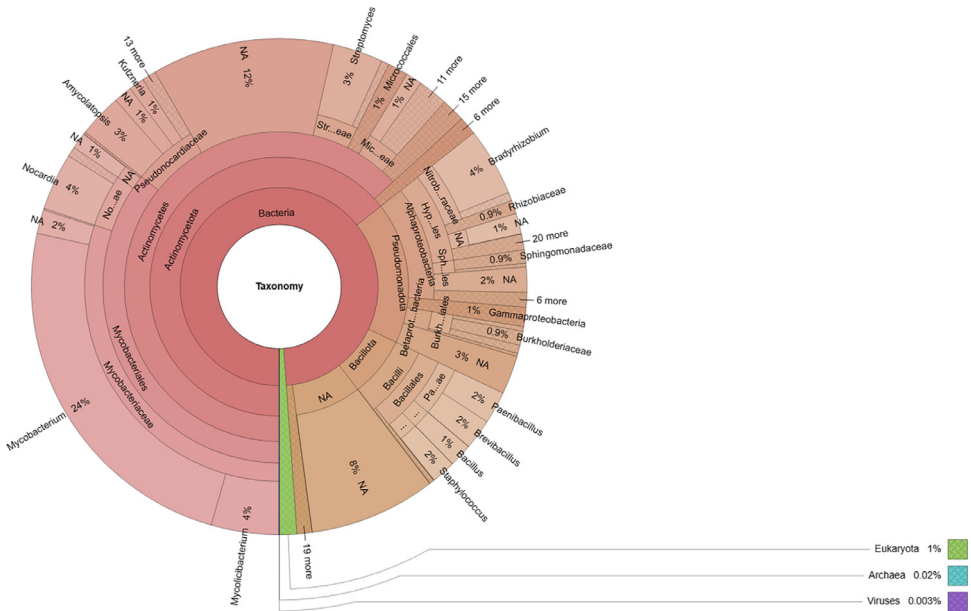
Fig. 1. The endophytic microbiome of *Citrus nobilis* L.

Mycobacterium (24.03 %), and *Mycobacterium marseillense* (2.39 %) were the most abundant phylum, class, order, family, genus, and species of the endophytic microbiome of *Citrus nobilis* L. Moreover, unidentified families and genera accounted for 29.55 % and 36.29 %, respectively. In addition, numerous bacteria, which were reported to play an important role in crop production, were identified from the microbiome, such as bacteria belonging to *Actinomyces, Streptomyces, Cellulomonas, Bacillus, Brevibacillus, Paenibacillus, Bradyrhizobium, Rhizobium, Azospirillum, and Pseudomonas* [8–14]. Of the eukaryota, *Ascomycota* (0.57 %), *Sordariomycetes* (0.54 %), *Hypocreales* (0.51 %), *Nectriaceae* (0.49 %), *Fusarium* (0.49 %), and *Fusarium falciforme* (0.87 %) were the most predominant phylum, class, order, family, genus, and species of the microbiome. The raw read sequences (fastq.gz files) have been deposited in Mendeley Data and are available at https://data.mendeley.com/datasets/khn2zv8jf8/1.

Fig. 2 shows that biosynthesis (55.64 %) was the primary function of the *Citrus nobilis* L. endophytic microbiome, followed by degradation/utilization/assimilation (30.53 %). Of the functions concerning biosynthesis, secondary metabolite biosynthesis (26.12 %) was the most abundant, followed by cofactor, prosthetic group, electron carrier, and vitamin biosynthesis (7.27 %); fatty acid and lipid biosynthesis (5.18 %); carbohydrate biosynthesis (4.03 %); amino acid biosynthesis (3.18 %); cell structure biosynthesis (2.89 %); and nucleoside and nucleotide biosynthesis (1.83 %). The raw read sequences (fastq.gz files) have been deposited in Mendeley Data and are available at https://data.mendeley.com/datasets/khn2zv8jf8/1.

4. Experimental Design, Materials and Methods

4.1. Sample collection

Three root samples (50–70 g each) of healthy *Citrus nobilis* L. were collected from three gardens (12°43′02″N, 107°57′35″E; 12°43′03″N, 107°57′36″E; 12°43′01″N, 107°57′31″E) in Buon Don District, Dak Lak Province, Vietnam, and the sample was then mixed and combined to gen-
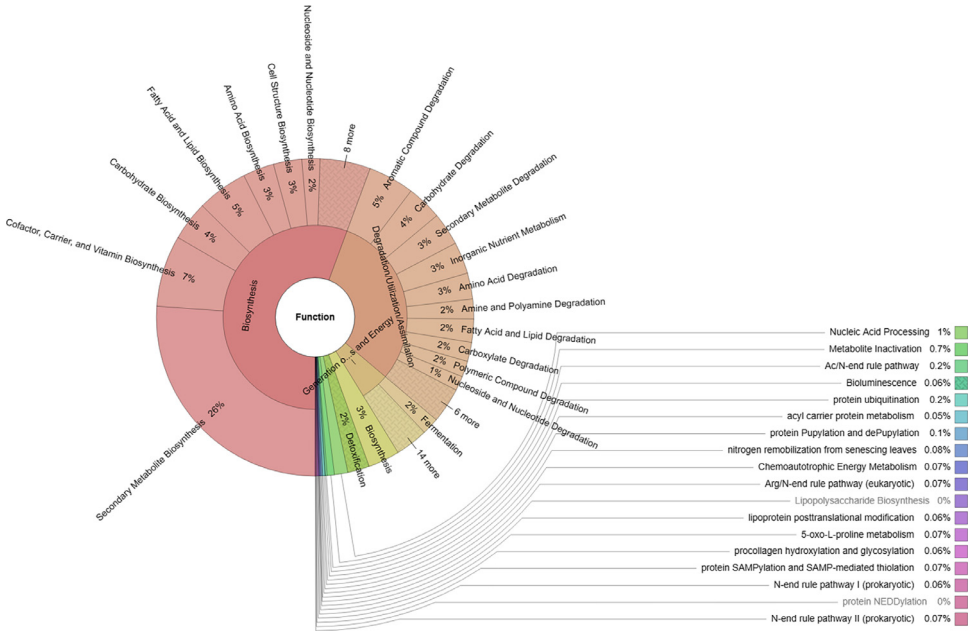
Fig. 2. Functional profiles of the endophytic microbiome of *Citrus nobilis* L.

erate the representative sample. The surface of the roots was sterilized, and then the sample was stored at −80 °C [5] until analysis.

4.2. Total genomic DNA isolation, shotgun metagenomic library preparation and sequencing

Total genomic DNA of the root sample was isolated using the DNeasy PowerSoil Pro kit (Qiagen, Germany). The shotgun metagenomic library was prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, USA). The Illumina platform (2 × 150 PE) was used to sequence the shotgun metagenomic library with DNBSeq-G99 (MGI) machine [15].

4.3. Bioinformatic analysis

Raw data were demultiplexed using bcl2fastq 2.20 [16]. Sequence data were then filtered using Trimmomatic 0.39 [17] and Cutadapt 2.10 [18]. The plant and human contaminating DNA were removed using Kraken2 [19]. Taxonomic distribution was analyzed using Kraken2 [19], and functional categories were analyzed using the MetaCyc database [20].

5. Limitations

Not applicable.

Ethics statement

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

CRediT author statement

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

Data Availability

Shotgun metagenomic dataset of root endophytic microbiome of citrus (Citrus nobilis L.) (Original data) (Mendeley Data)

Acknowledgements

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.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110777.

References

[1] General Statistics Office, in: Statistical Yearbook of Vietnam, Statistical Publishing House, Ha Noi, 2022, pp. 558–645.
[2] 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 Br. 42 (2022) 108108, doi:10.1016/j.dib.2022.108108.
[3] 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 Br. 43 (2022) 108372, doi:10.1016/j.dib.2022.108372.
[4] D.M. Tran, T.H. Nguyen, Rice (*Oryza sativa* L.) cultivated in the Central Highlands of Vietnam: dataset on the endophytic microbiome, Data Br. 50 (2023) 109551, doi:10.1016/j.dib.2023.109551.
[5] 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 Br. 48 (2023) 109103, doi:10.1016/j.dib.2023.109103.
[6] 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 Br. 52 (2024) 110039, doi:10.1016/j.dib.2024.110039.
[7] 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 Br. 52 (2023) 109863, doi:10.1016/j.dib.2023.109863.

[8] J. Gousia, S. Ishfaq, B. Uqab, S. Mudasir, *Actinomycetes* as biofertilisers for sustainable agriculture, in: G.H. Dar, R.A. Bhat, M.A. Mehmood (Eds.), Microbiomes For the Management of Agricultural Sustainability, Springer, Cham, 2023, pp. 183–192.

[9] O.S. Olanrewaju, O.O. Babalola, *Streptomyces*: implications and interactions in plant growth promotion, Appl. Microbiol. Biotechnol. 103 (2019) 1179–1188, doi:10.1007/s00253-018-09577-y.

[10] U. Ejaz, M. Sohail, A. Ghanemi, Cellulases: from bioactivity to a variety of industrial applications, Biomimetics 6 (2021) 44, doi:10.3390/biomimetics6030044.

[11] A.K. Saxena, M. Kumar, H. Chakdar, N. Anuroopa, D.J. Bagyaraj. *Bacillus* species in soil as a natural resource for plant health and nutrition. J. Appl. Microbiol.. 128 (20206) 1583–1594. 2020doi: 10.1111/jam.14506.

[12] A.K. Panda, S.S. Bisht, S. DeMondal, N. Senthil Kumar, G. Gurusubramanian, A.K. Panigrahi, *Brevibacillus* as a biological tool: a short review, Antonie Van Leeuwenhoek 105 (2014) 623–639, doi:10.1007/s10482-013-0099-7.

[13] S. Lal, L. Chiarini, S. Tabacchioni, New insights in plant-associated *Paenibacillus* species: biocontrol and plant growth-promoting activity, in: M. Islam, M. Rahman, P. Pandey, C. Jha, A. Aeron (Eds.), *Bacilli* and Agrobiotechnology, Springer, Cham, 2016, pp. 237–279.

[14] B.A. Bhat, L. Tariq, S. Nissar, S.T. Islam, S.U. Islam, Z. Mangral, N. Ilyas, R.Z. Sayyed, G. Muthusamy, W. Kim, T.U.H. Dar, The role of plant-associated rhizobacteria in plant growth, biocontrol and abiotic stress management, J. Appl. Microbiol 133 (2022) 2717–2741, doi:10.1111/jam.15796.

[15] L.A. Le, T.P. Nguyen-Hoang, V.P. Huynh, T.H. Nguyen, T.V. Nguyen, T.D. Ho-Huynh, Microbiome dataset analysis from a shrimp pond in Ninh Thuan, Vietnam using shotgun metagenomics, Data Br. 31 (2020) 105731, doi:10.1016/j.dib.2020.105731.

[16] Illuminabcl2fastq2 Conversion Software v2.20, 2017 https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html.

[17] 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.

[18] M. Martin, CUTADAPT removes adapter sequences from high-throughput sequencing reads, EMBnet J. 17 (2011) 10–12, doi:10.14806/ej.17.1.200.

[19] 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.

[20] 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.