



Data Article

Rhizosphere microbiome dataset of Robusta coffee (*Coffea canephora* L.) grown in the Central Highlands, Vietnam, based on 16S rRNA metagenomics analysis



Dinh Minh Tran

Institute of Biotechnology and Environment, Tay Nguyen University, Buon Ma Thuot, Dak Lak 630000, Vietnam

ARTICLE INFO

Article history:

Received 7 February 2022

Revised 21 March 2022

Accepted 25 March 2022

Available online 28 March 2022

Keywords:

Rhizosphere microbiome data

Robusta coffee

16S rRNA metagenomics

The Central Highlands region

ABSTRACT

Vietnam is the second-largest coffee producer in the world after Brazil. Of the two main coffee production species, namely, Arabica and Robusta, Vietnam is the largest producer of Robusta worldwide [1]. Based on previous reports, the planted coffee area in Vietnam was 695.600 ha and its production was 1.76 million tons in 2020, in which the Central Highlands region accounts for approximately 73% of the planted area and production [2]. Hence, this region is known as the capital of coffee plantations and production in Vietnam. Previous studies have focused on the diversity of rhizospheric bacteria from this plant species cultivated in this region based on cultivation methods [3–8]. However, no report has been found on the rhizospheric microbial diversity of this important plant in Vietnam. To our knowledge, a dataset of rhizospheric microbial communities of the coffee plant grown in the Central Highlands is still unclear.

This report presents a dataset of the rhizosphere microbiome from a representative sample obtained by mixing five rhizospheric soil samples of *Coffea canephora* L. cultivated in the Central Highlands region using metagenomic next-generation sequencing. This dataset provides information on the rhizospheric microbial diversity of Robusta coffee, particularly its functionality. Therefore, cultivation techniques for sustainable Robusta coffee production in the region could be developed by applying indigenous rhizospheric microbial resources.

E-mail address: tmdinh@ttn.edu.vn<https://doi.org/10.1016/j.dib.2022.108106>2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Microbiology: <i>Microbiome</i>
Specific subject area	Metagenomics
Type of data	Figures, Tables, and Fastq files
How the data were acquired	Illumina MiSeq platform
Data format	Raw and Analyzed
Description of data collection	A representative rhizospheric soil sample (500 g) of <i>Coffea canephora</i> L. was mixed from five samples (100 g each) collected from a 6-year-old coffee field in Khanh Xuan Ward, Buon Ma Thuot City, Dak Lak Province. Total microbial genomic DNA was extracted from the sample and the 16S rRNA metagenomic sequencing was performed using the Illumina MiSeq platform (2 × 150-bp paired ends)
Data source location	<ul style="list-style-type: none"> • Ward/City/Province: Khanh Xuan/Buon Ma Thuot/Dak Lak • Region: The Central Highlands • Country: Vietnam • Latitude and longitude coordinates for collected samples: 12°38'55.20"N,107°59'22.06"E
Data accessibility	Data are available at the NCBI with Bioproject PRJNA797920 (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA797920)

Value of the Data

- Data provide information on the rhizospheric microbial diversity of *Coffea canephora* L. and its functionality in the Central Highlands and other regions in Vietnam.
- Data could be used for the comparative analysis of the rhizospheric and endophytic microbiome profiles of *Coffea canephora* L. cultivated in the Central Highlands and other regions in Vietnam.
- Data will be useful for subsequent studies on the conservation of rhizospheric microbial genetic resources and the development of cultivation techniques for applying them for sustainable Robusta coffee production in the Central Highlands to achieve the nutrients required for various stages of development and growth.

1. Data Description

This dataset presents the taxonomic and functional profiles of the rhizosphere microbiome of *Coffea canephora* L. cultivated in the Dak Lak Province of the Central Highlands, Vietnam. The results showed that 256,357 reads were identified from 256,462 analyzed reads (Table 1). Taxonomic analysis (Fig. 1) showed that 28 phyla were identified from the sample. Among these phyla, Proteobacteria (26.4%) were the most abundant, followed by Actinobacteriota (19.83%), Acidobacteriota (15%), Gemmatimonadota (10.35%), Chloroflexi (9.24%), and Myxococcota (6.97%).

Table 1
 Summary of analyzed, classified, and unclassified reads in this study.

Reads	Count
Total analyzed reads	256,462
Classified reads	256,357
Unclassified reads	105

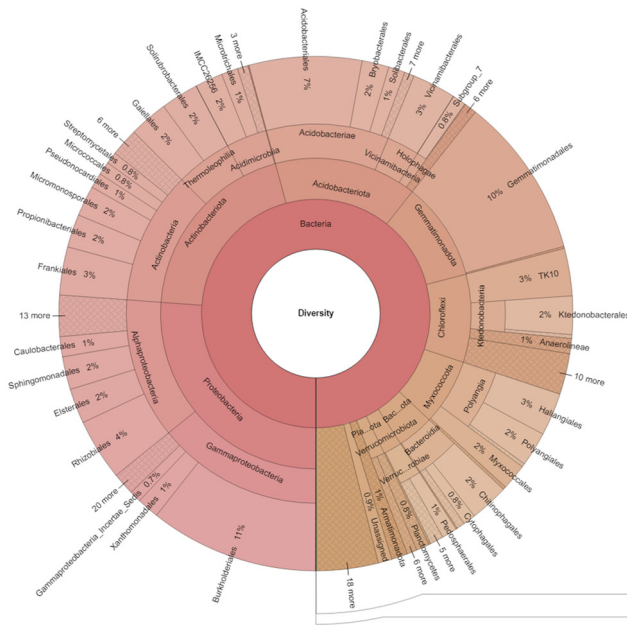


Fig. 1. Taxonomic profiles of rhizosphere microbiome of *Coffea canephora* L. in the Central Highlands region, Vietnam.

Of the 119 bacterial orders detected, Burkholderiales (10.74%) were shown to be the most dominant, followed by Gemmatimonadales (10.22%), Acidobacteriales (7.15%), Rhizobiales (3.91%), and Frankiales (3.04%). Moreover, 156 families were detected. Among these families, Gemmatimonadaceae (10.22%) were found to be the most abundant, followed by Xanthobacteraceae (2.96%), Haliangiaceae (2.71%), Nitrosomonadaceae (2.56%), and Chitinophagaceae (2.44%). Finally, 242 genera were identified from the rhizosphere of Robusta coffee.

Functional analysis (Fig. 2) showed that the primary function of the microbiome of Robusta coffee was biosynthesis (69.67%), followed by degradation/utilization/assimilation (13.92%) and the generation of precursor metabolites and energy (12.92%). Among the functions involved in biosynthesis, amino acid biosynthesis (16.91%) was the most abundant, followed by cofactor, prosthetic group, electron carrier, and vitamin biosynthesis (15.51%); nucleoside and nucleotide biosynthesis (14.88%); fatty acid and lipid biosynthesis (8.8%); carbohydrate biosynthesis (5.02%); cell structure biosynthesis (3.43%); and secondary metabolite biosynthesis (2.75%).

2. Experimental Design, Materials and Methods

2.1. Rhizospheric soil sampling

Five rhizospheric soil samples (approximately 100 g each, 5–30 cm in depth) of *Coffea canephora* L. were collected from five different sites of a 6-year-old coffee field in Khanh Xuan Ward, Buon Ma Thuot City, Dak Lak Province, on October 30, 2021. The samples were mixed well, combined into one representative sample, kept at 4°C in an ice box, and brought to the laboratory within 1 h after sampling. The sample was stored at –80°C until analysis.

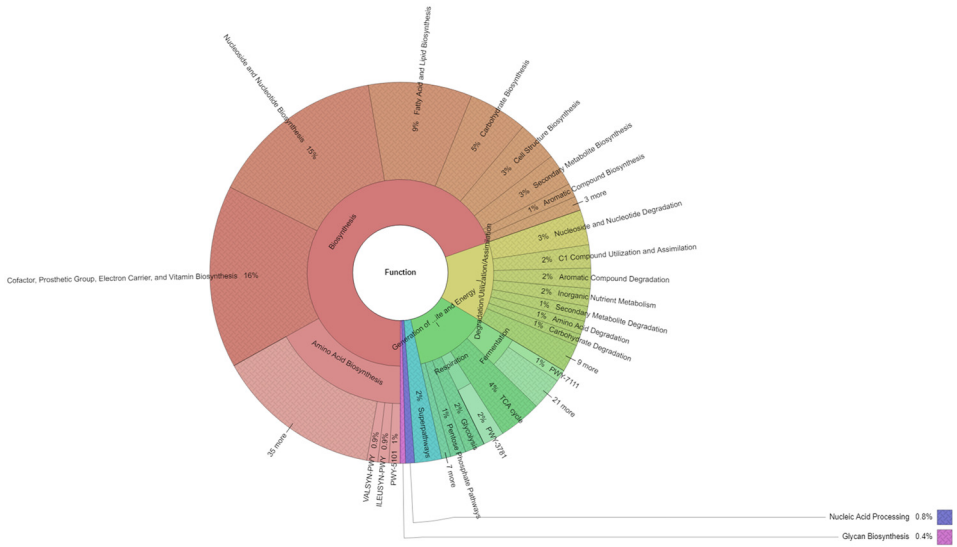


Fig. 2. Functional profiles of rhizosphere microbiome of *Coffea canephora* L. in the Central Highlands region, Vietnam.

2.2. Isolation of microbial genomic DNA

Microbial genomic DNA was extracted from 300 mg of soil sample using the DNeasy Power-Soil kit (Qiagen, Germany) in accordance with the supplier's instructions.

2.3. Library preparation and 16S rRNA metagenomic sequencing

The 16S rRNA gene (regions V1–V9) was amplified using primers [9], and then libraries of 16S rRNA gene amplicons were prepared using the Swift amplicon 16S plus ITS (internal transcribed spacer) panel kit (Swift Biosciences, USA) in accordance with the supplier's instructions. Finally, the Illumina MiSeq platform (2 × 150 bp paired ends) was used to perform the library 16S rRNA gene amplicon sequencing.

2.4. Taxonomic and functional analyses

Taxonomic and functional profiles of rhizospheric microbes were analyzed in accordance with the method of Tran et al. [9]. In brief, bcl2fastq was used to demultiplex raw basecall files. Adapters, primers, and low-quality sequences (average score of <20 and read length of <100 bp) were removed using Trimmomatic (version 0.39) [10] and Cutadapt (version 2.10) [11]. The q2-dada2 plugin and denoise-single method within the QIIME2 pipeline (version 2020.8) [12] were used to cluster and dereplicate the reads into amplicon sequence variants. QIIME2 aligned with the SILVA SSURef reference database (version 138) [13] was used for taxonomic analysis of amplicon sequence variants in accordance with the classify-consensus-blast method [14]. Finally, PICRUSt2 (version 2.3.0-b) [15] and MetaCyc databases [16] were used to deduce the functional profiles of rhizospheric microorganisms from the sample based on the results of 16S rRNA gene amplicon sequencing.

Ethics Statements

None

CRedit Author Statement

Dinh Minh Tran: Conceptualization, Methodology, Investigation, Formal analysis, Software, Data curation, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The author declares that it has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Rhizosphere microbiome dataset of the Robusta coffee (*Coffea canephora* L.) grown in the Central Highlands, Vietnam, based on analysis using 16S rRNA metagenomics (Original data).

Acknowledgments

The Vietnam National Foundation for Science and Technology Development (NAFOSTED) funded this research under grant number [106.04-2019.337](#).

References

- [1] S. Wirryadiaputra, L.K. Tran, Indonesia and Vietnam, in: *Plant-Parasitic Nematodes of Coffee*, Springer Netherlands, Dordrecht, 2008, pp. 277–292.
- [2] General statistics office, in: *Statistical Yearbook of Vietnam*, Statistical Publishing House, Ha Noi, 2020, pp. 559–561.
- [3] C. Nguyen, S. Subbotin, M. Madani, P. Trinh, M. Moens, *Radopholus duriophilus* sp. n. (Nematoda: Pratylenchidae) from Western Highland of Vietnam, *Nematology* 5 (2003) 549–558, doi:[10.1163/156854103322683265](#).
- [4] B. Duong, H.X. Nguyen, H.V. Phan, S. Colella, P.Q. Trinh, G.T. Hoang, T.T. Nguyen, P. Marraccini, M. Lebrun, R. Duponnois, Identification and characterization of Vietnamese coffee bacterial endophytes displaying in vitro antifungal and nematocidal activities, *Microbiol. Res.* 242 (2021) 126613, doi:[10.1016/j.micres.2020.126613](#).
- [5] H. Hoang, L.H. Tran, T.H. Nguyen, D.A.T. Nguyen, H.H.T. Nguyen, N.B. Pham, P.Q. Trinh, T. de Boer, A. Brouwer, H.H. Chu, Occurrence of endophytic bacteria in Vietnamese Robusta coffee roots and their effects on plant parasitic nematodes, *Symbiosis* 80 (2020) 75–84, doi:[10.1007/s13199-019-00649-9](#).
- [6] Q.P. Trinh, T.M.L. Le, T.D. Nguyen, H.T. Nguyen, G. Liebanas, T.A.D. Nguyen, *Meloidogyne daklakensis* n. sp. (Nematoda: Meloidogynidae), a new root-knot nematode associated with robusta coffee (*Coffea canephora* Pierre ex A. Froehner) in the Western Highlands, Vietnam, *J. Helminthol.* 93 (2019) 242–254, doi:[10.2478/helm-2019-0014](#).
- [7] D.N. Nguyen, S.L. Wang, A.D. Nguyen, M.D. Doan, D.M. Tran, T.H. Nguyen, V.A. Ngo, C.T. Doan, T.N. Tran, V.C. Do, V.B. Nguyen, Potential application of rhizobacteria isolated from the Central Highland of Vietnam as an effective biocontrol agent of Robusta coffee nematodes and as a bio-fertilizer, *Agronomy* 11 (2021) 1887, doi:[10.3390/agronomy11091887](#).
- [8] D.M. Tran, T.O. Huynh, T.H. Nguyen, T.O. Do, Q.V. Nguyen, A.D. Nguyen, Molecular analysis of genes involved in chitin degradation from the chitinolytic bacterium *Bacillus velezensis*, *Antonie van Leeuwenhoek* 115 (2022) 215–231, doi:[10.1007/s10482-021-01697-2](#).
- [9] 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](#).
- [10] 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](#).
- [11] M. Martin, CUTADAPT removes adapter sequences from high-throughput sequencing reads, *EMBnet J.* 17 (1) (2011) 10–12, doi:[10.14806/ej.17.1.200](#).
- [12] E. Bolyen, J.R. Rideout, M.R. Dillon, N.A. Bokulich, C.C. Abnet, et al., Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2, *Nat. Biotechnol.* 37 (9) (2019) 852–857, doi:[10.1038/s41587-019-0209-9](#).

- [13] C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F.O. Glöckner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools, *Nucleic Acids Res.* 41 (D1) (2013) D590–D596, doi:[10.1093/nar/gks1219](https://doi.org/10.1093/nar/gks1219).
- [14] N.A. Bokulich, B.D. Kaehler, J.R. Rideout, M. Dillon, E. Bolyen, R. Knight, G.A. Huttley, J.G. Caporaso. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome.* 6 (2018), doi:[10.1186/s40168-018-0470-z](https://doi.org/10.1186/s40168-018-0470-z).
- [15] G.M. Douglas, V.J. Maffei, J.R. Zaneveld, S.N. Yurgel, J.R. Brown, C.M. Taylor, C. Huttenhower, M.G.I. Langille, PICRUSt2 for prediction of metagenome functions, *Nat. Biotechnol.* 38 (6) (2020) 685–688, doi:[10.1038/s41587-020-0548-6](https://doi.org/10.1038/s41587-020-0548-6).
- [16] 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 (D1) (2016) D471–D480, doi:[10.1093/nar/gkv1164](https://doi.org/10.1093/nar/gkv1164).