

DATA NOTE

Open Access



Shotgun metagenomics dataset of the core rhizo-microbiome of monoculture and soybean-precedent carrot

Olubukola Oluranti Babalola^{1*} , Alaba Adewole Adebayo¹ and Ben Jesuorsemwen Enagbonma¹

Abstract

Objectives Carrot is a significant vegetable crop contributing to agricultural diversity and food security, but less is known about the core microbiome associated with its rhizosphere. More so, the effect of preceding crop and cropping history on the composition and diversity of carrot rhizo-microbiome remains largely unknown. With shotgun metagenomics, the study unveils how cropping systems direct rhizo-microbiome structure and functions, previously limited by other methods.

Data description Metagenomic-DNA molecule was extracted from four replicates each (12 samples) of a distant bulk soil and the rhizosphere soils from monoculture and soybean-precedent carrots, with the Power soil® DNA Isolation kit. The DNA samples were subjected to Next Generation Sequencing using the Illumina Novaseq X Plus (PE 150) platform. Raw sequencing reads were assembled and annotated with MEGAHIT and LCA algorithms in MEGAN software respectively, before a quality control check was done with FASTP. CD-Hit was used to de-replicate the sequences and the removal of host genomic-DNA and contaminant sequences was done with Bowtie2. The clean sequence data, in FastQ files, were analyzed for taxonomic classification and functional diversity of the rhizosphere microbiome using the Micro_NR and KEGG database respectively. The findings provide insights into microbiome dynamics, with potential implications for sustainable agricultural practices.

Keywords Agricultural sustainability, Cropping systems, Functional diversity, Illumina sequencing, Microbial structure

Objective

Carrot (*Daucus carota* L.) is among the most cultivated vegetable crops in the world, not only for its edibility but also for the health benefits associated with its nutritional composition. As a taproot crop, the rhizospheric microbiome of carrots contributes significantly to the growth and development of the crop. However, the community

structure, abundance, and diverse functionality of the core microbiome inhabiting carrot rhizosphere across cropping systems – mono-cropping and crop-rotation systems are less reported. To our knowledge, only a few studies on carrot microbiome have dealt with the rhizosphere [1], while most others focused on the endophytes [2, 3], studies on the impact of cropping systems on rhizo-microbiome are largely scarce. Majorly, previous assessments of the rhizosphere of carrots were based on culture-dependent and Sanger methods which offer inadequate information. These techniques are limited in unraveling the abundance of diverse functional genes, essential for unique metabolic pathways of agricultural

*Correspondence:

Olubukola Oluranti Babalola
olubukola.babalola@nwu.ac.za

¹Food Security and Safety Focus Area, Faculty of Natural and Agricultural Science, North-West University, Private Box X2046, Mmabatho 2735, North West Province, South Africa



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

importance. Therefore, this study presents a shotgun metagenomics dataset on the rhizosphere microbiome of monoculture and soybean-precedent carrot plants to better understand the microbial and functional shifts caused by the cropping systems with implications for sustainable agriculture. Data from this study will enable in-depth investigations into the potential of these microbes in improving soil fertility and health and enhancing the growth, yield, and productivity of plants – promising food security.

Data description

The dataset consists of the raw Next Generation Sequencing (NGS) data from a shotgun metagenomics sequencing of DNA molecules extracted from the rhizosphere soil of monoculture (MCR) and soybean-precedent carrot (SCR), and a distant bulk soil (BS) from a farm in Gauteng, South Africa Gauteng Province, South Africa (26°06'21.0" S 27°33'35.5" E; 26°06'21.8" S 27°33'41.8" E; and 26°06'21.1" S 27°33'36.9" E GPS location respectively).

Sampling

Four replicate soil samples at 0–15 cm depth were obtained from each sampling field. To capture a representative snapshot of the core microbial structure and functions, rhizosphere soils were sampled within the root development stage (specifically 50 days after sowing). The bulk soils were obtained from uncultivated fields approximately 10 m from planted fields. Samples were collected into sterile plastic bags with a hand trowel pre-sterilized with 70% ethanol before each sampling, and transported in a cooler box to the laboratory [4]. Samples were sieved (with a 2 mm sieve) [5], and stored in the refrigerator at –20°C before analysis [6].

Meta-DNA extraction and sequence

Metagenomic DNA in the soil samples was isolated by following the step-by-step procedure of the extraction kit (DNeasy® Power soil® DNA Isolation kit; Qiagen, Germany). The quality and concentration of the DNA extract were determined with a Nanophotometer® NP80 Touch (Implen,

Germany). A shotgun metagenomic library was prepared after fragmenting into segments (~ 350 bp) using a Covaris ultrasonic disruptor, and the quality of the constructed library was assessed with the Advanced Analytical Technologies Incorporated (AATI) analysis. The Illumina Novaseq X Plus (PE 150) system (Novogene, Singapore) was used to sequence the quality library [7].

Data processing

Poor or low-quality reads and adapters in the raw data were trimmed and filtered using the Fastp (v0.23.4) [8], while contaminations from the host genomic DNA were removed, using Bowtie2 (v2.5.4) alignment tool [9]. MEGAHIT (v1.1.2) was used to assemble the clean data [10], before ORF of the scaftig (> = 500 bp) of each sample was predicted from the MetaGeneMark (v2.10) platform [11], then, sequences with lengths less than 100 nt were removed. The data were demultiplexed with CD-Hit (v4.6.8) [12] to achieve unigene sequences. Information on the sequence data obtained from all the soil samples is presented in Data file 1. The clean data (in FastQ files) were deposited on the Sequence Read Archive of the NCBI database with accession-ID numbers; SRP539180 (Data set 1), SRP537154 (Data set 2), and SRP537537 (Data set 3) for MCR, SCR, and BS respectively. Notably, there is no other data relating to this project published earlier or elsewhere. However, the project is ongoing with core aim of exploring this metadata with implication for sustainable agriculture and food safety.

Further analysis on taxonomic classification and functional diversity through alignment of these unigene sequences was performed in DIAMOND (v2.1.9) with parameter settings: blastp, -e 1e- 5, using Micro_NR and KEGG databases respectively [13]. The Lowest Common Ancestor (LCA) algorithm in MEGAN (v6.8.20) was used to annotate the sequences [14]. Overview of the community structure and diverse functionality of the rhizomicrobiome in this study are represented in Data file 2 respectively (Table 1).

Limitations

Not applicable.

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	MS Excel file (.xlsx)	Figshare (https://doi.org/10.6084/m9.figshare.28357343) [15]
Data file 2	Overview of microbial community structure and diverse functions from dataset	PDF file (.pdf)	Figshare (https://doi.org/10.6084/m9.figshare.28357871) [16]
Data set 1	Core microbiome of monoculture carrot rhizosphere soil	FASTQ files (.fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRP539180) [17]
Data set 2	Core microbiome of soybean-precedent carrot rhizosphere soil	FASTQ files (.fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRP537154) [18]
Data set 3	Core microbiome of bulk soil from uncultivated field in a carrot farm	FASTQ files (.fastq)	NCBI Sequence Read Archive (https://identifiers.org/ncbi/insdc.sra:SRP537537) [19]

Abbreviations

DNA	Deoxyribonucleic Acid
MCR	Monoculture carrot rhizosphere
SCR	Soybean-precedent carrot rhizosphere
BS	Bulk soil
GPS	Global Positioning System
NGS	Next Generation Sequencing

Acknowledgements

The authors acknowledge the support of the management of Rosaly Farms, Tarlton, Gauteng, South Africa, for granting permission to access and sample their carrot farmland soil. AAA thanks the North-West University for the Doctoral bursary and research support. Also, OOB acknowledges the grant from the International Centre for Genetic Engineering and Biotechnology (ICGEB) (Grants number: CRP/ZAF22-93).

Authors' contributions

O.O.B: Conceptualization, Project administration, Supervision, Resource acquisition, Editing, and Review. AAA: Investigation, Data curation, Software validation, Formal analysis, Writing; original draft, Writing; review, and Editing. BJE: Conceptualization, Proofread original draft, Review, and Editing.

Funding

Open access funding provided by North-West University. This work was supported by the International Centre for Genetic Engineering and Biotechnology (ICGEB) (Grants number: CRP/ZAF22 - 93) awarded to OOB.

Data availability

The data described in this Data note can be freely and openly accessed on the NCBI SRA with accession-ID numbers; SRP539180 (Data set 1), SRP537154 (Data set 2), and SRP537537 (Data set 3), and the figshare databases under <https://doi.org/10.6084/m9.figshare.28357343> (Data file 1); <https://doi.org/10.6084/m9.figshare.28357871> (Data file 2). Please see Table 1 and references [15–19] for details and links to the data.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 28 February 2025 / Accepted: 8 April 2025

Published online: 12 April 2025

References

- Rachwał K, Gustaw K, Kazimierzczak W, Waśko A. Is soil management system really important? Comparison of microbial community diversity and structure in soils managed under organic and conventional regimes with some view on soil properties. *PLoS ONE*. 2021;16(9):e0256969. <https://doi.org/10.1371/journal.pone.0256969>.
- Abdelrazek S, Simon P, Colley M, Mengiste T, Hoagland L. Crop management system and Carrot genotype affect endophyte composition and *Alternaria dauci* suppression. *PLoS ONE*. 2020;15(6):e0233783. <https://doi.org/10.1371/journal.pone.0233783>.
- Abdelrazek S, Choudhari S, Thimmapuram J, Simon P, Colley M, Mengiste T, Hoagland L. Changes in the core endophytic mycobiome of Carrot taproots in response to crop management and genotype. *Sci Rep*. 2020;10(1):13685. <https://doi.org/10.1371/journal.pone.0233783>.
- Enagbonma BJ, Babalola OO. Metagenomics reveals the Microbiome multifunctionalities of environmental importance from termite mound soils. *Bioinform Biol Insights*. 2023;17:11779322231184025. <https://doi.org/10.1177/11779322231184025>.
- Enagbonma BJ, Ajilogba CF, Babalola OO. Metagenomic profiling of bacterial diversity and community structure in termite mounds and surrounding soils. *Arch Microbiol*. 2020;202(10):2697–709. <https://doi.org/10.1007/s00203-020-01994-w>.
- Enagbonma BJ, Babalola OO. Metagenomics shows that termite activities influence the diversity and composition of soil invertebrates in termite mound soils. *Appl Environ Soil Sci*. 2022;2022(1):7111775. <https://doi.org/10.1155/2022/7111775>.
- Babalola OO, Enagbonma BJ. Dataset of shotgun metagenomic evaluation of *Sorghum bicolor* rhizosphere Microbiome in soils preceded by *Glycine max*. *Data Brief*. 2025;111270. <https://doi.org/10.1016/j.dib.2025.111270>.
- Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using Fastp. *iMeta*. 2023;2(2):e107. <https://doi.org/10.1002/imt2.107>.
- Cheng WY, Liu WX, Ding Y, Wang G, Shi Y, Chu ES, Wong S, Sung JY, Yu J. High sensitivity of shotgun metagenomic sequencing in colon tissue biopsy by host DNA depletion. *Genomics Proteom Bioinf*. 2023;21(6):1195–205. <https://doi.org/10.1016/j.gpb.2022.09.003>.
- Deng C, Zhao R, Qiu Z, Li B, Zhang T, Guo F, Mu R, Wu Y, Qiao X, Zhang L, Cheng JJ, Ni J, Yu K. Genome-centric metagenomics provides new insights into the microbial community and metabolic potential of landfill leachate microbiota. *Sci Total Environ*. 2022;816:151635. <https://doi.org/10.1016/j.scitotenv.2021.151635>.
- Li C, Li X, Guo R, Ni W, Liu K, Liu Z, Dai J, Xu Y, Abduriyim S, Wu Z, Zeng Y, Lei B, Zhang Y, Wang Y, Zeng W, Zhang Q, Chen C, Qiao J, Liu C, Hu S. Expanded catalogue of metagenome-assembled genomes reveals resistome characteristics and athletic performance-associated microbes in horse. *Microbiome*. 2023;11(1):7. <https://doi.org/10.1186/s40168-022-01448-z>.
- Lian Q, Huettel B, Walkemeier B, Mayjonade B, Lopez-Roques C, Gil L, Roux F, Schneeberger K, Mercier R. A pan-genome of 69 *Arabidopsis thaliana* accessions reveals a conserved genome structure throughout the global species range. *Nat Genet*. 2024;56:982–91. <https://doi.org/10.1038/s41588-024-01715-9>.
- Hu B, Wang J, Li Y, Ge J, Pan J, Li G, He Y, Zhong H, Wang B, Huang Y, Han S, Xing Y, He H. Gut microbiota facilitates adaptation of the plateau Zokor (*Myosorex baileyi*) to the plateau living environment. *Front Microbiol*. 2023;14:1136845. <https://doi.org/10.3389/fmicb.2023.1136845>.
- Zhang X, Yu W, Shao W, He Z, Lu Y, Xin Y, Shen Z, Zhou Y, Hu X. Viral metagenome analysis uncovers hidden RNA Virome diversity in juvenile largemouth bass (*Micropterus salmoides*). *Aquaculture*. 2024;741571. <https://doi.org/10.1016/j.aquaculture.2024.741571>.
- Adebayo AA, Enagbonma BJ, Babalola OO. General information on the sequence data acquired from monoculture and soybean-precedent carrot rhizosphere, and a distant bulk soil. *figshare* <https://doi.org/10.6084/m9.figshare.28357343> (2025).
- Adebayo AA, Enagbonma BJ, Babalola OO, Babalola. Overview of the community structure and diverse functionality of the rhizo-microbiome in carrots under continuous and rotational cropping systems. *figshare* <https://doi.org/10.6084/m9.figshare.28357871> (2025).
- NCBI Sequence Read. Archive <https://identifiers.org/ncbi/insdc.sra:SRP539180> (2024).
- NCBI Sequence Read. Archive <https://identifiers.org/ncbi/insdc.sra:SRP537154> (2024).
- NCBI Sequence Read. Archive <https://identifiers.org/ncbi/insdc.sra:SRP537537> (2024).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.