

DATA NOTE

Open Access



Unveiling the *Allium ampeloprasum* rhizosphere microbiome and its functional dataset under different fertilization systems

Olubukola Oluranti Babalola^{1,2*} , Oluwaseun Emmanuel Shittu¹ and Ben Jesuorsemwen Enagbonma¹

Abstract

Objectives Leek (*Allium ampeloprasum*) is a nutritious vegetable popularly cultivated in South Africa and most regions of the world. It is generally recognised as a source of vitamins and vegetables. Nevertheless, little is known about its rhizosphere microbiome and its microbial functional dataset under various fertilization systems. Therefore, this study intended to unveil the rhizosphere microbiome of *Allium ampeloprasum* and their functional datasets through shotgun metagenomics sequencing analysis.

Keywords Food security, Leek, Functionality, Metabolism, Biofertiliser, Chemical fertiliser, Shotgun sequencing

Data description

Soil samples were collected from three different plots; chemical fertilizer (G1), biofertilizer plot (G2) and Bulk soil (G3) plots with each plot consisting of four replicates. The entire DNA was mined from soil samples with DNeasy® PowerSoil pro kit and shotgun metagenomic technique was then used to sequence the extracted DNA. The datasets revealed that biofertilizer enriched the rhizosphere microbiomes and their functional categories, and were predominated by bacteria at the kingdom level while metabolism dominated the functional categories (Table 1, Data file 5). Information on the clean metagenomic datasets is deposited in NCBI-SRA with accession numbers SRP537120, SRP537121, and SRP537188 (Table 1; Data files 1, 2, and 3). By providing insights into

microbial shifts and their functions driven by different fertilisation systems, these findings imply that biofertilizer can enhance soil health, leading to more productive and sustainable agricultural practices.

Objectives

Allium ampeloprasum L. is known to be classified under the genus *Allium* (Alliaceae) [1], which is widely distributed and consumed globally. Allium families are recognised to be a prodigious home of various vitamins, including vitamin A, C, and K, as well as a number of B vitamins, including B6 and folate. These plants also contain antioxidant properties and are high in dietary fibre when consumed as vegetables, making them important in food nutrition [2]. The viable host of microorganisms in their rhizosphere could actively impact the physiological growth and development of *Allium ampeloprasum* plants. Understanding the community structure of its rhizosphere microbiome and functional categories of agricultural importance under different fertilisation will help in agricultural sustainability. Profiling the community organization and microbiome diversity related to *Allium ampeloprasum*'s rhizosphere under different fertilisation

*Correspondence:

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

¹Food Security and Safety Focus Area, Faculty of Natural and Agricultural Sciences, North-West University, Private Mail Bag X2046, Mmabatho 2735, South Africa

²Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, Berkshire SL5 7PY, UK



© 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/>.

Table 1 An overview of the data sets or files

Label	ID of data set/data file	File extension (Type of files)	Data source and identification (access number or DOI)
Data file 1	Genome sequence data of <i>Allium ampeloprasum</i> rhizosphere soil from the biofertiliser plot.	FASTQ file (.fastq)	NCBI Sequence Read Archive (http://identifiers.org/insdc.sra:SRP537120) [17].
Data file 2	Genome sequence data of <i>Allium ampeloprasum</i> rhizosphere soil samples from the chemical fertiliser plot.	FASTQ file (.fastq)	NCBI Sequence Read Archive (http://identifiers.org/insdc.sra:SRP537121) [18].
Data file 3	Genome sequence data of uncultivated bulk soil.	FASTQ file (.fastq)	NCBI Sequence Read Archive (http://identifiers.org/insdc.sra:SRP537188) [19].
Data file 4	Nucleic acid information	Word document files (.docx)	Figshare (https://doi.org/10.6084/m9.figshare.28925831) [20].
Data file 5	Major microbiome and functional distributions	Word document files (.docx)	Figshare (https://doi.org/10.6084/m9.figshare.28930460) [21].

systems using a shotgun sequencing approach is essential in comprehending the complex impacts of agro-chemical fertilisers vs. biofertiliser on microbial ecosystems, which carry out an essential functions in plant health, sustainable agriculture and soil fertility. Among the various fertilisation methods utilised in agriculture, the treatment of plants with organic fertilisers (biofertilisers) is recognised for enhancing and boosting the diversity of microorganisms and the community structure within agricultural lands [3]. In this context, shotgun sequencing as used in this study offers a significant advantage over traditional microscopic techniques and 16 S rRNA sequencing, providing a detailed and comprehensive analysis of communities of microorganisms by capturing the full spectrum of microbial life, including archaea, fungi, bacteria and viruses, without being limited to specific regions of the genome [4]. Hence, this study profiled the functional and structural diversity of the microbiome from the rhizosphere of *Allium ampeloprasum* under different fertilisation systems.

Description of data

The soil samples utilised for the whole genome sequencing were sourced from Rosaly farm, located in the Gauteng province of South Africa. The farm covers an area of 15,000 m² and is divided into three plots treated with chemical fertiliser (G1), biofertilizer (G2) and uncultivated bulk soil (G3), with each plot at a distance of 20 meters apart from the next. The sampling sites (latitude: S 26°6'10.72764", longitude: E 27°34'41.7576) have an average temperature of 22 °C, and yearly rainfall of 794 mm. Soil samples were independently sourced from

the rhizosphere of *Allium ampeloprasum* crop (15 cm depth) in fertilisation plots during the flowering or bolting stage, on day 57 after sowing, and from uncultivated bulk soil. Within each plot, four replicate samples were taken: L1, L2, L3 and L4 for (G1); L9, L10, L11 and L12 for (G2); and LB1, LB2, LB3 and LB4 for (G3), resulting in a total of 12 samples. Plot A has a history of chemical fertiliser application, consisting of calcium, magnesium and potassium nitrate; ammonium and potassium sulphate; and N.P.K. (5:1:5). Plot B with organic fertilisers like Terramax (PGPRs), Humesoil (plant-derived materials), and Soluphos (PSM), while the bulk soil samples were taken from an uncultivated plot, 20 m away from fertilised plots, to prevent contamination from fertilised treatments. The soil samples were conveyed into the laboratory in an ice box and preserved at 4 °C.

20 g samples from each plot [5] were employed for the extraction of DNA through the PowerSoil DNeasy® pro kit. The DNA extracted were then sequenced via the Illumina NovaSeq X Plus (PE 150) platform (Table 1, Data file 4) [3]. A Covaris disruptor was used to fragment the DNA to about 350 bp, then PCR amplification, adapter ligation and end repair were performed to prepare the library. AATI analysis was used to confirm qPCR validated concentrations greater than 3 nM and the library's insert size. Using PE150 technology, quality-assured libraries were combined and sequenced. Preprocessing of data was carried out using Fastp [6] to eliminate poor-quality readings and adapters, and Bowtie2 [7] was utilised to remove impurities from the host. With the help of the meta-large preset MEGAHIT, the clean reads were put together [8], and 'N' junction scaffolds were thrown away in order to create scaffigs [9]. MetaGeneMark was used for ORF prediction [10]. A non-redundant gene catalog was produced by using CD-HIT to eliminate redundant ORFs [11]. Using the blastp algorithm (DIAMOND), the unigenes were aligned to the Micro_NR database as described by B Buchfink, C Xie and DH Huson [12], using 1e-5 as e-value cutoff [13]. LCA algorithm was employed for species annotations assignment in MEGAN software [14]. Relative abundances were determined using the R ade4 software [15]. Diversity indices were computed with "vegan" R software [16], with ANOSIM used to evaluate group differences.

Abbreviations

L1, L2, L3 and L4	Are replicated soil samples from <i>Allium ampeloprasum</i> rhizosphere under chemical fertilisation
L9, L10, L11, and L12	Are replicated soil samples from the <i>Allium ampeloprasum</i> rhizosphere under biofertiliser application
LB1, LB2, LB3, and LB4	Are replicated soil samples from bulk soil
PGPR	Plant growth-promoting rhizobacteria

Acknowledgements

OOB appreciates ICGB, Italy and the National Research Foundation, South Africa for the grants supporting research in her lab. OES expresses his gratitude to North-West University for a bursary. BJE is grateful to North-West University for postdoctoral offer.

Authors' contributions

OOB: Validation, conceptualisation, writing–review & editing, funding acquisition, supervision, project management and resources. OES: Data collection, methodology, presentation, drafting– original manuscript. BJE: Writing–review & editing software, formal analysis and data editing.

Funding

The South African National Research Foundation and ICGB, Italy, funded and supported this work based on the research grant from ICGB RESEARCH PROJECT (CRP/ZAF22-03) awarded to Prof. Olubukola Oluranti Babalola.

Data availability

This data note provides open access to all the associated datasets via the NCBI Sequence Read Archive SRP537121 (Data file 1), SRP537120 (Data file 2), and SRP537188 (Data file 3) and Figshare. <http://identifiers.org/insdc.sra:SRP537121> (Data file 1), <http://identifiers.org/insdc.sra:SRP537120> (Data file 2), <http://identifiers.org/insdc.sra:SRP537188> (Data file 3), <https://doi.org/10.6084/m9.figshare.28925831> (Data file 4), and <https://doi.org/10.6084/m9.figshare.28930460> (Data file 5). Details and links to the data can be found in Table 1 and references (Babalola et al., 2025a, 2025b; NCBI Sequence Read Archive, 2024; NCBI Sequence Read Archive, 2024a, 2024b).

Declarations

Ethics approval and consent to participate

The ethical guidelines are followed in this work as required for publishing inside *Springer (BMC)*. There are no human subjects or animal testing involved, nor the use of information from across social media or digital platforms.

Competing interests

The authors declare no competing interests.

Received: 27 March 2025 / Accepted: 21 July 2025

Published online: 06 August 2025

References

- Shahrajabian MH, Sun W, Cheng Q. A review of leek (*A. ampeloprasum* L.), an important vegetable and food ingredient with remarkable pharmaceutical activities. *Pharmacognosy Commun.* 2021;11(1):9–12.
- Xie T, Wu Q, Lu H, Hu Z, Luo Y, Chu Z, Luo F. Functional perspective of leeks: active components, health benefits and action mechanisms. *Foods.* 2023;12(17):3225.
- Babalola OO, Enagbonma BJ. Dataset of amplicon metagenomic assessment of barley rhizosphere bacteria under different fertilization regimes. *Data Brief.* 2024;52:109920.
- Madison JD, LaBumbard BC, Woodhams DC. Shotgun metagenomics captures more microbial diversity than targeted 16S rRNA gene sequencing for field specimens and preserved museum specimens. *PLoS One.* 2023;18(9):e0291540.
- Enagbonma BJ, Babalola OO. Metagenomics reveals the microbiome multifunctionalities of environmental importance from termite mound soils. *Bioinform Biol Insights.* 2023;17:1–10.
- Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta.* 2023;2(2): e107.
- Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9(4):357–9.
- Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods.* 2016;102:3–11.
- Lei Z, Zhang K, Li C, Jiao T, Wu J, Wei Y, Tian K, Li C, Tang D, Davis DI. Ruminal metagenomic analyses of goat data reveals potential functional microbiota by supplementation with essential oil-cobalt complexes. *BMC Microbiol.* 2019;19:1–10.
- Mandal SK, Kambhampati JM, Sharma VS, Sirisha VSL, Sharvani P, Reddy CN, Yadavalli R, Mishra B. Chapter 4 -Gene prediction through metagenomics. In: *Microbial Metagenomics in Effluent Treatment Plant*. Edited by Shah MP. India: Elsevier; 2024: 63–92.
- Commichaux S, Shah N, Ghurye J, Stoppel A, Goodheart JA, Luque GG, Cummings MP, Pop M. A critical assessment of gene catalogs for metagenomic analysis. *Bioinformatics.* 2021;37(18):2848–57.
- Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 2015;12(1):59–60.
- Liu W, Schmidt B, Liu Y, Voss G, Mueller-Wittig W. Mapping of BLASTP algorithm onto GPU clusters. In: 2011 *IEEE 17th International Conference on Parallel and Distributed Systems* 7–9 Dec. 2011. 2011. p. 236–243. <https://doi.org/10.1109/ICPADS.2011.79>.
- Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res.* 2007;17(3):377–86.
- Rao CR. The use and interpretation of principal component analysis in applied research. *Sankhyā: Indian J Stat Series A.* 1964;329–358.
- Oksanen J, Kindt R, Legendre P, O'Hara B, Simpson G, Solymos P, Stevens M, Wagner H. *Vegan: an R package for community ecologists* (R package Version). World Agrofor (ICRAF). 2016;14(41):37.
- NCBI Sequence Read Archive. Genome sequence data of *A. ampeloprasum* rhizosphere soil samples under biofertiliser treatment. In.; 2024. <http://identifiers.org/insdc.sra:SRP537120>.
- NCBI Sequence Read Archive. Genome sequence data of *A. ampeloprasum* rhizosphere soil samples under chemical fertiliser treatment. In.; 2024. [if gte mso 9]> Normal 0 false false EN-ZA X-NONE X-NONE. <http://identifiers.org/insdc.sra:SRP537121>.
- NCBI Sequence Read Archive. Genome sequence data of bulk soil samples. In.; 2024.
- Babalola OO, Shittu OE, Enagbonma BJ. The nucleic acid information of *Allium ampeloprasum* rhizosphere and uncultivated bulk soils samples. *Figshare* 2025. <https://doi.org/10.6084/m9.figshare.28925831.v1>.
- Babalola OO, Shittu OE, Enagbonma BJ. Major Microbiome and functional distributions. *Figshare*; 2025.

Publisher's Note

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