

University of Kentucky

UKnowledge

---

Horticulture Faculty Publications

Horticulture

---

8-2020

## Shotgun Metagenomic Sequencing Data of Sunflower Rhizosphere Microbial Community in South Africa

Olubukola Oluranti Babalola  
*North-West University, South Africa*

Temitayo Tosin Alawiye  
*North-West University, South Africa*

Carlos M. Rodriguez Lopez  
*University of Kentucky, carlos.rodriguezlopez@uky.edu*

Ayansina Segun Ayangbenro  
*North-West University, South Africa*

Follow this and additional works at: [https://uknowledge.uky.edu/horticulture\\_facpub](https://uknowledge.uky.edu/horticulture_facpub)



Part of the [Data Science Commons](#), and the [Horticulture Commons](#)

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

---

### Repository Citation

Babalola, Olubukola Oluranti; Alawiye, Temitayo Tosin; Rodriguez Lopez, Carlos M.; and Ayangbenro, Ayansina Segun, "Shotgun Metagenomic Sequencing Data of Sunflower Rhizosphere Microbial Community in South Africa" (2020). *Horticulture Faculty Publications*. 45.  
[https://uknowledge.uky.edu/horticulture\\_facpub/45](https://uknowledge.uky.edu/horticulture_facpub/45)

This Article is brought to you for free and open access by the Horticulture at UKnowledge. It has been accepted for inclusion in Horticulture Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

---

## Shotgun Metagenomic Sequencing Data of Sunflower Rhizosphere Microbial Community in South Africa

Digital Object Identifier (DOI)

<https://doi.org/10.1016/j.dib.2020.105831>

### Notes/Citation Information

Published in *Data in Brief*, v. 31, 105831.

© 2020 The Author(s)

This is an open access article under the CC BY license. (<https://creativecommons.org/licenses/by/4.0/>)



## Data Article

# Shotgun metagenomic sequencing data of sunflower rhizosphere microbial community in South Africa



Olubukola Oluranti Babalola<sup>a,\*</sup>, Temitayo Tosin Alawiye<sup>a</sup>,  
Carlos Rodriguez Lopez<sup>b</sup>, Ayansina Segun Ayangbenro<sup>a</sup>

<sup>a</sup> Food Security and Safety Niche, Faculty of Natural and Agricultural Sciences, Private Mail Bag X2046, North-West University, South Africa

<sup>b</sup> Environmental Epigenetics and Genetics Group, Department of Horticulture, College of Agriculture, Food and Environment, University of Kentucky, United States of America

## ARTICLE INFO

## Article history:

Received 21 April 2020

Revised 18 May 2020

Accepted 3 June 2020

Available online 11 June 2020

## Keywords:

Illumina HiSeq

Microbiota

Metagenome

MG RAST

Soil

## ABSTRACT

This dataset presents shotgun metagenomic sequencing of sunflower rhizosphere microbiome in Bloemhof, South Africa. Data were collected to decipher the structure and function in the sunflower microbial community. Illumina HiSeq platform using next generation sequencing of the DNA was carried out. The metagenome comprised 8,991,566 sequences totaling 1,607,022,279 bp size and 66% GC content. The metagenome was deposited into the NCBI database and can be accessed with the SRA accession number SRR10418054. An online metagenome server (MG RAST) using the subsystem database revealed bacteria had the highest taxonomical representation with 98.47%, eukaryote at 1.23%, and archaea at 0.20%. The most abundant genera were the *Conexibacter* (17%), *Nocardioides* (8%), *Streptomyces* (7%), *Geodermatophilus* (6%), *Methylobacterium* (5%), and *Burkholderia* (4%). MG-RAST assisted analysis also revealed functional annotation based on subsystem, carbohydrates sequence had 13.74%, clustering based subsystem 12.93%, amino acids and derivatives 10.30%

\* Corresponding author.

E-mail address: [olubukola.babalola@nwu.ac.za](mailto:olubukola.babalola@nwu.ac.za) (O.O. Babalola).

coupled with other useful functional traits needed for plant growth and health.

© 2020 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/>)

## Specification Table

Subject	Microbial Ecology, Genomics and Molecular Biology
Specific subject area	Microbial Biotechnology
Type of data	Raw metagenomics data
How the data were obtained	Bioinformatic analysis of DNA sequences performed with an online server MG-RAST
Data format	Raw data (fastq.gz.file)
Parameter for data collection	Environmental sample, Sunflower rhizospheric soil
Description of data collection	PowerSoil® isolation kit was used to extract DNA from the sunflower rhizospheric soil following the manufacture's instruction after which shotgun metagenomics sequencing was determined using Illumina HiSeq platform
Data source location	Institution: North-West University, Mafikeng, North West Province, South Africa. Samples GPS location (26.296138S: 26.972175E)
Accessibility of data	The raw sequences have been deposited in the NCBI repository with SRA accession number SRR10418054 <a href="https://www.ncbi.nlm.nih.gov/search/all/?term=SRR10418054">https://www.ncbi.nlm.nih.gov/search/all/?term=SRR10418054</a>

## Value of Data

- The data provides information on the rhizosphere microbiome of sunflower
- The microorganisms inhabiting the rhizosphere serves as hotspot for active biomolecules and novel genes
- Understanding the rhizosphere microbiome is important for plant growth and health
- These data are valuable, and this offers the possibilities of identifying new genes which could be an impetus for solving hunger and agricultural sustainability

## 1. Description of Data

The dataset contains a raw sequence data obtained using shotgun metagenomic of sunflower rhizosphere microbiome. The data files in FASTQ format were deposited at the National Center for Biotechnology Information (NCBI) with SRA accession number SRR10418054. The data are presented in [Figs. 1 and 2](#), respectively.

## 2. Experimental design, materials, and methods

Soil samples were obtained from sunflower rhizosphere soils in Bloemhof, South Africa (26.296138S: 26.972175E) and the DNA samples were extracted using the PowerSoil® (MO Bio labs, USA) isolation kit following the instructions of the manufacture. DNA concentration and purity were determined using a NanoDrop Lite Spectrophotometer (Thermo Fischer Scientific, CA, USA). Extracted DNA was sent for metagenome shotgun sequencing to the Molecular Research Laboratory ([www.mrdnalab.com](http://www.mrdnalab.com)), Texas, USA, using the Illumina HiSeq platform. Qubit® dsDNA HS Assay Kit (Life Technologies) was used to determine the initial concentration of DNA. Library preparation was done using the Nextera DNA Flex library preparation kit (Illumina) according to the manufacturer's guidelines. In brief, 50 ng of DNA were used for library preparation. After DNA fragmentation Illumina sequencing adapters were added and products amplified using six

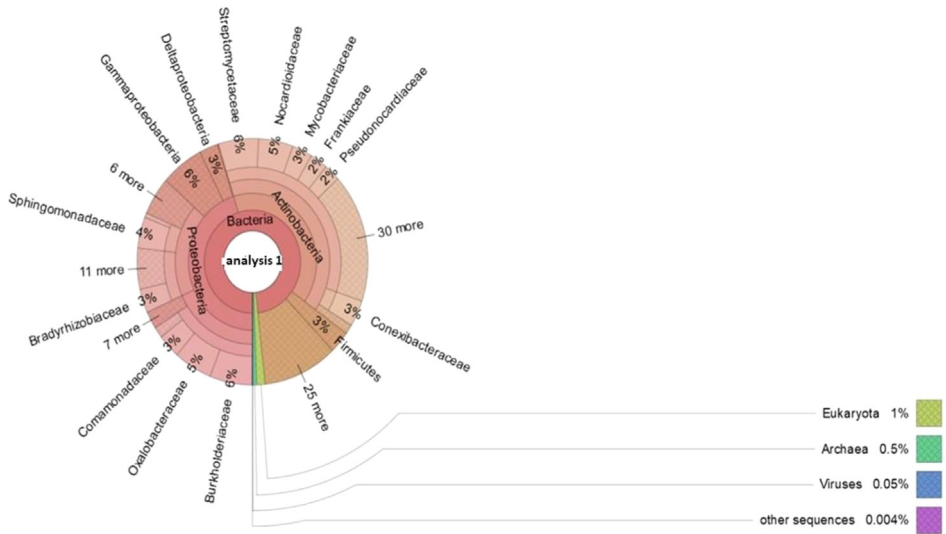


Fig. 1. Structure of the microbiome in sunflower rhizosphere.

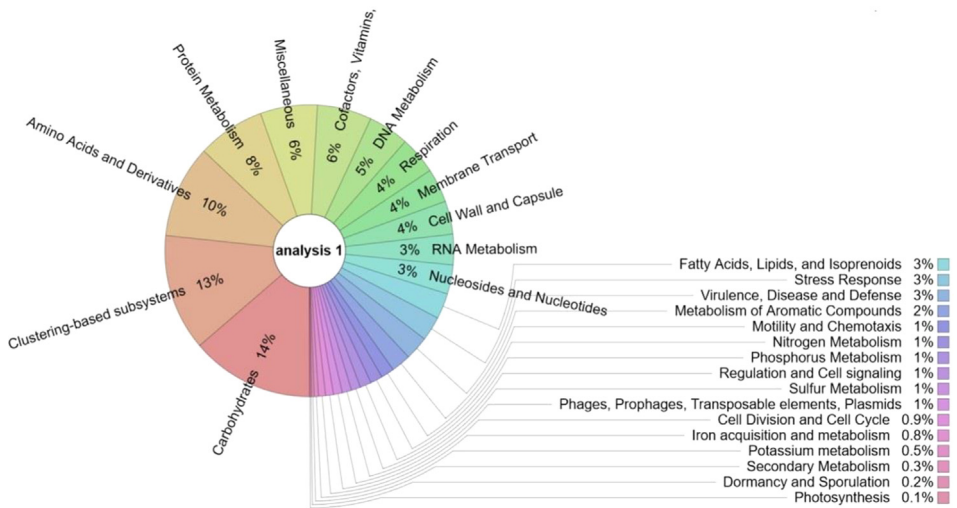


Fig. 2. Functional structure based on subsystem in sunflower rhizosphere soil metagenome.

cycles of PCR during which unique indices were added. After library amplification, their concentration was estimated using the Qubit® dsDNA HS Assay Kit (Life Technologies), while the average library fragment size was measured using the Agilent 2100 Bioanalyzer (Agilent Technologies). Libraries were then pooled in equimolar ratios of 0.7 nM and sequenced paired-end for 300 cycles using the NovaSeq 6000 platform (Illumina).

The online metagenomic rapid annotation server MG-RAST ([www.mg-rast.org](http://www.mg-rast.org)) was used for the quality control of the raw metagenome sequences [1]. After performing quality control (QC), BLAT (the BLAST-like alignment tool) algorithm was used to annotate the sequences [2] against the M5NR database [3], which encompasses non-redundant integration of many databases.

The dataset contains 8991,566 sequences totaling 1607,022,279 bp with an average length of 179 bp. 1432,434 sequences (15.93%) failed to pass the quality control pipeline. Of those, dereplication identified 1046,824 sequences as artificial duplicate reads. Of the sequences that passed quality control, 17,387 sequences (0.24%) contain ribosomal RNA genes, 3722,027 sequences (52.74%) contain predicted proteins with known functions, and 3317,360 sequences (47.01%) contain predicted proteins with unknown function.

## Declaration of Competing Interest

The authors declare that they have no conflict of interest, either financial or commercial wise.

## Acknowledgments

O.O.B. thanks the National Research Foundation, South Africa, for the grant (UID 123634) that has supported research in our laboratory. T.T.A. would like to thank the North-West University for a postgraduate bursary. C.R.L. would like to thank North-West University for the Postdoctoral bursary.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.dib.2020.105831](https://doi.org/10.1016/j.dib.2020.105831).

## References

- [1] F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes, *BMC Bioinform.* 9 (1) (2008) 386.
- [2] W. Kent, BLAT—The BLAST-like alignment tool, *Genome Res.* 12 (4) (2002) 656–664.
- [3] A. Wilke, T. Harrison, J. Wilkening, D. Field, E. Glass, N. Kyrpides, K. Mavrommatis, F. Meyer, The M5nr: a novel nonredundant database containing protein sequences and annotations from multiple sources and associated tools, *BMC Bioinform.* 13 (1) (2012) 141.