Metagenomics survey of major metabolic network of sunflower microbiota

TT ALAWIYE

orcid.org /0000-0002-2202-5749

Thesis accepted for the degree Doctor of Philosophy in Biology at the North-West University

Promoter: Prof OO Babalola

Graduation: May 2021

Student number: 31362834

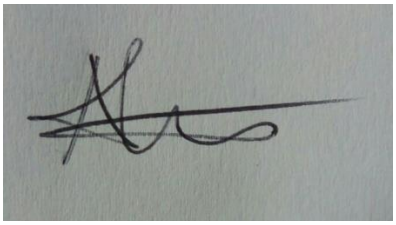
DECLARATION

I, with the signature below declare that this thesis submitted to the North-West University for the degree of Doctor of Philosophy in Biology in the Faculty of Natural and Agricultural Sciences, School of Biological Sciences, and the work herewith is my original work except the citations and that this work has not been submitted at any other University in part or whole for the award of any degree.

STUDENT NAME

Temitayo Tosin ALAWIYE

SIGNATURE



DATE: 17/2/2021

Promoter's name

Professor Olubukola Oluranti BABALOLA

DEDICATION

I dedicate this work to the Almighty God for His favour and kindness over my life and to all who have contributed to the success of this work.

ACKNOWLEDGMENTS

I have received a great deal of support and guidance through this journey. I would like to express my gratitude to people that have helped me thus far. I would like to thank my supervisor, Prof O.O Babalola for her support and persistent encouragement during the course of this study. I also want to thank all members and staff of the Department of Microbiology and staff of the Faculty of Natural and Agricultural Sciences for their support. I also want to thank the North-West University for the postgraduate bursary given to me, it is much appreciated.

I also express my gratitude to the past and current members of Microbial Biotechnology Research Group and friends for all their support. Thank you for the motivation that kept me going. Dr. B.R. Aremu, Dr. C.F. Ajilogba, Dr. O.B. Ojuederie, Dr. O. Emmanuel, Dr. A.S. Ayangbenro, Dr. A.E. Amoo, Dr. N. Igiehon, Mr. Kehinde Odelade, Mr. Atilade Adedeji, Mr. Benjamin Enagbonma, Mr. Mathew Enebe, Mr. Ayomide Fadiji, Mr. Batholomew Adeleke, Mrs. Seun Fasusi, Miss M. Khantsi and Mr. Saheed Akinola all of whose support and friendship are unforgettable. I would also like to thank Miss Oluwabukola Adediji and Mr. Oreoluwa Akano for always being there through thick and thin.

Lastly, I am sincerely grateful to my parents Engr. and Mrs. O. Alawiye. Thank you for your love and support. Also, to my lovely siblings Oyindamola, Adetoun, Olukayode, and Tolulope. I love you all. While this acknowledgment is finite, my thanks are not. I greatly appreciate you all for helping me succeed. I have learnt so much from all of you and without your support I would not be where I am today.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL ABSTRACT

The microbial communities inhabiting the root, termed the rhizosphere, are in a symbiotic association with their host. However, the plant-microbe interaction study at its current stage is still an evolving field of science. Though still largely unexplored, the soil consists of a metabolically active microbiome where microorganisms are abundant. They provide beneficial services to the host plant, such as protection against pathogens and mineral acquisition enhancement, which in the long run help in plant growth and health. Therefore, it is imperative to explore the microbes inhabiting this niche and what they do at the functional level. This study was designed to use high throughput sequencing and computational approaches to study the soil microbiome taxonomical and functional traits induced by the cultivation of sunflower. Sunflower rhizosphere and bulk soil samples collected from Palmietfontein (R1) Bloemhof rhizosphere (R2) and bulk soil Palmietfontein (B1) Bloemhof bulk (B2) in South Africa were used for the shotgun metagenomic sequencing using Illumina HiSeq platform. MG-RAST was used for the taxonomical and functional analysis of the metagenomic sequences. The metagenomes were then mapped against the SEED and KEEG subsystems database where gene descriptions were assigned. The most dominant domains were bacteria, eukaryote, and archaea. In the R1 sample bacteria accounted for 98.82% of the obtained sequences, followed by eukaryote, which accounted for 0.81% and archaea was 0.29%. 98.47% of R2 sequences belonged to bacteria, 1.23% to eukaryote and 0.2% to archaea. Sequences in bulk soil samples were assigned to bacteria 98.61% and 98.53%, eukaryote 0.82% and 1.05%, and archaea 0.48% and 0.34% in B1 and B2 samples, respectively. The most dominant phyla in the rhizosphere, which also shared same features as the bulk soil were the *Actinobacteria*, *Proteobacteria*, *Acidobacteria*, *Bacteroidetes*, *Planctomycetes*, and *Verrucomicrobia*. There was no significant difference (*p* value > 0.05) in the structural

composition of the alpha diversity of the sunflower rhizosphere and bulk soil and also the two locations sampled while there was a significant difference in beta diversity, which was visualized using principal coordinate analysis (PCoA), which explained axis 1 and 2 with a combined variation of 85.90%. The metabolic pathways and functional attributes in the sunflower microbial community were also determined. There was no significant difference (*p* value >0.05) within the functional diversity categories of the sunflower rhizosphere and bulk soils. However, PCoA, which explained the β functional diversity between the sunflower rhizosphere and bulk soil showed a clear separation between them. This was ascertained using the analysis of similarities (ANOSIM), which revealed that there was a significant difference in the functional categories of the sunflower rhizosphere and bulk soils and also there was significant difference in the two locations (*p* value = 0.01; R= 0.58). The PCoA showed that axis 1 and 2 had combined community variation of 52.39%. Principal component analysis (PCA) showed the distribution of the functional categories for both microhabitats. Motility and chemotaxis, photosynthesis, stress response, iron acquisition and metabolism, and cell wall and capsule placed the sunflower rhizosphere (R1) apart from B2, R2, and B1. Strikingly from our study, functional annotation of the genes revealed genes coding for nitrogen fixation (*nifH*), siderophore production, 1-aminocyclopropane-1-carboxylate (ACC) deaminase (*acdS*), mineral phosphate solubilization (*ppx/gppA*), exopolysaccharide biosynthesis glycosyltransferase (*epsF*, *exoF*), and high temperature stress response genes (*htrA*). From previous reports, these genes have been discovered to help in the growth and health of plants. Location R1 was more enriched in genes potentially needed for plant growth and development, which means sunflower would most likely thrive well there. Our canonical correspondence analysis to determine whether physicochemical variables drive the microbiome revealed calcium was the best predictor for structural diversity, while phosphorus was the best predictor for

functional gene categories. These results indicated that the microbial community structure and function were closely correlated with environmental factors. This study shows that microbial consortia have the potential to be efficiently used as bioinoculants to optimize crop growth and health.

LIST OF PUBLICATIONS

Chapter Two: Metagenomic insight into plant microbiota for sustainable agriculture. *This review chapter has been submitted for publication in Rhizosphere*

Authors: Temitayo Tosin Alawiye, Olubukola Oluranti Babalola, Luis Gabriel Wall, Kesen Ma

Candidate's Contribution: performed the literature search and wrote the first draft of the manuscript.

Chapter Three: Bacterial diversity and community structure in a typical plant rhizosphere. *This review is already published in Diversity.* (2019), 11, 179; DOI:10.3390/d11100179

Authors: Temitayo Tosin Alawiye, Olubukola Oluranti Babalola

Candidate's Contribution: performed the literature search and wrote the first draft of the manuscript.

Chapter Four: Metagenomic insight into the community structure and functional genes in the sunflower rhizosphere microbiome. *This chapter has been submitted for publication in PLOS ONE and is under review, has undergone first phase of revision*

Authors: Temitayo Tosin Alawiye, Olubukola Oluranti Babalola

Candidate's Contribution: designed the study, did the literature search, carried out the laboratory work, performed all the analyses, interpreted the results and wrote the first draft of the manuscript.

Chapter Five: Unveiling the metabolic pathways and functional diversity in sunflower rhizosphere microbiome as revealed by shotgun metagenomics. *This chapter has been submitted for publication in Journal of Applied Microbiology and is under review*

Authors: Temitayo Tosin Alawiye, Olubukola Oluranti Babalola

Candidate's Contribution: designed the study, did the literature search, carried out the laboratory work, performed all the analyses, interpreted the results, and wrote the first draft of the manuscript.

CHAPTER ONE

GENERAL INTRODUCTION

1.1 Introduction to this chapter

The microbial component of the plant holobiont, also termed plant microbiota (which is defined as the set of microorganisms in a habitat) or plant microbiome (the collective genomes of these microorganisms) especially in the rhizosphere has important functions in plant growth and health (Zhalnina et al., 2018; Sasse et al., 2018; Compant et al., 2019).

Microbiota play a significant role in soil biogeochemical processes and in turn have effect on soil functions. Therefore, there is a need to provide important insights into the functions of microbiota in the soil. This would comprise a systems-level understanding of community structure and function (Wagg et al., 2019). Plant-associated microbiota benefit from their host because plants release exudates through the root. The rhizosphere is a microbial hot spot considered as one of the most dynamic interfaces on earth and is rich in microorganisms. Thus, it comprises diverse microorganisms that secrete growth-promoting hormones that help in plant growth and health (Alawiye and Babalola, 2019; Trivedi et al., 2020).

Due to the fast and sharp reduction in the cost of next-generation sequencing, metagenomics became common at the turn of the last decade. It is a technology that gathers information on the genetic repertoires of microbial communities from different hosts (Jagadeesan et al., 2019). Metagenomics helps to understand the composition and functional diversity of the microbiome. Thus, shotgun metagenomic sequencing approach can directly sequence the collective genomes in

a given environmental sample, therefore giving important insights into the potential functions of microbial communities from the soil (Sabale et al., 2019).

With a global production of 20-40 million tonnes (List, 2014) and a substantial private sector-breeding effort, particularly for its oil, sunflower (*Helianthus annuus* L) is one of the most important oilseed crops (FAO, 2017). It belongs to the family *Asteraceae* and is grown for its edible oil (Addrah et al., 2020). It is next to soybean and groundnut at the global level in terms of edible oilseed crops (Shilpa et al., 2015). Sunflower is an annual, erect, broadleaf plant with a strong taproot and prolific lateral spread of surface roots. Stems are usually round early in the season and become angular and woody later in the season. It is cultivated for the high oil content of its seeds which represents up to 80% of its economic value. The oil is mostly used for cooking and frying. Sunflower seeds can be used in nut and flour diets, they have medicinal purposes (anti-inflammatory and diuretic effects), the flowers serve as ornamental plants and in some places are used in religious ceremonies (Jocic et al., 2015; Seiler et al., 2017). Sunflower has many industrial uses, such as raw materials used in several foods, fodder, and bioenergy production (Hussain et al., 2018).

Sunflower is a temperate zone crop but it can perform well under various climatic conditions (Iqbal et al., 2020). They are known for their hardiness and the ability to survive extreme heat. However, the optimal temperature for growing sunflower is between 22 and 26°C, with soil temperature of at least 12 to 16°C when the seeds are sown. Temperature during flowering should not exceed 38 to 40°C (Debaeke et al., 2017). The ideal pH for cultivation is between 6.5 and 8.0. Sunflower performs well on a wide range of soils such as sandy loam, black soils, and alluviums but it does best in a well-drained, fertile soil.

Plants and soil microbial communities interact in the rhizosphere for facilitation of nutrient uptake/utilization, stabilization of soil structure, reduction of disease prevalence by out-competing soil-borne pathogens or increase in disease prevalence by microbial pathogen invasion. Soil microbiomes are complex and the factors affecting their composition and biological function are still undescribed despite their economic importance (Rodriguez et al., 2019).

Under the current scenario of global warming and rising world population, it is imperative to gain a better understanding of the drivers of soil microbiome diversity and function (Backer et al., 2018). Microorganisms are found in all types of environment but only a few of them have been isolated and many more are yet to be identified, isolated, and characterized (Ma et al., 2018). The purpose of this study was to investigate the detailed mechanisms that drive the assembly and composition of the rhizosphere microbiota, thereby helping us understand how the rhizosphere microbiomes can be harnessed to improve food production. In this study, we gained insight into the community structures, composition, functionality, and metabolic processes microbial communities play in the sunflower rhizosphere using high through-put sequencing and computational approaches, which confers no biases.

1.2 Problem statement

The world anticipates a rise in its population to 10 billion by the year 2050 and feeding the teeming population amidst decline in the resources coupled with climate change requires optimizing the reliability, resource use, and environmental impacts of food production. One important way to achieve that is to integrate beneficial plant microbiomes (Busby et al., 2017). The soils, which house the microbial communities are among the most complex system known to science and knowledge of them is still limited despite their economic importance. The diversity in the soil

presents a massive but largely unexplored genetic, metabolic, and biological pool and can be exploited for the discovery of novel genes (Cowan et al., 2005).

The rhizosphere has high microbial diversity because of the rich exudates from plant roots. Roots recruit a specific microbial community in this zone which is helpful to the plant. These microorganisms can be used as bioinoculants such as fertilizers and pesticides because they are eco-friendly and not harmful compared to chemicals (Dubey and Sharma, 2019). Using conventional methods of isolating has not been successful since 99% of these microorganisms cannot be cultured in the laboratory. Therefore, this research work uses metagenomics to establish microbial knowledge of the rhizosphere microbiome for improved food security and will help to increase sustainable agriculture.

1.3 Research aim and objectives

The aim of this research is to use shotgun metagenomics to define the structural composition, abundance, beneficial functional genes, and metabolic pathways of the sunflower's rhizosphere microbial communities. To achieve this aim, the research objectives include:

1. Determine the physicochemical characteristics of the sunflower soil samples and the physicochemical properties that drive the sunflower microbial communities

2. Characterize the diversity and community structure of the sunflower rhizosphere microbiome

3. Determine the plant beneficial functional genes in the sunflower microbiome

4. Identify the metabolic pathways and functional diversity of sunflower rhizosphere microbiome

CHAPTER TWO

METAGENOMIC INSIGHT INTO PLANT MICROBIOTA FOR SUSTAINABLE AGRICULTURE

Abstract

Microorganisms are relatively abundant in nature occupying every ecological niche. Plants release exudates through the root, and this attracts microorganisms to themselves, which are often abundant in the rhizosphere. The microbiota thriving at the rhizosphere, which is the root-soil interface play an important role in supporting plant growth and health. Therefore, the interactions between plant, soil, and microorganisms in the rhizosphere are important to investigate. Over the years, with the advent of metagenomics comes an improvement in the study of the plant microbial community. Metagenomic study has revealed microbial knowledge of microbiota for plant growth, development, and health. This review discusses the usefulness of metagenomics approaches to study the rhizosphere microbiota, the importance of metagenomics to microbial ecology, and also highlights the sequencing techniques and their limitations. Metagenomics has developed methods to analyze the microbial communities and to identify numerous gene sequences with the potential for coding enzymes for agricultural purposes. When novel microbial species are discovered through metagenomics, they can be developed using microbial biotechnological techniques and can be used as fertilizers and pesticides, which pose no harm to the environment compared to chemical fertilizers and pesticides. Thus, plant-associated microbiomes can be a potential bio-prospect for agricultural purposes.

Keywords: Ecological niche, microbial ecology, omics techniques, rhizosphere, sustainable food production

2.1 Introduction

One of the several Next Generation Sequencing (NGS) applications is metagenomics. At the time of sequencing the samples, the phyla, genera, and species present there are largely unidentified and the aim of the sequencing is to determine the microbial composition as accurately as possible (Raiyani and Singh, 2020). Metagenomics is studying groups of genomes from a mixed populace of microorganisms with a non-culture based method (Ogier et al., 2020), and the genetic functions of genomes acquired directly from an environmental sample (Astudillo-García et al., 2019). One of the major uses of metagenomics is to generate microbial interaction knowledge which helps to improve human health, food security, and energy production. The heterogeneity, evolution, and phylogeny of an organism in a particular ecosystem can be determined with the use of metagenomics (Perazzolli et al., 2020). Metagenome studies help in the full characterization of a community in order to know essentially what organisms are present and their roles. They have also helped researchers create a scenario of an environment's microbial life with no need for isolating or culturing any organisms (Koutsandreas et al., 2019).

The rhizosphere is a complex ecosystem consisting of a narrow zone of nutrient rich soil that surrounds the plant root. It has high density of microorganisms which could be beneficial, harmful, or neutral. The release of exudates in the rhizosphere attracts and promotes beneficial microorganisms while combating pathogens or harmful ones (O'Banion et al., 2020). In agriculture, the rhizosphere is the most studied habitat due to its enormous potential for plant nutrition and health (Berg et al., 2020). The rhizosphere is rich in microbial species when compared

to the surrounding soils. Modern technologies such as metagenomics provides much deeper insights and expands our knowledge on rhizosphere microbial communities (Bhargava et al., 2019).

Rhizosphere microbial communities have a prominent impact on disease suppression, biotic stress resistance, and nutrition of the plant (Pascale et al., 2020). Beneficial microbiota compete favourably with microorganisms which cause disease or generates microbial agents which improve plant health (de Faria et al., 2020). The plant host habitat has a microbial community which includes the diverse functional gene pool emanating from viruses, prokaryotes and eukaryotes. Such plant habitat ranges from whole single plants to specific organs such as the shoots, flowers, roots, leaves and seeds (Singh et al., 2019a). Bacteria are the most abundant microorganisms in these communities but fungi, oomycetes, algae, protozoa, nematodes and viruses are also important contributors (Patra et al., 2018; Compant et al., 2019).

It is estimated that the world will have up to 10 billion people in the year 2050 (Gerten et al., 2020). It is imperative to feed this emerging population; thus, agricultural productivity will need to be increased within the next few decades. To produce more food for the populace, according to Ferreira et al. (2019), we would require more agricultural land, more capital for smallholder farmers, effective agricultural inputs (good planting materials, fertilizer, herbicides, pesticides etc), improved mechanization and extensive use of plant growth promoting microorganisms (PGPMs). This could be achieved through the means of microbial biotechnology. PGPM are organisms in the soil that enhance acquisition of nutrients, tolerance to abiotic stress, hormone production, and pathogen protection, which help in plant growth promotion.

PGPM encompass plant growth promoting rhizobacteria and plant growth promoting fungi (Jain et al., 2020). They can be used as microbial fertilizer, which is gaining a lot of attention due to the undesirable impacts of over dependence and improper use of chemical fertilizer, and also the increased consciousness in the relationship that exists between rhizosphere microorganisms and their host plants (Igiehon and Babalola, 2017; Elliott et al., 2020). Hence, the use of metagenomics is an additional powerful effective tool for discovering unknown microorganisms and to track the plant/soil microbiome, in order to choose the best and most sustainable agricultural practices. This review aims to explain the metagenomic approaches for studying the rhizosphere plant microbiota, the sequencing techniques, and their limitations, stating the benefit for agriculture.

2.2 Plant microbiome in the rhizosphere

The microbiome is the complete set of microorganisms of a particular habitat (Fierer, 2017; Saleem et al., 2019). Rhizospheric microorganisms consist of a diverse population including fungi, bacteria, and archaea (Lee et al., 2019). In the rhizosphere, the microbiota has different functions such as regulating soil fertility, impacting on plant performance and in biogeochemical cycling (Dubey et al., 2019). The microbiota found in the rhizosphere helps the plant in the acquisition of nutrients, production of phytohormones, phytotoxic compound degradation, biotic and abiotic stress tolerance, (Figure 2.1) and for pathogen suppression (Compant et al., 2019; Pandey and Gupta, 2019). The microbial community (microbiota) and plant host has a dual relationship, which is an essential determining factor in plant health and productivity (Tkacz et al., 2020).

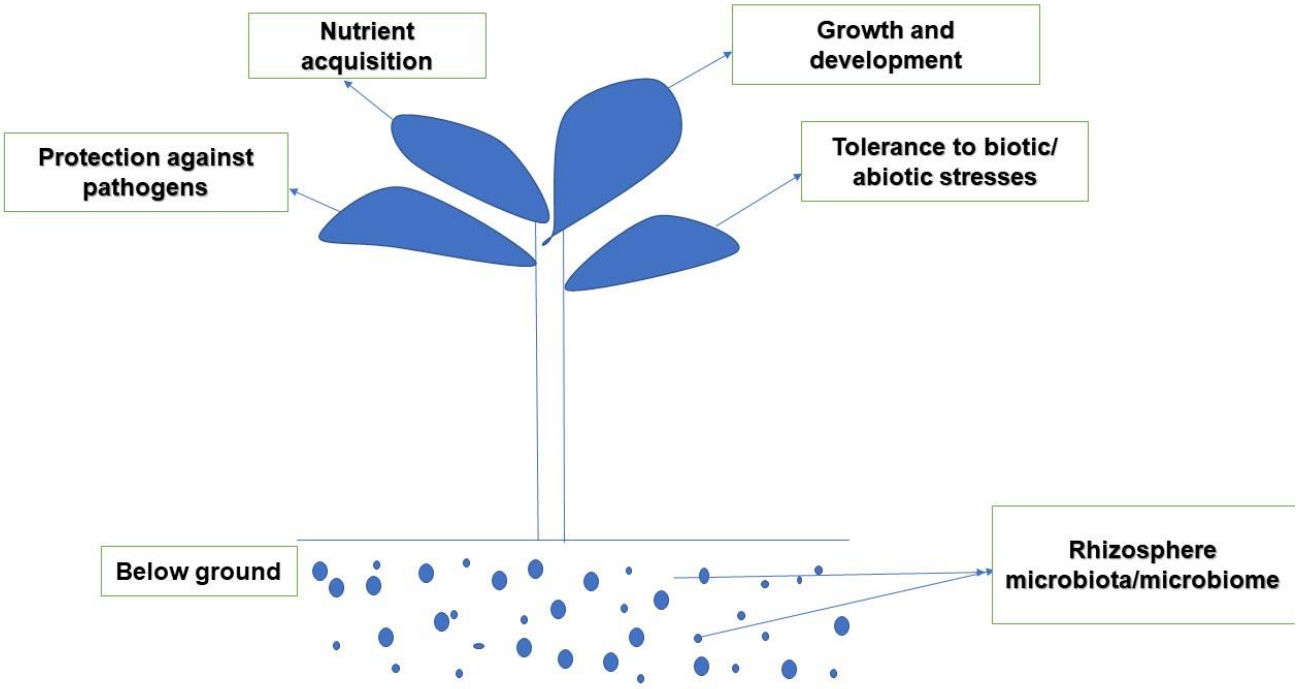


Figure 2.1: Effect of rhizosphere microbiota/microbiome on crop production

The plant microbiome contains a diverse yet uncharacterized microbial community. Insight into the microbial community can help enrich our knowledge of plant microbial ecology and their interaction within the community (Bhargava et al., 2019). Over the past two decades, phylogenetic surveys estimated that 99% of the microbial species are not yet cultured compared to the 1% of the known cultured prokaryotes in a single plant microbiome (Berg et al., 2020). The recent advances in metagenomics using next generation sequencing technologies have helped in the further investigation of the functional genetic diversity of different microbes. Metagenomics has the ability to discover microorganisms at the genetic level of diversity without biases that come with manual cultivation that does not give the wholesome representation of the microbiome (Malla et al., 2019). The soil microbial population structure and functions are likely to be influenced by environmental requirements and disturbance and these requirements could change the soil

characteristics (Igiehon and Babalola, 2018). Management practices in agriculture also modify soil microbial community structure (Zhao et al., 2016; Faggioli et al., 2019; Higo et al., 2020).

Almost all ecological niches are certain to have microorganisms in them. Microorganisms found on a plant surface are known as epiphytes, those that inhabit the inside of plant tissues are known as endophytes (de Faria et al., 2020). A plant has three distinct habitats for the microbiota: the rhizosphere is the first habitat, where the soil encloses the root of a plant in which the chemistry and microbiology are controlled by root growth, exchange of nutrients, and respiration (Dong et al., 2019). The rhizosphere can further be classified into three main zones namely the endorhizosphere, which is the microbiota colonizing the rhizosphere endophytic compartment, and the rhizoplane is the rhizosphere surface where microorganisms and soil particles adhere, and the third zone is the ectorhizosphere consisting of soil immediately adjoining the root. The relevance is because the microbiological structure of microorganisms is different from one plant compartment to the other (Scavo et al., 2019). The second habitat is inside the plant tissues known as the endosphere (Compant et al., 2020), and the third is the phyllosphere which includes the leaves and stem surface. Nutrient and water are limited in the phyllosphere which makes microorganisms not to be abundant, coupled with excessive sun radiation found on the leaves and stem surfaces. Hence, the phyllosphere is regarded as a hostile habitat for microorganisms (Dong et al., 2019). The habitat of the rhizosphere microbiota is of utmost importance because most activities take place there. In the rhizosphere, bacteria are usually the dominant class of microbiota with relative abundance of 97%, fungi and archaea are usually 1% and 0.1% respectively. Bacteria such as *Pseudomonas*, *Agrobacterium* (Naqqash et al., 2020), *Alcaligenes* (Hussain et al., 2020), *Bacillus* (Kumari et al., 2019) are some of the prevalent rhizospheric microbiota.

In a genomic study by Adegboye et al. (2018), new strains of *Streptomyces* were discovered from the rhizosphere of a maize plant in North West Province of South Africa. These strains are known for various metabolic potential, which include the ability to biosynthesize a wide range of antibiotics and ability to degrade certain compounds.

2.3 Metagenomic technology approach to the rhizosphere microbial communities

Plants can no longer be considered as stand-alone entities and a more holistic approach is needed in terms of the abundance, diversity, and beneficial effects of associated microbiota to the host plant (Berlanas et al., 2019). In the study of natural products from soil microorganisms, Handelsman and her colleagues were the first to invent the word metagenomics (Handelsman et al., 1998). This seemingly new aspect in life science has been used to study microbial populations in diverse ecosystems ranging from soil, water, oceans, mining sites, and even humans (Prayogo et al., 2020). Metagenomics gives access to microbial communities and this assists in knowing the functional genes responsible in that community (Ambrosino et al., 2019). It can also be used to detect pathogens in crops. Nhlapo et al. (2018) used NGS approach to detect viruses affecting sweet potato without previous knowledge.

The fast and relative reduction in the cost of next-generation sequencing has advanced the development of metagenomics (Malla et al., 2019). As stated by Sugathan et al. (2017), and (Cantalupo and Pipas, 2019), the first step in metagenomics is experimental design followed by sampling, extraction of DNA, sequencing of the DNA extracted, assembly, annotation (binning), statistical analysis, data storage and data sharing.

The most widely used NGS technology is the Illumina system. Though attention is shifting to PacBio and MinION (Nanopore) technologies (Loit et al., 2019). The new sequencing technology such as PacBio and Nanopore are the latest third generation sequencing approach. The major advantage of the third generation over Illumina system is that it completes its reads in 4-6 hours rather than days, with the ability to obtain ultra-long reads of up to hundred thousands of bases (Amarasinghe et al., 2020), it is relatively low cost, connects to a laptop, and does not require specialized computer training or equipment for data analysis. This sequencing technique is useful in generating microbiota profile derived from full length 16S rRNA gene sequence at a relatively high taxonomic resolution at the species level (Kumar et al., 2019b).

MinION (Nanopore) technology could expand microbiome analyses, including the promise it holds for plant agriculture due to its portability, and its affordability, although the error rate is high, which means that it may be limited to particular applications or need to be further refined (Loit et al., 2019). Metagenomics sequencing has two approaches which are amplicon sequencing also known as targeted sequencing, and whole-genome sequencing, which is also called shotgun metagenomics (Rausch et al., 2019).

2.3.1 Amplicon sequencing

Amplicon sequencing targets a specific marker gene which is found in various organisms (Brumfield et al., 2020). It helps to identify members in a given microbial community. Bacteria and archaea are targeted using the 16S rRNA marker genes while ITS (Internal Transcribed Spacer) and 18S rRNA are used to target fungal species (Boers et al., 2019; Papik et al., 2020). In a study by Chen et al. (2017), soybean and maize, which are energy crops were used in the remediation of cadmium polluted soils. The rhizosphere community was studied using 16S rRNA

gene sequence using the illumina MiSeq sequencing technique. The outcome showed the importance of bio-energy cropping in the remediation of cadmium polluted soils by altering the topological functions of individual OTUs (Operational Taxonomic Unit) and keystone population. *Proteobacteria* and *Acidobacteria* were the keystone bacteria connecting different co-expressed OTUs thereby explaining the effect of bioenergy cropping on microbial interactions in the remediation of soil contaminated with cadmium.

Several steps must be taken to effectively analyse targeted metagenomics data after extracting DNA: 1) fragments should be amplified by PCR, for example, 16S rRNA region marker genes or ITS marker genes; 2) the second step is a short DNA sequence barcode, which is added to each read during PCR amplification (Gołębiewski and Tretyn, 2020), while the last step is data analysis, which comprise generation of OUT, which is specie distinction in microbiology. OTU can be achieved mainly by using three strategies, which are de novo, closed-reference, and open-reference OTU picking. Sequences are clustered into OTU based on similarity, this method is called de novo OTU picking. In a situation where the diversity of microbiota is not accurate based on arbitrarily defined OTU, the OTU analysis may not reflect the true and biologically meaningful microbial composition. Hence, aligning sequences directly to a closed reference OTU picking is done. Closed-reference OTU picking is a potential choice to avoid the difficulty of de novo OTU picking. However, open-reference OTU picking is carried out because ideally, closed-reference OTU picking cannot identify novel phylotypes as a result of excluding sequences that are not matched to the reference database. Therefore, additional de novo OTU picking is performed for those sequences that failed to match the reference. Open-reference OTU picking can solve this problem while determining the taxonomic analysis using amplicon sequencing (Li et al., 2018). A reference database will be needed during analysis to find the closest match to an OUT, which will help infer

the taxonomic lineage. Some of the commonly used databases for 16S are Silva (Glöckner, 2019), Greengenes (Myer et al., 2020), while Unite (Nilsson et al., 2019) can be used for ITS. Some groups of microorganisms like protists and viruses are highly diverse therefore these databases are not available for them due to the fact that no sequence information is available on these databases, unlike for bacteria and fungi (Breitwieser et al., 2019).

Amplicon sequencing has some disadvantages which affect the outcome of metagenomics data (Rausch et al., 2019). Some of the disadvantages that can be encountered during PCR and sequencing include biases due to the failure to capture the most important part of the diversity in the community (Pérez-Cobas et al., 2020). Errors can occur during sequencing (e.g chimeras) and this can generate false sequences that are usually not easy to identify (Pfeifer, 2017). Quality trimming, error correction, and read overlapping can be used to handle sequencing errors. When quality trimming is taking place, the average quality scores are computed over a sliding window across the whole reads. At the beginning of the read, any quality score that is larger than a particular threshold is trimmed, likewise the same is done at the end of the read whereby any score that lies beneath the threshold is also trimmed. Error correction will undergo a test with a program known as BayesHammer (Schirmer et al., 2015; Hanussek et al., 2020). This method is a form of construction on a Hamming graph based on the k-mer general makeup of the reads. Another way in which sequencing errors can be corrected is by overlapping paired-end reads. The reads are aligned, and the optimal overlap is considered after that, error correction and the reads are assembled into a single sequence. Thus, the quality scores are used for aligning the reads as well as for the error correction (Hanussek et al., 2020). Therefore, there is no possibility of solving the biological functions of the taxa because expectedly, amplicon sequencing only gives insight into the taxonomic composition of the microbial community (Mallick et al., 2019).

2.3.2 Shotgun metagenomics

Shotgun metagenomics (SGM) gives insight into the whole microbial community as well as the potential functions of the microbiota in those communities (Berg et al., 2020). After extracting DNA from cells in a microbial community, it is then sheared into tiny fragments and sequencing will be carried out independently, unlike amplicon sequencing which targets a certain genomic locus for amplification. The resulting reads will align with diverse genomic locations for the myriad genomes available in the sample. Hence, SGM gives the understanding of two important functions of a microbial community which is ‘‘who is there’’ and ‘‘what they can do there’’ (Fadiji and Babalola, 2020).

A standard metagenomic study consists of several steps which include: sample collection, processing, and sequencing, sequenced reads which will be pre-processed, analysis of the sequenced reads which will profile the taxonomic functional and genomic features of the microbiome, post-processing analysis which includes statistical and biological analysis and lastly validation, this is an important follow up in analyses because conclusions from high-dimensional biological data are likely to be affected by study-driven biases which are the biases during extraction and microbiome data analysis, thereby necessitating follow-up analyses (Bharti and Grimm, 2019; Fadiji and Babalola, 2020).

In research conducted by (Chaudhari et al., 2020), the taxonomic and functional composition of pea rhizosphere microbiome (Table 2.1) was compared to the bulk soil. A specific microorganism was observed in the rhizosphere which is just a portion of the taxonomic and functional diversity found in the bulk soil. Studies through metagenomics have shown the detection of some functions that are predicted to be associated with growth promotion and nutrition which are of importance

to the plant (Chaudhari et al., 2020). Shotgun metagenomics can be grouped into two, namely read-based and assembly-based metagenomics. Read-based metagenomics gives a proper explanation into taxonomy makeup of a sample which determines the presence or absence of an organism, and genes or metabolic pathways. As a result, read-based metagenomics is only related to taxonomy and functions. In assembly-based metagenomics, the reads are *de novo* assembled into contigs, thereafter following a binning process they are clustered into genome bins which helps greatly in alignments to predict the functions of individual genes and operons (Nelkner et al., 2019). It is possible to identify organisms up to the strain level in shotgun metagenomics study (Buytaers et al., 2020).

Table 2.1: Analysis using NGS techniques on rhizosphere microbiomes

Sequencing techniques	Habitat	Plant type	Major observations	References
Illumina	Rhizosphere	Pea (*Pisum sativum)*	Higher abundance of genes beneficial for plant growth promotion and several other genes were reported in the rhizosphere samples in comparison to the bulk soils. Major predicted functional cores observed in the rhizosphere community compared to the bulk soil were iron complex outer membrane receptor, cobalt-zinc-cadmium resistance, sigma-70 factor, and ribonuclease E.	(Chaudhari et al., 2020).
Roche 454	Rhizosphere	Kodo millet (*Paspalum scrobiculatum*)	Metagenomics revealed taxonomic communities with functional capabilities that have the potential to support growth	(Prabha et al., 2019).

			and development of plants under nutrient-deprived, semi-arid, and dry region. The results revealed sixty-five taxonomically diverse phyla identified in the rhizobiome. *Actinobacteria* and *Proteobacteria* were the most abundant and the metabolic functional capabilities include carbon fixation, nitrogen, phosphorus, sulfur, iron and aromatic compound metabolism, stress response, secondary metabolite synthesis and virulence, disease, and defense.	
Illumina Miseq	Rhizosphere	Soybean (*Glycine max*)	Functional categories potentially needed for plant growth most like potassium metabolism was stable in the rhizosphere when compared to the bulk soil.	(Goss-Souza et al., 2019).
Illumina	Rhizosphere	Sorghum (*Sorghum* bicolor)	Genes encoding nitrogenase reductase *nifH* and dinitrogenase *nifDK* were assigned to *Bradyrhizobium*. These are important genes potentially needed for N2 fixation and nodulation	(Hara et al., 2019).
Illumina	Endophyte	Maize (*Zea mays*)	Twenty-eight different phyla and nutrient pathways were identified. The major phyla were *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Proteobacteria*, *Acidobacteria*, *Chloroflexi*, *Verrucomicrobia*, *Tenericutes*, *Planctomycetes*,	(Fadiji et al., 2020).

	Cyanobacteria, and *Chlorobi*. The nutrient pathways, nitrogen, phosphorous metabolism, secondary and iron metabolism relating to potential growth of plant were also identified

2.4 Analysis of metagenomics data

There are several tools available for researchers to analyse metagenomes to decipher and characterize microbial community diversity and functions (Almeida and De Martinis, 2019). Analysis of gene abundances of different samples will provide insights into the functional and taxonomy changes in a community. Therefore, in order to understand the functional diversity of a sample, quantification of genes, and statistical analysis of the resulting gene abundance is crucial. The quantification of genes is a crucial step to know, understand, and compare metagenomes. This process consists of three main steps: i) quality assessment and filtering of the input data; ii) alignment of the reads to a reference database; and iii) evaluation of the gene abundances. (Boulund et al., 2018).

Metagenomic data are usually large and require sophisticated bioinformatic tools to analyse and interpret them. Different sequencing approaches require different tools for their analysis (Almeida and De Martinis, 2019). Relative progress has been made in the field of metagenomics where new tools are developed frequently and are made available for analysis. For shotgun metagenomics, quality control is performed (Figure 2.2) on the sequenced data where the first step is trimming and filtering of the data using trimmomatic (Almeida and De Martinis, 2019). The reads will then be assembled with MEGAHIT (Metagenomics of Human Intestinal Tract) (Roux et al., 2017; Almeida et al., 2019), or MetaSPAdes (Tikariha and Purohit, 2019) to mention a few. When

assemblies are done, the outcome is sequence fragments which consist of varying length. The purpose of binning is to group the fragments into bins using various methods to cluster the related fragments (Zhang et al., 2019b). Binning, which is assigning metagenomic sequence to a taxonomic group has a significant role in the analysis of metagenome data because it provides knowledge on the availability of new genomes that are usually not easy to identify, and also determines the prominent members and types of taxa available in a microbial community (Nelkner et al., 2019). Some of the tools used for binning are MEGAN (Metagenome Analyzer) (Bağcı et al., 2019), CONCOCT (Kouchaki et al., 2019), and MetaBAT (Kang et al., 2019). Thereafter annotation is carried out using Bowtie (Thai et al., 2019). It can also be performed on a dedicated analysis platform like MG-RAST (Metagenomic Rapid Annotation using Subsystem Technology) and IMG/M (Integrated Microbial Genomes & Microbiomes) where annotation components are provided (Hong et al., 2019). The prepared reads will then be mapped against a reference genome in the database so as to allow the microbial population structure (taxonomy) and function (gene annotations) to be identified (Breitwieser et al., 2019). Some of the databases that can be used as reference genome are KEGG (Kyoto Encyclopaedia of Genes and Genomes) (Szklarczyk et al., 2019), UniProt (Consortium, 2019), and RefSeq (Rajput et al., 2019).

Metagenomics data analysis demands powerful state-of-the-art tools with large performance computing capacity. The absence of adequate facilities usually pose difficulty for researchers handling this field of study (Boers et al., 2019).

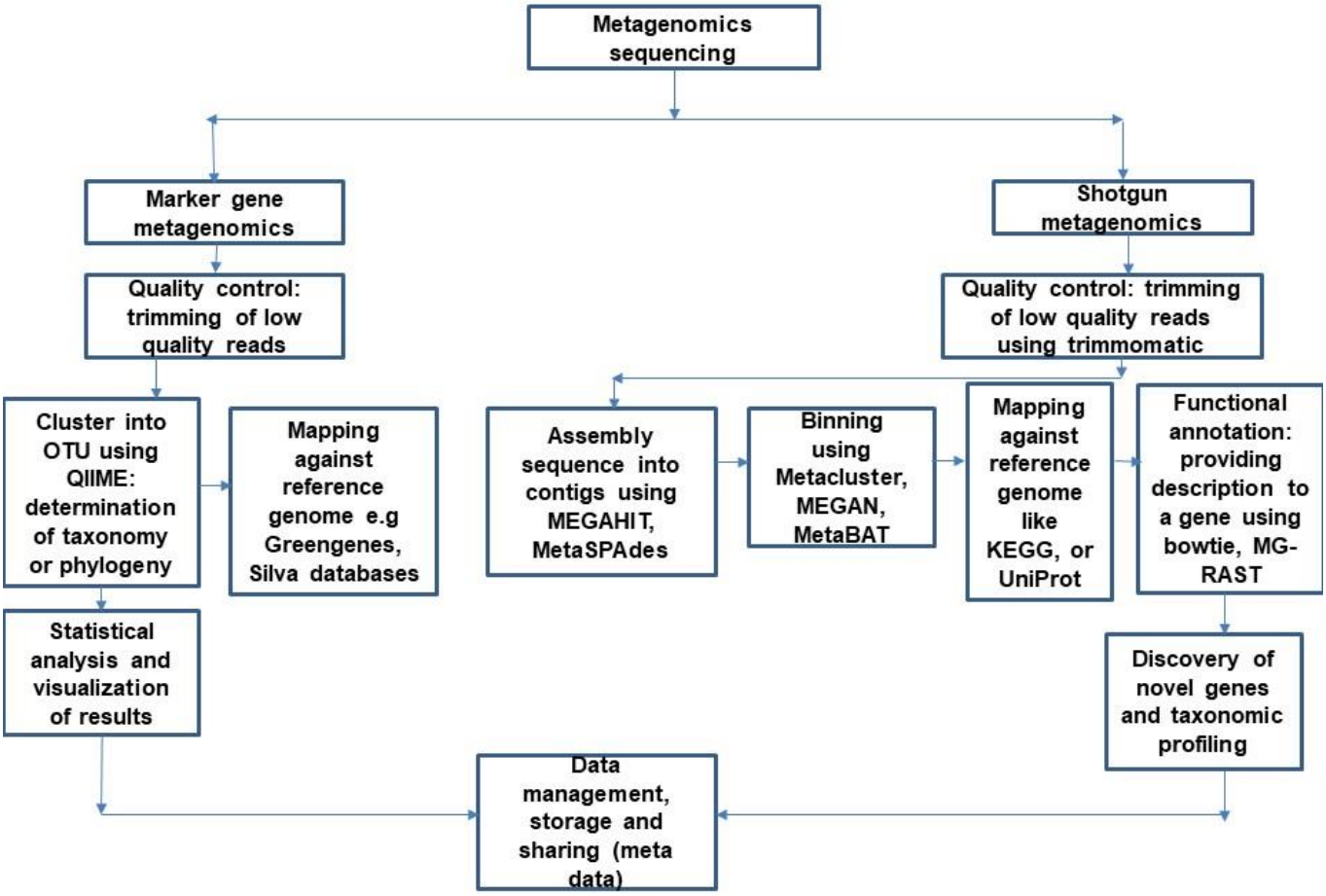


Figure 2.2: Schematic overview showing the major procedures in shotgun and 16S sequencing analysis

2.5 Metagenomics: the current innovation for increased food production

Sustainable and environmentally friendly food production is a long-term goal to take account of increasing needs in food production for the emerging populace. This makes agricultural researchers to continually develop new innovations such as alternative agriculture, beneficial microorganisms, soil quality, integrated pest management, conservative agriculture, integrated

nutrient management, and climate-smart agriculture (Mendoza-Arroyo et al., 2020; Meena et al., 2020), and in recent times genomic studies. Metagenomics sheds light on the underlying factors that determine the genes responsible for the transportation of signal molecules at the plant microbiota level. Hence, it has the potential to identify traits that maximize the benefits of modern agricultural practices and manipulate the plant-microbe towards achieving sustainable agriculture (Singh et al., 2019b). A metagenomic approach in agriculture has enabled us to determine and acknowledge the diversity of microorganisms available in the soil (Wolińska, 2019). These approaches are useful in determining the hidden diversity of microbiota inhabiting the soil and the surrounding rhizosphere. These methods not only identify the hidden microbiota in the community but they also elaborate the molecular bases of the plant-microbiome interactions (Kumari et al., 2019). Metagenomics has been used recently to determine changes in the environment in connection to fertilization and agricultural management; this revealed a shift in taxonomic composition and functional redundancy of microbial communities in the soil. Reports have shown that in the practice of different agricultural management in relation to fertilization, *Proteobacteria* represent the dominant phylum. The most prevailing classes were *Alphaproteobacteria* (51.1%), *Betaproteobacteria* (20.8%), *Deltaproteobacteria* (19.6%) and *Gammaproteobacteria* (8.55%) (Purohit et al., 2018). Until recently, the microbial communities have been difficult to study, but metagenomics has broadened our knowledge about soil microorganism structures and functions (Wolińska, 2019), thereby making us understand the crucial roles microorganisms play in the soil. These are roles such as growth (plant growth promoters), nitrogen fixation, nutrient assimilation, plant stress tolerance and health status of the plant (Lobo et al., 2019). Therefore, detailed understanding of the microbial diversity, composition, abundance, functional genes, and

metabolic pathways at genome level could be an impetus to understanding the contribution of the microbial community towards plant growth and health (Fadiji and Babalola, 2020).

2.5.1 Plant microbe interaction: Biofertilizer

Plant microbiota interaction is a symbiotic interaction which plays an important role in biofertilizer to boost plant growth. Plants need nitrogen for their growth, and this is why it is the most necessary nutrient. Plant growth promoting rhizobacteria (PGPR) are bacteria that enhance the growth of plants either in a direct or indirect fashion. Some processes of PGPR mechanism are nitrogen fixation, phosphate solubilisation, and phytohormones production (Verma et al., 2019). The direct effect of PGPR on plant growth includes the acquisition of nutrients such as iron, phosphate and fixed nitrogen from the environment or by regulating the growth of the plant through altering the levels of hormones such as ethylene, cytokinin, and auxin (Oleńska et al., 2020). PGPR secrete various hormones, siderophore, and hydrogen cyanide (HCN) production, all of these are used for plant growth and a better yield (Kumar et al., 2019a). The indirect outcome of PGPR on the growth of a plant is the process by which inhibitory substances are produced by bacteria to control pathogenic organisms which could either be fungi or bacteria and give the host plant a strong natural resistance preventing the plant from pathogen attacks (Alawiye and Babalola, 2019). There are two types of PGPR namely intercellular (iPGPR), and extracellular (ePGPR). The iPGPR are usually found in a distinct nodular structure of the root cell while the ePGPR may be located either on the rhizoplane, in the rhizosphere or in between the spaces of root cortex cell (Mondal and Sarkar, 2019). The bacterial genera that belong to the ePGPR group are *Agrobacterium, Azotobacter, Caulobacter, Azospirillum, Arthrobacter, Serratia, Bacillus, Erwinia, Flavobacterium, Burkholderia, Chromobacterium, Micrococcous,* and *Pseudomonas* (Lobo et al., 2019; Kenneth et al., 2019). *Mesorhizobium, Allorhizobium, Bradyrhizobium, Rhizobium,* and

Frankia species are the iPGPR that are classified into the Rhizobiaceae family. Both Endophytes and *Frankia* species possess the capability to fix atmospheric nitrogen symbiotically with higher plants (Adhya et al., 2018; Otlewska et al., 2020).

2.5.2 Plant microbe interaction: Plant health status and nutrient assimilation

Agricultural and biological input such as plant extracts, microbial inoculants, and beneficial insects are plant beneficial products which can be harnessed by growers to increase food production and improve the health of plants (Olanrewaju et al., 2019). Many areas in the world are faced with the unavailability of plant nutrients, which is a major restriction to plant growth. Organisms in the soil interact and produce exudates in which plants put into use most of the mineral nutrients produced (Pascale et al., 2020). The plant exudates and nutrients interact in the soil and thus exert an influence on the microclimate of the rhizosphere (Vives-Peris et al., 2020).

Interactions among the biotic components (soil, plant, and microbiota) is an important tool for improving agricultural systems which are overly dependent on chemicals (Berlanas et al., 2019). The interaction between the microbiota and the plants are usually diverse and are dependent on the chemical signals being exchanged between the two. The chemical signals are the metabolites and substrates produced by the plant and organisms (Vives-Peris et al., 2020). All species of plant, their different soil types and associated rhizosphere can be colonized by unique microbial communities.

Plant hormones, which include auxin, cytokinin, and volatile growth stimulants like ethylene and 2, 3-butanediol help to control a large number of pathogens (Verma et al., 2019). Siderophore is produced by *Pseudomonas spp* which possess antifungal antibiotic used in managing plant diseases (Tariq et al., 2020). Different species of *Pseudomonads* have produced various antibiotics

such as pyocyanin, phenazine-1-carboxylic acid (PCA), 2-acetamidophenol, pyrrolnitrin, pyoluteorin, 2,4-diacetylphloro-glucinol, viscosinamide and tensin. *Pseudomonas aeruginosa strain* MF-30 isolated from the rhizosphere soil of maize revealed beneficial metabolite such as indole-3-acetic acid (IAA), siderophores, phosphatase and protease, which show antifungal properties against *Rhizoctonia solani* (Singh et al., 2020). The characteristics of *Pseudomonas spp* are: production of a wide spectrum of metabolites that are active biologically (for instance, antibiotics, siderophores, volatile and growth- promoting substances).

On top of all of these considerations, there is a need to revise our models with which we understand plant nutrient acquisition in soil and natural conditions, including plant microbiome as part of the system that drives plant nutrition. Our actual model is based on plant absorption of mineral nutrients from the soil solution; thus, the mineralization process is a key step in the plant/crop nutrition in the soil. The plant microbiome is absolutely absent in that model because of historical reasons. Knowledge of nitrogen fixers and phosphorus solubilizers are key examples that can be managed to improve plant functioning and management of the plant microbiome in order to make the nutrient acquisition and fertilizer use more efficient and environmentally non-risky. This old model (which general agriculture still relies on) is based on plant physiology concepts developed considering the plant as a single and isolated organism. Our present knowledge of microbiomes and the view of the plant as a microbial ecosystem cannot be avoided any longer.

2.6 Future prospect

Metagenomics application can help improve our understanding in discovering new genes, bio-products, and plant growth promoting microorganisms, which when harnessed can further increase food production and sustainability (Rodriguez and Durán, 2020). This field of study will champion

the cause of agriculture by improving the ability of microorganisms to control pathogens, enhance the capability of microorganisms to ameliorate the negative effects of osmotic stressors such as drought and salinity, and also improve the ability of microorganisms in the phytoremediation of soil contaminants (Lobo et al., 2019).

Even with the novel phylogenetic and functional insight metagenomics brings, an information gap still exists in this field of science (Berg et al., 2020). The microorganism can be a very useful tool in solving the majority of our agricultural problems so the development of metagenomics will provide an opportunity for obtaining beneficial microbial resources from these unculturable microorganisms by identifying their pathways and developing suitable media for the yet to be cultured microbes (Ambrosino et al., 2019). Metagenomics is not without limitations despite its many benefits. Some of the limitations are generating millions of sequence reads which are difficult to analyze and also time consuming even with the availability of state of-the-art bioinformatic tools. None the less, it is still a good innovation for studying several complex ecosystems (Wolińska, 2019). Therefore, it is believed that scientists will find a way to solve these limitations in the near future. Application of this technology may lead to discovery of other novel microorganisms and promoting of significant increases in microbial culturability (Ma and Kim, 2018), which will shape new questions for the future.

Conclusion

There is a question that is frequently overlooked. The knowledge of the plant microbiome opens the question about revisiting the plant physiology regarding nutrition and acquisition of nutrients from soil, in the presence of the soil microbiome. Most probably this knowledge will change our way of understanding agriculture and the use of fertilizers for plant/crop nutrition. There are

various functional attributes used by microorganisms which help in the growth of the plant for improved agriculture. These microbiota are either beneficial or harmful to the plant. Hence, there is a need to study them extensively. Microbiota have various functional attributes in the growth of a plant which include nitrogen fixation, phosphate solubilization, potassium solubilization, production of lytic enzymes, antibiosis etc. Despite of all these encouraging facts being discovered by researchers, there are many microorganisms which have not been explored because they are unculturable. Thus, metagenomics can be used to determine what organisms are in the soil and what they are doing in the community. More functional analysis of microbes in the soil should be further harnessed to help improve plant yield through sustainable agriculture. Thus, this knowledge is important for industrial application such as bio-fuel and discovery of new bio-products including pesticides, fertilizers, and even antibiotics. Therefore, effective molecular techniques such as metagenomics should be encouraged to give a better insight into the structures and functions of the rhizosphere microbial diversity. Application of this technology may lead to discovery of other novel microorganisms that will shape the discussion for the future. We should always remember that all these metagenomic analyses with big and complex datasets are just single pictures at a particular moment of a dynamic process that will continue for a long time with changes at different scales, that is plant biology or just simply, life.

CHAPTER THREE

BACTERIAL DIVERSITY AND COMMUNITY STRUCTURE IN A TYPICAL PLANT RHIZOSPHERE

Abstract

Bacteria play a vital role in the quality of soil, health, and in the production of plants. This has led to several studies in understanding the diversity and structure in the plant rhizosphere. Over the years, there have been overwhelming advances in molecular biology, which have led to the development of omics techniques, which utilize RNA, DNA, or proteins as biomolecules; these have been gainfully used in plant-microbe interactions. The bacterial community found in the rhizosphere is known for its colonization around the roots due to availability of exudates, and their composition that affects plant growth directly or indirectly. Metabolic fingerprinting takes a snapshot of the metabolic composition at a given time. We review metabolites with ample information on their benefits to plants which are found in the rhizobacteria such as *Pseudomonas* spp. and *Bacillus* spp. Exploring plant-growth-promoting rhizobacteria using omics techniques can be a true success story for agricultural sustainability.

Keywords: food security, metabolite, omics techniques, plant-microbe interaction, plant growth

3.1 Introduction

Soils are home to a variety of bacteria, be it neutral, negative (pathogenic), allelopathic or beneficial (symbiotic) to plants, and the interactions of these bacteria happen within the soil matrix (Table 3.1). This gives elucidation as to why soils are the most diverse habitats on the planet to the extent that even in harsh soil conditions bacteria still thrive, as reported for the coldest, driest desert

on Earth (Babalola et al., 2009). The plant rhizosphere consists of diverse microorganisms which modulate the physiology and morphology of the plant roots and in the process, improve plant growth through the promotion of hormones, and also serve as protectants against plant pathogens (Philippot et al., 2013). It is a critical zone of soil encompassing the plant root thereby making it a hot spot of high abundance and diversity of microorganisms. The plant rhizosphere also attracts bacteria from the soil environment such as the plant growth promoting rhizobacteria (Zhang et al., 2017b).

The soil has a great effect on plants and bacteria and vice versa. This is a result of the influence of plants on the soil through rhizodeposits, water, plant litter, gas, and nutrient exchanges (Fanin and Bertrand, 2016). Where rhizodeposits are present, bacteria abound and so such environments are richer in microflora and available nutrients than the external environment, which could be termed as the bulk soil (Babalola et al., 2007). The interface between root and soil is a strategic entryway for plants to absorb water and mineral nutrients from the soil environment and release rhizodeposits into the soil. These rhizodeposits, which occur in different forms facilitate below-ground interactions between plants and microorganisms, and consequently affect the biodiversity of the region. Naturally, root-microbial interactions are very complex since such interactions involve myriads of microorganisms (Preece and Penuelas, 2016).

Plants, being the major sources of organic carbon in the soil, can be drivers of microbial growth and activity; hence, plants can affect the structure of the bacterial community in the soil (Lange et al., 2015). Additionally, plant root-soil and microbial interaction has a key influence on plant community dynamics and nutrient cycling. A transformation in the plant community structure affects litter composition, which changes nutrient turnover rates and soil characteristics. The

change in soil characteristics might further cause a change in plant community structure and composition (Zhang et al., 2017a).

Under unfavourable conditions, plant-growth-promoting rhizobacteria alleviate the effects in plants of such parameters as germination rate, drought tolerance and plant yield components. The use of microorganisms in agriculture for crop protection against plant pathogens and pests because of the metabolites they produce, and also biological control against diseases, may present an alternative for plant disease prevention (Babalola et al., 2007; Enebe and Babalola, 2018).

Since microorganisms found in the rhizosphere are important for plant health and biogeochemical cycles, engineering the rhizosphere may finally put an end to the use of agrochemicals by substituting their functions with beneficial microbes. Therefore, understanding the community structure and diversity of active microorganisms in the rhizosphere is key to enhancing plant growth and increasing agricultural productivity (Li et al., 2019a).

Table 3.1: Interactions occurring between the rhizosphere and bacteria can be positive or negative

Plant rhizosphere	Rhizobacteria	Effect	References
		Positive effect	
Tomato (*Lycopersicon esculentum* var. esculentum)	*Burkholderia cepacia; B. unamae; B. tropica; B. xenovorans*	N_2 fixation, plant growth promotion and bioremediation	(Caballero-Mellado et al., 2007)
Corn (*Zea mays*)	*Bacillus subtilis; Pseudomonas. fluorescens; Pantoea agglomerans*	Antifungal activity	(Petatan-Sagahon et al., 2011)

Wheat (*Triticum aestivum*)	*Azospirillum brasilense*	Promotes the uptake of NO_3^-, K^+ and $H_2PO_4^-$	(Saubidet et al., 2002)
Rice (*Oryza sativa*)	*Azotobacter vinelandii; Azotobacter chroococcum*	Plant growth promotion	(Yanni and El-Fattah, 1999)
Wheat	*Azospirillum lipoferum*	Promotes the development of the root system of wheat even under the contamination of crude oil	(Muratova et al., 2005)
		Negative effect	
Wheat	*Pseudomonas fluorescens* angstrom313	Plant growth reduction	(Åström et al., 1993)
Tomato	*Bacillus subtilis*	Causes sour skin, a bacterial soft-rotting disease of onion	(Murphy et al., 2000)
Castor oil plant (*Ricinus communis)*	*Bacillus cereus*	Causes rootlets rot	(Carvalho et al., 2007)
Rice paddies	*Burkholderia pseudomallei*	Causative agents of melioidosis	(Manivanh et al., 2017)

3.2 Rhizosphere bacterial community

The rhizosphere is home to enormous numbers of diverse bacterial species, some of which are culturable and some are as yet unculturable. Culturable bacteria form an important part of the rhizosphere, the majority of which are Gram negative (Babalola and Akindolire, 2011; Hou and Babalola, 2013) Among such bacteria are the rhizobacteria, which are characterized by aggressive colonization and subsequent establishment on plant roots. Many studies have characterized

rhizobacteria in the rhizosphere of different plant species. Rhizobacterial genera viz., *Acetobacter, Acinetobacter, Alcaligenes, Arthrobacter, Azoarcus, Azospirillum, Azotobacter, Bacillus, Beijerinckia, Burkholderia, Derxia, Enterobacter, Gluconacetobacter, Herbaspirillum, Klebsiella, Ochrobactrum, Pantoae, Pseudomonas, Rhodococcus, Serratia, Stenotrophomonas* and *Zoogloea* have been subjected to extensive research for decades (Babalola, 2010; Olanrewaju et al., 2017) and have been implicated for beneficial effects on plant yield components. Of special interest are those that are common and peculiar to a locality (Hou and Babalola, 2013) and those that have plant-growth-promoting traits such as *Pseudomonas fluorescens* NWU65, *Vibrio fluvialis* NWU37, *Ewingella americana* NWU59, *P. putida* NWU12 (Hou and Babalola, 2013), the Actinobacteria (Adegboye et al., 2012; Adegboye and Babalola, 2013) and many Proteobacteria (Babalola and Glick, 2012; Babalola, 2003; Singh et al., 2007).

The bacterial population in the soil can grow very rapidly and make use of a wide range of different substances as nutrient sources. Over the last decade, the role of plant growth promoting rhizobacteria (PGPR) cannot be overemphasized due to their positive effect on plants. About 2-5% of bacteria found in the rhizosphere have plant-growth-promoting traits. Thus, they are potential tools for sustainable agriculture in the future (Goswami et al., 2016). The PGPR can promote growth either in direct or indirect ways by means of various mechanisms, which include providing a nitrogen source for plants through nitrogen fixation, applying biological control measures in combating soil-borne pathogens, producing plant growth substances (phytohormones) that improve plant development and producing metabolites such as siderophores, antibiotics, cyanides, and ammonia (Bumunang et al., 2013).

It is conventional to observe cultivar effects on the bacterial composition of the rhizosphere because of the different compounds released by the root, collectively termed root exudate (Figure

3.1). Root exudates and other root deposits confer physical and chemical changes to the soil rhizosphere when compared to the bulk soil. The root exudates released are quickly digested by the root-associated microbes and modified before being discharged into the rhizosphere soil by microorganisms (Zhang et al., 2017a). The absorption of root deposits by rhizosphere microbes enhances soil quality as seen in a study by Beauregard et al. (2013) where root polysaccharides caused the biofilm matrix production of *Bacillus subtilis*, a microorganism beneficial to plants. These microbes do not only perceive secreted signals from the plant root but also release different signalling molecules to control their host plant by improving biotic and abiotic stress tolerance or resistance, plant growth, and root development (Zhang et al., 2017a).

Bacteria in the rhizosphere respond differently to root exudates which stimulate bacterial biomass and activity. A typical cultivar effect was reported among the cowpea cultivars-Rabuor (Kenya) local, IT95K-286-4 and IT94D-437-1 (Babalola et al., 2007). The pod characteristics of some cultivars was significantly higher which eventually lead to the high yield of cowpea due to the application of microbial inoculants, *Pseudomonas* 44MS8, and *Pseudomonas* 10M3, and *Enterobacter sakazakii* 8MR5 to the soil.

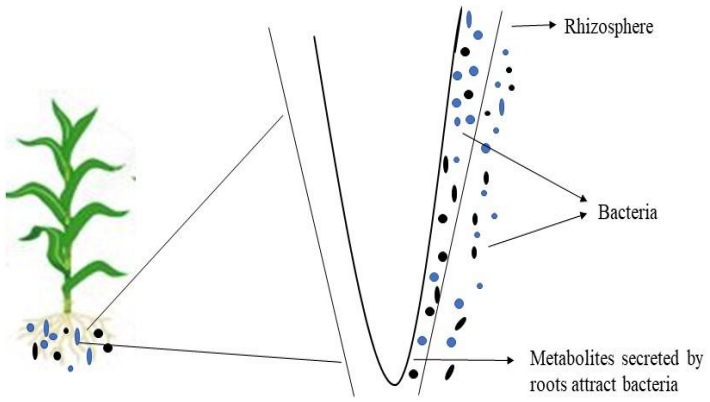


Figure 3.1: Young roots shape the bacterial population in the rhizosphere community by secreting more organic materials when compared to the older roots

3.3 Factors controlling the distribution and abundance of bacterial communities in the rhizosphere

Bacterial functional diversity is the ability of bacterial communities to use a wide spectrum of different compounds such as cellulose, sugars, and lignin that occur in plant tissues and whose proportional abundance changes during decomposition (Lahav and Steinberger, 2001). Other investigations indicate that the type and amount of available organic substrates in the rhizosphere strongly influence the abundance of bacterial groups and their functional diversity (Grayston et al., 2001; Lejon et al., 2007). Minerals such as nitrogen and iron can also affect the number of bacteria that can be found in the rhizosphere and change the composition of the rhizobacterial community

(Zhang et al., 2019a). It may be possible that potential nitrogen fixing and phosphate solubilizing bacteria add to plant nutrients by improving nitrogen and phosphorus intake by the plants, which could lead to an increase in plant yield (Asari, 2015).

Rhizosphere-inhabiting bacteria are most likely to emanate from surrounding bulk soil and often times from the seeds, then thrive under conditions that prevail in the plant roots. Plant type, plant age, soil pH, plant growth stages, organic compounds, and nutrient availability are some of the factors that influence the abundance and distribution of rhizobacteria in the soil. In other words, plant species, plant cultivars, plant developmental stages and soil characteristics have thus been identified as major factors that contribute to the determination of composition of the rhizosphere community (Schlemper et al., 2017).

3.4 How plant growth stages affect bacterial community dynamics

The bacterial communities in the rhizosphere region benefit from the nutrients that are secreted by younger roots and are under constant influence from their environment thus bringing changes to the community. Rhizobacteria are not static; they are subjected to changing environmental conditions of temperature, water content and nutrient availability. It has been discovered that in young plant roots, bacterial communities are controlled by r-strategists, which are species having rapid growth rates and abilities to make use of simple substrates (Brimecombe et al., 2001; Zhou et al., 2016). As the roots of these young plants mature, there are specific elements of root exudates that can have selective control in the rhizosphere by spurning some species and strengthening the competitive activity and population of other species. As a result, there is a slight movement in dominance to bacterial communities with relatively slow growth rates and those that have the ability to degrade more complex substrates (k-strategists). As the root tips of the plant mature, soil

microorganisms will colonize the roots and the population densities will increase rapidly a few centimeters from the root tips, and that is where insoluble, soluble and volatile root exudates are used by the rhizobacteria for growth and metabolism (Franke et al., 2018). The different root zones in a plant can support distinct bacterial communities showing the differences in root exudation qualitatively and quantitatively (Reinhold-Hurek et al., 2015). Soil type also plays a crucial role in the determination of the unique dominant bacteria that colonize the rhizosphere (Mendes et al., 2018). Therefore, the rhizosphere bacterial population belonging to the same plant species may differ both spatially and temporally.

3.5 Plant-root interactions in the rhizosphere

The key roles of plant roots are to anchor the plant, absorb water and essential nutrients as well as store the nutrients and accumulate and secrete a diverse array of compounds including several primary and secondary metabolites, proteins, and peptides (Weisskopf et al., 2006). There is therefore a relationship between the plant roots and the soil. The interface between plant root and soil is influenced by different interactions triggered by soil microorganisms and plants roots.

Plants release nutrients and other organic substances in the rhizosphere region, which attract different kinds of bacteria (Adegboye et al., 2012; Adegboye and Babalola, 2013). There is a signal in the rhizosphere when a plant releases exudate from its root. These signals can recruit nitrogen-fixing and growth-promoting bacteria, for instance rhizobia. Plant species belonging to the family Fabaceae mainly benefit in this association. In addition, plant-produced flavonoids are also involved in the establishment of these associations (Rasmann and Turlings, 2016). Bacteria colonize the rhizoplane and the rhizosphere, which is an important interaction between the plant roots and the surrounding soil. There is a regular flow of organic substrates from the plant which

the bacteria take advantage of and as a result promote the growth of the plant through provision of soluble inorganic nutrients and production of growth-promoting substances (Kai et al., 2016). Hence, the level of interaction with the plant root is shaped by the nature of rhizodeposits and soil properties. In other words, rhizodeposition varies depending on the growth stages and species of the plant, in addition to the conditions of the environment (Bulgarelli et al., 2013).

The soil physical structure supports the above ground part of the plant by the root system. The plant also requires certain essential elements from the soil such as N, P, K, S, Mg, Ca (macronutrients) and B, Cu, Fe, Mn, Mo, Ni, Zn (micronutrients). The presence of toxic elements in the soil can also limit plant growth.

3.6 Influence of plant exudates on rhizosphere microbial dynamics

In a study by Mendes et al. (2014), on soybean plants, certain bacterial groups with specific nutritional functions were more in abundance in the rhizosphere than in the bulk soil. The bacterial groups include those that were involved in iron uptake and metabolism, membrane transport, nitrogen, phosphorus, and potassium metabolism. These nutritional traits might be of benefit to the plant and this implies that soybean plants may attract specific microbial groups from the bulk soil into the rhizosphere on the basis of the functional qualities that enhance their productivity. "Secretion system type IV" which falls under the membrane transport is involved in the mutualistic associations between bacteria and other organisms. This membrane transport is present in *Acidobacteria*, *Proteobacteria*, *Actinobacteria* and *Bacteroidetes* as reported by (Mendes et al., 2014). Thus, it is expected that those plants that need the secretion system for their optimal performance will have more bacterial communities associated with the above listed phyla (Table 3.2). Microorganisms implicated in the metabolism of potassium are *Rhizobium*, *Pseudomonas*,

Bacillus, and *Flavobacterium*. These bacteria are able to make potassium more available for plant use (Raghavendra et al., 2016), while the rhizobia are involved in nitrogen metabolism including nitrogen fixation.

The exudate from plants is also a determining factor of species richness and abundance in the rhizosphere. Carbon is one of the photosynthates fixed by plants, which is partly transported into the root zone and excreted from root tissues. Various organic acids such as aliphatic and aromatic acids, amides, carbohydrates (glucose and xylose), fructose, lactic, malic, oxalic, pyruvic, succinic, and amino acids secreted from the root play roles in chemotaxis and are also referred to as PGPR bioactive factors (Babalola, 2010). Some examples of the amino acids are alanine, asparagine, glutamic acid, glutamine, isoleucine, leucine, serine and valine. These acids, as well as other compounds, are released into the rhizosphere and thus serve as nutrients for rhizobacteria.

Table 3.2: Interactions of metabolites released by the rhizobacteria and their importance to plants

Rhizobacteria	Metabolites	Function of the metabolites	References
Pseudomonas spp.	Phenazines, pyrrolnitrin, pyoluteorin, viscosinamide	Viscosinamide found to prevent the infection of sugarbeet by *Pythium ultimum*	(Bloemberg and Lugtenberg, 2001)
Streptomyces spp	Siderophores	Alleviate metal-contamination stress on plants	(Dimkpa et al., 2008)
Bacillus amyloliquefaciens strain FZB 42	lipopeptides, surfactins, bacillomycin D, fengycins	Antifungal activity	(Koumoutsi et al., 2004)
P. aeruginosa	Pyoverdine, pyochelin, salicylic acid	Induces resistance to plant diseases caused by *Botrytis cinerea* on bean and tomato, *Coletotrichum lindemuthianum*	(Höfte and Bakker, 2007)
Bacillus amyloliquefaciens	Bacillomycin D, fengycins, surfactins	Enhance plant growth and suppress plant pathogenic organisms	(Chen et al., 2007)
Serratia spp.	Siderophores, pyrrolnitrin, prodigiosin	Antifungal activity against different phytopathognic fungi *Verticillium dahlia, Rhizoctonia solani, Sclerotinia sclerotiorum*	(Kalbe et al., 1996)
P. fluorescens	Pyoluteorin, pyrrolnitrin, phenazine-1-carboxylic acid, 2,4-diacetylphloroglucinol	Suppress soil borne plant pathogens, including fungi and nematodes	(Raaijmakers et al., 2002)
Rhizobium meliloti	Siderophores	Antifungal activity against *Macrophomina phaseolina*	(Arora et al., 2001)

P. chlorophis PCL1391	Phenazine-1-carboxamide	Antifungal	(Chin-A-Woeng et al., 2005)
P. fluorescens	Siderophore pyoverdin	Antibacterial and antifungal activity. Enhanced plant growth	(Boruah and Kumar, 2002)
P. fluorescens PfMDUS	Chitinase, beta-1,3-glucanase, siderophores, salicylic acid, hydrogen cyanide	Inhibit mycelial growth of *R. solani*	(Nagarajkumar et al., 2004)
P. fluorescens Pf-5	Antibiotics, siderophores	Antibiotics toxic to soilborne fungi and oomycetes that infect plant roots/siderophores involved in iron acquisition	(Loper and Gross, 2007)
Klebsiella oxytoca C1036	Butyl 2-pyrolidone-5-carboxylate	Active against soft-rot disease pathogen in tobacco	(Park et al., 2009)
Streptomyces AcH 505	Auxofuran	Stimulate the growth of fly agaric, suppresses growth of ectomycorrhizal fungi	(Riedlinger et al., 2006)
P. syringae strain 366	Phenazine-1-carboxylic acid, 2-amino phenoxazone, 2-amino phenol	Inhibits downy brome root growth	(Gealy et al., 1996a)
P. fluorescens D7	Uncharacterized phytotoxins	Inhibits downy brome root growth	(Gealy et al., 1996b)
P. fluorescens CHAO	Antibiotics, hydrogen cyanide and an exoprotease	Protists growth inhibition, encystation, paralysis and cell lysis	(Jousset et al., 2006)

P. aeruginosa PUPa3	Phenazine-1-carboxamide	Broad-spectrum antifungal activity and biofertilizing traits	(Kumar et al., 2005)

3.7 Metabolites

The rhizosphere possesses a detailed dynamic ecological relationship between plant and microbes. A wide range of chemicals are usually secreted, and can be grouped as signalling compounds, growth regulators, and nutrient solubilizers. The determination of chemicals involved in this signalling can help in sustainable agriculture, most especially as biocontrol agents (Verma et al., 2018; Beneduzi et al., 2012).

Some of the metabolites secreted by plant roots are (Table 3.3) glucoberin, 4-methyl-2-pentanone, alkaloids, amino acids, benzaldehyde, biochanin A, calystegine, cyanidin, dehydrotomatine flavonoids, formononetin, furfural, glycoalkaloids, glycosides, kaempferol, lectins, maeckiain, medicarpin, okundoperoxide, p-coumaric acid, peptides, phylloquinone, polyphenols, quercetin glycosides, rhamnoside, salicylic acid, terpenoids, tryptophan, ϒ- butyrolatone, ϒ-terpinene, α-tomatine and α-tocopherol (Bais et al., 2004). Besides plant metabolites, bacteria also secrete bacterial metabolites. Some of the bacterial metabolites are 2,4-diacetylphloroglucinol, 2-amino phenol, 2-amino phenoxazone, antibiotics, auxofuran, bacillomycin D, beta-1,3-glucanase, butyl 2-pyrolidone-5-carboxylate, chitinase, exoprotease, fengycins, hydrogen cyanide (HCN), indole-acetic-acid (IAA), lipopeptides, phenazine-1-carboxamide, phenazines, phosphate, phytotoxins, prodigiosin protease, pyochelin, pyoluteorin, pyoverdine, pyrrolnitrin, salicylic acid, siderophores, surfactins, and viscosinamide (Santoyo et al., 2012).

Accumulations of secondary metabolites often occurs in plants when they undergo various stresses, elicitors, or signal molecules. Synthesis of secondary metabolites increases when certain factors are involved. These factors are the physical, chemical, and microbial factors, which can function as biotic/abiotic elicitors leading to an increase in the production of secondary metabolites. The biotic elicitors have biological origin and are derived from microorganisms such as fungi, bacteria, viruses, or plant cell wall components and chemicals that are released by plants against plant pathogens. (Thakur et al., 2018).

Table 3.3: Plant roots secrete a range of metabolites into the rhizosphere

Plant	**Metabolites**	**Function of the metabolites**	**References**
Sweet chestnut (*Castanea sativa* Mill.)	ϒ-butyrolactone, ϒ-terpinene, furfural, benzaldehyde, 4-methyl-2-pentanone	Peculiar aroma, the most typical organoleptic characteristic of chestnut-based products	(Cirlini et al., 2012)
Tobacco (*Nicotiana tabacum)*	Salicylic acid, alkaloids, flavonoids	Plays a role in plant growth and development, photosynthesis, ion uptake and transport. Resistant to pathogens by inducing production of pathogenesis-related proteins. UV filtration and symbiotic nitrogen fixation	(Choi et al., 2004)

Wheat (*Triticum* spp)	Tryptophan	For the biosynthesis of indole-3-acetic acid by associative bacteria	(Sachdev et al., 2009)
Legume seeds	Flavonoids, alkaloids, terpenoids, peptides, amino acids	Serve as eco-sensing signals for suitable rhizobacteria and arbuscular mycorrhizal fungal towards symbiotic mutualisms. Defense molecules against pathogen and insect pests.	(Ndakidemi and Dakora, 2003)
Chickpea (*Cicer arietinum*)	Biochanin A, medicarpin, formononetin, maeckiain	Not stated	(Armero et al., 2001)
Potato (*Solanum tuberosum*)	Glycoalkaloids, calystegine, lectins	They protect plants against phytopathogens	(Friedman, 2006)
Spice (*Scleria stiatinux*)	Okundoperoxide	It contains antiviral and antifungal properties, also used as herbal tea for fevers	(Efange et al., 2009)
Tomato (*Solanum lycopersicum*)	α-tomatine, dehydrotomatine	Defend plants against attack by microorganisms and herbivores due to their insecticidal activity and have allelopathic effects on many weeds.	(Kozukue et al., 2004)

3.8 Metabolic fingerprinting

Biochemical fingerprinting can be of great benefit for scientific and commercial purposes in relation to plant breeding, response of plants to biotic and abiotic stresses, and characterization of genetic mutants. The study of metabolic fingerprinting in the plant rhizosphere is important,

because metabolites are low-molecular-weight compounds synthesized by plants for essential functions, especially growth and development, and specific functions such as defense mechanisms. Metabolites can be analysed using various approaches, which can be classed into five categories (Garcia-Perez et al., 2008): (i) Metabolite target analysis: this approach is target-driven where one or a few single compounds are analysed (ii) metabolic profiling: an assortment of metabolites already defined according to a class of compounds or based on their correlation with a particular pathway. Examples of such class of compounds are carbohydrates, fatty acids, organic phosphates, or amino acids (iii) metabolic fingerprinting: fast analysis of samples through pattern recognition using high throughput technology (iv) metabolomics: non-biased identification and quantification of all the metabolites in a biological system, and (v) metabonomics: the measurement of dynamic changes across the metabolome of living systems with respect to time in response to physiological stimuli or genetic alteration (Garcia-Perez et al., 2008). Multiple methods are required to analyze different subsets of metabolites due to differences in volatility, polarity, solubility and chromatographic behaviour (Jorge et al., 2016).

In the metabolic fingerprinting approach, the goal is not to classify each detected metabolite but to correlate patterns, and fingerprint the metabolites that change in response to disease, toxin exposure, and environmental or genetic alteration (Garcia-Perez et al., 2008).

3.9 The use of omics techniques to analyse the rhizosphere

Next generation sequencing (NGS) has been used to infer the microbial community structure of the rhizosphere and, the core microbial community. This has led to better understanding of the structure, abundance, spatial distribution diversity and important members of the rhizosphere community (White III et al., 2017). Plants in which this platform has been studied include pea,

soybean, corn, wheat, and oat (Newman et al., 2016). The NGS tools are metagenomics, metatranscriptomics, metaproteomics, and metabolomics. A review on this has been done by White III et al. (2017). Although each omics study has its advantages and disadvantages, to better capture the structure and diversity in the rhizosphere, it is advisable to carry out the multi-omics analysis which is a modern approach in systems biology.

Conclusions and future outlooks

This review showed that different metabolites are found in the rhizobacteria. They have a relevant role to play in agriculture as these metabolites are responsible for stimulating plant growth, suppressing diseases such as fungal phytopathogens and plant-parasitic nematodes by production of cyanide, siderophores, ammonia and other volatile metabolites. Another benefit is that they have a positive effect on the growth of plants under abiotic stress such as salinity or drought through mechanisms such as molecular, physiological, and biochemical. In crop production, these rhizobacteria can be used to improve crop yield by acting as biofertilizers, thereby helping to reduce reliance on chemical fertilizers and pesticides, which pollute the environment. Furthermore, biofertilizers have the advantage of being eco-friendly and cheap.

Research in soil microbiology has largely been focused on the analysis of microbial processes that take place in the soil. We need to understand the microbial populations because they are the basic assemblages that drive these processes. Soil Microbiology is undergoing a rapid transition where interdisciplinary approaches are required. It involves the collaboration in many areas of science like biology, physics, bioinformatics, mathematics, statistics, and computer science. Thus, to better understand the rhizosphere and develop new natural products, in-depth research will be needed in future where techniques, state-of-the-art technologies, along with robust data software and analysis.

The study of the rhizosphere is a rapidly advancing area of research and there has been tremendous progress in the usage of genetic fingerprinting techniques to study rhizobacterial communities that are found in the rhizosphere. However, quantitative assessment is still a huge challenge, due to the biases, which are associated with the isolation of nucleic acids (DNA or RNA) and PCR, and none of these techniques provide full understanding and access of the genetic diversity of the bacterial community. Thus, omics studies are important tools, which can be used to identify and characterize the microbial genes and functions that help microorganisms thrive in the plant rhizosphere. Such information will help improve the ability to fight plant diseases, and promote beneficial bacterial functions for agriculture, which will have direct positive effects on global food production, thereby enhancing food security.

CHAPTER FOUR

METAGENOMIC INSIGHT INTO THE COMMUNITY STRUCTURE AND FUNCTIONAL GENES IN THE SUNFLOWER RHIZOSPHERE MICROBIOME

Abstract

The rhizosphere microbial communities consist of a diverse set of microorganisms that can be beneficial to plant. These beneficial microorganisms are key determinants of plant productivity and health. In this study, we used shotgun metagenomics to explore and characterize the microbiome of the sunflower rhizosphere and bulk soil. Rhizosphere shared features with the bulk soil with dominant phyla such as *Actinobacteria*, *Proteobacteria*, *Acidobacteria, Bacteroidetes, Planctomycetes*, and *Verrucomicrobia*. There was no significant difference in the alpha diversity between the two locations. Likewise, there was no significant difference in the alpha diversity of the sunflower rhizosphere and bulk soils though diversity was lower in the rhizosphere, suggesting a selection of microorganisms by sunflower rhizosphere to the bulk soil community. The genes present with their corresponding proteins as observed in our study conferred potential plant beneficial properties such as siderophore production, nitrogen fixation, phosphate solubilizing, 1-aminocyclopropane-1-carboxylate (ACC) deaminase. Furthermore, other genes such as exopolysaccharides producing, high-temperature stress response, heat and cold shock response genes, which help withstand environmental stresses were also identified. Of note from our study is the gene phenazine biosynthesis protein, which confers biocontrol. With the current indiscriminate use of pesticides that are considered harmful to the ecosystem, this potential plant

beneficial functional genes can be further exploited and used as a biotechnological application for sustainable agriculture.

Keywords: Food security, microbial community, microorganisms, omics study, sustainable agriculture

4.1 Introduction

The rhizosphere microbiome, which includes the totality of microorganisms, their genomes, and interactions in the rhizosphere has been the focus of several studies in recent times. The microorganisms that inhabit the rhizosphere have numerous beneficial functions, which range from nutrient acquisition, stress tolerance, and protection against pathogens (Mendes et al., 2018). Soil is not just a single environment but consists of many environments that have distinct microbial communities. The distinct soil environments may range from a few micrometers to millimeters apart, their microbial abundances, rate of microbial activity, abiotic characteristics, and the composition of the microbial community differ (Fierer, 2017). Soil microorganisms have the largest reservoir of biodiversity (Qiao et al., 2019), and these microorganisms perform different activities, although these biological activities mostly take place in the rhizosphere where there is communication between the plant roots and the microorganisms. Plants modify the rhizosphere directly through rhizodeposition and root exudates leading to the changes in the composition and function of the rhizosphere microbiome (Qiao et al., 2019). Plant species growing in a particular environment can recruit significant different microbial communities in both the endosphere and rhizosphere (Compant et al., 2019).

Plant-associated microorganisms affect the fitness and physiology of the host plant by altering nutrient availability to the plant thereby improving plant resistance to biotic and abiotic stressors (Berg et al., 2015). Microbial community activity, especially in the rhizosphere, is often connected to variation in the physical or chemical properties of the soil such as soil texture, pH, or changes in land use (Durrer et al., 2017). Biotic and abiotic factors determine the structure and composition of the soil microbiome. Thus, highlighting the strong relationship between plant and microorganism using the impact of soil type, soil age, soil pH, and mineralogy (Uroz et al., 2016). In addition, plant location and plant species are determinants to which microorganisms can grow and thrive in the rhizosphere (Philippot et al., 2013; Oberholster et al., 2018).

The microbiome can be explored because it has been postulated to be one of the key components of technological innovation to improve plant health and production (Durrer et al., 2017). Bacteria being the most dominant domain in the rhizosphere possess plant-growth-promoting-traits. Therefore, they have the potential of being used as a viable tool for sustainable agriculture (Alawiye and Babalola, 2019).

Reports have shown that the genes that have been identified to have plant beneficial attributes range from genes that confer plant-beneficial properties such as *phl* (phloroglucinol synthesis) (Vacheron et al., 2016) or *nif* (nitrogen fixation) (Nouioui et al., 2019) to genes involved in numerous indirect functions or secondary plant-beneficial ones like pqq (pyrroloquinoline quinone synthesis) (Bruto et al., 2014). Other plant-beneficial function contributing genes that have been identified to enhance the growth of plants in the rhizosphere are ACC deaminase, auxin synthesis, hydrogen cyanide synthesis, phosphate solubilization, and 2,4-diacetylphloroglucinol synthesis genes (Bruto et al., 2014; del Carmen Orozco-Mosqueda et al., 2020).

Understanding the rhizosphere microbiome through the means of taxonomic, genomic, and functional components is important for sustainable crop production. There has been a good level of success toward the characterization of the rhizosphere microbiome in some crop plants such as rice, soybean, corn, barley and wheat by exploring the structure, functional genes and factors that drives the microbiome assemblages (Xu et al., 2018).

We hypothesized that the rhizosphere microbiome would be more diverse than the bulk soil because of the activities that take place there and that the soil physicochemical parameters will affect the microbial diversity and community structure. We also hypothesized that genes beneficial to plant growth would be higher in the rhizosphere than in the bulk soil due to plant selectivity. Previously, to determine the diversity, structural, and functional genes present in the rhizosphere, conventional approach (culture dependent method) have been employed, but these come with biases and omit organisms and functional elements for analysis (McIntyre et al., 2017). Therefore, in this study, we used shotgun metagenomics to determine the diversity, characterize the structure, and identify the plant beneficial functional genes in sunflower rhizosphere soils thereby necessitating our ability to predict and harness microbiome dynamics and functionality by providing information about belowground effects on microbial communities in order to enhance crop productivity.

4.2 Materials and methods

4.2.1 Soil sampling and analyses

Rhizosphere soils, which were the soils firmly attached to the root and bulk soils, which were loose soils away from the plant (Oberholster et al., 2018), were collected from two sunflower fields in

South Africa: Palmietfontein (26º19'39.94"S: 26º52'52.57"E), and Bloemhof (26º17'46.1"S: 26º58'19.83"E). Palmietfontein has an average rainfall of 200 mm, average annual temperature of 22°C and the month of collection, June 2018, had annual average temperature of 11°C while Bloemhof has an average rainfall of 350 mm, average annual temperature of 20°C and annual temperature at the month of collection was 9.9°C. Rhizosphere soil samples from Palmietfontein (R1), Bloemhof (R2) and bulk soil samples from Palmietfontein (B1), Bloemhof (B2) were collected in triplicate at 0-20 cm depth using 2 cm soil auger (Mendes et al., 2014). Samples were collected in sterile bags, after which, they were placed in a cooler box and brought to the laboratory for analysis.

Phosphorus (P) was analyzed using the P Bray method while Total Carbon (C) and Nitrogen (N) were determined using a TruSpec elemental determinator (Oberholster et al., 2018). pH meter using the ratio 1:2.5 (soil/water) was used to measure the soil pH according to Enagbonma et al. (2019). Organic matter content was measured using Walkley-Black method. The soil moisture was determined by a procedure described by Cui et al. (2019) while the hydrometer technique was used for the particle size analysis (Kettler et al., 2001). The particle size classes used to assign texture was that of the United States Department of Agriculture (USDA) with sand (2.0-0.5 mm), silt (0.05-0.002 mm), and clay (< 0.002 mm). Organic carbon was analysed using the dichromate digestion (Shi et al., 2011). Potassium (K) and calcium (Ca) were analysed by methods described by Deke et al. (2016).

4.2.2 DNA extraction from soil samples, sample preparation, and sequencing

Total genomic DNA was extracted from each soil sample using PowerSoil® isolation kit (MO Bio labs, USA), following the manufacturer's user guide. DNA concentration and purity were

determined using a NanoDrop Lite Spectrophotometer (Thermo Fischer Scientific, CA, USA). Extracted DNA was sent for Metagenome Shotgun Sequencing at the Molecular Research Laboratory (www.mrdnalab.com) Texas. Qubit® dsDNA HS Assay Kit (Life Technologies) was used to determine the 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 from each sample were used for library preparation. After DNA fragmentation Illumina sequencing adapters were added and products amplified using 6 cycles of PCR during which unique indices were added to each sample. After library amplification, their concentration was estimated using the Qubit® dsDNA HS Assay Kit (Life Technologies), while 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).

4.2.3 Data processing and statistical analysis

The online metagenomics rapid annotation server MG-RAST (www.mg-rast.org) was used for the quality control of the raw shotgun metagenomic sequences (Meyer et al., 2008). Low quality sequences were removed using a modified DynamicTrim (where sequences with >5 ambiguous base pairs with 15 phred score cutoff were removed) and a length filtering (sequences with a length of >2 standard deviation from the mean were removed) according to Cox et al. (2010). Artificial replication reads that occurred during sequencing were removed (Gomez-Alvarez et al., 2009). After performing quality control (QC), BLAT (the BLAST-like alignment tool) algorithm was used to annotate the sequences (Kent, 2002) against the M5NR database (Wilke et al., 2012), which encompasses nonredundant integration of many databases. The microbial community groupings

were identified using the SEED subsystem. An e-valve of 5, with maximum identity of 60%, maximum alignment length of 15 base pairs, and maximum abundance of 1 were used when mapped against the subsystems database which integrated SILVA, Greengenes, and RDP databases (Garcia-Mazcorro et al., 2017). The relative abundance of the four samples were determined and used for statistical analysis.

After the dataset for each replicates of the rhizosphere and bulk soil samples of the two locations was normalized by finding the average and relative abundance, One-Way Analysis of variance (ANOVA) was used to determine if there was significant difference in rhizosphere and bulk soil samples and Tukey's pairwise comparison was used for the means at significant level (P valve < 0.05). The diversity indices (alpha and beta diversity) of the structural and beneficial functional genes were also determined. The analysis of similarity (ANOSIM) with 999 random permutations was employed to test for significance for beta diversity. These analyses were done using the PAST software version 2.17c (Hammer et al., 2001). The distribution of microbial communities' diversity and structure between the sunflower rhizosphere and the bulk soil was ascertained using principal component analysis (PCA) based on a Bray-Curtis dissimilarity matrix. Likewise, principal coordinates analysis (PCoA) based on Bray Curtis dissimilarity matrix was used to visualize the beta diversity of the structural and functional genes diversity from both rhizosphere and bulk soil habitats.

To determine the environmental variable that best shape the structures of the microbiome, canonical correspondence analysis (CCA) was performed based on a forward selection of environmental variables and the Monte Carlo permutation test, with 999 random permutations, was employed to test for significance. The relative abundance of the plant beneficial functional gene was plotted using circos (http://circos.ca/), while CANOCO 5 (Microcomputer Power, Ithaca,

NY) was used to view the PCA and PCoA. The raw sequence has been deposited into the GenBank database (https://www.ncbi.nlm.nih.gov/search/all/?term=SRR10418054) where SRA accession number for R1 is SRR10426233, R2 SRR10418054, while B1 and B2 have accession number of SRR10426310 and SRR10418081, respectively.

4.3 Results

4.3.1 Physicochemical characteristics of the soil

The geographical map of the sample location is shown in Figure 4.1. The pH is in the range of 5.78-6.6, which shows slightly acidic to nearly neutral. The range of the particle size is 72-84% for sand, 0-6% for silt, and 16-24% for clay. Calcium ranged between 246-536 mg/kg, with R1 having the highest and R2 having the lowest. The physical and chemical properties of the soils showed phosphorus (P) in R2 was predominantly higher than R1, B1, and B2. Potassium (K) measured ranged between 220-349 mg/kg. Org. C (Organic carbon) ranged between 0.29-1.77%, with R1 having the highest. The total N of our samples was very low (0.047-0.134%), while R1 has the highest for total C with 1.60%, which ranges between 0.505-1.60% for the rhizosphere and bulk soils. The parameters such as Org C, OM, Total C, Total N, P^{3-}, CA^{2+}, K^{+}, were significant different for the two locations likewise there was significant difference in the rhizosphere and bulk soil samples. Thus, our two locations are different based on the physicochemical parameters (Table 4.1).

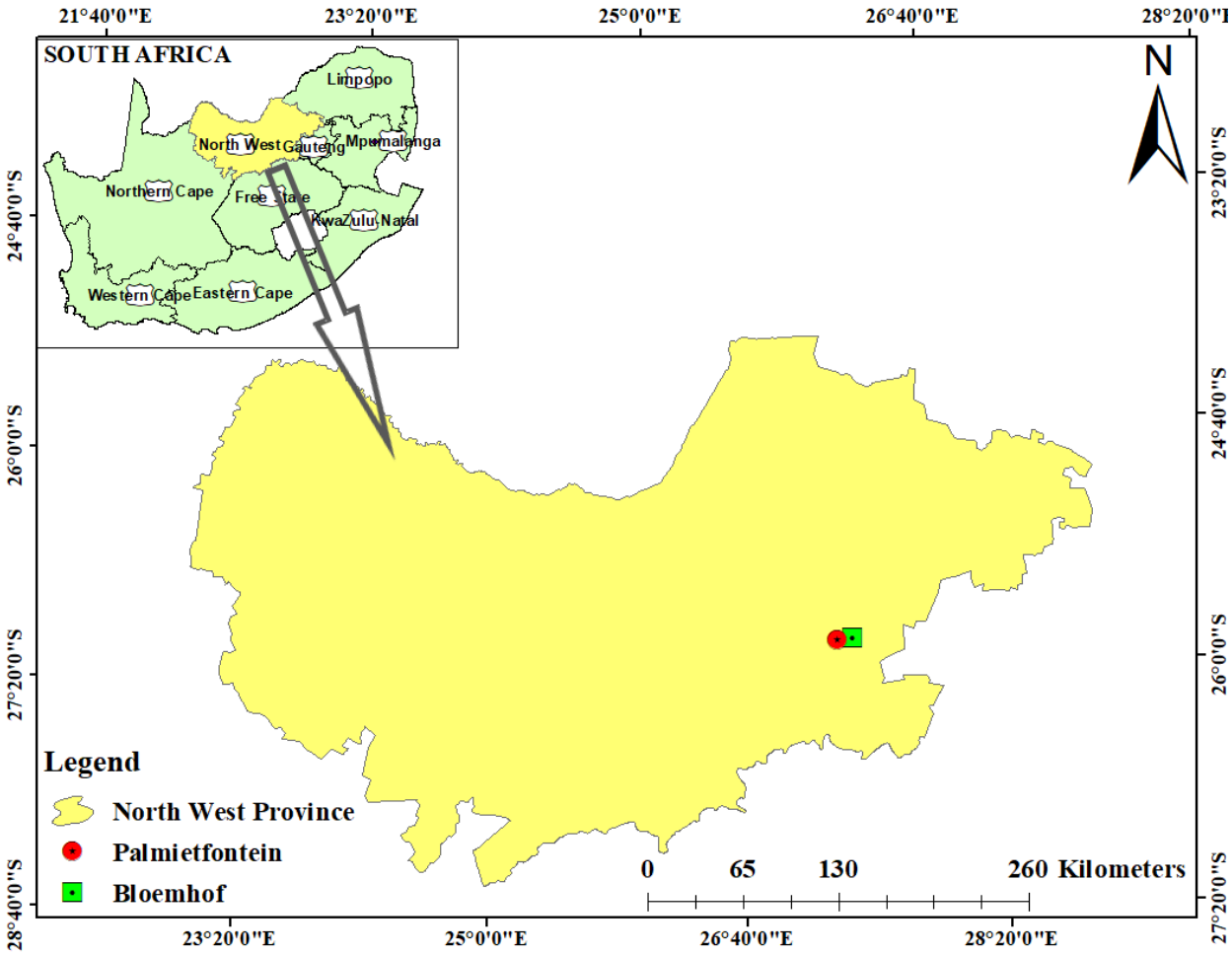


Figure 4.1: Map showing the sampling location

Table 4.1: Physicochemical properties of the soil rhizosphere (R1 and R2) and bulk soil samples (B1 and B2) of two sunflower fields in South Africa, Palmietfontein (R1, B1) and Bloemhof (R2, B2)

Physicochemical values	Sample Locations			
	R1	R2	B1	B2
Org C (%)	1.77 ± 0.06^{a}	0.49 ± 0.08^{b}	0.47 ± 0.21^{b}	0.98 ± 0.21^{c}
OM (%)	1.83 ± 0.41^{a}	0.62 ± 0.07^{b}	0.70 ± 0.05^{b}	1.27 ± 0.54^{ab}
$N\text{-}NH_4$ (mg/kg)	19.67 ± 2.30^{a}	16.23 ± 1.73^{a}	15.67 ± 2.79^{a}	18.37 ± 1.23^{a}

Total C (%)	1.60 ± 0.29^{a}	0.51 ± 0.09^{b}	0.52 ± 0.06^{b}	1.01 ± 0.07^{c}
Total N (%)	0.13 ± 0.03^{a}	0.05 ± 0.01^{b}	0.05 ± 0.01^{b}	0.09 ± 0.02^{ab}
pH (H_20)	6.00 ± 0.26^{a}	6.50 ± 0.30^{a}	6.60 ± 0.35^{a}	5.78 ± 0.59^{a}
P^{3-} (mg/kg)	19.33 ± 3.47^{a}	74.43 ± 10.69^{b}	28.31 ± 2.38^{ac}	16.95 ± 2.40^{ac}
Ca^{2+} (mg/kg)	536.00 ± 34.51^{a}	246.00 ± 13.89^{b}	462.00 ± 24.56^{c}	385.00 ± 21.79^{d}
K^{+} (mg/kg)	349.00 ± 32.23^{a}	220.00 ± 17.00^{b}	223.00 ± 41.33^{b}	342.00 ± 32.70^{a}
Sand (%)	72.00 ± 7.00^{a}	80.00 ± 9.64^{a}	84.00 ± 2.00^{a}	72.00 ± 9.85^{a}
Silt (%)	6.00 ± 1.50^{a}	0.00 ± 0.00^{b}	0.00 ± 0.00^{b}	4.00 ± 1.00^{a}
Clay (%)	22.00 ± 4.00^{a}	20.00 ± 3.40^{a}	16.00 ± 0.00^{a}	24.00 ± 3.61^{a}

Each value is expressed as mean ± standard deviation (n=3). $^{<a\text{-}z>}$ indicates significant difference in values of samples.

4.3.2 Genomic overview of metagenomic sequences

After undergoing quality control, the sample reads for the four soil samples showed there are 10,000, 7,000, 8,000, and 9,000 species count in the R1, R2, B1, B2 microbiome respectively (Figure 4.2). R1 location had the most species count when compared to R2 location indicating there are more variation of species in R1.

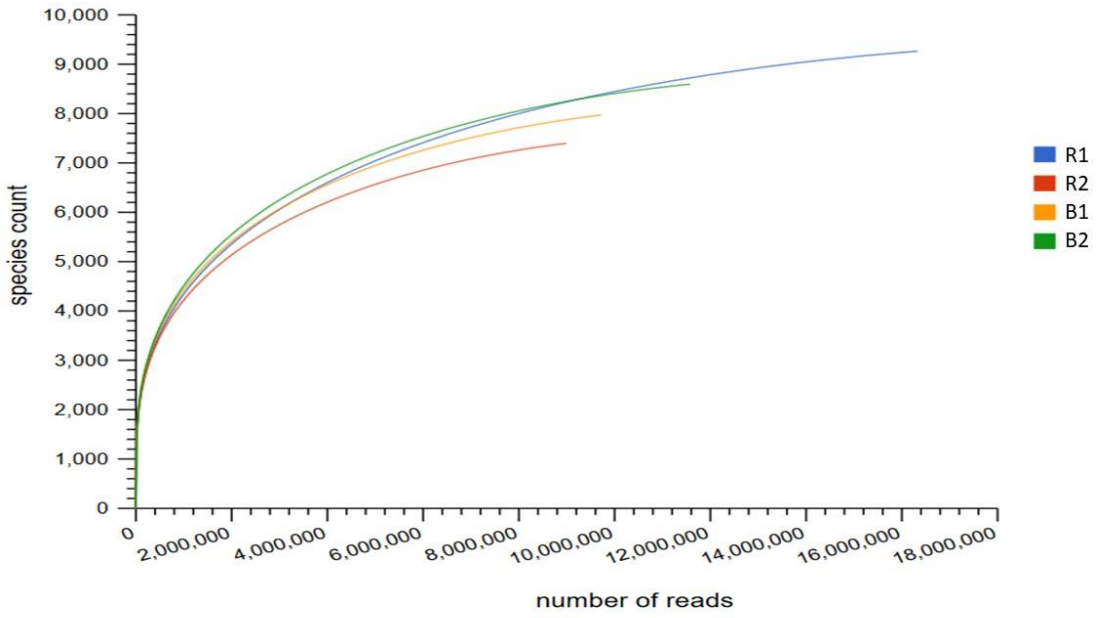


Figure 4.2: Rarefaction curve showing species richness in the sunflower rhizosphere and its bulk soil samples

The DNA sequencing from the four soil samples (that is R1, R2, B1, B2) using Illumina technology resulted into the bp count for R1 as 2,388,568,202 bp, sequence count of 13,736,515, mean square length of 174bp, mean GC percent of 66. R2 had bp count of 1,390,678,255, sequence count of 7,559,132 mean square length of 184, and GC content of 67. B1 and B2 had bp count of 1,551,871,377 bp and 1, 789,010,308 bp respectively, sequence count of 8,282,394 and 9,753,315, mean sequence length of 187 bp and 183 bp, and GC content of 66% and 65% respectively (Table 4.2). The metagenomes of R1, R2, B1, and B2 comprised of predicted protein features of 12,015,321, 6,253,966 7,408,001, and 8,657,907 respectively while possessing identified protein features of 4,985,022, 2,953,873, 3,160,887, and 3,688,032 respectively for metagenomes R1, R2, B1, and B2.

Table 4.2: Sequence data of the four soil samples R1, R2, B1, B2 obtained using MG-RAST pipeline

	R1	R2	B1	B2
Upload: bp Count	2,754,944,067bp	1,607,022,279bp	1,790,884,389bp	2,080,916,475bp
Upload: Sequences Count	16,326,676	8,991,566	9,722,700	11,575,112
Upload: Mean Sequence Length	169 ± 71 bp	179 ± 73 bp	184 ± 70 bp	180 ± 69 bp
Upload: Mean GC percent	66 ± 11 %	66 ± 11 %	65 ± 10 %	64 ± 11 %
Artificial Duplicate Reads: Sequence Count	1,894,346	1,046,824	1,160,489	1,472,880
Post QC: bp Count	2,388,568,202 bp	1,390,678,255bp	1,551,871,377bp	1,789,010,308bp
Post QC: Sequences Count	13,736,515	7,559,132	8,282,394	9,753,315
Post QC: Mean Sequence Length	174 ± 67 bp	184 ± 69 bp	187 ± 67 bp	183 ± 66 bp
Post QC: Mean GC percent	66 ± 9 %	67 ± 8 %	66 ± 9 %	65 ± 9 %

Processed: Predicted Protein Features	12,015,321	6,253,966	7,408,001	8,657,907
Processed: Predicted rRNA Features	22,881	13,097	16,307	19,980

4.3.3 Taxonomy diversity and community structure

In the R1 sample bacteria accounted for 98.82% of the obtained sequences, followed by eukaryota, which accounted for 0.81% and archaea was 0.29%. 98.47% of R2 sequences belonged to bacteria, 1.23% to eukaryota and 0.2% to archaea. Sequences in bulk soil samples were assigned to bacteria 98.61% and 98.53%, eukaryote 0.82% and 1.05%, and archaea 0.48% and 0.34% in B1 and B2 samples, respectively but there was no significant difference ($p > 0.05$) across the two locations. Also, the statistical value for the soil compartment (R1, B1 and R2, B2) is more than 0.05 hence, there was no significant difference. The small percentage of sequences unaccounted for in each sample were from viruses or unclassified.

The most dominant phyla were the *Actinobacteria* accounting for 44.19% and 45.37% for R1 and B1 respectively, while *Proteobacteria* represented 37.99% and 34.43% of the total. *Acidobacteria*, *Bacteroidetes*, and *Planctomycetes* had a relative abundance of 3.32, 2.48 and 2.42% for R1 and 2.76, 3.19, and 2.34% for B1, respectively (Figure 4.3). Whereas, for sample R2, the most dominant phyla were *Actinobacteria* 44.96%, *Proteobacteria* 44.17%, *Bacteroidetes* 1.58%, and *Firmicutes* 1.39% while B2 had *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, and *Bacteroidetes* as the dominant phyla with a relative abundance of 41.05, 38.95, 4.67, and 2.42% respectively. There was no significant difference ($p > 0.05$) in the phylum domain between the location R1, B1

and R2, B2. In addition, there was no significant difference in the phylum domain of the rhizosphere and the bulk soil samples.

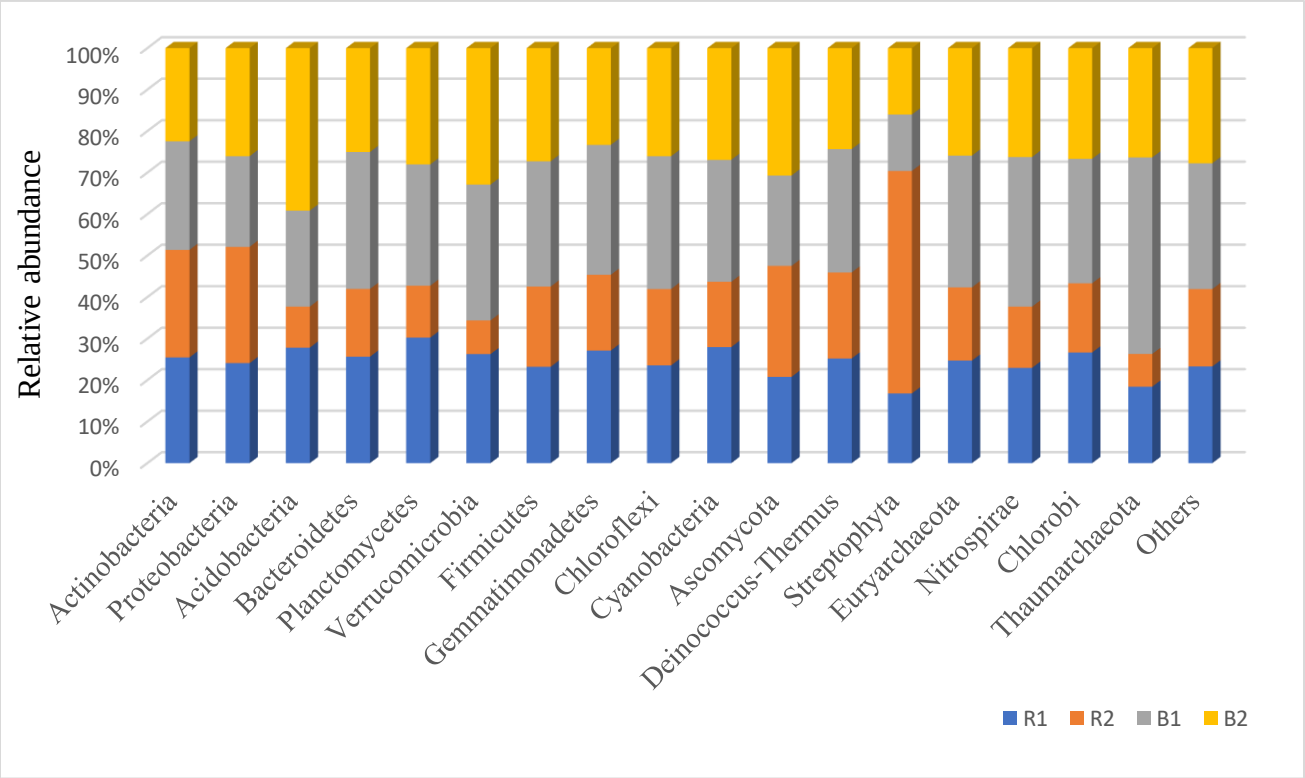


Figure 4.3: Classification of the most abundant phyla observed using shotgun metagenomic sequences in the rhizosphere (R1 and R2) and bulk soil samples (B1 and B2) of two sunflower fields in South Africa, Palmietfontein (R1, B1) and Bloemhof (R2, B2).

At the genus level, there was no significant difference ($p > 0.05$) between the rhizosphere and the bulk soil samples likewise between the locations. The dominant genera for R1 are *Conexibacter* 15%, *Streptomyces* 8%, *Mycobacterium* 7%, and *Geodermatophilus* 5% respectively while for the R2 habitat, *Conexibacter*, *Nocardioides*, and *Streptomyces* accounted for 17, 8, and 7% of the metagenomic sequence respectively. *Conexibacter* 15%, *Streptomyces* 10%, *Arthrobacter* 5%, and

Nocardioides 5% dominated the B1 habitat while B2 had *Conexibacter* 12%, *Streptomyces* 8%, and *Mycobacterium* 7% were the dominant genera (Figure 4.4).

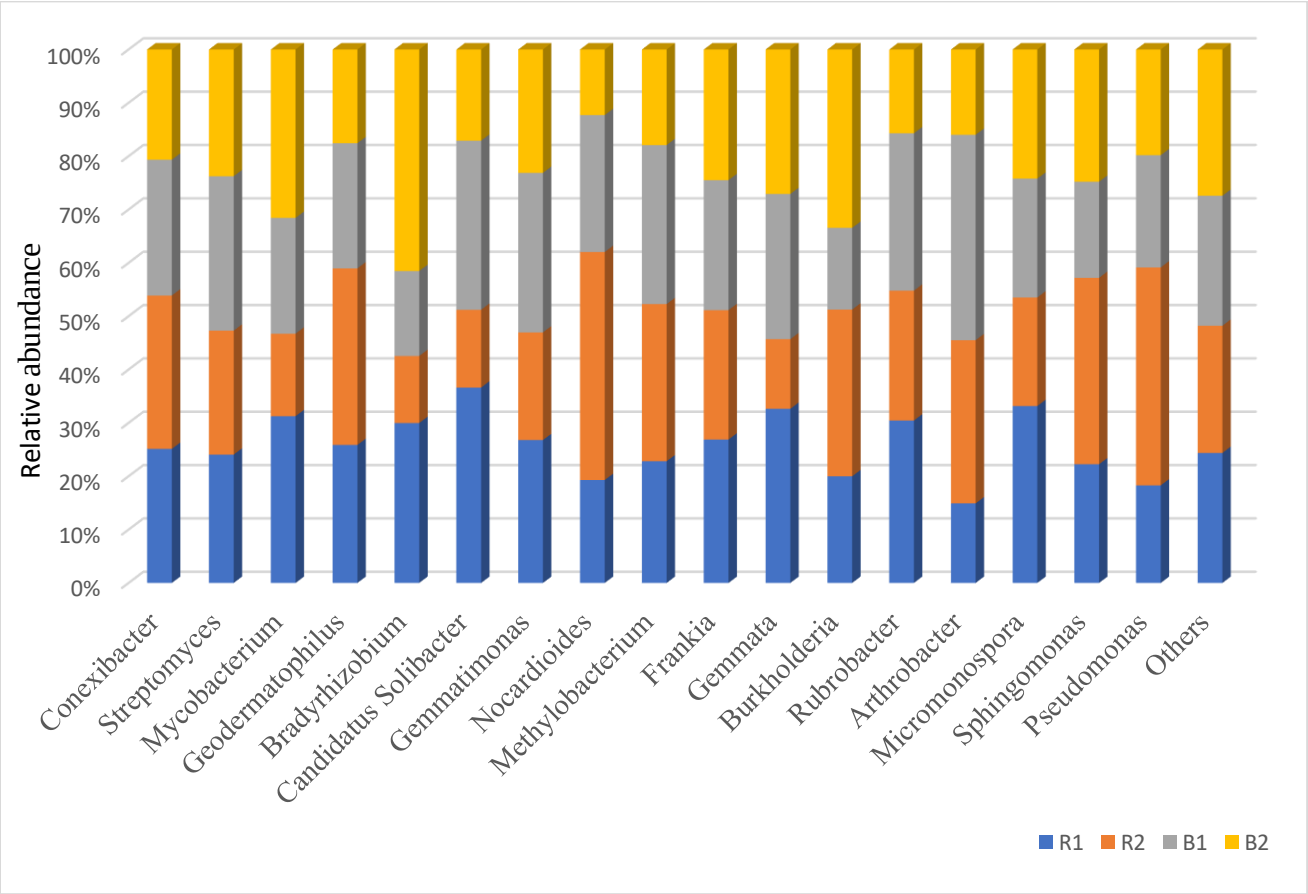


Figure 4.4: Classification of the most abundant genus observed using shotgun metagenomic sequences in the rhizosphere (R1 and R2) and bulk soil samples (B1 and B2) of two sunflower fields in South Africa, Palmietfontein (R1, B1) and Bloemhof (R2, B2).

4.3.4 Microbial community diversity richness and evenness of sunflower rhizosphere R1, R2, and bulk soils B1, B2 examined

α diversity indices (Shannon and evenness) at the genus level presented no significant difference ($p > 0.05$) between the rhizosphere and bulk soil samples. R2 had the highest Shannon diversity but no significance differences were detected ($p > 0.05$) between the other microhabitats. In addition, there was no significant difference ($p > 0.05$) between the microbial diversity of the two locations. The results showed that B2 had the least microbial community diversity (Table 4.3). To test for a significant difference between the β diversity of the rhizosphere soil samples and the bulk soil samples, ANOSIM showed there was significant difference between β diversity of the microbial community (p value = 0.01; R = 0.58). The principal coordinates analysis (PCoA) was used to visualize the β diversity of samples based on the relative abundances of metagenomes obtained from the sunflower rhizosphere and bulk samples (Figure 4.5). Comparing the microbial communities, PCoA indicated no distinct clustering by the rhizosphere and bulk soils and the locations. For example, samples R2 (a-c) were distinct and far from samples B2 (a-c), which explains that microbial structures are unique to those of the rhizosphere R2. Samples R1 (a-c) were slightly away from samples B1 (a-c) which means the community structure are different. In addition, samples B1 (a-c) and B2 (a-c) were close together, which means that the two soil samples are similar. The two locations are also distinct from one another. PCA was used to determine how microorganisms were spread out in the microbial communities (Figure 4.6). The location of the metagenomes shows the total frequency of each sequence that were associated with the structural composition with the vector arrow indicating the organism that most strongly determines the distribution. For instance, microorganisms such as *Oerskovia*, *Telluria*, *Massilia*, *Xanthomonas*,

Capnocytophaga, and *Janthinobacterium* placed the rhizosphere soil microorganism (R2) apart from the microorganisms found in B1, B2, and R1 (Figure 4.6).

Table 4.3: Diversity indices indicating the Simpson, Shannon, and Evenness of microorganism at the genus level

	R1	**R2**	**B1**	**B2**	**P value**
Simpson-1-D	0.6891	0.7804	0.7137	0.6817	0.54
Shannon-H	1.673	2.096	1.745	1.637	
Eveness-e^H/S	0.2538	0.3537	0.2604	0.2235	

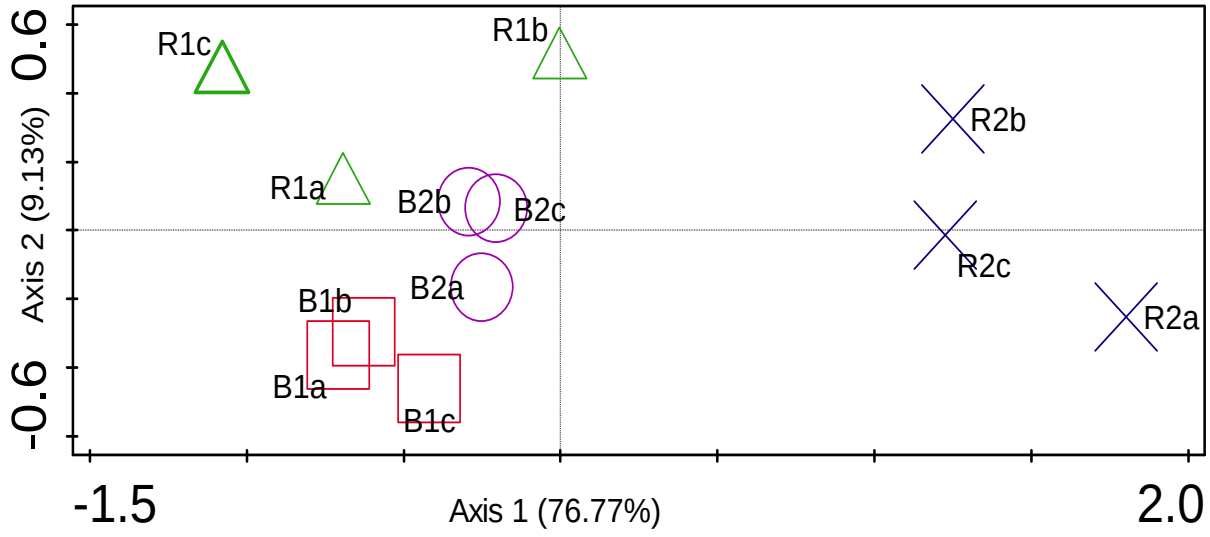


Figure 4.5: Principal coordinates analysis (PCoA) for genus derived from rhizosphere (R1 and R2) and bulk soil samples (B1 and B2), in Palmietfontein (R1, B1) and Bloemhof (R2, B2).

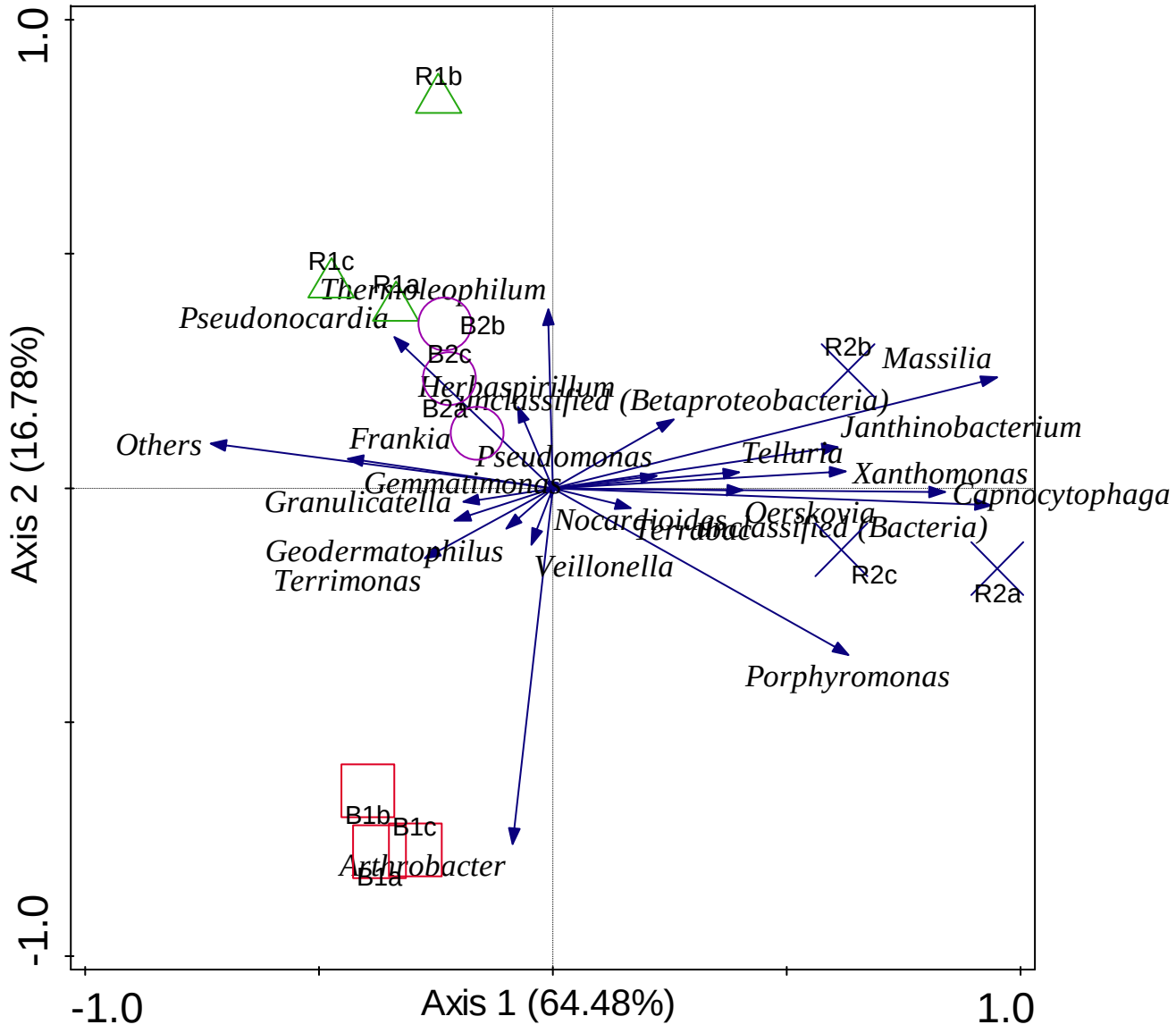


Figure 4.6: PCA of microorganisms (RDP classifier) at the genus level. The length of the vectors reveals the strength of influence of the microorganisms at both rhizosphere and the bulk soil sample.

4.3.5 Diversity indices of functional genes observed from sunflower rhizosphere and bulk soils

The alpha indices which depict the Simpson, Shannon, and evenness (Table 4.4) showed there was no significant difference in the alpha diversity, which is the gene diversity within the habitats. Furthermore, it was checked if there was significant difference in the beta diversity, which indicates the diversity of the functional genes responsible for plant growth in the sunflower rhizosphere compared to the bulk soils. ANOSIM, which provides a way to test statistically whether there is a significant difference between two or more groups of sampling locations showed there was significant difference between the gene orthologs identified in the rhizosphere soil of the two locations (*p* value = 0.01; R = 0.58) as displayed using PCoA (Figure 4.7) where the rhizosphere (R1, R2) is far apart from bulk soils (B1, B2) and the locations, Palmietfontein (R1, B1) and Bloemhof (R2, B2) are not close together.

Table 4.4: Diversity indices of the functional genes observed in our R1, R2, B1, B2 metagenomes

	R1	R2	B1	B2	*P* value
Simpson-1-D	0.7774	0.7841	0.7748	0.7755	0.76
Shannon-H	2.0110	2.0630	2.0040	2.0030	
Eveness-e^H/S	0.2576	0.2714	0.2649	0.2646	

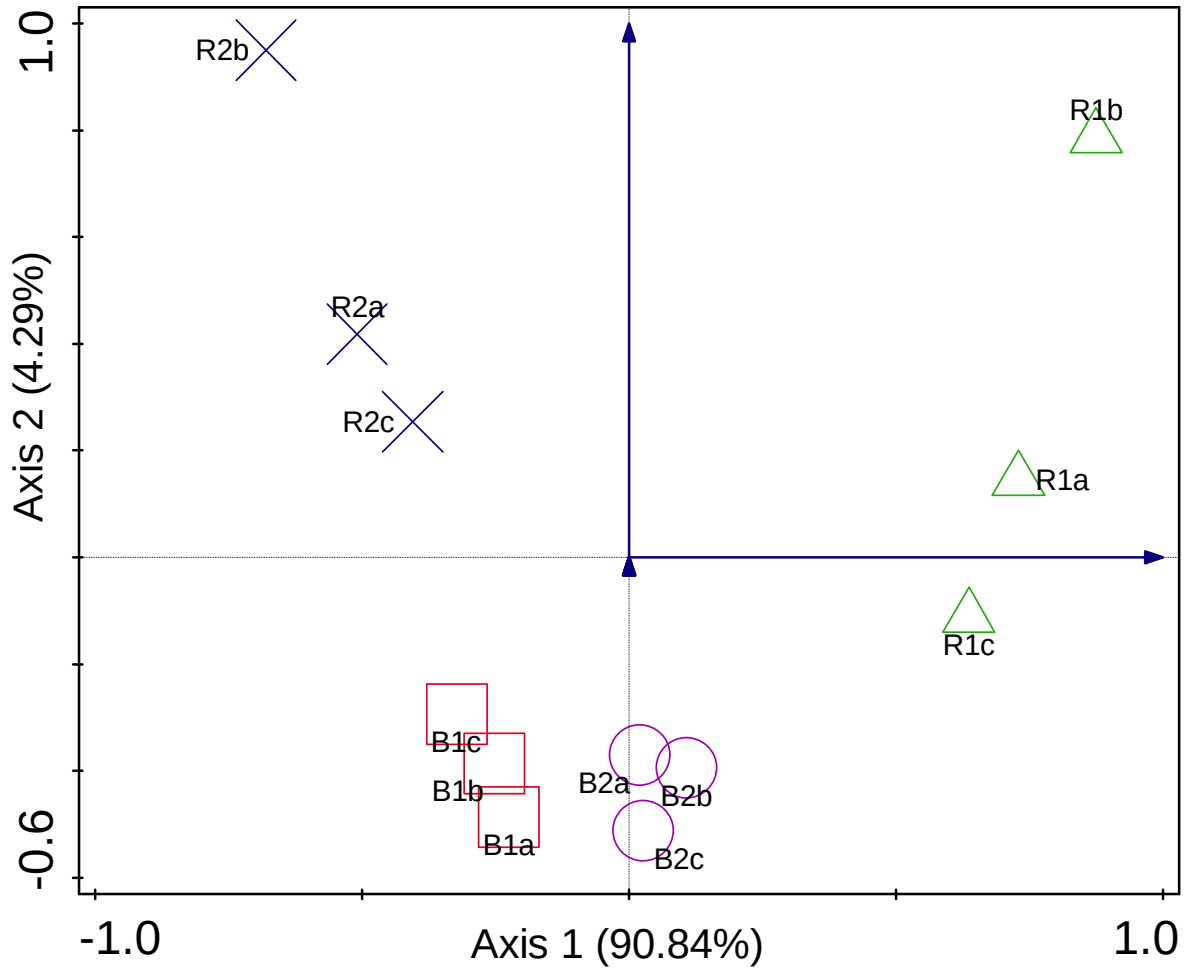


Figure 4.7: The gene diversity of the soil sample locations R1, R2, B1, B2 as displayed using principal coordinates analysis (PCoA)

4.3.6 Microbial diversity and community structure are influenced by environmental variables

The canonical correspondence analysis indicated that richness and distribution of all the environmental variables have influence on microbial diversity and community structure. The correlation between the soil microbial community taxonomic composition and the soil physicochemical parameters was analysed using canonical correspondence analysis (Figure 4.8).

Calcium, Org Carbon, OM, N-NH4, and N positively correlated with abundance of genus such as *Thermoleophilum*, *Frankia*, *Pseudonocardia*, *Nocardioides*, *Pseudomonas*, *Xanthomonas*, and *Herbspirillum*. Similarly, Phosphorus concentration positively correlated with genus *Oerskovi*, *Porphyromonas*, *Veillonella*, *Geodermatophilus* but negatively correlated with *Gemmatimonas*, *Terrimonas*, *Granulicatella*, and *Arthrobacter*. The R2 soil type also correlated with Phosphorus which could explain why diversity was higher in that environment (Figure 4.9).

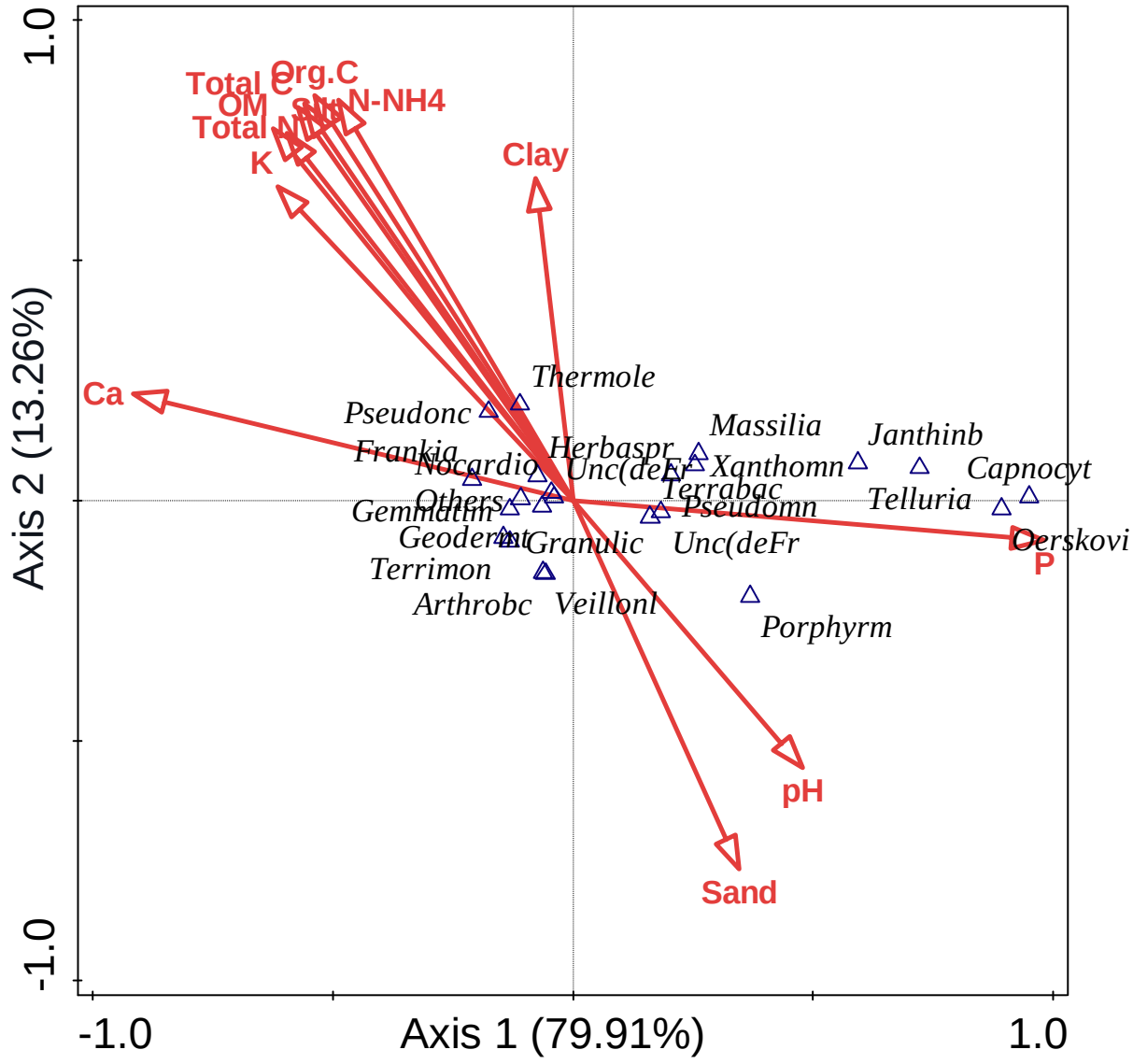


Figure 4.8: Canonical correspondence analysis of the microorganisms at the genus level and soil chemical parameters for both sunflower rhizosphere and bulk soil samples.

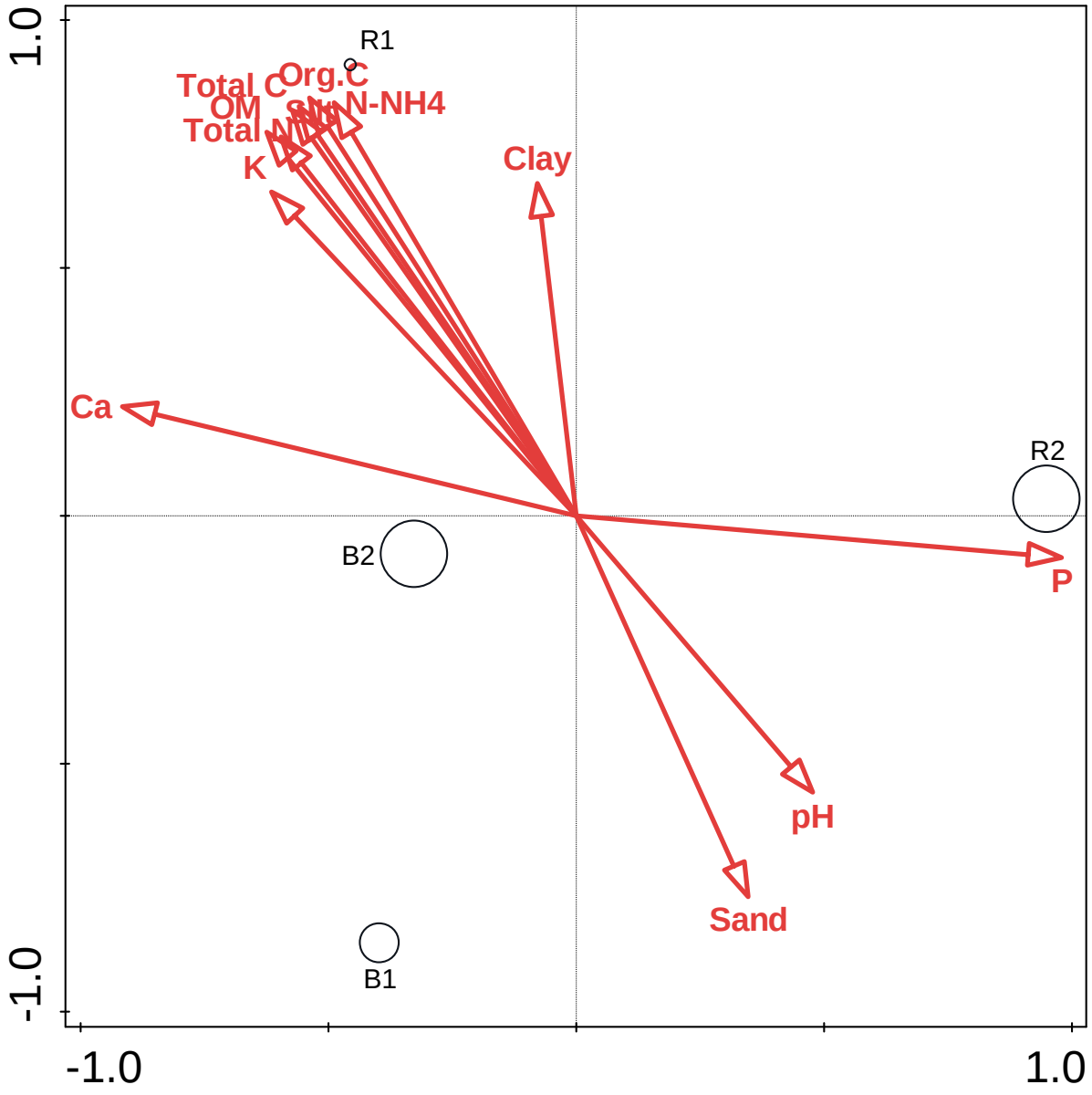


Figure 4.9: Canonical correspondence analysis (CCA) of the major soil chemical parameters based on soil types (R1, R2, B1, B2) explaining combined variation of 93.17%

4.3.7 Plant growth promoting properties from sunflower rhizosphere and bulk soils

Shotgun metagenomics sequencing revealed genes that possess plant-beneficial potential, plant-growth promoting and biocontrol traits with different functions in provision of nutrients, antagonist

tendencies against plant pathogens, synthesis of plant hormones, and hormone level modulation (Figure 4.10). There was no significant difference (*p* value = 0.405) in the genes from the two sampling locations (that is Palmietfontein and Bloemhof). In addition, the genes obtained from the rhizosphere and bulk soil samples was also not significant (*p* value = 0.852).

4.3.7.1 Nitrogen fixing genes

Nitrogen fixing genes are important due to their involvement in the conversion of atmospheric nitrogen to a form plants can utilize. Two types of *nif* genes belonging to the Cysteine desulfurase (EC 2.8.1.7) *nifS* subfamily and Cysteine desulfurase (EC 2.8.1.7) *sufS* subfamily were identified. Sample R1 and R2 has *sufS* gene with relative abundance of 395 and 211 respectively while B1 and B2 had relative abundance of 262 and 272 (Table 4.5, Figure 4.10). The *nifS* subfamily possess relative abundance of 94 and 44 for the R1 and B1 soils, respectively. Another gene that confers fixation of nitrogen observed was the iron-sulfur cluster assembly scaffold protein *nifU* containing relative abundance of 25 in the whole samples. Generally, the relative abundance of *sufS* and *nifS* were much more enriched in the location R1 than R2 with rhizosphere having more abundance than the bulk soil samples.

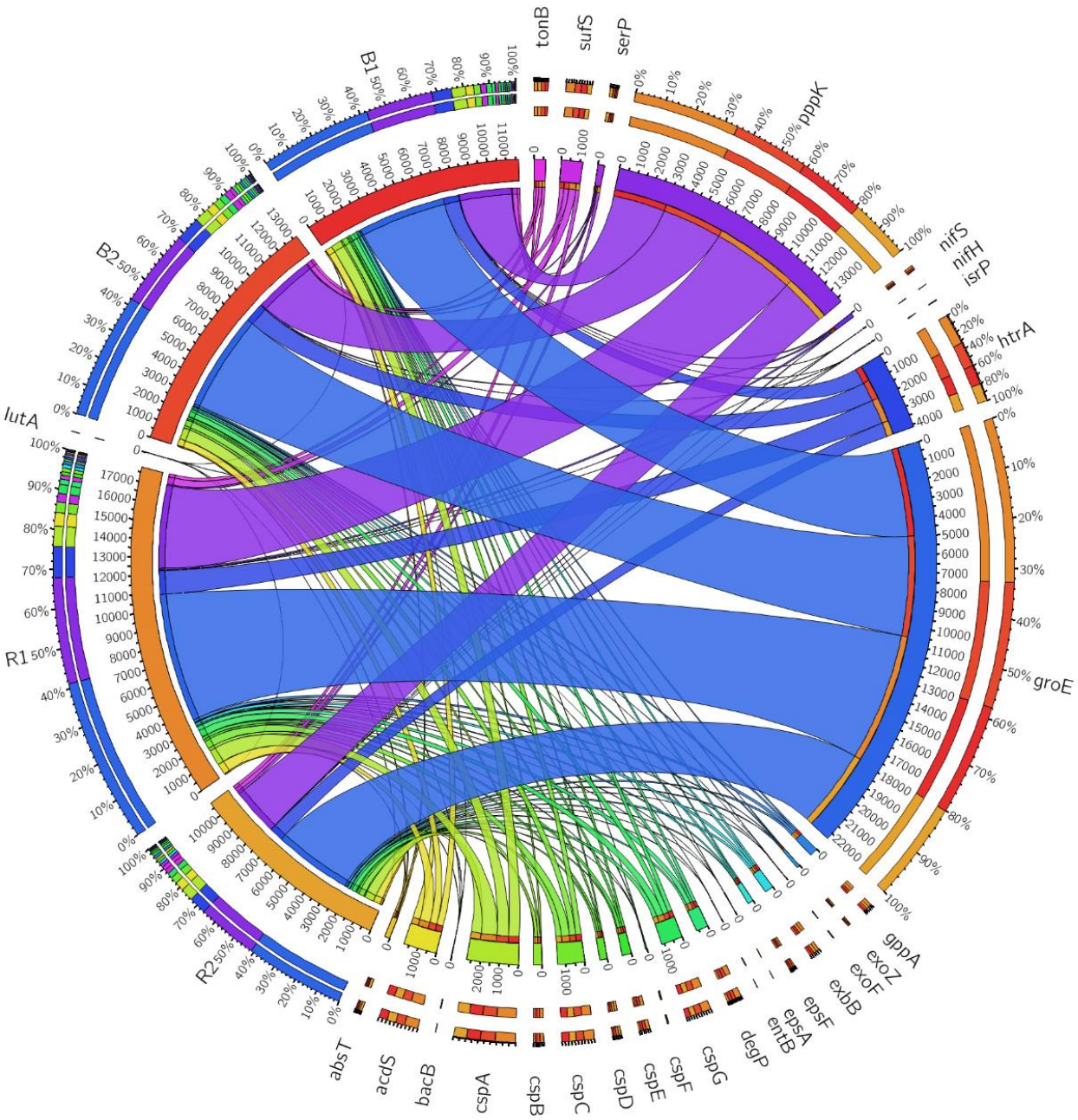


Figure 4.10: Circos showing the relative abundance of beneficial functional genes obtained from sunflower rhizosphere (R1, R2) and bulk (B1, B2) soil samples

For circos visualization purpose, we coded the genes with no aliases as: Siderophore bacillibactin: bacB; Siderophore enterobactin: entB; ABC Fe^{3+} siderophore transporter: absT**;** Iron siderophore receptor protein: isrP**;** Serine protease: serP; Phosphate solubilization polyphosphate kinase: pppK

4.3.7.2 Siderophore producing genes

Two types of 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase genes were found belonging to the (EC 1.3.1.28) bacillibactin and (EC 1.3.1.28) enterobactin. Other key functions of the siderophore producing gene are the ABC Fe3+ siderophore transporter inner membrane subunit which had relative abundance of 73 and 112 for R1 and R2 metagenomes respectively while B1 and B2 possessed relative abundance of 58 and 43 respectively (Table 4.5, Figure 4.10). These genes are responsible for transmembrane transport. Also, R1 was discovered to have 146 producing siderophore of Ferric siderophore transport system with aliases (alternative name for genes) *exbB* while R2 had 116. B1 and B2 had relative abundance of 110 and 103 of Ferric siderophore transport system, respectively. In addition, the biological function dependent siderophore receptor with alias *tonB* had relative abundance of 176 for R1, 163 for R2, 135 for B1 and 161 for B2. From our results, Palmiefontein (R1) location possess high relative abundance of siderophore producing genes of EC 1.3.1.28 bacillibactin, *tonB*, *exbB* and Iron siderophore receptor protein compared to Bloemhof location (R2). The bulk soil samples contain a smaller number of siderophore producing genes when compared to the rhizosphere soils.

4.3.7.3 ACC deaminase producing genes

ACC deaminase possesses enzymatic activities that potentially help in the growth of a plant either in a stressed or normal condition. From our study, only one type of ACC deaminase gene was

discovered belonging to the EC 3.5.99.7 aliases. The sunflower rhizosphere soil samples R1 and R2 possessed relative abundance of 599 and 395 of ACC deaminase while the bulk soil samples B1 and B2 had relative abundance of 371 and 400, respectively. In addition, the location R1 had the higher relative abundance of *acdS* gene than R2 location.

4.3.7.4 Exopolysaccharide producing genes

R1 metagenome with biological function of exopolysaccharide biosynthesis glycosyltransferase with alias of *epsF* (EC 2.4.1.) had relative abundance of 80 while R2, B1, and B2 had relative abundance of 48, 74, and 80, respectively. Other exopolysaccharide producing genes with biological functions such as exopolysaccharide production protein precursor with aliases *exoF* were also found containing abundance of 17 for R1, 3 for R2, while B1 and B2 had relative abundance of 4 and 2 respectively. Exopolysaccharide production protein *exoZ* gene had 38 for R1 metagenome, 45 for R2 metagenome, 24 and 39 for B1 and B2 metagenomes respectively (Figure 4.10). The proteins produced by these genes are potentially responsible for biofilm formation and plant colonization. Our results showed that the genes are abundant in the rhizosphere soils than the bulk soil samples with R1 possessing more abundance than R2 for most genes.

4.3.7.5 Genes potentially contributing to phosphate solubilisation

Phosphorus is the second most important nutrient needed for plant growth. It is found in insoluble form in the soil but can be converted to a soluble form and can then function as biofertilizer for plant use. From our study, several genes potentially responsible for solubilizing phosphate to release soluble form of phosphorus for plant utilization were identified. These genes are polyphosphate kinase with the alias of EC 2.7.4.1 and phosphatase having alias *ppx*/*gppA* family. The R1 location had the most abundant gene of phosphate kinase with 4725 when compared to the

R2 location relative abundance of 2691 (Figure 4.10). Regarding biological function phosphatase *ppx/gppA*, number of potentially phosphate solubilizing genes identified were 145 for R1, 108, 81, and 106 for R2, B1, and B2 metagenomes, respectively. In addition, the rhizosphere was more enriched in genes potentially responsible for solubilizing phosphate than bulk soils.

4.3.7.6 High temperature stress response genes

Strikingly, two types of high temperature stress response genes were discovered in our study. They are involved in the production of protease/chaperone protein and serine protease with aliases *htrA* and *degP/htrA*, do-like EC.3.4.21 respectively. The relative abundance of *htrA* protease/chaperone gene were 1379 and 798 for R1 and R2 location respectively while B1 and B2 had relative abundance of 874 and 1100, respectively. At the same time, serine protease *degP/htrA* EC.3.4.21 had 246 for R1 while R2, B1, and B2 had 104, 148, 151, respectively. These genes confer plants with adaptive mechanisms that help survive stressed environmental conditions. From our results, the rhizosphere is more enriched than the bulk soil samples although, R1 location is the most enriched across the location.

4.3.7.7 Heat and cold stress genes

Well represented from our study are the cold and heat stress genes. *cspABCDEFG*, *groEL* and *groES* (Table 4.5, Figure 4.10). *cspA* had relative abundance of 898 for R1, 557 for R2, 616 and 653 for B1, and B2, respectively. *cspB* had relative abundance of 166 and 63 for both R1 and R2 while B1 and B2 had relative abundance of 93 and 128, respectively. *cspC* had 456 for R1, 319 for R2, 312 for B1 and B2 had 341. The heat shock protein chaperone with aliases *groEL* had relative abundance of 7318 in R1, 4453 in R2 metagenome, 4977 in B1 and 5603 in B2 while co-chaperone *groES* gene had 1409 for R1, 814 for R2, 1016 for B1 and the number for B2 was 1134.

CspABCEG were abundant in the rhizosphere than the bulk soils although rhizosphere location R1 was more enriched than R2.

Table 4.5: Relative abundance of plant growth functional genes as observed in sunflower rhizosphere (R1, R2) and bulk (B1, B2) soil samples

Genes	Biological functions	Aliases	Relative abundance of genes			
			R1	R2	B1	B2
Nitrogen fixation	Cysteine desulfurase *nifS*	EC 2.8.1.7 *nifS*	94	19	44	69
	Cysteine desulfurase *sufS*	EC 2.8.1.7 *sufS*	395	211	262	272
	Nitrogenase (molybdenum-iron) reductase and maturation protein *nifH*	*nifH*	6	5	1	0
Siderophore	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	EC 1.3.1.28 bacillibactin	8	2	1	3
	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase	EC 1.3.1.28 enterobactin	4	4	0	3
	ABC Fe3+ siderophore transporter		73	112	58	43
	Ferric siderophore transport system, biopolymer transport protein *exbB*	*exbB*	146	116	110	103

	Aerobactin siderophore receptor *IutA*	*IutA*	2	7	3	2
	Iron siderophore receptor protein		9	2	6	10
	TonB-dependent siderophore receptor	*tonB*	176	163	135	161
ACC deaminase	1-aminocyclopropane-1-carboxylate deaminase *acdS*	EC 3.5.99.7	599	395	371	400
Exopolysaccharides	Exopolysaccharide biosynthesis glycosyltransferase *epsF*	*epsF* EC 2.4.1.-	80	48	74	80
	Exopolysaccharide production protein *exoF* precursor	*exoF*	17	3	4	2
	Exopolysaccharide production protein *exoZ*	*exoZ*	38	45	24	39
	Exopolysaccharide biosynthesis transcriptional activator *epsA*	*epsA*	4	3	3	3
High temperature stress response	Protease/chaperone protein *htrA*	*htrA*	1379	798	874	1100
	Serine protease, *degP/htrA*, do-like	*degP/htrA* EC 3.4.21.-	246	104	148	151
	Serine protease	EC 3.4.21.-	78	147	67	52
Phosphate solubilization	Polyphosphate kinase	EC 2.7.4.1	4725	2691	2966	3388
	Phosphatase, *ppx/gppA* family	*ppx/gppA*	145	108	81	106
Heat and cold stress shock	Cold shock protein *cspA*	*cspA*	898	557	616	653

	Cold shock protein *cspB*	*cspB*	166	63	93	128
	Cold shock protein *cspC*	*cspC*	456	319	312	341
	Cold shock protein *cspD*	*cspD*	110	116	77	92
	Cold shock protein *cspE*	*cspE*	167	75	115	120
	Cold shock protein *cspF*	*cspF*	18	9	15	21
	Cold shock protein *cspG*	*cspG*	388	232	226	291
	Heat shock protein 60 family chaperone *groEL*	*groEL*	7318	4453	4977	5603
	Heat shock protein 60 family co-chaperone *groES*	*groES*	1409	814	1016	1134

4.4 Discussion

Soil is a very complex ecosystem in which microorganisms play an important role therein. High throughput sequencing is the current gold standard to characterize soil microbial communities. Here, the diversity, structural composition, and functional genes of microbial communities in sunflower rhizosphere and bulk soils were examined using metagenome sequencing. α-diversity was used to estimate the number of taxa (richness) and distribution (evenness) within the microbial community using Shannon and Simpson indices. Contrary to previous results, our study show no significant difference in α-diversity between rhizosphere and the bulk soil communities, which answers our first hypothesis that rhizosphere microbiome would be more diverse than the bulk soil

because of the activities that take place there. Guo et al. (2016) and Cui et al. (2019) reports that plant selects microorganisms and as result diversity of the microbial community is usually low in the rhizosphere when compared to the bulk soils. β-diversity of the microbiome for the two habitats was estimated using canonical correspondence analysis based on Bray-Curtis dissimilarity matrix. PCoA of difference vectors showed clear separations between the sunflower rhizosphere and bulk soil samples and also the two locations sampled.

PCA was used to visualize how distributed and predominant the microorganisms were between the sunflower rhizosphere in comparison to the bulk soil samples and this explains 85.90% of the total variation. The results show that microorganisms such as *Oerskovia*, *Telluria*, *Pseudomonas*, *Massilia*, *Xanthomonas*, *Capnocytophaga*, and *Janthinobacterium* were more in R2 than in B1, B2, and R1. The abundance and distribution around R2 could be as a result of factors such as soil type, management practices, and soil properties and this is in agreement with previous studies (Schlemper et al., 2017; Jiang et al., 2017). On the other hand, genus such as *Terrimonas*, *Veillonella*, *Arthrobacter*, *Granulicatella*, *Geodermatophilus*, *Gemmatimonas*, and *Frankia* are present in higher abundance in bulk soil samples than in rhizosphere samples (Yang et al., 2017).

The canonical correspondence analysis found that Ca^{2+}, Org C, OM, N-NH4, N, P, K^{+}, and pH influenced the microbial diversity and community structure. However, the microbial diversity of each location was directly proportional to the richness and distribution of the physicochemical parameters present. It has been reported that soil microbial diversity is highly influenced by plants through the exudation of carbohydrates, carboxylic acids, and amino acids (García-Salamanca et al., 2013; Oberholster et al., 2018). Other factors such as soil type, nutrition, management practices, soil properties, plant age and crop species affect diversity (Mendes et al., 2014; Yang et al., 2016; Mahoney et al., 2017). In our study, from the soil analysis, we noticed that R2 with

phosphorus value of 74.43 mg/kg was higher in the rhizosphere in comparison to the bulk soil (Figure 4.7). That could explain why diversity is higher in that habitat and most likely due to soil properties or soil type (Hol et al., 2015; Trivedi et al., 2016). The soil samples in this study are slightly acidic (with an interval of 5.78-6.60) and this is also a major part of microbial community predictor. This correlates with a similar finding reported by Rousk et al. (2010), which indicates bacterial diversity and composition have a positive relationship with pH between 4-7 interval. pH is said to have an effect on the composition of other domains, but the influence was far weaker in fungi than for the bacterial community (Rousk et al., 2010).

Profiling the functional microbiome unravelled many potential plant beneficial, plant growth promoting and biocontrol attributes implicated to be involved in functions such as synthesis of plant hormones, provision of nutrients, and signal molecule synthesis necessary for plant microbe interaction (Nelkner et al., 2019). Strikingly from our study, functional annotation of the genes revealed genes coding for nitrogen fixation, siderophore production, 1-aminocyclopropane-1-carboxylate (ACC) deaminase, mineral phosphate solubilization, exopolysaccharide biosynthesis glycosyltransferase, and high temperature stress response genes.

Regarding the genomic insight into the R1, R2, B1, and B2 metagenomes, they all possess genes that encode nitrogen fixation *nifS* and *sufS* with biological function of cysteine desulfurase. The enzymatic reduction of atmospheric nitrogen to ammonium, which is most needed for plant growth, is an important part of the nitrogen cycle and this process is known as biological nitrogen fixation. Biological nitrogen fixation derived nitrogen plays a crucial role in crop production by substituting chemical fertilizers with biological nitrogen (Li et al., 2019b).

However, our metagenomes consist of other types of genes such as *iscU/nifU*, *nifU*, *nifM*, *nifT*, *nifX*, *nifH*, *nifA*, *nifE*, *nifN*, *nifB*, *nifQ*, *nifW*, and *nifO*. Diazotrophic prokaryotes in the ecosystem

have the ability of fixing molecular nitrogen using a reaction known as nitrogenase enzymes. The *nifH* gene encodes the Fe protein, which acts as nitrogenase reductase, therefore the presence of this gene in all diazotrophic bacteria makes it an ideal molecular marker for N-fixation pathway (Zehr et al., 2003; Cobo-Díaz et al., 2015). Nitrogen is one of the most important nutrients needed in crop production. Interestingly, microbial nitrogen cycling was completely represented in our rhizosphere metagenomic samples resulting in functional genes related to nitrogen fixation (*nifH*) and denitrification (*nirV*) (Schmidt et al 2019). In addition, genes responsible for nitrite transporter *nirC*, expression of nitric oxide and nitrite reductase *nir* and *nor* were also discovered (Sessitsch et al., 2012).

In a study by Glick (2012), 1-aminocyclopropane-1-carboxylate (ACC) deaminase that encodes for *acdS* genes was discovered. Plant growth promoting bacteria contain enzymes ACC deaminase that improve the growth of plants by lowering the ethylene level under stressed condition. Ethylene is an important plant hormone because its synthesis is accelerated by biotic and abiotic stresses (Melnikava et al., 2017; Igiehon et al., 2019). The coping mechanism of a plant to abiotic stresses, for example salt stress, can be attributed to the reduction of stress ethylene using ACC deaminase enzymes effected by plant growth promoting rhizobacteria (Singh et al., 2015; Qin et al., 2016). From previous studies, ACC deaminase have also been discovered and characterized in certain plant associated fungi, for example the biocontrol strain *Trichoderma asperellum* (Viterbo et al., 2010). Interestingly, our results displayed phylum *Ascomycota* from the eukaryote domain, which *Trichoderma asperellum* belongs to. Thus, microbial *acdS* can be an approach to withstand stressed conditions.

The metagenomes also contain quite large numbers of siderophore producing genes, which assist in iron acquisition. Siderophores are small molecules synthesized and secreted by a wide range of

microorganisms to scavenge iron (Thode et al., 2018). There are two ways plants acquire iron. The first is the acidification of the rhizosphere and the reduction of Fe^{3+} ions by membrane bound Fe^{3-} chelate reductase, after which the root cells absorb Fe^{2+}, while the second approach in the acquisition of iron is the secretion of low molecule plant siderophore in order to solubilize the bound, iron after which the membrane proteins transport it into the root cells (Altomare and Tringovska, 2011; Radzki et al., 2013). Kumar et al. (2018) described some siderophore producing microorganisms obtained from the rhizosphere, which enhance plant growth in wheat cultivation and at the same time inhibiting plant pathogen *Fusarium solani*. Siderophores act as virulence factors in several pathogenic microorganisms and it has been proven that siderophores of mutualistic and commensal species can help minimize the increase in plant pathogens (Kramer et al., 2019). The siderophore genes observed in our study are of the aliases EC 1.3.1.28, which are involved in the production of 2,3-dihydroxybenzoate, an important precursor of the siderophores bacillibactin and enterobactin (Liu et al., 2017; Xu et al., 2019). Of note are the siderophore receptors gene like *lutA,* which also play a key role in virulence in pathogens as mutation in siderophore receptors alter the transport of siderophore-iron complex (Khan et al., 2018). In addition, the metagenomes revealed genes Ferric hydroxamate ABC transporter (TC 3.A.1.14.3), *fhuC, fhuD,* and *fhuB,* which help in ATP-binding protein, periplasmic substrate binding protein, and permease component, all coding for Ferric hydroxamate ABC transporter. In a study by Dimkpa et al. (2009) hydroxamate siderophore produced by *Streptomyces tendae* F4 strain was proven to enhance cadmium and iron uptake in sunflower plants, consequently promoting the growth of the plant.

Phosphorus is a key component in plant nutrition, playing an important function in metabolic processes such as photosynthesis, nutrient transport, cell division, and signal transduction. Most

organic and inorganic phosphates in the soil are in insoluble form and a sizeable number of them cannot be used by plants (Alaylar et al., 2019). Phosphate solubilizing genes were well represented in our study. *pqqABCDEF* gene, a protein coenzyme that is involved in the biosynthesis of pyrroloquinoline quinone (PQQ), plays an important function in solubilizing phosphates in the soil thereby making it useful for plants (Enagbonma and Babalola, 2020). In a study conducted by You et al. (2020), bacterial isolate *Burkholderia cenocepacia* CR318 isolated from the rhizosphere of maize was confirmed to have a mechanism that can solubilize inorganic phosphate and subsequently improved the growth of the maize plant. Phosphate-solubilizing microorganisms have been well established in the rhizosphere, where they possess the ability of solubilizing soil insoluble phosphate by releasing organic acid, most especially gluconic acid (GA), 2-ketogluconic acid, tartaric acid, oxalic acid, citric acid and other carboxylic acids (Farhat et al., 2009). The organic acid chelates divalent cations from complex phosphorus mineral-like tricalcium phosphate and hydroxyapatite and releases free P, which the plant eventually takes up (Bharwad and Rajkumar, 2020). Findings have shown the important role PQQ synthesis-using phosphate solubilizing microorganisms play, thus establishing microbial phosphate solubilization as one of the important requirements for plant growth (Misra et al., 2012).

Serine protease (EC 3.4.21.-) are enzymes cleaving peptide bonds normally found in proteins. They are ubiquitous in nature, present in both prokaryotes and eukaryotes, with important biological functions (Neitzel, 2010). Proteases that are found in soil originate from different sources, which includes plants, microorganisms, and animals. For instance, the proteases from plant biomass or microbial processes have been discovered to be relevant in the way a plant responds to environmental circumstances across several ecosystems (Vranova et al., 2013). We discovered two types of high temperature stress response genes *degP/htrA* in our study. *htrA* serves

as a stress response protease that helps in degrading proteins that are damaged due to environmental stresses, most especially high temperature, and also helps against oxidative damage (Phillips and Roop, 2001).

Furthermore, exopolysaccharide producing genes were represented in our study. *epsF* had biological functions of exopolysaccharide biosynthesis glycosyltransferase while *exoF* had biological function of exopolysaccharide production protein precursor. These EPS genes have shown evidence that they are responsible for biofilm formation. These plant associated biofilms have biological functions such as protection against numerous environmental stresses which include pH changes, osmotic shock, desiccation, and salinity (Asari, 2015). Other important functions include reduction in microbial competition and protecting the host plant thereby increasing growth and yield of the crop (Asari, 2015). In a study by Kasim et al. (2016) it was shown that biofilm formation was increased with increasing salt concentration, which explains why it is most useful during salinity stress.

Another notable gene represented in our study is the *phzF* encoding gene, which is responsible for phenazine synthesis. The *phzF* proteins are implicated in the production of phenazine derivative antibiotic and antifungal compounds against bacteria and fungi (Stephan et al., 2019). For instance, phenazines produced in *Pseudomonas* are involved in the biocontrol of many diseases including soil-borne pathogens (Arseneault et al., 2016). In addition, we discovered genes responsible for stress alleviation from our metagenomic samples of sunflower rhizosphere microbiome origin. Genes *cspABCDEFG*, *groES*, and *groEL* that are responsible for cold shock and heat shock protein production were observed in our study. These results were in accordance with a study by Enagbonma and Babalola (2020) where these genes were responsible for cold and heat shock response. These genes come into play during drought periods when water is a problem. Drought

is a major limiting factor in agricultural production and a plant is said to undergo drought stress either due to difficulty in the water supply to the root or when the transpiration rate becomes extremely high (Ghobadi et al., 2013). In addition, chaperones encoding gene for heat shock are not only expressed when the temperature is high but also in response to other environmental stresses such as salinity and osmotic water, cold, and oxidative stress (Wang et al., 2004).

Conclusion

This study gives insight into the structural diversity, composition, and plant growth functional genes in sunflower rhizosphere and bulk soil from two different locations. Our shotgun metagenomics approach showed bacteria as the most dominant domain followed by eukaryote, and archaea. We found no significant difference in the α-diversity (the microbial diversity within the habitats) but there was a clear separation in the β-diversity (the microbial diversity between the habitats), which confers significant difference between the sunflower rhizosphere and bulk soil samples and also the two locations. The outcomes from this research suggest that sunflower would thrive in both locations because compositionally the microbial diversity was not significant in the two locations although there was difference in the microbial structure of both R1 (Palmietfontein) and R2 (Bloemhof) sites. Metagenomes of the agricultural sunflower soil revealed the phyla *Actinobacteria* and *Proteobacteria* as the dominant members of the community and these phyla have been established to produce several enzymes with growth promoting traits on plants. The metagenomes also have an infinitesimal number of phyla *Euryarchaeota*, *Thaumarchaeota*, *Ascomycota*. The effect of physicochemical parameters on the microbial structural composition was also a determining factor in driving sunflower microbiome. These findings suggest that edaphic conditions drive microbial communities. We identified genes that confer fixation of

nitrogen, siderophore production, ACC deaminase producing genes, heat and cold shock genes, phosphate solubilizing producing genes, among others. There was no significant difference in the genes responsible for plant growth in the sunflower rhizosphere microbiome in comparison to the bulk soils, likewise there was no significant difference between the two locations. Although, for most of our identified genes, the location R1 possessed more enriched genes for potential plant growth than the location R2. Conclusively, these genes can be harnessed for biotechnological applications in producing bio-products such as fertilizers and pesticides to enhance food production. Since some of these microbes are still yet unculturable, micro-cultivation technology can be potentially harnessed where other cultured microorganisms can be co-cultured with uncultivable microorganisms as recreating this method in the laboratory will increase throughput and access to unique species for commercial production.

CHAPTER FIVE

UNVEILING THE METABOLIC PATHWAYS AND FUNCTIONAL DIVERSITY IN SUNFLOWER RHIZOSPHERE MICROBIOME AS REVEALED BY SHOTGUN METAGENOMICS

Abstract

Profiling the metabolic processes in the microbiome is important in understanding and engineering the plant microbiome for beneficial use. The metagenomics approach can reveal the metabolic potentials and functional groups of microorganisms, thereby reducing the complexity of the rhizosphere microbiome. In this study, we describe a functional characterization of the microbiome from the rhizosphere and bulk soil samples of sunflowers using shotgun metagenomics. Soil samples from the rhizosphere and bulk of sunflower were sequenced using Illumina HiSeq platform. Our results revealed there was no significant difference in the α diversity of the functional diversity between the two locations. In addition, there was no significant difference within the sunflower rhizosphere and bulk soil samples while β diversity varied significantly between the two microhabitats. Our metagenomic dataset identified pathways like nitrogen, phosphorus, and iron that have contributed to nutrient cycle. From our study, physicochemical properties of the soil such as PO_4^{3-}, pH and Total N were the major drivers of the microbiome of sunflower functional categories. These results provide insights into the metabolic potential of microorganisms in the ecosystem, which when harnessed could be useful in biotechnological application for sustainable agriculture.

Keywords: Biotechnology, Illumina platform, metabolic potential, nutrient pathways, soil ecosystem

5.1 Introduction

Microorganisms found in many habitats do not live in isolation but coexist in a complex system in which interactions occur among members. Thus, they are important in community assembly and ecosystem function (Hallam and McCutcheon, 2015). In a plant, roots release exudates in the rhizosphere. Because of intense microbial interaction and activities taking place there, it is crucial to identify and define the interactions among soil microorganisms to understand structural and functional microbial diversity (Hallam and McCutcheon, 2015; Ren et al., 2015).

The rhizosphere microbiome is a central determinant of plant performance, providing important functions such as mineralization of nutrients, manipulation of plant hormonal balance, and suppression of potential pathogens (Gao et al., 2019). The microorganisms in the rhizosphere support plant physiological processes such as growth and energy metabolism. In contrast, plants secrete metabolites and mucilage that can influence the growth of microorganisms (Fonseca et al., 2018).

A total of approximately 5-21% of the carbon fixed through photosynthesis is transported to the soil in the form of root exudates, and this serves as a major attraction for soil microorganisms. Exudates released from the roots such as carbohydrates and amino acids are chemical attractants, resulting in an increase in the number of certain microorganisms in the rhizosphere in comparison to the bulk soil (Jain et al., 2020).

The metabolic activity of individual organisms is linked to their community and ecosystem. Organisms take up energy and resources from the environment, convert them into other forms, and excrete them back into the environment (Ofaim et al., 2017). Several important enzymatic activities such as hydrolysis of C-substrates and organic forms of nutrients such as N, P and S take place to a greater extent in the root region than in bulk soil (Wei et al., 2019), implying that some specific microbes, which are active in the rhizosphere have a relationship with plants. Metagenomics can identify and decipher functional genes and metabolic processes and, most importantly, present information on the biogeochemical cycles and the functional activities of the microbiome (Mahoney et al., 2017; Mattarozzi et al., 2020).

Metagenomes from shotgun sequencing of environmental samples reveal a sizable amount of information on the microbiome. To better understand the activities of the microbiome and their impact on the environment, protein prediction and taxonomic analysis of the metagenome is determined. However, due to the metabolic function and inter dependency between the microorganisms, it is imperative to study the connections between the composition of the microbiome and their functional activity in the ecosystem (Ruiz-Moreno et al., 2019).

Because the rhizosphere is more dynamic and microorganisms there provide basic ecosystem functioning required for plant growth and health, we hypothesized that relative abundance of nutrient uptake and processing genes relating to cell wall and division, iron acquisition and metabolism, photosynthesis, cofactor, vitamins, prosthetic group, and pigment should be higher in the rhizosphere than in the bulk soil.

The study aimed to investigate the sunflower microbiome using next generation sequencing to elucidate microbial functional annotation and mapping into metabolic pathways and to determine

the main environmental factors driving the functional diversity of the microbial community of the sunflower rhizosphere and bulk soils. Understanding the linkage between the soil microbiota and their metabolic functions is important for defining agricultural and ecological processes and could also be used as a biomarker of soil health (Salam and Obayori, 2019).

5.2 Materials and methods

5.2.1 Soil sampling and analysis

Rhizosphere soil samples from Palmietfontein (R1), Bloemhof (R2) and bulk soil samples from Palmietfontein (B1), Bloemhof (B2) were collected in triplicate at 0-20 cm depth using 2cm soil auger (Mendes et al., 2014). The location GPS are Palmietfontein (26º19'39.94"S: 26º52'52.57"E), and Bloemhof (26º17'46.1"S: 26º58'19.83"E). After collection, samples were placed in a cooler box and brought to the laboratory where they were stored at 4°C before being used.

Phosphorus (P) was analyzed using the P Bray method, while Total Carbon (C) and Nitrogen (N) were determined using a TruSpec elemental determinator (Oberholster et al., 2018). A pH meter using the ratio 1:2.5 (soil/water) was used to measure the soil pH, according to Enagbonma et al. (2019). Organic matter content was measured using Walkley-Black method. The soil moisture was determined by the procedure described by Cui et al. (2019) while the hydrometer technique was used for the particle size analysis (Kettler et al., 2001). The particle size classes used to assign texture was of the United States Department of Agriculture (USDA) with sand (2.0-0.5 mm), silt (0.05-0.002 mm), and clay (< 0.002 mm). Organic carbon was analyzed using the dichromate digestion (Shi et al., 2011). Potassium (K) and calcium (Ca) were analyzed by methods described by Deke et al. (2016).

5.2.2 DNA extraction from soil samples, sample preparation, and sequencing

Total genomic DNA was extracted from each soil sample using PowerSoil® isolation kit (MO Bio labs, USA), following the manufacturer's user guide. 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) Texas, USA.

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 from each sample was used for library preparation. After DNA fragmentation Illumina sequencing adapters were added, and products amplified using 6 cycles of PCR during which unique indices were added to each sample. After library amplification, their concentration was estimated using the Qubit® dsDNA HS Assay Kit (Life Technologies). Right before libraries were pooled in equimolar ratios of 0.7nM, Agilent 2100 Bioanalyzer (Agilent Technologies) was used to measure the average library fragment size. It was then sequenced at paired-end for 300 cycles using NovaSeq 6000 Illumina platform.

5.2.3 Data processing and statistical analysis

The online metagenomics rapid annotation server MG-RAST (www.mg-rast.org) was used for the quality control of the raw metagenomic sequences (Meyer et al., 2008). Low-quality sequences were removed according to the procedure of Cox et al. (2010) where a modified DynamicTrim was used to remove sequences having base pair >5 with 15 phred score cutoff and also length filtering, which is sequences having length >2 standard deviation from the mean. Artificial replication reads that occurred during sequencing were also removed (Gomez-Alvarez et al., 2009). After performing quality control (QC), BLAT (the BLAST-like alignment tool) algorithm

was used to annotate the sequences (Kent, 2002) against the M5NR database (Wilke et al., 2012), which encompasses nonredundant integration of many databases. The SEED and KEGG subsystems were used for the microbial community groupings, functional, and gene analysis. SEED subsystem level 1, and 3 databases were used for the functional analysis while KEGG (functions) was used for the gene analysis (Mitra et al., 2011). An e-valve of 5, with the maximum identity of 60%, maximum alignment length of 15 base pairs, and maximum abundance of 1 were used when mapped against the subsystems databases which integrated SILVA, Greengenes, and RDP databases (Garcia-Mazcorro et al., 2017).

After the dataset for each replicates of the rhizosphere and bulk soil samples of the two locations was normalized by finding the average and relative abundance, One-Way Analysis of variance (ANOVA) was used to determine if there was a significant difference in functional analysis in the rhizosphere and bulk soil samples, and Tukey's pairwise comparison was used for the means at a significant level (P valve < 0.05). Diversity indices (alpha and beta diversity) of the functional analysis at SEED subsystem level 1 was also determined. These analyses were done using the PAST software version 2.17c (Hammer et al., 2001). Principal coordinates analysis (PCoA) based on Bray Curtis dissimilarity matrix was used to visualize the functional diversity from both rhizosphere and bulk soil habitats while the principal component analysis (PCA) based on a Bray-Curtis dissimilarity matrix was used to view the distribution of the functional diversity between the two microhabitats.

CANOCO 5 (Microcomputer Power, Ithaca, NY) was used to view the PCA and PCoA. The Shinyheatmap was used to plot the heatmap with transformed z-score displaying a relative abundance of functional diversity (Khomtchouk et al., 2017). The raw sequence has been deposited into the GenBank database (https://www.ncbi.nlm.nih.gov/search/all/?term=SRR10418054),

where SRA accession number for R1 is SRR10426233, R2 SRR10418054), while B1 and B2 have accession number of SRR10426310 and SRR10418081, respectively.

5.3 Results

5.3.1 Sequencing statistics and soil physicochemical analysis

After undergoing quality control, the bp count for R1 was 2,388,568,202 bp, sequence count of 13,736,515, mean square length of 174 bp, and mean GC percent of 66. R2 had bp count of 1,390,678,255, sequence count of 7,559,132 mean square length of 184, and GC content of 67%. B1 and B2 had bp count of 1,551,871,377 bp and 1,789,010,308 bp respectively, sequence count of 8,282,394 and 9,753,315, mean sequence length of 187bp and 183bp, and GC content of 66% and 65% respectively.

From our results, Potassium (K) measured ranged between 220-349 mg/kg. Org. Carbon ranged between 0.29-1.77%. The total N was rather low (0.047-0.134%), while total C ranges between 0.505-1.60%. The pH of our samples is between 5.78-6.60, which shows slightly acidic to nearly neutral. The range of the particle size is 72-84% for sand, 0-6% for silt, and 16-24% for clay.

5.3.2 Microbial taxa group (family) found in sunflower soil samples

Bacteria dominated the metagenome for both rhizosphere and bulk soil followed by eukaryote and then archaea, although there were small numbers of unclassified sequences, viruses, and other sequences. In the rhizosphere soils, families *Pseudomonadaceae*, *Micromonosporaceae, Thermoleophilaceae, Frankiaceae, Oxalobacteraceae*, *Flavobacteriaceae, Xanthomonadaceae,* and *Sphingomonadaceae* were dominant, whereas the bulk soils were dominated by *Micrococcaceae, Veillonellaceae, Gemmatimonadaceae,* and *Carnobacteriaceae*. There was no

significant difference in the relative abundance of the family between the two locations ($p > 0.05$) and also no significant difference between the rhizosphere and bulk soils.

5.3.3 Functional analysis of the metagenomes

The SEED subsystem was used to detect the functional categories. At SEED subsystem level 1, twenty-eight functional categories were observed with carbohydrate having the highest with 14.19% and 13.74% for R1 and R2 respectively (Figure 5.1), followed by clustering-based subsystems of 12.96, 12.93, 13.03, and 12.91% for R1, R2, B1, and B2 metagenomes, while amino acids and derivatives had 10.45 and 10.30% for R1 and R2 metagenomes, and 10.46 and 10.26% for B1 and B2 metagenomes. There was significant difference in the functional categories between the rhizosphere and the bulk soil samples and the two locations also showed significant difference (p value < 0.05). Principal component analysis (PCA) was conducted to display the level of distribution of functional categories between the sunflower rhizosphere and bulk soil samples (Figure 5.2). The vector arrows of motility and chemotaxis, photosynthesis, stress response, and iron acquisition and metabolism were more concentrated in the sunflower rhizosphere location (R1) than other locations (B2, R1, and B1).

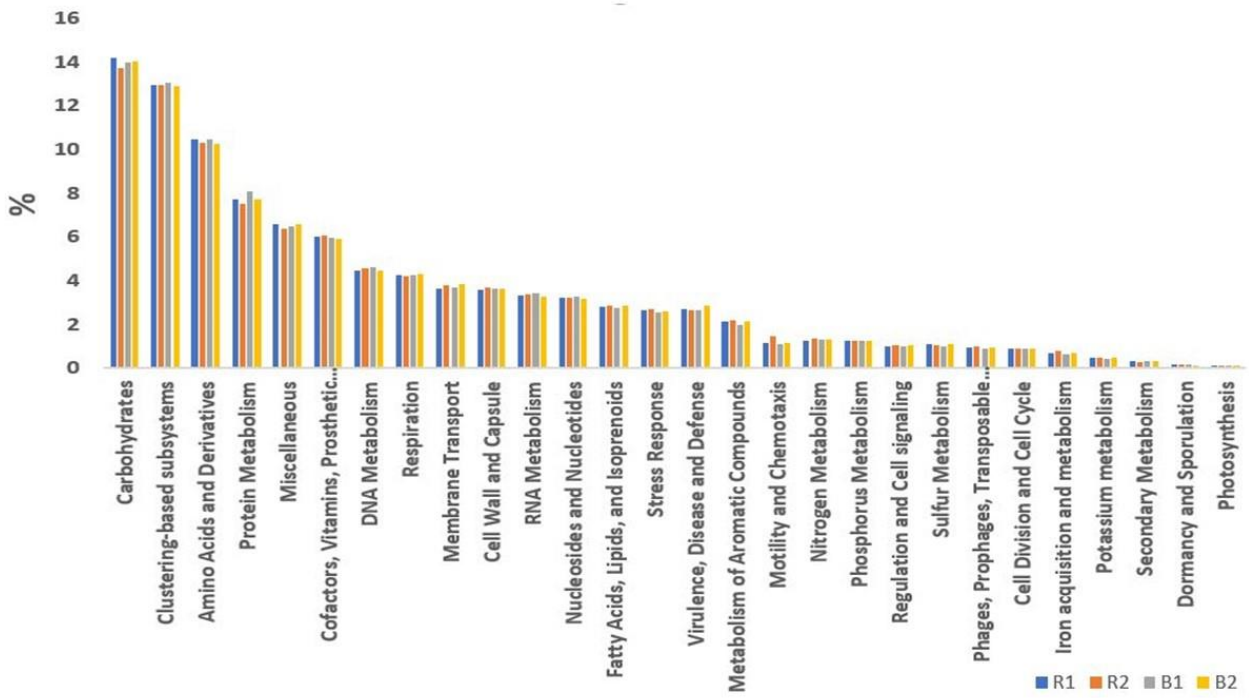


Figure 5.1: Classification of the functional categories using SEED subsystem level 1 in the sunflower rhizosphere and bulk soil samples.

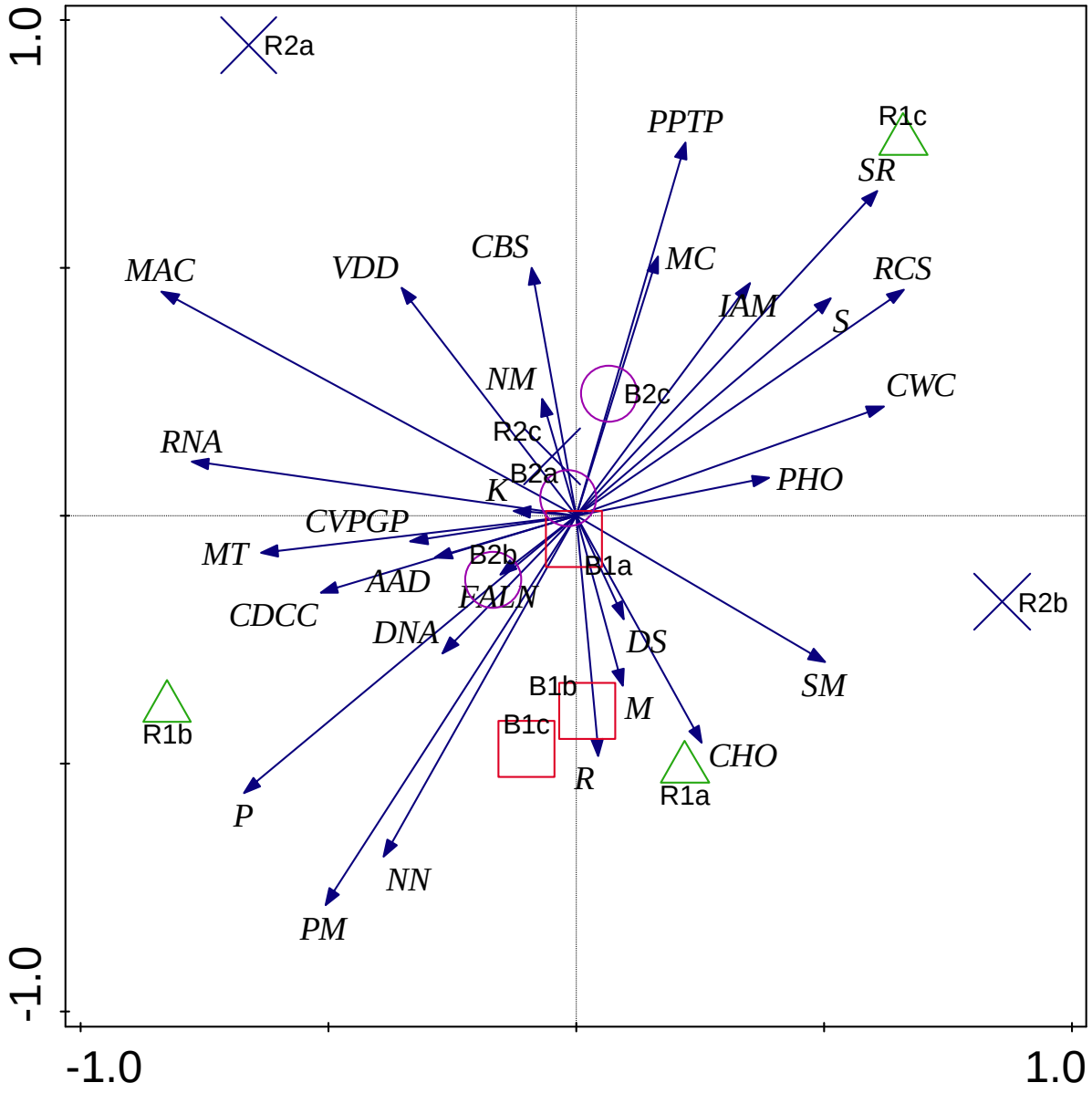


Figure 5.2: Functional analysis at subsystem level 1 as displayed by PCA. The length of the vectors reveals the strength of metabolic processes. Axis one explains variation of 30.12% while axis two explains variation of 22.27%. **Abbreviations:** carbohydrates (CHO), clustering-based subsystems (CBS), amino acids and derivatives (AAD), protein metabolism (PM), miscellaneous (M), cofactors, vitamins, prosthetic groups, pigments (CVPGP), DNA metabolism (DNA), respiration (R), membrane transport (MT), cell wall and capsule (CWC), RNA metabolism (RNA),

nucleosides and nucleotides (NN), fatty acids, lipids, and isoprenoids (FALN), stress response (SR), virulence, disease and defense (VDD), metabolism of aromatic compounds (MAC), motility and chemotaxis (MC), nitrogen metabolism (NM), phosphorus metabolism (P), regulation and cell signaling (RCS), sulfur metabolism (S), phages, prophages, transposable elements, plasmids (PPTP), cell division and cell cycle (CDCC), iron acquisition and metabolism (IAM), potassium metabolism (K), secondary metabolism (SM), dormancy and sporulation (DS), and photosynthesis (PHO).

5.3.4 Functional diversity indices observed in sunflower microbiome

The alpha indices, which depict the Simpson, Shannon, and evenness (Table 5.1) showed there was no significant difference in the alpha diversity, which is the diversity within the habitats. Using ANOSIM (*p* value = 0.01; R = 0.58), there was a significant difference in the beta diversity, which indicates the diversity of the functional categories between the sunflower rhizosphere and bulk soils and there was also significant difference in the two locations as displayed using PCoA (Figure 5.3) whereby the rhizosphere (R1, R2) is apart from bulk soils (B1, B2) and the locations, Palmietfontein (R1, B1) and Bloemhof (R2, B2) showed slight separation.

Table 5.1: Diversity indices in the functional attributes in the rhizosphere and bulk soil samples

	R1	R2	B1	B2	P value
Simpson-1-D	0.9262	0.9278	0.9259	0.9269	0.99
Shannon-H	2.875	2.89	2.869	2.881	
Eveness-e^H/S	0.6329	0.6425	0.6291	0.6368	

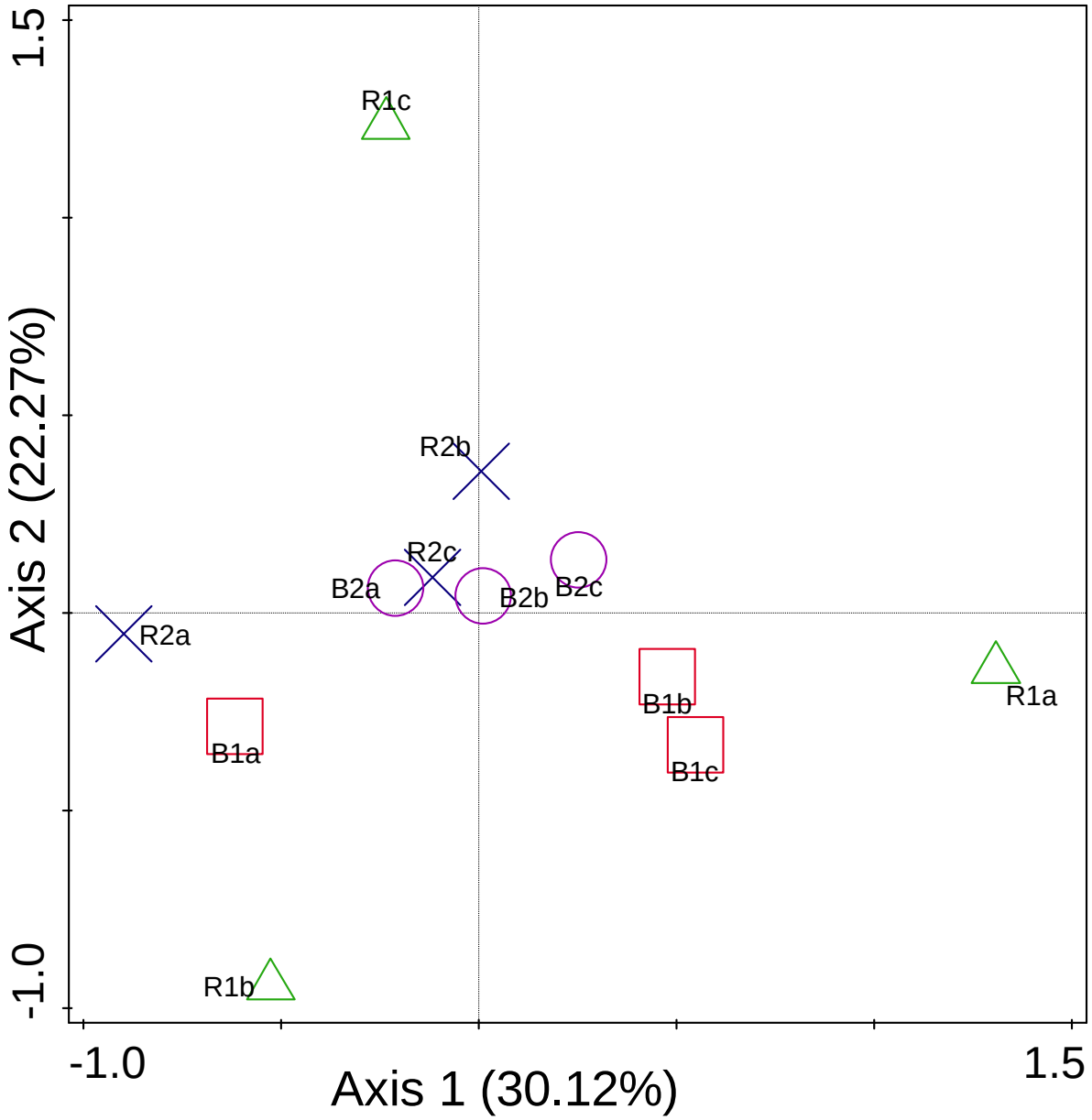


Figure 5.3: Functional analysis (subsystem level 1) derived from sunflower rhizosphere and the bulk soils being displayed using principal coordinates analysis (PCoA).

5.3.5 Relationship between functional analysis categories and environmental variables

To test our hypothesis on whether the functional diversity of the microbiome is driven by the physicochemical parameters, the impact of the measured physicochemical parameters of the soil and relative abundance of the functional analysis categories at subsystem level 1 was evaluated using canonical correspondence analysis (CCA). Based on significant test analysis through forward selection, the results (Table 5.2) revealed P, pH, and Total N were the environmental factors that best explained variation in the functional categories (Figure 5.4). Considering the length of the vector of pH, it was a strong determinant in functional categories. pH on the axis 2 positively correlated with the likes of nitrogen metabolism, DNA metabolism, dormancy and sporulation, cell wall and capsule while P on the axis 1 positively correlated with iron acquisition and metabolism, photosynthesis, motility and chemotaxis but negatively correlated with secondary metabolism.

Table 5.2: Environmental parameters that strongly explain the variation in functional category structure

Environmental variables	Contribution %	Pseudo-F	p
P	67.0	4.1	0.499
pH	28.9	6.9	0.776
Total N	4.2	<0.1	1

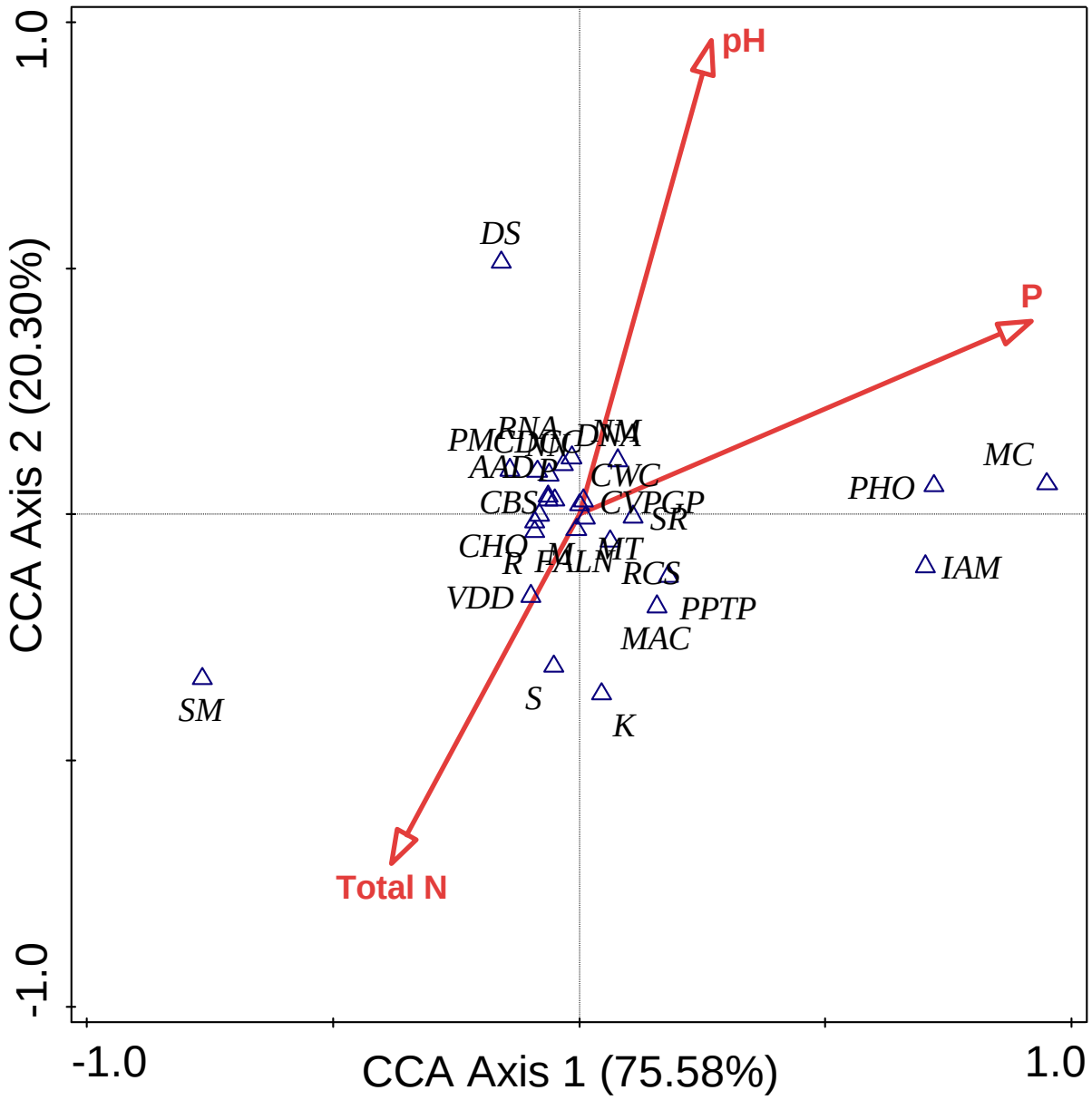


Figure 5.4: Physicochemical parameters and functional categories of sunflower rhizosphere and bulk soils as observed by canonical correspondence analysis (CCA). **Abbreviations:** carbohydrates (CHO), clustering-based subsystems (CBS), amino acids and derivatives (AAD), protein metabolism (PM), miscellaneous (M), cofactors, vitamins, prosthetic groups, pigments (CVPGP), DNA metabolism (DNA), respiration (R), membrane transport (MT), cell wall and capsule (CWC), RNA metabolism (RNA), nucleosides and nucleotides (NN), fatty acids, lipids,

and isoprenoids (FALN), stress response (SR), virulence, disease and defense (VDD), metabolism of aromatic compounds (MAC), motility and chemotaxis (MC), nitrogen metabolism (NM), phosphorus metabolism (P), regulation and cell signaling (RCS), sulfur metabolism (S), phages, prophages, transposable elements, plasmids (PPTP), cell division and cell cycle (CDCC), iron acquisition and metabolism (IAM), potassium metabolism (K), secondary metabolism (SM), dormancy and sporulation (DS), and photosynthesis (PHO).

5.3.6 Pathways involved in nutrient cycling obtained from sunflower rhizosphere (R1, R2) and bulk soils (B1, B2)

Sequences associated with nitrogen pathways, phosphorus pathways and iron metabolic pathways which are associated with nutrient cycling were observed in the sunflower rhizosphere and bulk soil samples. The relative abundance of the nitrogen (Figure 5.5), phosphorus (Figure 5.6), and iron metabolic pathways (Figure 5.7) was higher in sample R1 compared to other samples. The rhizosphere had more relative abundance to the bulk soil samples. Furthermore, there was no significant difference between the two locations and between the rhizosphere and bulk soil samples ($p > 0.05$).

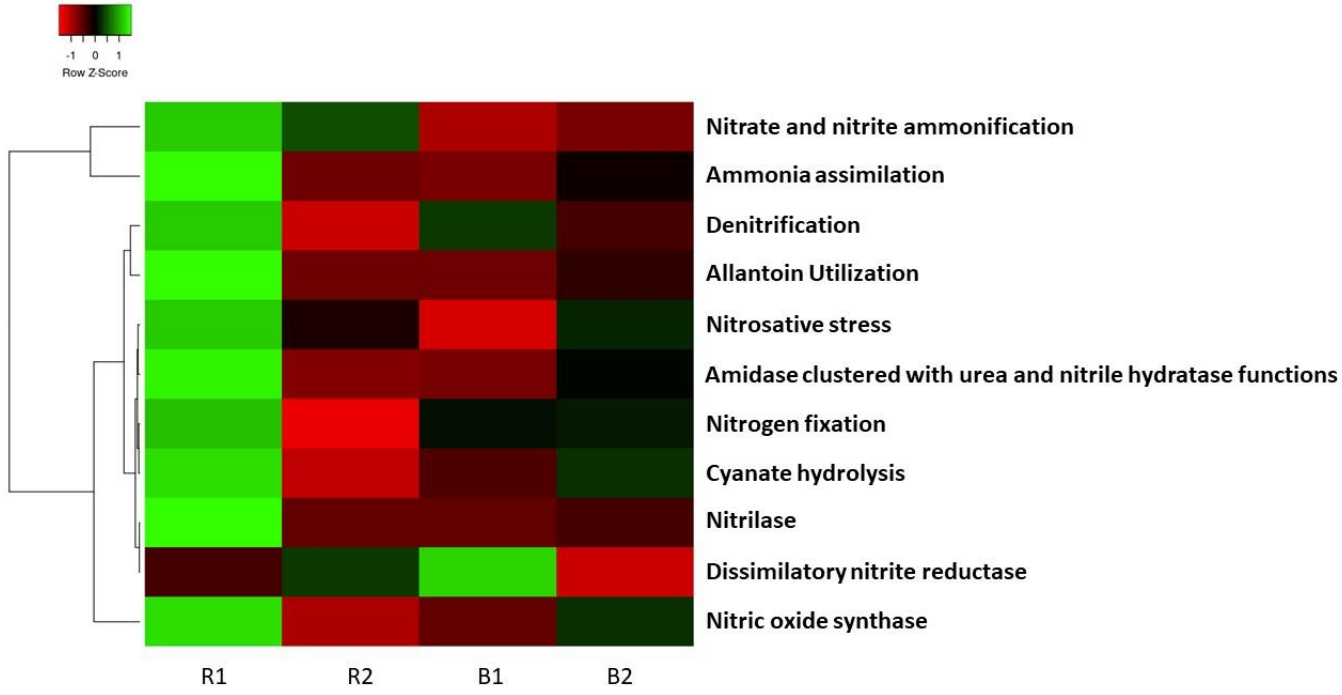


Figure 5.5: Relative abundance of nitrogen metabolism as observed in our sunflower rhizosphere (R1, R2) and bulk (B1, B2) soil samples. The scale bar represents the colour saturation gradient based on the relative abundances with a z-score transformed relative abundance of the nitrogen pathways

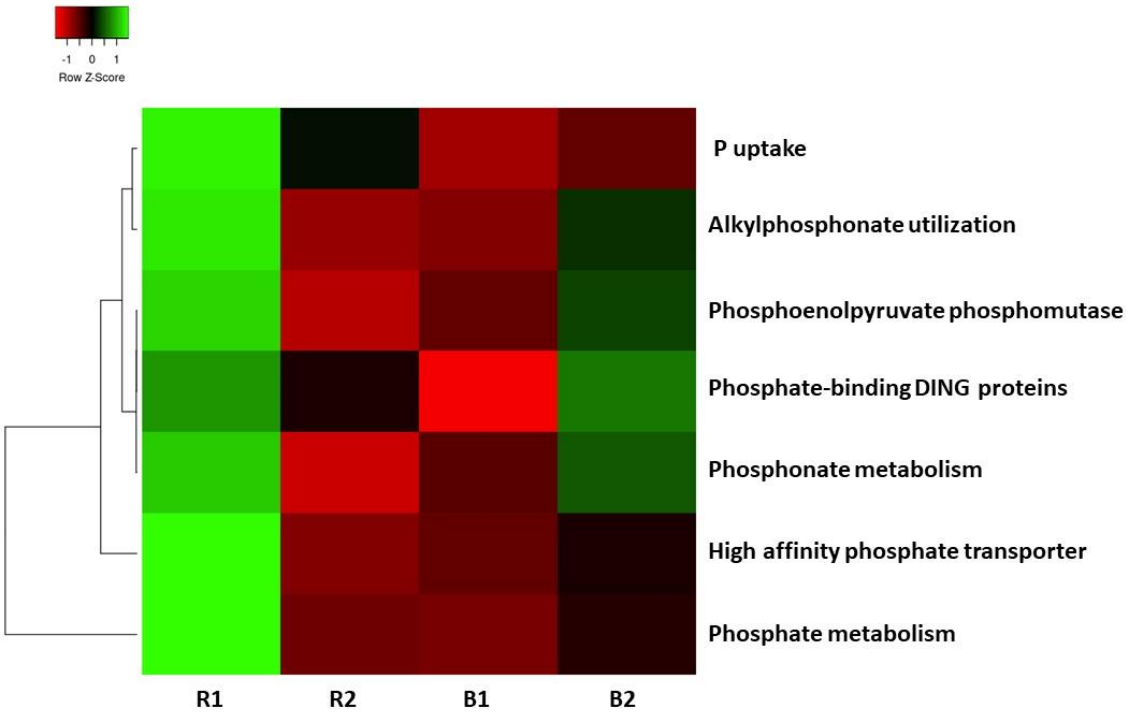


Figure 5.6: Relative abundance of phosphorus metabolism as observed in our sunflower rhizosphere (R1, R2) and bulk (B1, B2) soil samples. The scale bar represents the colour saturation gradient based on the relative abundances with a z-score transformed relative abundance of the phosphorus pathways

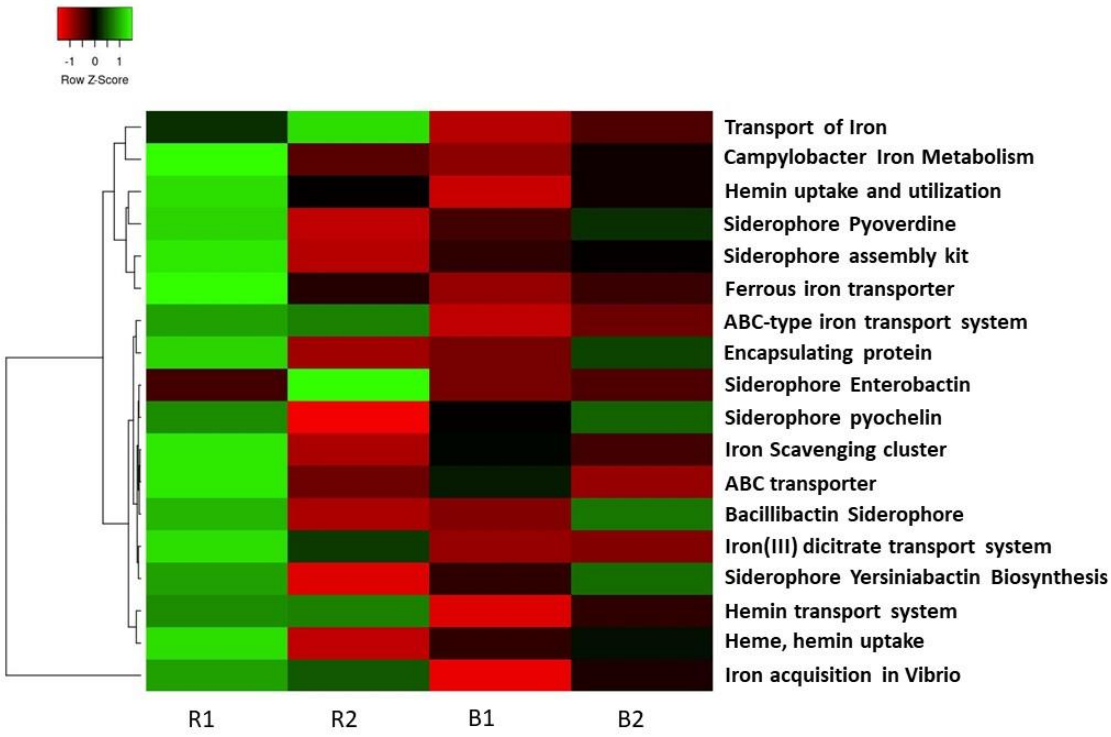


Figure 5.7: Relative abundance of iron metabolism as observed in our sunflower rhizosphere (R1, R2) and bulk (B1, B2) soil samples. The scale bar represents the colour saturation gradient based on the relative abundances with a z-score transformed relative abundance of the iron pathways

5.4 Discussion

Microorganisms found in natural ecosystems are responsible for driving various biogeochemical cycles of elements, mineralization of dead organic matter, fixation of carbon and nitrogen and most especially the protection of plants from both biotic and abiotic stresses (Salam and Obayori, 2019). Here, we used the shotgun metagenomic approach to decipher the metabolic pathways and functional attributes in a sunflower microbial community. The most dominant classes were the *Actinobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, *Deltaproteobacteria*, and *Planctomycetacia*. The theoretical limit of 2.81 as described by Dinsdale et al. (2008) showed our

α diversity of the metagenomes subsystem of sunflower rhizosphere and bulk soil was within the limit. We have some dominant metabolisms, for instance, carbohydrates and amino acids and derivatives in our environmental samples, which explains why the evenness of our metagenomes is low, around 0.62 (Table 5.1) (Enagbonma et al., 2019).

To test our hypothesis on whether each microhabitat has distinct metabolic characterization, a PCA was carried out. The analysis showed a combined variation of 52.39% between the sunflower rhizosphere and bulk soils and very high predictive metabolic potential within the ecosystem. From our study, the position that each location occupies further explains which metabolism take place there. Making use of this information, it is plausible to determine which metabolism is relevant for the microorganisms in that habitat in comparison to other habitats. For example, the vector arrows of motility and chemotaxis, photosynthesis, stress response, and iron acquisition and metabolism were more concentrated in the sunflower rhizosphere location (R1) than other locations (B2, R1, and B1). The roots release different types of exudates, mucilage and other compounds in the rhizosphere via rhizodeposition, which serves as a carbon and energy source for microorganisms (Goss-Souza et al., 2019).

Each microhabitat has a dominant functional gene category as observed in our results. The predominance of metabolism related to cell division and cell cycles, protein metabolism, amino acid and derivatives, and phosphorus metabolism were expected because they serve as source of nutrient for the growth and multiplication of bacteria (Enagbonma et al., 2019). In addition, our study also revealed functional categories such as RNA metabolism, photosynthesis, nitrogen metabolism, phages, prophages, transposable elements, plasmids, regulation and cell signaling, and cofactors, vitamins, prosthetic groups, and pigments dominate the rhizosphere soils. Nutrients such as nitrogen, sulfur, and phosphorus that help in nutrient cycling were evident in both our

sunflower rhizosphere and bulk soil samples. Mendes et al. (2014) reported bacterial groups involved in specific nutritional functions such as nitrogen, phosphorus, and iron were present in greater amounts in the rhizosphere, our study shows there was high relative abundance of metabolic processes in phosphorus, iron and nitrogen cycles. Nitrogen, phosphorous, and iron metabolisms are some of the important mechanisms microorganisms use in inhabiting and protecting the host plant (Hong et al., 2019).

Our results revealed that functional categories are apparently influenced by soil properties. Soil physicochemical parameters are drivers of soil microbial communities that are involved in many essential biogeochemical processes and play an important role in plant nutrient acquisition (Jeanbille et al., 2016). Soil phosphorus helps in the abundance of soil microorganisms and an essential element for the activities of many key enzymes (Cocozza et al., 2019; Liang et al., 2020). Bacterial functional categories have been reported to be dependent on soil physical and chemical properties (Pan et al., 2014). Our results confirm that edaphic factors drive the functional categories in both sunflower rhizosphere and bulk soils. Thus, this explains that soil and climatic parameters remain significant in the composition, diversity, and metabolic potential of microbiomes across diverse environments (Lladó et al., 2018).

Conclusion

This study describes the metabolic pathways and unveils the functional profile of the microbiome in the sunflower rhizosphere and bulk soil using the next generation sequencing approach. Our shotgun metagenomics analysis revealed metabolic processes of the likes of nitrogen, phosphorus, and iron and the detection of functional metabolism in our microhabitats. There was no significant difference exhibited in the alpha diversity of the functional categories within the rhizosphere and

bulk soil samples but that cannot be said for our beta diversity which revealed that the functional categories between our two microhabitats were significant and likewise the functional categories of the two sample locations differs significantly from one another. This study also presented proof that environmental variables are important drivers of functional categories in both habitats. Going by the results obtained, it showed that the rhizosphere is more enriched in nutrient cycling pathways than the bulk. In addition, location R1 is enriched than the location R2. Metagenomics has been a useful tool for studying the diversity and metabolic potential of microorganisms that can be use in biotechnological application for sustainable agriculture. A better understanding of the ecosystem functioning is of utmost importance in achieving a sustainable world when we consider the beneficial role of the rhizosphere microbiome in soil environments, therefore necessitating further studies of functional gene categories for plant-microbe interactions.

CHAPTER SIX

6.0 SUMMARY CONCLUSION AND RECOMMENDATION

Soil hosts a variety of living organisms comprising plants, animals, and microorganisms, which do not live in isolation but form a complex interactive network. A very important ecological niche known as the rhizosphere is found at the root of a plant and houses metabolically diverse microorganisms. Plants provide different types of nutrients through the root exudates and this attracts a lot of microorganisms at the root regions, thereby necessitating extensive study to discover their functions, structures, and diversities. These microorganisms play useful roles, such as conferring adaptive advantage to plants, protecting plants from invading pathogens, improving plant health, growth, and production. With the reduction in the cost of next generation sequencing, metagenomic studies have received considerable attention in recent times. Metagenomics gives insight into their structure and functions, thereby presenting information on microorganisms that cannot be easily cultured in the laboratory. Understanding the microbial community in which the microbes resides in association with the plant is important. Therefore, identifying and determining their functional gene information and application of biotechnological techniques is necessary for sustainable agriculture.

This study showed detailed microbial communities of the sunflower rhizosphere and bulk soils with the help of shotgun metagenomics. Bacteria dominated the metagenomes, which explains why plant growth promoting bacteria have attracted much attention when compared to other groups of organisms. Bacteria phyla includes *Actinobacteria*, *Proteobacteria*, *Acidobacteria*, *Bacteroidetes*, *Planctomycetes*, and *Verrucomicrobia.* Of interest from our study is the domain archaea with phyla *Thaumarchaeota* and *Euryarchaeota*, which are ammonia oxidizing archaea

useful in nitrogen and carbon cycling. These nutrients are largely required for plant growth and development. Therefore, archaeal microbes would be useful in metabolizing these important nutrients for plant survival, especially in the arid and semi-arid regions.

For the alpha diversity, there was variation in the microbial diversity between the two locations. However, there was no significant difference in the microbial diversity between the two locations ($p > 0.05$). In addition, there was no significant difference in the alpha diversity (that is the diversity within the habitat) of rhizosphere and bulk soil samples while the beta diversity (which is the microbial diversity between the habitats) varied significantly ($p < 0.05$). Twenty-eight (28) functional categories were observed with carbohydrates having the highest with 14.19% and 13.74% for R1 and R2 while B1 and B2 had 14 and 14.02% respectively, followed by clustering based subsystems of 12.96, 12.93, 13.03, and 12.91% for R1, R2, B1, and B2 metagenomes while amino acids and derivatives had 10.45 and 10.30% for R1 and R2 metagenomes, and 10.46 and 10.26% for B1 and B2 metagenomes. There was significant difference in the functional categories between the rhizosphere and the bulk soil samples likewise the two locations showed significant difference for the functional categories (p value < 0.05). Some of the genes observed in our study were the 1-aminocyclopropane-1-carboxylate (ACC) deaminase, nitrogen fixation genes, high temperature response genes and genes that confer biocontrol measures against pathogens through the mechanism of synthesis and excretion of antibiotics such as proteases and siderophores. The findings showed that location Palmietfontein (R1) was more enriched in gene abundance than the location Bloemhof (R2), this could be as a result of the differences in their physicochemical parameters. Therefore, R1 is a better location for planting sunflower in order to get a better yield. The microbes expressing these genes in the rhizosphere are potentially important in the growth and development of plant. These genes can then be biotechnologically harnessed for agricultural

improvement and sustainability. With the ever-increasing human population and a task to produce more agricultural crops, there is a need to tackle this quest without compromising the safety of the environment and human health. With the injudicious use of chemical fertilizers and pesticides worldwide, there is a need to provide alternatives. The rhizosphere has useful microbiota which are capable of fixing nitrogen, solubilizing phosphate, and control phytopathogens. Hence, they can potentially be used as biofertilizers and biopesticides, which would reduce our dependency on chemical products.

The use of shotgun metagenomics has provided valuable insight into the microbial communities of sunflower. It revealed functional genes such as *nifH*, *acdS*, *ppx/gppA*, *epsF*, *exoF*, and *htrA* that have the potential to serve as biofertilizer and biocontrol tools for sustainable agriculture. Additional investigation is needed to develop more media that would accommodate culturing of the yet uncultured microbes. Micro-culturing technology can also be potentially harnessed where difficult microbes can be co-cultured with culturable microbes. Alternatively, when these novel genes are identified and their metabolic pathways are explained, the microorganisms can be isolated and further investigated, or the genes encoding the compound can be isolated and then expressed in heterologous hosts for controlled production. This would facilitate the optimization of soil microbial technology to benefit the growth, development, and health of plants in their natural environments. Therefore, there is a need for further investigation to unveil the full potential of these plant beneficial functional genes and exploit their metabolic pathways by conducting experimental trials to test the efficacies and ascertain their use for biotechnological purposes. This is critical if we are to translate microbiome knowledge into applications for sustainable agriculture.

Going forward, plant can be engineered to release hormones or exudates that attract and maintain beneficial microbiomes for higher yields and more resilient crops. Future research should be

focused on customizing plant growth promoting microbes for different cropping systems whereby microbe-friendly plants are selected to attract beneficial microbes and how plants can be successfully inoculated with useful consortia or transplant beneficial microbiomes while maintaining their activity over time. Similarly, metatranscriptomics can be used to determine which of the isolate in the metagenome is expressing a particular gene and the quantity being expressed. The organisms expressing a particular gene of interest can be optimised further to boost sustainable agriculture.

The ratio of sequences assigned to unknown genes or unknown organisms is a limitation in metagenomic studies because it will not allow us to infer further information and likelihood of getting novel genes for research purposes. Therefore, an expansion of reliable genome sequence databases should be encouraged.

REFERENCES

Addrah, M. E., Zhang, Y., Zhang, J., Liu, L., Zhou, H., Chen, W. & Zhao, J. (2020). Fungicide Treatments to Control Seed-borne Fungi of Sunflower Seeds. *Pathogens,* 9**,** 29.

Adegboye, M. F. & Babalola, O. O. (2013). Phylogenetic characterization of culturable antibiotic producing *Streptomyces* from rhizospheric soils. *Mol. Biol. S,* 1.

Adegboye, M. F., Babalola, O. O., Ngoma, L. & Okoh, A. I. (2012). Analysis of *Streptomyces spp*. native to Mahikeng soils in South Africa. *Journal of Pure and Applied Microbiology,* 6**,** 1001-1010.

Adegboye, M. F., Lobb, B., Babalola, O. O., Doxey, A. C. & Ma, K. (2018). Draft Genome Sequences of Two Novel Cellulolytic Streptomyces Strains Isolated from South African Rhizosphere Soil. *Genome Announcment.,* 6**,** e00632-18.

Adhya, T. K., Lal, B., Mohapatra, B., Paul, D. & Das, S. (2018). *Advances in soil microbiology: recent trends and future prospects*, Springer. Soil Microbe-Plant Interaction. pp 1-238.

Alawiye, T. T. & Babalola, O. O. (2019). Bacterial Diversity and Community Structure in Typical Plant Rhizosphere. *Diversity,* 11**,** 179.

Alaylar, B., Güllüce, M., Karadayi, M. & Isaoglu, M. (2019). Rapid Detection of Phosphate-Solubilizing Bacteria from Agricultural Areas in Erzurum. *Current Microbiology,* 76**,** 804-809.

Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., Lawley, T. D. & Finn, R. D. (2019). A new genomic blueprint of the human gut microbiota. *Nature,* 568**,** 499-504.

Almeida, O. G. & De Martinis, E. C. (2019). Bioinformatics tools to assess metagenomic data for applied microbiology. *Applied Microbiology and Biotechnology,* 103**,** 69-82.

Altomare, C. & Tringovska, I. (2011). Beneficial soil microorganisms, an ecological alternative for soil fertility management. *Genetics, biofuels and local farming systems.* Springer, 161-214.

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E. & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology,* 21**,** 1-16.

Ambrosino, L., Tangherlini, M., Colantuono, C., Esposito, A., Sangiovanni, M., Miralto, M., Sansone, C. & Chiusano, M. L. (2019). Bioinformatics for Marine Products: An Overview of Resources, Bottlenecks, and Perspectives. *Marine Drugs,* 17**,** 576.

Armero, J., Requejo, R., Jorrín, J., López-Valbuena, R. & Tena, M. (2001). Release of phytoalexins and related isoflavonoids from intact chickpea seedlings elicited with reduced glutathione at root level. *Plant Physiology and Biochemistry,* 39**,** 785-795.

Arora, N., Kang, S. & Maheshwari, D. (2001). Isolation of siderophore-producing strains of *Rhizobium meliloti* and their biocontrol potential against *Macrophomina phaseolina* that causes charcoal rot of groundnut. *Current Science***,** 673-677.

Arseneault, T., Goyer, C. & Filion, M. (2016). Biocontrol of potato common scab is associated with high *Pseudomonas fluorescens* LBUM223 populations and phenazine-1-carboxylic acid biosynthetic transcript accumulation in the potato geocaulosphere. *Phytopathology,* 106**,** 963-970.

Asari, S. Y. (2015). *Studies on plant-microbe interaction to improve stress tolerance in plants for sustainable agriculture*. No. 76.

Åström, B., Gustafsson, A. & Gerhardson, B. (1993). Characteristics of a plant deleterious rhizosphere *Pseudomonad* and its inhibitory metabolite (s). *Journal of Applied Bacteriology,* 74**,** 20-28.

Astudillo-García, C., Hermans, S. M., Stevenson, B., Buckley, H. L. & Lear, G. (2019). Microbial assemblages and bioindicators as proxies for ecosystem health status: potential and limitations. *Applied Microbiology and Biotechnology,* 103**,** 6407-6421.

Babalola, O. & Akindolire, A. (2011). Identification of native rhizobacteria peculiar to selected food crops in Mmabatho municipality of South Africa. *Biol Agric Hort,* 27**,** 294-309.

Babalola, O. O. (2003). Molecular techniques: An overview of methods for the detection of bacteria. *African journal of biotechnology,* 2**,** 710-713.

Babalola, O. O. (2010). Beneficial bacteria of agricultural importance. *Biotechnology Letters,* 32**,** 1559-1570.

Babalola, O. O. & Glick, B. R. (2012). The use of microbial inoculants in African agriculture: current practice and future prospects. *J Food Agri Environ,* 10**,** 540-549.

Babalola, O. O., Kirby, B. M., Roes-Hill, L., Cook, A. E., Cary, S. C., Burton, S. G. & Cowan, D. A. (2009). Phylogenetic analysis of actinobacterial populations associated with Antarctic Dry Valley mineral soils. *Environmental microbiology,* 11**,** 566-576.

Babalola, O. O., Sanni, A. I., Odhiambo, G. D. & Torto, B. (2007). Plant growth-promoting rhizobacteria do not pose any deleterious effect on cowpea and detectable amounts of ethylene are produced. *World Journal of Microbiology and Biotechnology,* 23**,** 747-752.

Backer, R., Rokem, J. S., Ilangumaran, G., Lamont, J., Praslickova, D., Ricci, E., Subramanian, S. & Smith, D. L. (2018). Plant growth-promoting rhizobacteria: context, mechanisms of action, and roadmap to commercialization of biostimulants for sustainable agriculture. *Frontiers in Plant Science,* 9**,** 1473.

Bağcı, C., Beier, S., Górska, A. & Huson, D. H. (2019). Introduction to the analysis of environmental sequences: metagenomics with MEGAN. *Evolutionary Genomics.* Springer, 591-604.

Bais, H., Park, S., Weir, T., Callaway, R. & Vivanco, J. (2004). How plants communicate using the underground information superhighway. *Trends in Plant Science,* 9**,** 26-32.

Beauregard, P., Chai, Y., Vlamakis, H., Losick, R. & Kolter, R. (2013). *Bacillus subtilis* biofilm induction by plant polysaccharides. *Proceedings of the National Academy of Sciences,* 110**,** E1621-E1630.

Beneduzi, A., Ambrosini, A. & Passaglia, L. M. (2012). Plant growth-promoting rhizobacteria (PGPR): their potential as antagonists and biocontrol agents. *Genetics and Molecular Biology,* 35**,** 1044-1051.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K. & Corral, G. H. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome,* 8**,** 1-22.

Berg, G., Rybakova, D., Grube, M. & Köberl, M. (2015). The plant microbiome explored: implications for experimental botany. *Journal of Experimental Botany,* 67**,** 995-1002.

Berlanas, C., Berbegal, M., Elena, G., Laidani, M., Cibriain, J. F., Sagües, A. & Gramaje, D. (2019). The fungal and bacterial rhizosphere microbiome associated with grapevine rootstock genotypes in mature and young vineyards. *Frontiers in Microbiology,* 10**,** 1142.

Bhargava, P., Khan, M., Verma, A., Singh, A., Singh, S., Vats, S. & Goel, R. (2019). Metagenomics as a Tool to Explore New Insights from Plant-Microbe Interface. *Plant Microbe Interface.* Springer, 271-289.

Bharti, R. & Grimm, D. G. (2019). Current challenges and best-practice protocols for microbiome analysis. *Briefings in Bioinformatics*. 1-16.

Bharwad, K. & Rajkumar, S. (2020). Modulation of PQQ-dependent glucose dehydrogenase (mGDH and sGDH) activity by succinate in phosphate solubilizing plant growth promoting *Acinetobacter sp*. SK2. *3 Biotech,* 10**,** 5.

Bloemberg, G. & Lugtenberg, B. (2001). Molecular basis of plant growth promotion and biocontrol by Rhizobacteria. *Current Opinion in Plant Biology,* 4**,** 343-350.

Boers, S. A., Jansen, R. & Hays, J. P. (2019). Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical

microbiological diagnostic laboratory. *European Journal of Clinical Microbiology & Infectious Diseases,* 38**,** 1059-1070.

Boruah, H. & Kumar, B. (2002). Plant disease suppression and growth promotion by a *Fluorescent Pseudomonas* strain. *Folia Microbiologica*, 47**,** 137-143.

Boulund, F., Pereira, M., Jonsson, V. & Kristiansson, E. (2018). Computational and statistical considerations in the analysis of metagenomic data. *Metagenomics.* Elsevier, 81-102.

Breitwieser, F. P., Lu, J. & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics,* 20**,** 1125-1136.

Brimecombe, M. J., Leij, F. A. & Lynch, J. M. (2001). Nematode community structure as a sensitive indicator of microbial perturbations induced by a genetically modified Pseudomonas fluorescens strain. *Biology and Fertility of Soils,* 34**,** 270-275.

Brumfield, K. D., Huq, A., Colwell, R. R., Olds, J. L. & Leddy, M. B. (2020). Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *Plos One,* 15**,** e0228899.

Bruto, M., Prigent-Combaret, C., Muller, D. & Moënne-Loccoz, Y. (2014). Analysis of genes contributing to plant-beneficial functions in plant growth-promoting rhizobacteria and related *Proteobacteria*. *Scientific Reports,* 4**,** 6261.

Bulgarelli, D., Schlaeppi, K., Spaepen, S., Van Themaat, E. V. L. & Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annual Review of Plant Biology,* 64**,** 807-838.

Bumunang, E. W., Babalola, O. O. & Barros, E. (2013). Bacterial community profiling in the rhizosphere of field grown GM and non-GM maize. *Journal of Pure and Applied Microbiology***,** 1-21.

Busby, P. E., Soman, C., Wagner, M. R., Friesen, M. L., Kremer, J., Bennett, A., Morsy, M., Eisen, J. A., Leach, J. E. & Dangl, J. L. (2017). Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biology,* 15**,** e2001793.

Buytaers, F. E., Saltykova, A., Denayer, S., Verhaegen, B., Vanneste, K., Roosens, N. H., Piérard, D., Marchal, K. & De Keersmaecker, S. C. (2020). A Practical Method to Implement Strain-Level Metagenomics-Based Foodborne Outbreak Investigation and Source Tracking in Routine. *Microorganisms,* 8**,** 1191.

Caballero-Mellado, J., Onofre-Lemus, J., Estrada-De Los Santos, P. & Martínez-Aguilar, L. (2007). The tomato rhizosphere, an environment rich in nitrogen-fixing *Burkholderia species* with capabilities of interest for agriculture and bioremediation. *Applied Environental Microbiology,* 73**,** 5308-5319.

Cantalupo, P. G. & Pipas, J. M. (2019). Detecting viral sequences in NGS data. *Current Opinion in Virology,* 39**,** 41-48.

Carvalho, D., Oliveira, D., Corrêa, R., Campos, V., Guimarães, R. & Coimbra, J. (2007). Rhizobacteria able to produce phytotoxic metabolites. *Brazilian Journal of Microbiology,* 38**,** 759-765.

Chaudhari, D., Rangappa, K., Das, A., Layek, J., Basavaraj, S., Kandpal, B. K., Shouche, Y. & Rahi, P. (2020). Pea (*Pisum sativum* l.) Plant Shapes Its Rhizosphere Microbiome for

Nutrient Uptake and Stress Amelioration in Acidic Soils of the North-East Region of India. *Frontiers in Microbiology,* 11**,** 968.

Chen, X., Koumoutsi, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., Morgenstern, B., Voss, B., Hess, W. & Reva, O. (2007). Comparative analysis of the complete genome sequence of the plant growth–promoting bacterium *Bacillus amyloliquefaciens* FZB42. *Nature Biotechnology,* 25**,** 1007.

Chen, Z., Zheng, Y., Ding, C., Ren, X., Yuan, J., Sun, F. & Li, Y. (2017). Integrated metagenomics and molecular ecological network analysis of bacterial community composition during the phytoremediation of cadmium-contaminated soils by bioenergy crops. *Ecotoxicology and Environmental Safety,* 145**,** 111-118.

Chin-A-Woeng, T., van den Broek, D., Lugtenberg, B. & Bloemberg, G. (2005). The *Pseudomonas chlororaphis* PCL1391 sigma regulator psrA represses the production of the antifungal metabolite phenazine-1-carboxamide. *Molecular Plant-microbe Interactions,* 18**,** 244-253.

Choi, H., Choi, Y., Verberne, M., Lefeber, A., Erkelens, C. & Verpoorte, R. (2004). Metabolic fingerprinting of wild type and transgenic tobacco plants by 1H NMR and multivariate analysis technique. *Phytochemistry,* 65**,** 857-864.

Cirlini, M., Dall'Asta, C., Silvanini, A., Beghè, D., Fabbri, A., Galaverna, G. & Ganino, T. (2012). Volatile fingerprinting of chestnut flours from traditional Emilia Romagna (Italy) cultivars. *Food Chemistry,* 134**,** 662-668.

Cobo-Díaz, J. F., Fernández-González, A. J., Villadas, P. J., Robles, A. B., Toro, N. & Fernández-López, M. (2015). Metagenomic assessment of the potential microbial nitrogen pathways in the rhizosphere of a Mediterranean forest after a wildfire. *Microbial Ecology,* 69**,** 895-904.

Cocozza, C., Brilli, F., Miozzi, L., Pignattelli, S., Rotunno, S., Brunetti, C., Giordano, C., Pollastri, S., Centritto, M. & Accotto, G. P. (2019). Impact of high or low levels of phosphorus and high sodium in soils on productivity and stress tolerance of *Arundo donax* plants. *Plant Science,* 289**,** 110260.

Compant, S., Cambon, M. C., Vacher, C., Mitter, B., Samad, A. & Sessitsch, A. (2020). The plant endosphere world–bacterial life within plants. *Environmental Microbiology*. 1-18.

Compant, S., Samad, A., Faist, H. & Sessitsch, A. (2019). A review on the plant microbiome: ecology, functions and emerging trends in microbial application. *Journal of Advanced Research,* 19**,** 29-37.

Consortium, U. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research,* 47**,** D506-D515.

Cowan, D., Meyer, Q., Stafford, W., Muyanga, S., Cameron, R. & Wittwer, P. (2005). Metagenomic gene discovery: past, present and future. *TRENDS in Biotechnology,* 23**,** 321-329.

Cox, M., Peterson, D. & Biggs, P. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics,* 11**,** 485.

Cui, Y., Bing, H., Fang, L., Wu, Y., Yu, J., Shen, G., Jiang, M., Wang, X. & Zhang, X. (2019). Diversity patterns of the rhizosphere and bulk soil microbial communities along an altitudinal gradient in an alpine ecosystem of the eastern Tibetan Plateau. *Geoderma,* 338**,** 118-127.

de Faria, M. R., Costa, L. S. A. S., Chiaramonte, J. B., Bettiol, W. & Mendes, R. (2020). The rhizosphere microbiome: functions, dynamics, and role in plant protection. *Tropical Plant Pathology***,** 1-13.

Debaeke, P., Casadebaig, P., Flenet, F. & Langlade, N. (2017). Sunflower crop and climate change: vulnerability, adaptation, and mitigation potential from case-studies in Europe. *OCL,* 24**,** D102.

Deke, A., Adugna, W. & Fite, A. (2016). Soil physic-chemical properties in termite mounds and adjacent control soil in Miyo and Yabello Districts of Borana Zone, Southern Ethiopia. *American Journal of Agriculture and Forestry,* 4**,** 69.

del Carmen Orozco-Mosqueda, M., Glick, B. R. & Santoyo, G. (2020). ACC deaminase in plant growth-promoting bacteria (PGPB): An efficient mechanism to counter salt stress in crops. *Microbiological Research,* 235**,** 126439.

Dimkpa, C., Merten, D., Svatoš, A., Büchel, G. & Kothe, E. (2009). Siderophores mediate reduced and increased uptake of cadmium by *Streptomyces tendae* F4 and sunflower (*Helianthus annuus*), respectively. *Journal of Applied Microbiology,* 107**,** 1687-1696.

Dimkpa, C., Svatoš, A., Dabrowska, P., Schmidt, A., Boland, W. & Kothe, E. (2008). Involvement of siderophores in the reduction of metal-induced inhibition of auxin synthesis in *Streptomyces spp*. *Chemosphere,* 74**,** 19-25.

Dinsdale, E. A., Edwards, R. A., Hall, D., Angly, F., Breitbart, M., Brulc, J. M., Furlan, M., Desnues, C., Haynes, M. & Li, L. (2008). Functional metagenomic profiling of nine biomes. *Nature,* 452**,** 629-632.

Dong, C.-J., Wang, L.-L., Li, Q. & Shang, Q.-M. (2019). Bacterial communities in the rhizosphere, phyllosphere and endosphere of tomato plants. *PloS One,* 14**,** e0223847.

Dubey, A., Malla, M. A., Khan, F., Chowdhary, K., Yadav, S., Kumar, A., Sharma, S., Khare, P. K. & Khan, M. L. (2019). Soil microbiome: a key player for conservation of soil health under changing climate. *Biodiversity and Conservation,* 28**,** 2405-2429.

Dubey, S. & Sharma, S. (2019). Rhizospheric Microbiome Engineering as a Sustainable Tool in Agriculture: Approaches and Challenges. *Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications.* Springer, 257-272.

Durrer, A., Gumiere, T., Taketani, R., da Costa, D., e Silva, M. & Andreote, F. (2017). The drivers underlying biogeographical patterns of bacterial communities in soils under sugarcane cultivation. *Applied Soil Ecology,* 110**,** 12-20.

Efange, S., Brun, R., Wittlin, S., Connolly, J., Hoye, T., McAkam, T., Makolo, F., Mbah, J., Nelson, D. & Nyongbela, K. (2009). Okundoperoxide, a bicyclic cyclofarnesylsesquiterpene endoperoxide from Scleria striatinux with antiplasmodial activity. *Journal of Natural Products,* 72**,** 280-283.

Elliott, A. J., Daniell, T. J., Cameron, D. D. & Field, K. J. (2020). A commercial arbuscular mycorrhizal inoculum increases root colonization across wheat cultivars but does not increase assimilation of mycorrhiza-acquired nutrients. *Plants, People, Planet*. 1-12.

Enagbonma, B., Aremu, B. & Babalola, O. (2019). Profiling the functional diversity of termite mound soil bacteria as revealed by shotgun sequencing. *Genes,* 10**,** 637.

Enagbonma, B. J. & Babalola, O. O. (2020). Unveiling Plant-Beneficial Function as Seen in Bacteria Genes from Termite Mound Soil. *Journal of Soil Science and Plant Nutrition***,** 1-10.

Enebe, M. C. & Babalola, O. O. (2018). The influence of plant growth-promoting rhizobacteria in plant tolerance to abiotic stress: a survival strategy. *Applied Microbiology and Biotechnology,* 102**,** 7821-7835.

Fadiji, A. E., Ayangbenro, A. S. & Babalola, O. O. (2020). Metagenomic profiling of the community structure, diversity, and nutrient pathways of bacterial endophytes in maize plant. *Antonie van Leeuwenhoek,* 113**,** 1559-1571.

Fadiji, A. E. & Babalola, O. O. (2020). Metagenomics methods for the study of plant-associated microbial communities: A review. *Journal of Microbiological Methods***,** 105860.

Faggioli, V. S., Cabello, M. N., Grilli, G., Vasar, M., Covacevich, F. & Öpik, M. (2019). Root colonizing and soil borne communities of arbuscular mycorrhizal fungi differ among soybean fields with contrasting historical land use. *Agriculture, Ecosystems & Environment,* 269**,** 174-182.

Fanin, N. & Bertrand, I. (2016). Aboveground litter quality is a better predictor than belowground microbial communities when estimating carbon mineralization along a land-use gradient. *Soil Biology and Biochemistry,* 94**,** 48-60.

FAO (2017). Food and Agriculture Organisation of the United Nations: Food and agriculture data. FAOSTAT, Rome, Italy.

Farhat, M. B., Farhat, A., Bejar, W., Kammoun, R., Bouchaala, K., Fourati, A., Antoun, H., Bejar, S. & Chouayekh, H. (2009). Characterization of the mineral phosphate solubilizing activity of *Serratia marcescens* CTM 50650 isolated from the phosphate mine of Gafsa. *Archives of Microbiology,* 191**,** 815-824.

Ferreira, C. M., Soares, H. M. & Soares, E. V. (2019). Promising bacterial genera for agricultural practices: An insight on plant growth-promoting properties and microbial safety aspects. *Science of the Total Environment,* 682**,** 779-799.

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology,* 15**,** 579.

Fonseca, J. P., Hoffmann, L., Cabral, B. C. A., Dias, V. H. G., Miranda, M. R., de Azevedo Martins, A. C., Boschiero, C., Bastos, W. R. & Silva, R. (2018). Contrasting the microbiomes from forest rhizosphere and deeper bulk soil from an Amazon rainforest reserve. *Gene,* 642**,** 389-397.

Franke, A., Van den Brand, G., Vanlauwe, B. & Giller, K. (2018). Sustainable intensification through rotations with grain legumes in Sub-Saharan Africa: A review. *Agriculture, Ecosystems & Environment,* 261**,** 172-185.

Friedman, M. (2006). Potato glycoalkaloids and metabolites: roles in the plant and in the diet. *Journal of Agricultural and Food Chemistry,* 54**,** 8655-8681.

Gao, Z., Karlsson, I., Geisen, S., Kowalchuk, G. & Jousset, A. (2019). Protists: Puppet masters of the rhizosphere microbiome. *Trends in Plant Science,* 24**,** 165-176.

Garcia-Mazcorro, J. F., Castillo-Carranza, S. A., Guard, B., Gomez-Vazquez, J. P., Dowd, S. E. & Brigthsmith, D. J. (2017). Comprehensive molecular characterization of bacterial communities in feces of pet birds using 16S marker sequencing. *Microbial Ecology,* 73**,** 224-235.

Garcia-Perez, I., Vallejo, M., Garcia, A., Legido-Quigley, C. & Barbas, C. (2008). Metabolic fingerprinting with capillary electrophoresis. *Journal of Chromatography A,* 1204**,** 130-139.

García-Salamanca, A., Molina-Henares, M., van Dillewijn, P., Solano, J., Pizarro-Tobías, P., Roca, A., Duque, E. & Ramos, J. (2013). Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil. *Microbial Biotechnology,* 6**,** 36-44.

Gealy, D., Gurusiddaiah, S. & Ogg, A. (1996a). Isolation and characterization of metabolites from *Pseudomonas syringae*-strain 3366 and their phytotoxicity against certain weed and crop species. *Weed Science,* 44**,** 383-392.

Gealy, D., Gurusiddaiah, S., Ogg, A. & Kennedy, A. (1996b). Metabolites from *Pseudomonas fluorescens* strain D7 inhibit downy brome (*Bromus tectorum*) seedling growth. *Weed Technology,* 10**,** 282-287.

Gerten, D., Heck, V., Jägermeyr, J., Bodirsky, B. L., Fetzer, I., Jalava, M., Kummu, M., Lucht, W., Rockström, J. & Schaphoff, S. (2020). Feeding ten billion people is possible within four terrestrial planetary boundaries. *Nature Sustainability,* 3**,** 200-208.

Ghobadi, M., Taherabadi, S., Ghobadi, M.-E., Mohammadi, G.-R. & Jalali-Honarmand, S. (2013). Antioxidant capacity, photosynthetic characteristics and water relations of sunflower (*Helianthus annuus* L.) cultivars in response to drought stress. *Industrial Crops and Products,* 50**,** 29-38.

Glick, B. R. (2012). Plant growth-promoting bacteria: mechanisms and applications. *Scientifica,* 2012.

Glöckner, F. O. (2019). The SILVA Database Project: An ELIXIR core data resource for high-quality ribosomal RNA sequences. *Biodiversity Information Science and Standards,* 3**,** e36125.

Gołębiewski, M. & Tretyn, A. (2020). Generating amplicon reads for microbial community assessment with next-generation sequencing. *Journal of Applied Microbiology,* 128**,** 330-354.

Gomez-Alvarez, V., Teal, T. & Schmidt, T. (2009). Systematic artifacts in metagenomes from complex microbial communities. *The ISME Journal,* 3**,** 1314.

Goss-Souza, D., Mendes, L. W., Borges, C. D., Rodrigues, J. L. & Tsai, S. M. (2019). Amazon forest-to-agriculture conversion alters rhizosphere microbiome composition while functions are kept. *FEMS Microbiology Ecology,* 95**,** fiz009.

Goswami, D., Thakker, J. & Dhandhukia, P. (2016). Portraying mechanics of plant growth promoting rhizobacteria (PGPR): A review. *Cogent Food & Agriculture,* 2**,** 1127500.

Grayston, S., Griffith, G., Mawdsley, J., Campbell, C. & Bardgett, R. D. (2001). Accounting for variability in soil microbial communities of temperate upland grassland ecosystems. *Soil Biology and Biochemistry,* 33**,** 533-551.

Guo, J., McCulley, R., Phillips, T. & McNear Jr, D. (2016). Fungal endophyte and tall fescue cultivar interact to differentially affect bulk and rhizosphere soil processes governing C and N cycling. *Soil Biology and Biochemistry,* 101**,** 165-174.

Hallam, S. J. & McCutcheon, J. P. (2015). Microbes don't play solitaire: how cooperation trumps isolation in the microbial world. *Environmental Microbiology Reports,* 7**,** 26-28.

Hammer, Ø., Harper, D. A. & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica,* 4**,** 9.

Handelsman, J., Rondon, M., Brady, S., Clardy, J. & Goodman, R. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology,* 5**,** R245-R249.

Hanussek, M., Bartusch, F. & Krüger, J. (2020). BOOTABLE: Bioinformatics benchmark tool suite for applications and hardware. *Future Generation Computer Systems,* 102**,** 1016-1026.

Hara, S., Morikawa, T., Wasai, S., Kasahara, Y., Koshiba, T., Yamazaki, K., Fujiwara, T., Tokunaga, T. & Minamisawa, K. (2019). Identification of nitrogen-fixing Bradyrhizobium

associated with roots of field-grown sorghum by metagenome and proteome analyses. *Frontiers in Microbiology,* 10**,** 407.

Higo, M., Tatewaki, Y., Iida, K., Yokota, K. & Isobe, K. (2020). Amplicon sequencing analysis of arbuscular mycorrhizal fungal communities colonizing maize roots in different cover cropping and tillage systems. *Scientific Reports,* 10**,** 1-13.

Höfte, M. & Bakker, P. (2007). Competition for iron and induced systemic resistance by siderophores of plant growth promoting rhizobacteria. *Microbial Siderophores.* Springer, 121-133.

Hol, W., De Boer, W., de Hollander, M., Kuramae, E., Meisner, A. & van der Putten, W. (2015). Context dependency and saturating effects of loss of rare soil microbes on plant productivity. *Frontiers in Plant Science,* 6**,** 485.

Hong, C. E., Kim, J. U., Lee, J. W., Bang, K. H. & Jo, I. H. (2019). Metagenomic analysis of bacterial endophyte community structure and functions in Panax ginseng at different ages. *3 Biotech,* 9**,** 300.

Hou, M. P. & Babalola, O. O. (2013). Evaluation of plant growth promoting potential of four rhizobacterial species for indigenous system. *Journal of Central South University,* 20**,** 164-171.

Hussain, A., Ahmad, M., Mumtaz, M. Z., Ali, S., Sarfraz, R., Naveed, M., Jamil, M. & Damalas, C. A. (2020). Integrated application of organic amendments with *Alcaligenes sp*. AZ9 improves nutrient uptake and yield of maize (*Zea mays*). *Journal of Plant Growth Regulation***,** 1-16.

Hussain, M., Farooq, S., Hasan, W., Ul-Allah, S., Tanveer, M., Farooq, M. & Nawaz, A. (2018). Drought stress in sunflower: Physiological effects and its management through breeding and agronomic alternatives. *Agricultural Water Management,* 201**,** 152-166.

Igiehon, N. O. & Babalola, O. O. (2017). Biofertilizers and sustainable agriculture: exploring arbuscular mycorrhizal fungi. *Applied Microbiology and Biotechnology,* 101**,** 4871-4881.

Igiehon, N. O. & Babalola, O. O. (2018). Rhizosphere microbiome modulators: contributions of nitrogen fixing bacteria towards sustainable agriculture. *International Journal of Environmental Research and Public Health,* 15**,** 574.

Igiehon, N. O., Babalola, O. O. & Aremu, B. R. (2019). Genomic insights into plant growth promoting rhizobia capable of enhancing soybean germination under drought stress. *BMC Microbiology,* 19**,** 159.

Iqbal, J., Irshad, J., Bashir, S., Khan, S., Yousaf, M. & Shah, A. (2020). Comparative study of water extracts of Moringa leaves and roots to improve the growth and yield of sunflower. *South African Journal of Botany,* 129**,** 221-224.

Jagadeesan, B., Gerner-Smidt, P., Allard, M. W., Leuillet, S., Winkler, A., Xiao, Y., Chaffron, S., Van Der Vossen, J., Tang, S. & Katase, M. (2019). The use of next generation sequencing for improving food safety: translation into practice. *Food Microbiology,* 79**,** 96-115.

Jain, A., Chakraborty, J. & Das, S. (2020). Underlying mechanism of plant–microbe crosstalk in shaping microbial ecology of the rhizosphere. *Acta Physiologiae Plantarum,* 42**,** 1-13.

Jeanbille, M., Buée, M., Bach, C., Cébron, A., Frey-Klett, P., Turpault, M. P. & Uroz, S. (2016). Soil parameters drive the structure, diversity and metabolic potentials of the bacterial communities across temperate beech forest soil sequences. *Microbial Ecology,* 71**,** 482-493.

Jiang, J., Song, Z., Yang, X., Mao, Z., Nie, X., Guo, H. & Peng, X. (2017). Microbial community analysis of apple rhizosphere around Bohai Gulf. *Scientific reports,* 7**,** 1-9.

Jocic, S., Miladinovic, D. & Kaya, Y. (2015). Breeding and genetics of sunflower. *Sunflower*. Elsevier, 1-25.

Jorge, T., Rodrigues, J., Caldana, C., Schmidt, R., van Dongen, J., Thomas-Oates, J. & António, C. (2016). Mass spectrometry-based plant metabolomics: Metabolite responses to abiotic stress. *Mass Spectrometry Reviews,* 35**,** 620-649.

Jousset, A., Lara, E., Wall, L. & Valverde, C. (2006). Secondary metabolites help biocontrol strain *Pseudomonas fluorescens* CHA0 to escape protozoan grazing. *Applied Environmental Microbiology,* 72**,** 7083-7090.

Kai, M., Effmert, U. & Piechulla, B. (2016). Bacterial-plant-interactions: approaches to unravel the biological function of bacterial volatiles in the rhizosphere. *Frontiers in Microbiology,* 7**,** 108.

Kalbe, C., Marten, P. & Berg, G. (1996). Strains of the genus *Serratia* as beneficial rhizobacteria of oilseed rape with antifungal properties. *Microbiological Research,* 151**,** 433-439.

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H. & Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ,* 7**,** e7359.

Kasim, W. A., Gaafar, R. M., Abou-Ali, R. M., Omar, M. N. & Hewait, H. M. (2016). Effect of biofilm forming plant growth promoting rhizobacteria on salinity tolerance in barley. *Annals of Agricultural Sciences,* 61**,** 217-227.

Kenneth, O. C., Nwadibe, E. C., Kalu, A. U. & Unah, U. V. (2019). Plant growth promoting rhizobacteria (PGPR): A novel agent for sustainable food production. *Am J Agric Biol Sci,* 14**,** 35-54.

Kent, W. (2002). BLAT—the BLAST-like alignment tool. *Genome Research,* 12**,** 656-664.

Kettler, T., Doran, J. & Gilbert, T. (2001). Simplified method for soil particle-size determination to accompany soil-quality analyses. *Soil Science Society of America Journal,* 65**,** 849-852.

Khan, A., Singh, P. & Srivastava, A. (2018). Synthesis, nature and utility of universal iron chelator–Siderophore: A review. *Microbiological Research,* 212**,** 103-111.

Khomtchouk, B. B., Hennessy, J. R. & Wahlestedt, C. (2017). Shinyheatmap: Ultra fast low memory heatmap web interface for big data genomics. *PloS One,* 12.

Kouchaki, S., Tapinos, A. & Robertson, D. L. (2019). A signal processing method for alignment-free metagenomic binning: Multi-resolution genomic binary patterns. *Scientific Reports,* 9**,** 1-10.

Koumoutsi, A., Chen, X., Henne, A., Liesegang, H., Hitzeroth, G., Franke, P., Vater, J. & Borriss, R. (2004). Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *Journal of Bacteriology,* 186**,** 1084-1096.

Koutsandreas, T., Ladoukakis, E., Pilalis, E., Zarafeta, D., Kolisis, F. N., Skretas, G. & Chatziioannou, A. A. (2019). ANASTASIA: an automated metagenomic analysis pipeline for novel enzyme discovery exploiting next generation sequencing data. *Frontiers in Genetics,* 10**,** 469.

Kozukue, N., Han, J., Lee, K. & Friedman, M. (2004). Dehydrotomatine and α-tomatine content in tomato fruits and vegetative plant tissues. *Journal of Agricultural and Food Chemistry,* 52**,** 2079-2083.

Kramer, J., Özkaya, Ö. & Kümmerli, R. (2019). Bacterial siderophores in community and host interactions. *Nature Reviews Microbiology***,** 1-12.

Kumar, A., Patel, J., Meena, V. S. & Ramteke, P. (2019a). Plant growth-promoting rhizobacteria: strategies to improve abiotic stresses under sustainable agriculture. *Journal of Plant Nutrition,* 42**,** 1402-1415.

Kumar, K. R., Cowley, M. J. & Davis, R. L. Next-generation sequencing and emerging technologies. Seminars in thrombosis and hemostasis, 2019b. Thieme Medical Publishers, 661-673.

Kumar, P., Thakur, S., Dhingra, G., Singh, A., Pal, M. K., Harshvardhan, K., Dubey, R. & Maheshwari, D. (2018). Inoculation of siderophore producing rhizobacteria and their

consortium for growth enhancement of wheat plant. *Biocatalysis and Agricultural Biotechnology,* 15**,** 264-269.

Kumar, R., Ayyadurai, N., Pandiaraja, P., Reddy, A., Venkateswarlu, Y., Prakash, O. & Sakthivel, N. (2005). Characterization of antifungal metabolite produced by a new strain *Pseudomonas aeruginosa* PUPa3 that exhibits broad-spectrum antifungal activity and biofertilizing traits. *Journal of Applied Microbiology,* 98**,** 145-154.

Kumari, B., Mallick, M., Solanki, M. K., Solanki, A. C., Hora, A. & Guo, W. (2019). Plant growth promoting rhizobacteria (PGPR): modern prospects for sustainable agriculture. *Plant health under biotic stress.* Springer, 109-127.

Lahav, I. & Steinberger, Y. (2001). Soil bacterial functional diversity in a potato field. *European journal of soil biology,* 37**,** 59-67.

Lange, M., Eisenhauer, N., Sierra, C., Bessler, H., Engels, C., Griffiths, R., Mellado-Vázquez, P., Malik, A., Roy, J. & Scheu, S. (2015). Plant diversity increases soil microbial activity and soil carbon storage. *Nature Communications,* 6**,** 6707.

Lee, S. A., Kim, Y., Kim, J. M., Chu, B., Joa, J.-H., Sang, M. K., Song, J. & Weon, H.-Y. (2019). A preliminary examination of bacterial, archaeal, and fungal communities inhabiting different rhizocompartments of tomato plants under real-world environments. *Scientific reports,* 9**,** 1-15.

Lejon, D. P., Sebastia, J., Lamy, I., Chaussod, R. & Ranjard, L. (2007). Relationships between soil organic status and microbial community density and genetic structure in two agricultural soils submitted to various types of organic management. *Microbial ecology,* 53**,** 650-663.

Li, F., Neves, A. L. & Ghoshal, B. (2018). Symposium review: mining metagenomic and metatranscriptomic data for clues about microbial metabolic functions in ruminants. *Journal of Dairy Science,* 101**,** 5605-5618.

Li, H., Su, J., Yang, X. & Zhu, Y. (2019a). Distinct rhizosphere effect on active and total bacterial communities in paddy soils. *Science of the Total Environment,* 649**,** 422-430.

Li, Y., Pan, F. & Yao, H. (2019b). Response of symbiotic and asymbiotic nitrogen-fixing microorganisms to nitrogen fertilizer application. *Journal of Soils and Sediments,* 19**,** 1948-1958.

Liang, J.-L., Liu, J., Jia, P., Yang, T.-t., Zeng, Q.-w., Zhang, S.-c., Liao, B., Shu, W.-s. & Li, J.-t. (2020). Novel phosphate-solubilizing bacteria enhance soil phosphorus cycling following ecological restoration of land degraded by mining. *The ISME Journal***,** 1-14.

List, G. (2014). Sunflower seed and oil. *Lipid Technology,* 26**,** 24-24.

Liu, G., Kong, Y., Fan, Y., Geng, C., Peng, D. & Sun, M. (2017). Whole-genome sequencing of *Bacillus velezensis* LS69, a strain with a broad inhibitory spectrum against pathogenic bacteria. *Journal of Biotechnology,* 249**,** 20-24.

Lladó, S., López-Mondéjar, R. & Baldrian, P. (2018). Drivers of microbial community structure in forest soils. *Applied Microbiology and Biotechnology,* 102**,** 4331-4338.

Lobo, C. B., Tomás, M. S. J., Viruel, E., Ferrero, M. A. & Lucca, M. E. (2019). Development of low-cost formulations of plant growth-promoting bacteria to be used as inoculants in beneficial agricultural technologies. *Microbiological Research,* 219**,** 12-25.

Loit, K., Adamson, K., Bahram, M., Puusepp, R., Anslan, S., Kiiker, R., Drenkhan, R. & Tedersoo, L. (2019). Relative Performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) Third-Generation Sequencing Instruments in Identification of Agricultural and Forest Fungal Pathogens. *Applied and Environmental Microbiology,* 85**,** e01368-19.

Loper, J. & Gross, H. (2007). Genomic analysis of antifungal metabolite production by *Pseudomonas fluorescens* Pf-5. *New Perspectives and Approaches in Plant Growth-Promoting Rhizobacteria Research.* Springer, 265-278.

Ma, K. & Kim, S. (2018). Development of techniques for isolating microorganisms. *Enliven: Microbiology Microbial Technology,* 5**,** 1-5.

Ma, M., Zhou, J., Ongena, M., Liu, W., Wei, D., Zhao, B., Guan, D., Jiang, X. & Li, J. (2018). Effect of long-term fertilization strategies on bacterial community composition in a 35-year field experiment of Chinese Mollisols. *AMB Express,* 8**,** 20.

Mahoney, A. K., Yin, C. & Hulbert, S. H. (2017). Community structure, species variation, and potential functions of rhizosphere-associated bacteria of different winter wheat (*Triticum aestivum*) cultivars. *Frontiers in Plant Science,* 8**,** 132.

Malla, M. A., Dubey, A., Kumar, A., Yadav, S., Hashem, A. & Abd_Allah, E. F. (2019). Exploring the human microbiome: The potential future role of next-generation sequencing in disease diagnosis and treatment. *Frontiers in Immunology,* 9**,** 2868.

Mallick, H., Franzosa, E. A., Mclver, L. J., Banerjee, S., Sirota-Madi, A., Kostic, A. D., Clish, C. B., Vlamakis, H., Xavier, R. J. & Huttenhower, C. (2019). Predictive metabolomic

profiling of microbial communities using amplicon or metagenomic sequences. *Nature Communications,* 10**,** 1-11.

Manivanh, L., Pierret, A., Rattanavong, S., Kounnavongsa, O., Buisson, Y., Elliott, I., Maeght, J., Xayyathip, K., Silisouk, J. & Vongsouvath, M. (2017). *Burkholderia pseudomallei* in a lowland rice paddy: seasonal changes and influence of soil depth and physico-chemical properties. *Scientific Reports,* 7**,** 3031.

Mattarozzi, M., Di Zinno, J., Montanini, B., Manfredi, M., Marengo, E., Fornasier, F., Ferrarini, A., Careri, M. & Visioli, G. (2020). Biostimulants applied to maize seeds modulate the enzymatic activity and metaproteome of the rhizosphere. *Applied Soil Ecology,* 148**,** 103480.

McIntyre, A. B., Ounit, R., Afshinnekoo, E., Prill, R. J., Hénaff, E., Alexander, N., Minot, S. S., Danko, D., Foox, J. & Ahsanuddin, S. (2017). Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. *Genome Biology,* 18**,** 182.

Meena, R. S., Kumar, S., Datta, R., Lal, R., Vijayakumar, V., Brtnicky, M., Sharma, M. P., Yadav, G. S., Jhariya, M. K. & Jangir, C. K. (2020). Impact of agrochemicals on soil microbiota and management: A review. *Land,* 9**,** 34.

Melnikava, A., Volkava, D. & Khramtsova, E. (2017). Characteristics of bacterial acdS-gene from the strain *Pseudomonas putida* B-37 and the creation of a genetic construct for determining its transient expression in the plant cells *Nicotiana benthamiana*. *Proceedings of the National Academy of Sciences of Belarus, Biological Series***,** 61-68.

Mendes, L., Raaijmakers, J., de Hollander, M., Mendes, R. & Tsai, S. (2018). Influence of resistance breeding in common bean on rhizosphere microbiome composition and function. *The ISME Journal,* 12**,** 212.

Mendes, L. W., Kuramae, E. E., Navarrete, A. A., Van Veen, J. A. & Tsai, S. M. (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. *The ISME Journal,* 8**,** 1577-1587.

Mendoza-Arroyo, G. E., Chan-Bacab, M. J., Aguila-Ramírez, R. N., Ortega-Morales, B. O., Canché Solís, R. E., Chab-Ruiz, A. O., Cob-Rivera, K. I., Dzib-Castillo, B., Tun-Che, R. E. & Camacho-Chab, J. C. (2020). Inorganic Phosphate Solubilization by a Novel Isolated Bacterial Strain *Enterobacter sp*. ITCB-09 and Its Application Potential as Biofertilizer. *Agriculture,* 10**,** 383.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R. & Wilke, A. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics,* 9**,** 386.

Misra, H., Rajpurohit, Y. & Khairnar, N. (2012). Pyrroloquinoline-quinone and its versatile roles in biological processes. *Journal of Biosciences,* 37**,** 313-325.

Mitra, S., Rupek, P., Richter, D. C., Urich, T., Gilbert, J. A., Meyer, F., Wilke, A. & Huson, D. H. (2011). Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. *BMC Bioinformatics,* 12**,** S21.

Mondal, S. & Sarkar, S. (2019). Role of plant growth promoting rhizobacteria in sustainable agriculture: A review. *Journal of Pharmacognosy and Phytochemistry,* 8**,** 2821-2831.

Muratova, A., Turkovskaya, O., Antonyuk, L., Makarov, O., Pozdnyakova, L. & Ignatov, V. (2005). Oil-oxidizing potential of associative rhizobacteria of the genus *Azospirillum*. *Microbiology,* 74**,** 210-215.

Murphy, J., Zehnder, G., Schuster, D., Sikora, E., Polston, J. & Kloepper, J. (2000). Plant growth-promoting rhizobacterial mediated protection in tomato against Tomato mottle virus. *Plant Disease,* 84**,** 779-784.

Myer, P. R., McDaneld, T. G., Kuehn, L. A., Dedonder, K. D., Apley, M. D., Capik, S. F., Lubbers, B. V., Harhay, G. P., Harhay, D. M. & Keele, J. W. (2020). Classification of 16S rRNA reads is improved using a niche-specific database constructed by near-full length sequencing. *Plos One,* 15**,** e0235498.

Nagarajkumar, M., Bhaskaran, R. & Velazhahan, R. (2004). Involvement of secondary metabolites and extracellular lytic enzymes produced by *Pseudomonas fluorescens* in inhibition of *Rhizoctonia solani*, the rice sheath blight pathogen. *Microbiological Research,* 159**,** 73-81.

Naqqash, T., Imran, A., Hameed, S., Shahid, M., Majeed, A., Iqbal, J., Hanif, M. K., Ejaz, S. & Malik, K. A. (2020). First report of diazotrophic *Brevundimonas spp*. as growth enhancer and root colonizer of potato. *Scientific Reports,* 10**,** 1-14.

Ndakidemi, P. & Dakora, F. (2003). Legume seed flavonoids and nitrogenous metabolites as signals and protectants in early seedling development. *Functional Plant Biology,* 30**,** 729-745.

Neitzel, J. J. (2010). Enzyme catalysis: the serine proteases. *Nature Education,* 3**,** 21.

Nelkner, J., Henke, C., Lin, T. W., Pätzold, W., Hassa, J., Jaenicke, S., Grosch, R., Pühler, A., Sczyrba, A. & Schlüter, A. (2019). Effect of long-term farming practices on agricultural soil microbiome members represented by Metagenomically Assembled Genomes (MAGs) and their predicted plant-beneficial genes. *Genes,* 10**,** 424.

Newman, M., Lorenz, N., Hoilett, N., Lee, N., Dick, R., Liles, M., Ramsier, C. & Kloepper, J. (2016). Changes in rhizosphere bacterial gene expression following glyphosate treatment. *Science of the Total Environment,* 553**,** 32-41.

Nhlapo, T., Rees, D., Odeny, D., Mulabisana, J. & Rey, M. (2018). Viral metagenomics reveals sweet potato virus diversity in the Eastern and Western Cape provinces of South Africa. *South African Journal of Botany,* 117**,** 256-267.

Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O. & Tedersoo, L. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Research,* 47**,** D259-D264.

Nouioui, I., Cortés, C. J., Carro, L., Castro, J. F., Gtari, M., Ghodhbane-Gtari, F., Klenk, H.-P., Tisa, L. S., Sangal, V. & Goodfellow, M. (2019). Genomic Insights into plant-growth-promoting potentialities of the genus *Frankia*. *Frontiers in Microbiology,* 10**,** 1457.

O'Banion, B. S., O'Neal, L., Alexandre, G. & Lebeis, S. L. (2020). Bridging the gap between single-strain and community-level plant-microbe chemical interactions. *Molecular Plant-Microbe Interactions,* 33**,** 124-134.

Oberholster, T., Vikram, S., Cowan, D. & Valverde, A. (2018). Key microbial taxa in the rhizosphere of sorghum and sunflower grown in crop rotation. *Science of the Total Environment,* 624**,** 530-539.

Ofaim, S., Ofek-Lalzar, M., Sela, N., Jinag, J., Kashi, Y., Minz, D. & Freilich, S. (2017). Analysis of microbial functions in the rhizosphere using a metabolic-network based framework for metagenomics interpretation. *Frontiers in Microbiology,* 8**,** 1606.

Ogier, J.-C., Pagès, S., Frayssinet, M. & Gaudriault, S. (2020). Entomopathogenic nematode-associated microbiota: from monoxenic paradigm to pathobiome. *Microbiome,* 8**,** 1-17.

Olanrewaju, O. S., Ayangbenro, A. S., Glick, B. R. & Babalola, O. O. (2019). Plant health: feedback effect of root exudates-rhizobiome interactions. *Applied Microbiology and Biotechnology,* 103**,** 1155-1166.

Olanrewaju, O. S., Glick, B. R. & Babalola, O. O. (2017). Mechanisms of action of plant growth promoting bacteria. *World Journal of Microbiology and Biotechnology,* 33**,** 197.

Oleńska, E., Małek, W., Wójcik, M., Swiecicka, I., Thijs, S. & Vangronsveld, J. (2020). Beneficial features of plant growth-promoting rhizobacteria for improving plant growth and health in challenging conditions: A methodical review. *Science of The Total Environment***,** 140682.

Otlewska, A., Migliore, M., Dybka-Stępień, K., Manfredini, A., Struszczyk-Świta, K., Napoli, R., Białkowska, A., Canfora, L. & Pinzari, F. (2020). When Salt Meddles Between Plant, Soil, and Microorganisms. *Frontiers in Plant Science,* 11**,** 1429.

Pan, Y., Cassman, N., de Hollander, M., Mendes, L. W., Korevaar, H., Geerts, R. H., van Veen, J. A. & Kuramae, E. E. (2014). Impact of long-term N, P, K, and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil. *FEMS Microbiology Ecology,* 90**,** 195-205.

Pandey, S. & Gupta, S. (2019). ACC deaminase producing bacteria with multifarious plant growth promoting traits alleviates salinity stress in French bean (*Phaseolus vulgaris*) plants. *Frontiers in Microbiology,* 10**,** 1506.

Papik, J., Folkmanova, M., Polivkova, M., Suman, J. & Uhlik, O. (2020). The invisible life inside plants: Deciphering the riddles of endophytic bacterial diversity. *Biotechnology Advances***,** 107614.

Park, M., Kim, Y., Lee, S. & Kim, I. (2009). Identification of an ISR-related metabolite produced by rhizobacterium *Klebsiella oxytoca* C1036 active against soft-rot disease pathogen in tobacco. *Pest Management Science: Formerly Pesticide Science,* 65**,** 1114-1117.

Pascale, A., Proietti, S., Pantelides, I. S. & Stringlis, I. A. (2020). Modulation of the root microbiome by plant molecules: the basis for targeted disease suppression and plant growth promotion. *Frontiers in Plant Science,* 10**,** 1741.

Patra, J. K., Vishnuprasad, C. N. & Das, G. (2018). *Microbial Biotechnology: Volume 1. Applications in Agriculture and Environment*, Springer, 1-197.

Perazzolli, M., Nesler, A., Giovannini, O., Antonielli, L., Puopolo, G. & Pertot, I. (2020). Ecological impact of a rare sugar on grapevine phyllosphere microbial communities. *Microbiological Research,* 232**,** 126387.

Pérez-Cobas, A. E., Gomez-Valero, L. & Buchrieser, C. (2020). Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial Genomics,* 6**,** e000409.

Petatan-Sagahon, I., Anducho-Reyes, M., Silva-Rojas, H., Arana-Cuenca, A., Tellez-Jurado, A., Cárdenas-Álvarez, I. & Mercado-Flores, Y. (2011). Isolation of bacteria with antifungal activity against the phytopathogenic fungi *Stenocarpella maydis* and *Stenocarpella macrospora*. *International Journal of Molecular Sciences,* 12**,** 5522-5537.

Pfeifer, S. (2017). From next-generation resequencing reads to a high-quality variant data set. *Heredity,* 118**,** 111-124.

Philippot, L., Raaijmakers, J. M., Lemanceau, P. & Van Der Putten, W. H. (2013). Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiology,* 11**,** 789-799.

Phillips, R. W. & Roop, R. M. (2001). *Brucella abortus* HtrA functions as an authentic stress response protease but is not required for wild-type virulence in BALB/c mice. *Infection and Immunity,* 69**,** 5911-5913.

Prabha, R., Singh, D. P., Gupta, S., Gupta, V. K., El-Enshasy, H. A. & Verma, M. K. (2019). Rhizosphere Metagenomics of *Paspalum scrobiculatum* L.(Kodo Millet) Reveals Rhizobiome Multifunctionalities. *Microorganisms,* 7**,** 608.

Prayogo, F. A., Budiharjo, A., Kusumaningrum, H. P., Wijanarka, W., Suprihadi, A. & Nurhayati, N. (2020). Metagenomic applications in exploration and development of novel enzymes from nature: a review. *Journal of Genetic Engineering and Biotechnology,* 18**,** 1-10.

Preece, C. & Penuelas, J. (2016). Rhizodeposition under drought and consequences for soil communities and ecosystem resilience. *Plant and Soil,* 409**,** 1-17.

Purohit, H., Kalia, V. & More, R. (2018). *Soft Computing for Biological Systems*, Springer. 1-300.

Qiao, C., Penton, C. R., Xiong, W., Liu, C., Wang, R., Liu, Z., Xu, X., Li, R. & Shen, Q. (2019). Reshaping the rhizosphere microbiome by bio-organic amendment to enhance crop yield in a maize-cabbage rotation system. *Applied Soil Ecology,* 142**,** 136-146.

Qin, Y., Druzhinina, I. S., Pan, X. & Yuan, Z. (2016). Microbially mediated plant salt tolerance and microbiome-based solutions for saline agriculture. *Biotechnology Advances,* 34**,** 1245-1259.

Raaijmakers, J., Vlami, M. & De Souza, J. (2002). Antibiotic production by bacterial biocontrol agents. *Antonie Van Leeuwenhoek,* 81**,** 537.

Radzki, W., Mañero, F. G., Algar, E., García, J. L., García-Villaraco, A. & Solano, B. R. (2013). Bacterial siderophores efficiently provide iron to iron-starved tomato plants in hydroponics culture. *Antonie Van Leeuwenhoek,* 104**,** 321-330.

Raghavendra, M., Nayaka, S. & Nuthan, B. (2016). Role of rhizosphere microflora in potassium solubilization. *Potassium solubilizing microorganisms for sustainable agriculture.* Springer, 43-59.

Raiyani, N. M. & Singh, S. P. (2020). Taxonomic and functional profiling of the microbial communities of Arabian Sea: A metagenomics approach. *Genomics,* 112**,** 4361-4369.

Rajput, B., Pruitt, K. D. & Murphy, T. D. (2019). RefSeq curation and annotation of stop codon recoding in vertebrates. *Nucleic Acids Research,* 47**,** 594-606.

Rasmann, S. & Turlings, T. (2016). Root signals that mediate mutualistic interactions in the rhizosphere. *Current Opinion in Plant Biology,* 32**,** 62-68.

Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., Domin, H., Fraune, S., Von Frieling, J. & Hentschel, U. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome,* 7**,** 1-19.

Reinhold-Hurek, B., Bünger, W., Burbano, C., Sabale, M. & Hurek, T. (2015). Roots shaping their microbiome: global hotspots for microbial activity. *Annual Review of Phytopathology,* 53**,** 403-424.

Ren, D., Madsen, J. S., Sørensen, S. J. & Burmølle, M. (2015). High prevalence of biofilm synergy among bacterial soil isolates in cocultures indicates bacterial interspecific cooperation. *The ISME Journal,* 9**,** 81-89.

Riedlinger, J., Schrey, S., Tarkka, M., Hampp, R., Kapur, M. & Fiedler, H. (2006). Auxofuran, a novel metabolite that stimulates the growth of fly agaric, is produced by the mycorrhiza helper bacterium *Streptomyces* strain AcH 505. *Applied Environmental Microbiology*, 72**,** 3550-3557.

Rodriguez, P. A., Rothballer, M., Chowdhury, S. P., Nussbaumer, T., Gutjahr, C. & Falter-Braun, P. (2019). Systems biology of plant-microbiome interactions. *Molecular Plant,* 12**,** 804-821.

Rodriguez, R. & Durán, P. (2020). Natural holobiome engineering by using native extreme microbiome to counteract the climate change effects. *Frontiers in Bioengineering and Biotechnology,* 8.

Rousk, J., Bååth, E., Brookes, P., Lauber, C., Lozupone, C., Caporaso, J., Knight, R. & Fierer, N. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal,* 4**,** 1340.

Roux, S., Emerson, J. B., Eloe-Fadrosh, E. A. & Sullivan, M. B. (2017). Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ,* 5**,** e3817.

Ruiz-Moreno, H. A., López-Tamayo, A. M., Caro-Quintero, A., Husserl, J. & Barrios, A. F. G. (2019). Metagenome level metabolic network reconstruction analysis reveals the microbiome in the Bogotá River is functionally close to the microbiome in produced water. *Ecological Modelling,* 399**,** 1-12.

Sabale, S. N., Suryawanshi, P. P. & Krishnaraj, P. (2019). Soil Metagenomics: Concepts and Applications. *Metagenomics-Basics, Methods and Applications.* IntechOpen.

Sachdev, D., Chaudhari, H., Kasture, V., Dhavale, D. & Chopade, B. (2009). Isolation and characterization of indole acetic acid (IAA) producing *Klebsiella pneumoniae* strains from rhizosphere of wheat (*Triticum aestivum*) and their effect on plant growth. *Indian Journal of Experimental Biology,* 47**,** 993-1000.

Salam, L. B. & Obayori, O. S. (2019). Structural and functional metagenomic analyses of a tropical agricultural soil. *Spanish Journal of Soil Science,* 9.

Saleem, M., Hu, J. & Jousset, A. (2019). More than the sum of its parts: microbiome biodiversity as a driver of plant growth and soil health. *Annual Review of Ecology, Evolution, and Systematics,* 50**,** 145-168.

Santoyo, G., Orozco-Mosqueda, M. & Govindappa, M. (2012). Mechanisms of biocontrol and plant growth-promoting activity in soil bacterial species of *Bacillus* and *Pseudomonas*: a review. *Biocontrol Science and Technology,* 22**,** 855-872.

Sasse, J., Martinoia, E. & Northen, T. (2018). Feed your friends: do plant exudates shape the root microbiome? *Trends in Plant Science,* 23**,** 25-41.

Saubidet, M., Fatta, N. & Barneix, A. (2002). The effect of inoculation with *Azospirillum brasilense* on growth and nitrogen utilization by wheat plants. *Plant and Soil,* 245**,** 215-222.

Scavo, A., Abbate, C. & Mauromicale, G. (2019). Plant allelochemicals: agronomic, nutritional and ecological relevance in the soil system. *Plant and Soil,* 442**,** 23-48.

Schirmer, M., Ijaz, U., D'Amore, R., Hall, N., Sloan, W. & Quince, C. (2015). Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Research,* 43**,** e37-e37.

Schlemper, T. R., Leite, M. F., Lucheta, A. R., Shimels, M., Bouwmeester, H. J., van Veen, J. A. & Kuramae, E. E. (2017). Rhizobacterial community structure differences among sorghum cultivars in different growth stages and soils. *FEMS Microbiology Ecology,* 93.

Seiler, G. J., Qi, L. L. & Marek, L. F. (2017). Utilization of sunflower crop wild relatives for cultivated sunflower improvement. *Crop Science,* 57**,** 1083-1101.

Sessitsch, A., Hardoim, P., Döring, J., Weilharter, A., Krause, A., Woyke, T., Mitter, B., Hauberg-Lotte, L., Friedrich, F. & Rahalkar, M. (2012). Functional characteristics of an endophyte community colonizing rice roots as revealed by metagenomic analysis. *Molecular Plant-Microbe Interactions,* 25**,** 28-36.

Shi, J., Yuan, X., Lin, H., Yang, Y. & Li, Z. (2011). Differences in soil properties and bacterial communities between the rhizosphere and bulk soil and among different production areas of the medicinal plant *Fritillaria thunbergii*. *International Journal of Molecular Sciences,* 12**,** 3770-3785.

Shilpa, M., Vikas, H., Sowmya, P., Srikantaiah, M. & Dorajeerao, A. (2015). *In vitro* and *in vivo* evaluation of antagonistic microorganisms on the incidence of leaf blight of sunflower caused by *Alternaria helianthi*. *Plant Archives,* 15**,** 411-416.

Singh, B. K., Munro, S., Potts, J. M. & Millard, P. (2007). Influence of grass species and soil type on rhizosphere microbial community structure in grassland soils. *Applied Soil Ecology,* 36**,** 147-155.

Singh, D., Raina, T. K., Kumar, A., Singh, J. & Prasad, R. (2019a). Plant microbiome: a reservoir of novel genes and metabolites. *Plant Gene,* 18**,** 100177.

Singh, P., Kujur, A., Yadav, A., Kumar, A., Singh, S. & Prakash, B. (2019b). Mechanisms of Plant-Microbe Interactions and its Significance for Sustainable Agriculture. *PGPR Amelioration in Sustainable Agriculture.* Elsevier, 17-39.

Singh, R. P., Shelke, G. M., Kumar, A. & Jha, P. N. (2015). Biochemistry and genetics of ACC deaminase: a weapon to "stress ethylene" produced in plants. *Frontiers in Microbiology,* 6**,** 937.

Singh, S., Singh, U. B., Malviya, D., Paul, S., Sahu, P. K., Trivedi, M., Paul, D. & Saxena, A. K. (2020). Seed Biopriming with Microbial Inoculant Triggers Local and Systemic Defense Responses against *Rhizoctonia solani* Causing Banded Leaf and Sheath Blight in Maize (Zea mays L.). *International Journal of Environmental Research and Public Health,* 17**,** 1396.

Stephan, B. I., Colombo, R. P., Silvani, V. A., Pérgola, M., Godeas, A. M., Pardo, A. G. & Bidondo, L. F. (2019). Short-term effects of genetically modified potato on arbuscular mycorrhizal fungal communities. *Journal of Soil Science and Plant Nutrition,* 19**,** 352-356.

Sugathan, S., Pradeep, N. & Abdulhameed, S. (2017). *Bioresources and Bioprocess in Biotechnology: Volume 2: Exploring Potential Biomolecules*, Springer. 2, 1-379.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H. & Bork, P. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research,* 47**,** D607-D613.

Tariq, M., Khan, A., Asif, M., Khan, F., Ansari, T., Shariq, M. & Siddiqui, M. A. (2020). Biological control: a sustainable and practical approach for plant disease management. *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science,* 70**,** 507-524.

Thai, B. T., Lee, Y. P., Gan, H. M., Austin, C. M., Croft, L. J., Trieu, T. A. & Tan, M. H. (2019). Whole genome assembly of the snout otter clam, Lutraria rhynchaena, using Nanopore and Illumina data, benchmarked against bivalve genome assemblies. *Frontiers in Genetics,* 10**,** 1158.

Thakur, M., Bhattacharya, S., Khosla, P. & Puri, S. (2018). Improving production of plant secondary metabolites through biotic and abiotic elicitation. *Journal of Applied Research on Medicinal and Aromatic Plants*. 12, 1-12.

Thode, S. K., Rojek, E., Kozlowski, M., Ahmad, R. & Haugen, P. (2018). Distribution of siderophore gene systems on a Vibrionaceae phylogeny: Database searches, phylogenetic analyses and evolutionary perspectives. *PloS One,* 13.

Tikariha, H. & Purohit, H. J. (2019). Assembling a genome for novel nitrogen-fixing bacteria with capabilities for utilization of aromatic hydrocarbons. *Genomics,* 111**,** 1824-1830.

Tkacz, A., Pini, F., Turner, T. R., Bestion, E., Simmonds, J., Howell, P., Greenland, A., Cheema, J., Emms, D. M. & Uauy, C. (2020). Agricultural selection of wheat has been shaped by plant-microbe interactions. *Frontiers in Microbiology,* 11**,** 132.

Trivedi, P., Delgado-Baquerizo, M., Anderson, I. & Singh, B. (2016). Response of soil properties and microbial communities to agriculture: implications for primary productivity and soil health indicators. *Frontiers in Plant Science,* 7**,** 990.

Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T. & Singh, B. K. (2020). Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology***,** 1-15.

Uroz, S., Oger, P., Tisserand, E., Cébron, A., Turpault, M., Buée, M., De Boer, W., Leveau, J. & Frey-Klett, P. (2016). Specific impacts of beech and Norway spruce on the structure and diversity of the rhizosphere and soil microbial communities. *Scientific Reports,* 6**,** 27756.

Vacheron, J., Moënne-Loccoz, Y., Dubost, A., Gonçalves-Martins, M., Muller, D. & Prigent-Combaret, C. (2016). *Fluorescent Pseudomonas* strains with only few plant-beneficial properties are favored in the maize rhizosphere. *Frontiers in Plant Science,* 7**,** 1212.

Verma, A., Kumar, S., Kumar, G., Saini, J., Agrawal, R., Satlewal, A. & Ansari, M. (2018). Rhizosphere metabolite profiling: An opportunity to understand plant-microbe interactions for crop improvement. *Crop improvement through microbial biotechnology.* Elsevier, 343-361.

Verma, M., Mishra, J. & Arora, N. K. (2019). Plant growth-promoting rhizobacteria: diversity and applications. *Environmental biotechnology: for sustainable future.* Springer, 129-173.

Viterbo, A., Landau, U., Kim, S., Chernin, L. & Chet, I. (2010). Characterization of ACC deaminase from the biocontrol and plant growth-promoting agent *Trichoderma asperellum* T203. *FEMS Microbiology Letters,* 305**,** 42-48.

Vives-Peris, V., de Ollas, C., Gómez-Cadenas, A. & Pérez-Clemente, R. M. (2020). Root exudates: from plant to rhizosphere and beyond. *Plant Cell Reports,* 39**,** 3-17.

Vranova, V., Rejsek, K. & Formanek, P. (2013). Proteolytic activity in soil: a review. *Applied Soil Ecology,* 70**,** 23-32.

Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E. E. & van der Heijden, M. G. (2019). Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nature Communications,* 10**,** 1-10.

Wang, W., Vinocur, B., Shoseyov, O. & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science,* 9**,** 244-252.

Wei, X., Ge, T., Zhu, Z., Hu, Y., Liu, S., Li, Y., Wu, J. & Razavi, B. S. (2019). Expansion of rice enzymatic rhizosphere: temporal dynamics in response to phosphorus and cellulose application. *Plant and Soil,* 445**,** 169-181.

Weisskopf, L., ABOU-MANSOUR, E., Fromin, N., Tomasi, N., Santelia, D., Edelkott, I., Neumann, G., Aragno, M., Tabacchi, R. & Martinoia, E. (2006). White lupin has developed a complex strategy to limit microbial degradation of secreted citrate required for phosphate acquisition. *Plant, Cell & Environment,* 29**,** 919-927.

White III, R., Rivas-Ubach, A., Borkum, M., Köberl, M., Bilbao, A., Colby, S., Hoyt, D., Bingol, K., Kim, Y. & Wendler, J. (2017). The state of rhizospheric science in the era of multi-omics: A practical guide to omics technologies. *Rhizosphere,* 3**,** 212-221.

Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E., Kyrpides, N., Mavrommatis, K. & Meyer, F. (2012). The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics,* 13**,** 141.

Wolińska, A. (2019). Metagenomic achievements in microbial diversity determination in croplands: A review. *Microbial Diversity in the Genomic Era.* Elsevier, 15-35.

Xu, J., Zhang, Y., Zhang, P., Trivedi, P., Riera, N., Wang, Y., Liu, X., Fan, G., Tang, J. & Coletta-Filho, H. D. (2018). The structure and function of the global citrus rhizosphere microbiome. *Nature Communications,* 9**,** 1-10.

Xu, W.-f., Ren, H.-s., Ou, T., Lei, T., Wei, J.-h., Huang, C.-s., Li, T., Strobel, G., Zhou, Z.-y. & Xie, J. (2019). Genomic and functional characterization of the endophytic *Bacillus subtilis* 7PJ-16 strain, a potential biocontrol agent of mulberry fruit sclerotiniose. *Microbial Ecology,* 77**,** 651-663.

Yang, Y., Wang, N., Guo, X., Zhang, Y. & Ye, B. (2017). Comparative analysis of bacterial community structure in the rhizosphere of maize by high-throughput pyrosequencing. *PLoS One,* 12**,** e0178425.

Yang, Z., Yang, W., Li, S., Hao, J., Su, Z., Sun, M., Gao, Z. & Zhang, C. (2016). Variation of bacterial community diversity in rhizosphere soil of sole-cropped versus intercropped wheat field after harvest. *PLoS One,* 11.

Yanni, Y. & El-Fattah, F. (1999). Towards integrated biofertilization management with free living and associative dinitrogen fixers for enhancing rice performance in the Nile delta. *Symbiosis,* 27**,** 319-331.

You, M., Fang, S., MacDonald, J., Xu, J. & Yuan, Z.-C. (2020). Isolation and characterization of *Burkholderia cenocepacia* CR318, a phosphate solubilizing bacterium promoting corn growth. *Microbiological Research,* 233**,** 126395.

Zehr, J. P., Jenkins, B. D., Short, S. M. & Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environmental Microbiology,* 5**,** 539-554.

Zhalnina, K., Louie, K. B., Hao, Z., Mansoori, N., da Rocha, U. N., Shi, S., Cho, H., Karaoz, U., Loqué, D. & Bowen, B. P. (2018). Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nature Microbiology,* 3**,** 470-480.

Zhang, L., Jiang, M., Ding, K. & Zhou, S. (2019a). Iron oxides affect denitrifying bacterial communities with the *nirS* and *nirK* genes and potential N2O emission rates from paddy soil. *European Journal of Soil Biology,* 93**,** 103093.

Zhang, L., Loh, K.-C., Lim, J. W. & Zhang, J. (2019b). Bioinformatics analysis of metagenomics data of biogas-producing microbial communities in anaerobic digesters: A review. *Renewable and Sustainable Energy Reviews,* 100**,** 110-126.

Zhang, R., Vivanco, J. & Shen, Q. (2017a). The unseen rhizosphere root–soil–microbe interactions for crop production. *Current Opinion in Microbiology,* 37**,** 8-14.

Zhang, X., Zhang, R., Gao, J., Wang, X., Fan, F., Ma, X., Yin, H., Zhang, C., Feng, K. & Deng, Y. (2017b). Thirty-one years of rice-rice-green manure rotations shape the rhizosphere microbial community and enrich beneficial bacteria. *Soil Biology and Biochemistry,* 104**,** 208-217.

Zhao, S., Li, K., Zhou, W., Qiu, S., Huang, S. & He, P. (2016). Changes in soil microbial community, enzyme activities and organic matter fractions under long-term straw return in north-central China. *Agriculture, Ecosystems & Environment,* 216**,** 82-88.

Zhou, D., Huang, X., Chaparro, J., Badri, D., Manter, D., Vivanco, J. & Guo, J. (2016). Root and bacterial secretions regulate the interaction between plants and PGPR leading to distinct plant growth promotion effects. *Plant and Soil,* 401**,** 259-272.

APPENDICES

Appendix 1: A published data article from the PhD study

Data in Brief 31 (2020) 105831

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

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

Olubukola Oluranti Babalola [a,*], Temitayo Tosin Alawiye [a], Carlos Rodriguez Lopez [b], Ayansina Segun Ayangbenro [a]

[a] Food Security and Safety Niche, Faculty of Natural and Agricultural Sciences, Private Mail Bag X2046, North-West University, South Africa
[b] 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 (O.O. Babalola).

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

Appendix 2: A published review article from this PhD study

Biologia
https://doi.org/10.2478/s11756-020-00574-z

REVIEW

Metabolomics: current application and prospects in crop production

Temitayo Tosin Alawiye[1] · Olubukola Oluranti Babalola[1]

Received: 2 April 2020 / Accepted: 27 July 2020

Abstract
Necessary steps need to be taken in order to cushion the effects of hunger, famine, and drought in the world. To ameliorate these challenges, biotechnological innovations like meta-omics techniques have been employed in crop improvement programs. Metabolomics is a viable technology to fully detect the functional network of metabolites with far-reaching biological uses in agriculture, medicine, and pharmaceutical disciplines. This review is grouped into four sections discussing 1) plant and microbial metabolomics, 2) metabolomics and its application in crop production, 3) metabolomics workflow and techniques explaining the analytical techniques and instrumentation with merits and demerits, and 4) metabolomic analysis of metabolites in a metabolic network. This technology has been applied in plant-microbe interaction, biological control measures, and abiotic stress tolerance, where various crops like maize, sunflower, soybean, and wheat have been used. In view of all these, further, development is needed to increase our understanding of microbial metabolites in order to develop bioproducts that will increase growth and eventual yield, protecting plants from pathogens and in the process providing nutritious food for the teeming populace. Although metabolomics can be applied to a wide range of scientific areas, we focus this review on plant improvement, which is a driver for improved food security.

Keywords Biotechnological innovations · Develop bioproducts · Food security · Metabolic network · Plant metabolomics

Abbreviations

PMN	Plant Metabolic Network
GMD	Golm Metabolome Database
AMF	Arbuscular Mycorrhizal Fungi
QTLs	Quantitative Trait Loci
MQTLs	Methylation Qualitative Trait Loci
GWAS	Genome Wide Association Studies
NMR	Nuclear Magnetic Resonance spectroscopy (NMR)
LC and GC-MS	Liquid and Gas Chromatography with Mass Spectrometry
CE-MS	Capillary Electrophoresis with Mass Spectrometry
GEMs	Genome-scale Metabolic Models
TCA	Tricarboxylic Acid
RNA	Ribonucleic Acid

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

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

Introduction

Metabolomics, also known as metabolite profiling, aims at simultaneously detecting the overall changes of metabolites in a biological system. It is a powerful concept with an inclusive observation of how metabolic network is regulated (Weckwerth and Fiehn 2002). The terms and ideas of metabolomics were introduced less than two decades ago, focusing on better understanding of biological networks through the precise and extensive study of metabolism (Fiehn 2008). Metabolism produces energy and building blocks which are used for reproduction, protection, maintenance, communication, and structure of a cell (Barupal and Fiehn 2017).

When metabolism takes place, metabolites are produced, they are small biological molecules involved in energy conversion and biosynthesis, and they perform important functions such as studying the interaction between the organism and the environment. It can also be used to determine the functions and health of an organism at the molecular level (Lu et al. 2017; Volkova and Geras' kin 2018). These metabolites do not only help in plant growth, development and adaptation to environmental changes but are also useful in the fields of food and agriculture, medicine, and energy (Chen et al. 2016). Metabolomics comprises the study of a wide

Published online: 05 August 2020