UNIVERSIDAD
NACIONAL
DE COLOMBIA

CARACTERIZACIÓN DEL METAGENOMA DE LA COMUNIDAD MICROBIANA EDÁFICA ASOCIADA A UN CULTIVO DE ARROZ *(Oryza sativa)* BAJO UN ESQUEMA AGRONÓMICO DE MANEJO DE AGRICULTURA POR AMBIENTES

Juan David Saavedra Correa

Universidad Nacional de Colombia

Facultad de Ciencias, Posgrado Interfacultades en Microbiología

Bogota D.C., Colombia

2023

CARACTERIZACIÓN DEL METAGENOMA DE LA COMUNIDAD MICROBIANA EDÁFICA ASOCIADA A UN CULTIVO DE ARROZ *(Oryza sativa)* BAJO UN ESQUEMA AGRONÓMICO DE MANEJO DE AGRICULTURA POR AMBIENTES

Juan David Saavedra Correa

Tesis presentada como requisito parcial para optar al título de:
Magister en Ciencias-Microbiología

Director:
Ph.D Sandra Milena Gonzalez Sayer
Co-director:
Ph.D Fabio Ancizar Aristizabal Gutierrez

Grupo de investigación:
“Bioprocesos y Bioprospección”

Universidad Nacional de Colombia
Facultad de Ciencias, Posgrado Interfacultades en Microbiología
Bogota D.C., Colombia
2023

"We know more about the movement of celestial bodies, than about the soil underfoot."

-Leonardo Da Vinci

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

_____

Juan David Saavedra Correa

Fecha 31/01/2023

Acknowledgements

Special thanks to my advisors Sandra Milena Gonzalez, Ibonne Aydee Garcia, Carolina Isabel Cuellar, Fabio Ancizar Aristizabal. This research would not have been possible without the support and learning provided by the “Universidad Nacional de Colombia”, the professors of the Master's Program in Sciences-Microbiology, and finally the support from different advisors and co-workers of the IBUN “Instituto de Biotecnología de la Universidad Nacional de Colombia”. I would also like to thank FEDEARROZ-FNA (National Federation of Rice Farmers in Colombia) and its researchers, for allowing the development of this work, and for the logistical and academic support provided during the completion of the thesis.

Abstract

Given the importance of rice cultivation in Colombia, many strategies have been used to increase yield per hectare, which seek to directly and indirectly encourage the promotion of ecosystem services such as nutrient cycling, with the soil microbiome being a key factor in the modulation of many nutrients present in the soil and having an effect on plant productivity, it was proposed as the objective of this research, to study the edaphic microbiomes of a commercial field of rice and its relationship with the physico-chemical properties of the soil. For this, bulk and rhizosphere soil samples were taken in a 33-hectare rice plot, previously characterized according to the yield data history, in three management zones (high, medium, and low yield). The soil samples were taken before planting the crop and seventy days after plants germination; moreover, physicochemical analyzes of the soil and DNA extraction were performed. Initially, a strategy for the study of microbiomes through 16s rRNA and ITS amplicons was proposed, however, the results obtained with this methodology were not reliable enough, for which the decision was made to carry out the analysis from the perspective of metagenomics, for subsequent shotgun sequencing and metagenome analysis. The microbial communities from the rice field reported a low diversity in general, the samples were found to be dominated by the phyla *Proteobacteria, Acidobacteria*, and *Actinobacteria*. Even though, there were variations in the composition and structure of the bulk soil microbiomes across time and between the microbiomes associated with the three management zones, no significant differences were discovered. The diversity, composition, and predicted function of rhizosphere microbiomes were found to be significantly different from the bulk soil microbiomes. Moreover, it was identified that these soils had a particularly acid pH, and it was also possible to detect that organic matter, as well as management practices had an impact on the diversity of the microbiomes.

Keywords: Rice soil metagenomics, Bulk soil Microbiomes, Rhizosphere microbiomes, Amplicon sequencing, Shotgun sequencing.

Resumen

Dada la importancia del cultivo del arroz en Colombia, se han utilizado muchas estrategias para incrementar el rendimiento por hectárea, las cuales buscan incentivar directa e indirectamente la promoción de servicios ecosistémicos como el ciclaje de nutrientes, siendo el microbioma del suelo un factor clave en la modulación de muchos nutrientes presentes en el suelo, con un efecto en la productividad de las plantas. Se propuso como objetivo de este trabajo de investigación, estudiar los microbiomas edáficos de un campo comercial de arroz y su relación con las propiedades físico-químicas del suelo. Para ello, se tomaron muestras de suelo de soporte y rizosférico en un lote de arroz de 33 hectáreas, previamente caracterizadas según el historial de datos de rendimiento, en tres zonas de manejo (rendimiento alto, rendimiento medio y rendimiento bajo). Las muestras de suelo se tomaron antes de la siembra del cultivo y después de la última fertilización química; Además, se realizaron análisis fisicoquímicos del suelo y extracción de ADN. Inicialmente se planteó una estrategia para el estudio de microbiomas a través de 16s rRNA y amplicones ITS, sin embargo, los resultados obtenidos con esta metodología no fueron lo suficientemente confiables, por lo que se tomó la decisión de realizar el análisis desde la perspectiva de la metagenómica, para su posterior secuenciación por "*shotgun*" y análisis de metagenoma. Las comunidades microbianas del campo de arroz reportaron una baja diversidad en general, se encontró que las muestras estaban dominadas por los filos *Proteobacteria, Acidobacteria* y *Actinobacteria*. Aunque hubo variaciones en la composición y estructura de los microbiomas del suelo de soportea lo largo del tiempo y entre los microbiomas asociados con las tres zonas de manejo, no se encontrarondiferencias significativas. Se encontróque la diversidad, la composición y la función predicha de los microbiomas de la rizosfera eran significativamente diferentes de los microbiomas del suelo de soportel. Además, se identificó que estos suelos tenían un

pH particularmente ácido, y también se pudo detectar que la materia orgánica incidía en la diversidad de los microbiomas, asi como las prácticas de manejo.

Keywords: Metagenómica del suelo de arroz, Microbiomas de suelo de soporte, Microbiomas de rizosfera, Secuenciacion shotgun.

Contents

List of figures

List of tables

List of Symbols and Abbreviations

abbreviations

Abbreviation	Term
DNA	Deoxyribonucleid Acid
RNA	Ribonucleid Acid
rRNA	Ribosomal RNA
ASV	Amplicon Sequence Variant
MAG	Metagenome-Assembled Genome
PCR	Polymerase chain reaction
PCoA	Principal Coordinate Analysis
N	Nitrogen
P	Phosphorus
K	Potassium
S	Sulfur
Ca	Calcium
Mg	Magnesium
Mn	Manganese
Al	Aluminium
B	Boron

Na	Sodium
Fe	Iron
Cu	Copper
Zn	Zinc
SOM	Soil Organic Matter
SOC	Soil Organic Carbon
ECEC	Effective cation exchange capacity
EC	Soil Electric conductivity
ATP	Adenosine triphosphate
MG-RAST	Metagenomic Rapid Annotations using Subsystems Technology
QIIME2	Quantitative Insights Into Microbial Ecology.
ITS	Internal transcribed spacer
NGS	Next Generation Sequencing

Introduction

Rice *(Oryza sativa)* is a crop of great global importance, being the primary food source in different countries, especially in the Asian continent, where the most significant production and planted area worldwide is concentrated (FAO, 2022). In the national context, rice represents the transitory crop with the largest planted area in the country, with 392,648 hectares in the first half of 2021 (DANE, 2021), contributing about 5.1% to agricultural GDP and providing about 82,500 direct jobs and 330,000 indirect jobs (FEDEARROZ, 2020). For these reasons, rice is considered a fundamental crop of Colombian development in the cultural, social, agricultural, and economic fields.

Traditional agronomic management practices have considered agricultural ecosystems as homogeneous places, ignoring the existing variability in the soil's physical, chemical, and biological properties (Moharana et al., 2020). An alternative that arose due to the need to increase the productivity of agricultural areas has been the characterization of management zones, which are areas within a field that share homogeneous properties (Breunig et al., 2020; Chen et al., 2021), through the collection and analysis of information on soils intended for agriculture, to understand their spatial and temporal heterogeneity (Doerge, 2005; Williams et al., 2016). Accordingly, it is essential to mention that modern agriculture presents various challenges for a constantly growing population. In order to generate and maintain food security, in recent years, alternatives and solutions have been sought that, from a sustainable approach, seek to improve the yield of the planet's arable areas.

The biological component of the soil is a preponderant factor in the identification of agricultural management zones. Thanks to the vast taxonomic and functional diversity of the edaphic microbial populations, important processes such as nutrient cycling are carried out. These processes generate ecosystem services directly and indirectly, including support, regulation, and provisioning (Jing et al., 2022; Lukac et al., 2017; S. Prasad et al., 2021). Various authors have demonstrated the effect of fluctuations in soil physical and

chemical variables on the composition of microbial communities, concluding that there is a correlation between soil variability and the structure and composition of the microbiome (Chang et al., 2017; De Gannes et al., 2015; Doni et al., 2022; Perera & Tirimanne, 2021; Tarafdar, 2022; Wieneke, 2017). Consequently, these microbial ecology studies have contributed to a better understanding of the edaphic microbiological potential by defining its function and relationship with agricultural yield.

The study of edaphic communities was limited by culture-dependent techniques, where only a minimal portion (1%) of the total diversity of soil microorganisms could be tracked under *in vitro* conditions (Prosser et al., 2007). However, in recent decades the development of culture-independent technologies has given a glimpse of the enormous microbial diversity hidden in the terrestrial environment. For example, metataxonomy is one tool that implements phylogenetic marker gene sequences for tracking microbial communities (Kim & Lee, 2020). This consists of isolating all the genomes found in an environmental sample from the extraction of their DNA and subsequent amplification of the marker gene. On the other hand, the emergence of next-generation sequencing technologies has allowed the complete sequencing of the diversity of genomes within an environmental sample (Reddy et al., 2021). These tools make it possible to establish an approximation to organism's functional and phylogenetic relationships in a given space and time (Baldrian, 2019; Jansson, 2013; Reddy et al., 2021).

The knowledge of the composition of the microbiomes associated with the soil in different management zones of a rice field through metataxonomic and metagenomic analysis may guarantee in the future the design of new agronomic management strategies that allow the conservation and promotion of the microbial population diversity that is linked with the production of different ecosystem services. This will help improve the yield and efficiency of rice cultivation. For this reason, the objective of this research is to characterize the edaphic microbial populations found in the different management zones of rice fields. Our study will consider the 3 different management zones present before planting and after chemical fertilization of a rice crop cycle in a rice field at the department of Meta, Colombia, to understand and establish the relationship between these microbial populations with the physicochemical soil characteristics.

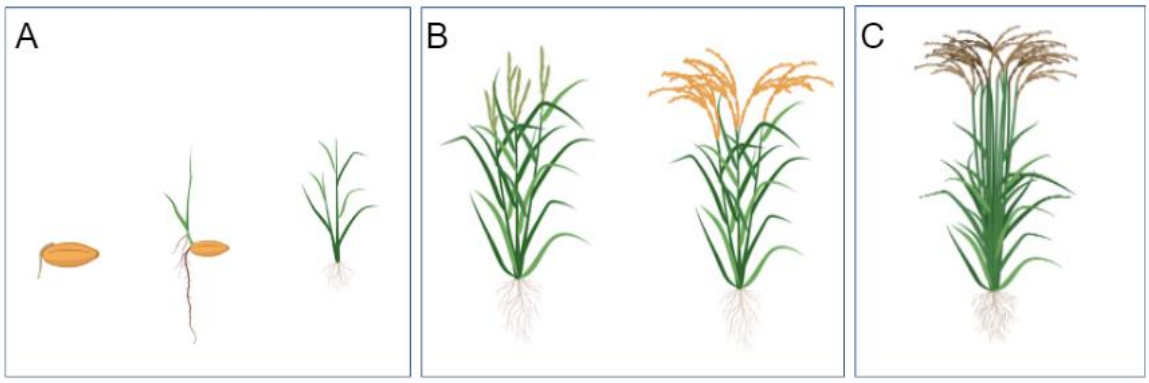
1. Chapter 1: Theoretical framework, problem statement, and objectives

1.1. Rice Cultivation (*Oryza sativa*)

Oryza sativa is an annual monocotyledonous plant of the *Poaceae* family, native to the Asian continent. Its cultivation represents significant relevance as it has positioned itself as one of the crops with the highest production globally. By 2021, around 517.1 million tons were produced in an area close to 160 million hectares (FAO, 2022). Rice is cultivated in Asia, Africa, America, Europe, and Oceania in a wide range of environments, at different altitudes that vary from zero to 2500 meters above sea level, in tropical, subtropical, and temperate climates (Chauhan et al., 2017), with different precipitation conditions and cultivation techniques (S. Prasad et al., 2021).

The life cycle of the crop is constituted of three main phases: (1) the vegetative phase, which begins with the germination of the seed and ends with the differentiation of the floral primordium, which generally lasts between 35 and 50 days; (2) followed by the reproductive phase that begins with the differentiation of the floral primordium and ends with flowering with an occurrence of 30 to 35 days and finally (3) maturation that begins with the flowering stage and ends with grain maturity with a duration of 30 to 45 days (Fig. 1-1) (Chauhan et al., 2017; Garcés-Varón & Medina-Rubio, 2018), however, the duration of the crop cycle is dependent on the variety of the plant and the various environmental factors.

Figure 1-1: Phenological stages of rice cultivation, (A) Vegetative stage (B) Reproductive stage, (C) Maturity stage.



The International Rice Research Institute has classified different agroecosystems in which rice is grown: these are the irrigated rice ecosystem which makes up 75% of the world's crops, the lowland rainfed rice ecosystem, the upland rice, and the flood-prone rice ecosystem (IRRI, 1993). Like the great variety of environments in which rice is grown, there is a variety of soils since it is grown in sandy loam soils up to heavy clay or clay loam, although it is usually sown in soils of fine and medium texture, in addition, It is cultivated in soils below a pH of 5 to soils that have a pH greater than 9. However, the optimum conditions for cultivation are found with a pH between 5.0 and 6.6 (Chauhan et al., 2017; R. Prasad et al., 2017).

Regarding the nutritional requirement of the crop, it has been well described that nitrogen (N), phosphorus (P), and potassium (K), are coming from mineralization and available components of fertilizer, contributing greatly to soil fertility (FAO, 2022). The presence of these elements is imperative to ensure the plant's productivity; their concentrations are good indicators of soil quality and productivity due to their favorable effects on the physical, chemical, and biological properties of soil (Dong et al., 2012). Nitrogen plays a vital role in all living organisms and is a constituent of protein, enzymes, hormones, vitamins, alkaloids, chlorophyll etc. Plant growth is adversely affected in its absence (Osman, 2013). Rice crops use around 21-25% of the total nitrogen fertilizers consumed globally. However, the efficiency of N use in rice is very low. The low efficiency of nitrogen use in rice and even

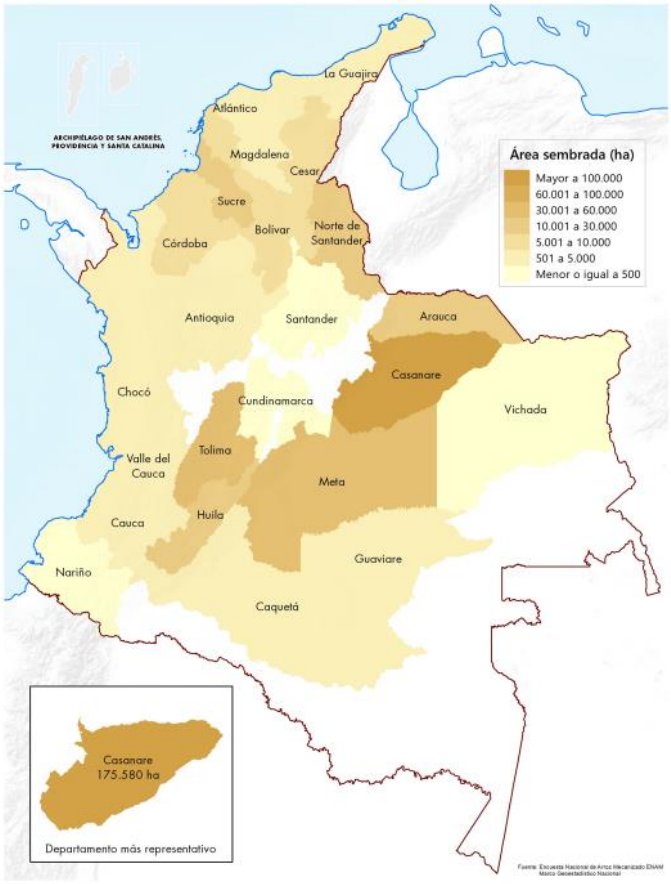
other crops is due to four mechanisms of N loss: surface run-off on sloping land, ammonia volatilization, leaching, and denitrification (Fahad et al., 2018; FAO, 2006).

Likewise, Phosphorus is an innate constituent of the phospholipids, DNA, RNA, ATP, and other crucial biomolecules. Consequently, an insufficient supply of P from soil to biological systems can constrain growth, and productivity, and affect crop metabolism (McLaren et al., 2020). However, excessive application of P fertilizers can cause P accumulation in the soil and can lead to P losses from leaching (Jiang et al., 2021). The other essential macronutrient is potassium, which is the main intracellular ion for all types of cells, and it is necessary for their normal function; it is an important factor in the resilience to stress, metabolism, development, and reproduction in plants. Depletion of plant-available K in soils results in a variety of negative impacts, including preventing optimum utilization of applied nitrogen and phosphorus fertilizers; which threatens the yields of the rice cropping systems (Cakmak, 2010; Olaniyan et al., 2022)

1.1.1. Rice cultivation in Colombia

In Colombia, rice cultivation is the third largest planted area, with more than 500,000 hectares (Fig 1-2) after coffee and corn cultivation. In 2021, a production of 2,373,193 tons was obtained (DANE, 2021). Production is located especially in four geographical areas, the Central Zone (Tolima, Huila, and Valle del Cauca), the Eastern Plains area (Meta and Casanare), the Caribbean (Antioquia, Bolívar, Córdoba, Cesar, Guajira, Santanderes, and Sucre) and Meta (FEDEARROZ, 2019, 2020). Colombian rice production has concentrated in the departments of Casanare and Meta, located in the region of the eastern plains (FEDEARROZ, 2019, 2020), which contribute around 50% of the country's rice production.

Figure 1-2: Rice crop planted area by departments in Colombia in the first semester of the year 2021 (DANE, 2021).



Rice is a transitory crop that is sown every six months under two systems in the country: the mechanized system that consists of the use of machines to carry out the different agricultural tasks, this being the most common, and the manual system, which has a state associated with small producers and family farming (DANE, 2014; FEDEARROZ, 2019). Likewise, depending on the adequacy of the crop, it can be classified as rainfed rice that is planted in systems, where water only comes from the rains and irrigated rice where water is provided throughout the crop cycle (Garcés-Varón & Medina-Rubio, 2018).

On the other hand, the National Federation of Rice Growers of Colombia is an organization that aims to benefit the rice producer, promoting the technological development of their

crops and seeking economic efficiency and greater competitiveness. FEDEARROZ has implemented the massive technology adoption model (AMTEC) that intends to continuously transfer the technology to the Integrated Management of Rice cultivation, through the integration of all the actors (Producers - Extensionists - Researchers, and official entities), to generate new production alternatives with the improvement of current agronomic practices, involving related institutions and groups in the sector to obtain greater benefits from the crop. Within this framework, sustainable management practices are promoted for the preservation of the different resources on which rice cultivation is dependent (FEDEARROZ, 2018).

1.2. Agriculture by management zones

The growing demand for agricultural products related to the increase in the human population on the planet has led to the search for new alternatives in the development of agriculture, which seek to sustain the production of agricultural systems, avoiding the degradation and loss of natural resources on which they depend. For example, practices such as precision agriculture, defined as the set of techniques aimed at optimizing the use of agricultural inputs (seeds, fertilizers, agrochemicals) based on the quantification of the spatial, and temporal variation of agricultural production, includes the use of new technologies, such as global positioning systems (GPS), improved machinery, sensors and satellite and aerial images, together with geographic information systems (GIS) plus the necessary management to optimize production and crop yields taking into account the variability and existing uncertainty within agricultural systems (Breunig et al., 2020; Gebbers & Adamchuk, 2010; Moharana et al., 2020; Williams et al., 2016)

Agriculture by management zones has been used in precision agriculture to diversify the agronomic management of a crop (Scudiero et al., 2018). This is based on understanding agricultural areas as places with heterogeneous zones in the conditions that compose them through the analysis of quantitative and qualitative data and historical factors. The analysis of these components that result from very marked differences in the soil’s physical, chemical, and biological properties, thus determines management zones with different

productive aptitudes in the same area (Chen et al., 2021; Doerge, 2005; Justo & Scianca, 2011)). This variability of soil properties is correlated with the fact that they operate with different intensities and on different space-time scales, also due to some extrinsic factors such as crop management practices, fertilization, and irrigation (Tripathi et al., 2015).

Moreover, this kind of crop management also tries to promote greater soil biodiversity in order to expand ecosystem services, which can influence plant growth and crop yield; promoting, in particular, the development of soil structure, carbon storage, nutrient availability, and moisture regulation (Williams et al., 2016).

1.3. Soil properties

Soils are the terrestrial support of life on earth and are essential for sustaining it. The soil covers about $1.43 \times 10^{8}\ km^{2}$ of the total area of the earth calculated in $5.1 \times 10^{8}\ km^{2}$. Soils change remarkably in different places on the planet, generating a vast diversity in their physical, chemical, and biological properties (Kutílek & Nielsen, 2017).

The soil contains three natural components: (1) the solid phase of the soil matrix (formed by mineral particles and solid organic particles); (2) the liquid phase, which is often represented by water, and which could more correctly be called the soil solution; and (3) the gas phase, which contains air and other gases (Rai et al., 2017).

The capacity to support life is mainly given by the structure of the soil, which refers to how the components of the soil (sand, clay, and silt) are grouped according to size, shape, and arrangement, forming aggregates (Bronick & Lal, 2005). The stability of the aggregates affects the movement and storage of water, aeration, biological activity, and plant growth, therefore influences a wide range of physical and biogeochemical processes (Amézketa, 1999).

The amount of organic matter, macronutrients, and micronutrients are key drivers of soil chemical properties; similarly, the effective cation exchange capacity (ECEC), which is a measure of soils' capacity to retain positively charged ions, is critical because it affects soil structure stability, nutrient availability, and pH (Brown & Lemon, 2008). Another determining characteristic in the soil is the pH, which is a measure that determines its acidity or alkalinity. It is defined as the negative logarithm of the concentration of hydrogen ions in a solution, which usually fluctuates between 3.0 and 9.0 on average in the planet's soil. Besides, is considered that soils with a pH greater than 7 are alkaline, and soils with a pH less than 7 are acidic. The pH of a solution affects nutrient solubilization, microbial populations, and the rate of humification and mineralization processes (Osorio, 2012; Soon & Hendershot, 2006).

1.3. Soil microbiomes

The study of microbial organisms in their natural environments is known as microbial ecology, which aims to distinguish the function and full diversity of microbial organisms that cohabit in a given space and time (Prosser et al., 2007; Reddy et al., 2021). Considering the fact that all environments on the planet (soil, ocean depths, acidic hydrothermal springs, tundra, etc.) are inhabited by microorganisms, which play a crucial role in the balance and regulation of the different processes of the ecosystems of the planet (Harwood, C. & Buckley, 2008). It is important to define that a microbiome within an ecological context is a community of microorganisms with various roles and ecological relationships within an environment (Berg et al., 2020), it is also understood as a microbiome as the sum of microorganisms and elements of the genome in a particular environment.

With the recognition of soil as an important component of the maintenance of terrestrial life, the study of microorganisms within the soil environment has grown substantially. (Baldrian, 2019). However, it is not enough to identify the composition of microbial communities, their potential, and their traits in order to comprehend soil activities, instead relationships

between physical and chemical parameters must also be established (De Gannes et al., 2015).

Soil microorganisms, especially rhizosphere microorganisms, are important mediators of plant nutrition, growth, and resistance to biotic and abiotic stress, so understanding the composition and dynamics of associated microbial communities is essential for crop management. (Andrew & Edwards, 2011; Doni et al., 2022). The diversity of microorganisms translates into metabolic diversity, which allows the formation of different ecological niches in certain spaces and times (Berg et al., 2020). The immense metabolic diversity of these organisms (autotrophs, heterotrophs, phototrophs, chemoorganotrophs, and lithotrophs and mixotrophs), plus the ecological interactions allows the maintenance of soil homeostasis, promoting the generation of multiple ecosystem services from which we benefit (Lopes, 2013; Lukac et al., 2017).

Among the main ecosystem services are microorganisms considered as "ecosystem engineers" responsible for physically and chemically modifying the properties of the soil. Similarly, there are other groups considered "biological regulators" which, based on ecological relationships such as predation, parasitism, etc., control different populations with the aim of maintaining them in their habitat (Lukac et al., 2017; S. Prasad et al., 2021; Tarafdar, 2022). Biodiversity, in turn, contributes to the productivity and resilience of agroecosystems, since it has been observed that the accelerated loss of microbial biodiversity is linked to a reduction in functions, stability, resilience, and productivity (Nielsen et al., 2015).

Knowledge of biodiversity is a significant and essential factor to ensure and maintain adequate soil conditions in accordance with the different ecosystem services provided by microorganisms (Chandra, 2021; Wolińska, 2018). For this reason, different methods have been defined for the study of communities, for example, α and β diversity, which are measures of the complexity of a community (Bolyen et al., 2019; Whittaker, 1972), both terms were introduced by Whittaker (1972)., who defined α diversity as the diversity within

a particular area and/or ecosystem, expressed as the number of species in this case of OTUS (species richness), also defined as local diversity. Meanwhile, β-diversity is the variation in species composition between different sites within a region or area of interest, the composition is determined by the abundance, identity, and number of species in a community.

1.3.1 Study of edaphic microbiomes

Soil is the most biodiverse environment on Earth, it is estimated that it contains approximately 1000 Gbp of microbial genome sequences per gram of soil (Trevors, 2009). Taking the above into account, the development of massive sequencing tools in the last two decades has allowed the study of microbial communities from the genome of all the individuals existing in an environment or niche (Marchesi & Ravel, 2015; Reddy et al., 2021). These sequencing techniques allow us to understand the evolution, ecology, and genetics of the microbial diversity present in the soil.

The development of culture-independent analysis of microorganisms has adopted two approaches: Metataxonomic analysis with amplicons and Shotgun metagenomics. The amplification and DNA sequencing of gene sequences function as phylogenetic molecular markers of microorganisms since they allow the taxonomic identification of the organisms present in an environmental sample (Garrido-Cardenas & Manzano-Agugliaro, 2017). Sequencing for metataxonomic analysis in microbial ecology studies has focused on using the 16S rRNA gene and the internal transcript interspacer region (ITS), which have highly conserved regions and variable regions between species.

The analysis of genome sequences and marker genes such as ribosomal RNA is in favor since sequencing costs have shown a decreasing trend and its capacity is increasing. In addition, the development of new bioinformatics tools has allowed for more robust and truthful analyses, which have been facilitated by the depth of sequencing and the ability to sequence multiple samples. This has made sequencing technologies very attractive to explore the diversity of microbial Operational Taxonomic Units (OTUS) in the world

(Nannipieri et al., 2020). The success of soil microbiome analysis depends on a combination of good sample selection, efficient DNA extraction methods, amplification, if necessary, detection strategies, and sequencing approaches, along with the management and sharing of data (Vogel et al., 2009).

Additionally, shotgun sequencing, which targets the entirety of the microbial genetic information contained in an environmental sample, offers another method for studying microbial communities. It can be used to retrieve the entire genome sequences from an environmental sample and assess the taxonomic composition and functional potential of microbial communities. The decreasing cost of sequencing and the development of better computational tools have made metagenomics more widely used (Quince et al., 2017).

NGS (Next Generation Sequencing) platforms are massive and parallel sequencing technologies, they allow the sequencing of nucleotides in metagenomes, genomes or in regions of interest of these. Currently, the platform most used in research is Illumina®, due to its low error rates (<1%), costs, and high performance. The Illumina platform is dependent on PCR (Polymerase Chain Reaction) and is carried out by binding DNA fragments to immobilized primers on a solid surface called "Flow Cell", here the DNA fragments will generate clusters through amplification. clonal by bridges thus amplifying the reading signal. Following this, the nucleotides that are added in the DNA polymerization are chemically modified so that when they are incorporated, termination occurs, which is reversible in the reaction. These nucleotide analogs are fluorescently labeled so that when they are incorporated into DNA sequencing, they emit a fluorescence signal that is read by the sequencer (Illumina, 2015).

Massive sequencing technologies include a series of universal steps that are broadly grouped into library preparation, sequencing reaction, and data analysis. The unique combination of specific protocols distinguishes each of the technologies and determines the type and amount of data produced from one platform to another (Chang et al., 2017;

Reddy et al., 2021). The differences between platforms are notable when comparing the data obtained in quality, quantity, and cost of sequencing.

1.3.2. Edaphic microbiomes associated with rice crops *(Oryza sativa)*

Before the appearance of NGS platforms, other molecular techniques were used, such as gene cloning such as 16s rRNA, polymerase chain reaction denaturing gradient gel electrophoresis (PCR-DGGE), and restriction fragment analysis. length polymorphism (RFLP), to evaluate the composition of the microbial communities in the bulk soil, rhizosphere soil, and tissues of the rice plant (Kim & Lee, 2020). Even so, the appearance of new massive and parallel sequencing technologies has allowed considerable progress in research on the composition and structure of microbial communities associated with rice cultivation.

Various studies have sought to demonstrate how different conditions, practices, and agricultural systems influence the composition and structure of the soil microbial community. Well, it has intended to understand in greater depth the existing correlation of microbial diversity in rice crops with the different ecosystem conditions and services, such as the increase in the availability of organic and inorganic nutrients, that promote plant development as a result of the multiple biochemical reactions carried out by these communities (Kim & Lee, 2020; Yuan et al., 2019).

It has generally been found that the bacterial phyla Proteobacteria, Chloroflexi, Actinobacteria, and Acidobacteria tend to be dominant in these bulk and rhizosphere soil microbiomes associated with rice crops (Kim & Lee, 2020; Lopes, 2013; J. Wang et al., 2017; Yuan et al., 2019). Regarding the most dominant fungal phylum, it has been determined that Ascomycota and Basidiomycota represent the highest species richness in rice crop soils (Chang et al., 2017; Ding et al., 2019; Maguire et al., 2020). Similarly, a pattern has been observed in the abundance of these phyla in the support and rhizosphere soil of rice fields in different parts of the planet, and it has been possible to understand that

the variability in the edaphic microbial communities is highly correlated with the heterogeneity of soil and plant characteristics.

In a study carried out by Yuan et al. (2019) through sequencing of the 16S rRNA gene, of soil samples from rice fields from 9 different regions of China, they showed that the abundance of archaea was mostly correlated with temperature and pH, while the abundance of bacteria was mainly affected by the availability of soil organic matter and the total nitrogen content. The dominant bacterial phyla were Proteobacteria (32.4%), Acidobacteria (17.8%), Bacteroidetes (9.3%), and Verrucomicrobia (6.0%).

In the same way, the influence of the application of inorganic fertilizer and organic fertilizer has been evaluated, where they found that both types of fertilization induce changes in the composition of the microbial community in both prokaryotes and eukaryotes, for example, in chemical fertilization the abundance increased. of some oligotrophic bacteria such as Bacteroidetes and Acidobacteria and increased the relative abundance of the fungal phylum Zygomycota. The application of organic fertilization increased the abundance of copiotrophic bacteria such as the Proteobacteria phylum, and the relative abundance of the Agaricomycetes and Orbiliomycetes classes of fungi, likewise it was observed that the total amount of nitrogen and ammonium in the soils was a critical determinant in the composition of the communities (Wang et al., 2017).

The search for the existing relationship between the physical and chemical characteristics with the biological component of the soils in agricultural areas with high yields is also relevant to understand crop yields, in research carried out in China it was possible to determine the changes in the composition of the soil microbiome related to rice fields that presented high levels of yield, analyzing the soil metagenome during four key stages of rice growth, it was found that the diversity of bacterial taxa is greater in soils with high rice yields in the same way they were related to functions of nitrogen metabolism and a greater abundance of genes involved in the nitrification process, for example, the hydroxylamine oxidoreductase gene and nitric oxide di-oxygenase, promoting the effective transformation

of nitrogen allowing the availability of this element for plants, being able to generate (Zhong et al., 2020).

In another investigation carried out on a rice crop in the department of Tolima - Colombia, with different management of the rice straw, it was also found that the predominant phylum in rice fields during a productive cycle was Proteobacteria, Acidobacteria, Chloroflexi, Actinobacteria, and Firmicutes, representing more than 60% of the sequences, these results allowed to show that the addition of organic matter influences the microbial edaphic community in the rice fields (Carreño, 2019).

1.4. Statement of the Problem and Justification

Much of the productivity of crops is determined by the fertility or quality of the soil, which can be evaluated according to its physical, chemical, and microbiological characteristics. Soil microorganisms play a fundamental role in the provision of different ecosystem services. Within the sustainable management of soils with the aim of improving crop yields, it is essential to involve practices that contribute to the conservation and increase of edaphic microbial populations, since, as is well known, microorganisms participate in the processes of soil formation and in all fundamental elementary cycles (Doni et al., 2022; Zhong et al., 2020).

Currently in Colombia, the productivity of rice crops presents multiple challenges that have been difficult to solve since the implementation of traditional agriculture, such as high production costs, depletion of water resources, erosion and desertification, climate change, low yields. and the negative impact on the environment due to agrochemicals that lead to the degradation of ecosystems that are vital in the balance of the planet. Technological renewal is needed, based on the use of new tools and strategies for agronomic management, which contribute to the agricultural sector guaranteeing the country's food security and boosting the national economy in the post-COVID pandemic era.

Within this framework, FEDEARROZ-FNA has promoted the Mass Adoption of Technology (AMTEC) program, seeking competitiveness for rice farmers, for which it has set itself the challenge of increasing yields by one ton under sustainable production, going beyond the average current national from 4.9 tons/hectare to 5.9. To meet this objective, FEDEARROZ has made available to all farmers a series of tools that include agriculture by management zones (FEDEARROZ, 2018).

Within the rice plots, areas have been identified where there are lower yields and in which the cost of production is generally higher; Likewise, there is great variability in production between lots, having cases with yields that vary between 4 to 10 tons per hectare in the same region of the country, which indicates that correcting the factors that affect yield by specific site leads to an increase in the productivity of the crop (Garcés-Varón & Medina-

Rubio, 2018). A determining factor to address changes in agronomic management by the environment is to understand the relationships between the soil components in each differentiated area of the lot, with the microbiological component being a priority for sustainable production.

Agriculture by management zones is part of precision agriculture, a set of strategies aimed at optimizing agricultural inputs based on the identification and quantification of the spatial and temporal variability of production. Unlike traditional agriculture, this modality moves away as far as possible from fixed or uniform management, depending on the analysis of the information collected. The observation of differences in the determining factors of production in agroecosystems is not new; change is the possibility of identifying, quantifying, and mapping that variability (Breunig et al., 2020). In order to identify the existing spatial variability in a rice field, environmental maps are created and each of these is characterized to find those factors that are limiting or enhancing production.

To ensure that these management practices are correct, it is essential to understand microbial activity and its influence on the properties of the soil and the plant species themselves. The understanding and management of these interactions allow agroecosystems to also contribute to the stability and conservation of the soil as an integral component, whose energy flow is balanced, favoring its functionality and self-regulation to lead to lower requirements in the application of inputs (fertilizers and other agrochemicals). With the development of nucleic acid sequencing technologies and biological data analysis, knowledge of the soil microbiome has expanded to such an extent that it is possible to establish the microorganisms present in a given condition, as well as an approximation to the metabolic activity that these are exerting and their relationship with the variety of rice cultivated.

These types of studies with high-throughput sequencing technologies have become a tool that allows establishing in a holistic way the influence that agronomic practices can have on microbial populations and helping to establish their effect on crop health and yield. In

order not only to conserve the soil through sustainable agronomic practices but also to achieve greater productivity for Colombian rice farmers, which contributes to food security. This project presented in alliance between the academy and the productive sector, aims to join efforts to generate basic knowledge about the interaction of soil components and their relationship with crop yield under agronomic management by management zones, aimed at solving real problems. .

On the other hand, in this situation generated by the pandemic caused by SARS-CoV-2, rice cultivation is definitely a catalyst for the regional and national economy in times of pandemic and post-pandemic, not only mitigating problems such as unemployment, but also businesses are also reinvented and ventures arise taking advantage of current conditions, for this reason it is necessary to continue developing research to increase the productivity and competitiveness of this cereal, guaranteeing that the sector remains in the country, so it is not ruled out that knowledge The basic principle of rice soil microbiomes will promote agribusiness around biological products for the soil in the near future, by demonstrating its massive adoption in the agronomic practices of the crop.

1.5. Research Question

Are there temporal and spatial differences in the composition of the soil microbiomes from a commercial field of rice, characterized for the management of agriculture by management zones?

1.6. Hypothesis

The composition of the edaphic microbiomes is different within the determined management zones, and in two evaluated moments of a productive cycle of the rice field, which allows the establishment of associations between the physicochemical characteristics of the soil and the microbial diversity evaluated.

1.7. Objectives

1.7.1 General Objective

To study the edaphic microbiomes of a commercial rice field and its relationship with the physical-chemical properties of the soil, in two moments of a productive cycle characterized for the management of agriculture by management zones.

1.7.2 Specific Objectives

Obtain the sequences of the 16s rRNA and ITS1/ITS2 genomic regions of the microorganisms associated with the soil from three management zones established within a rice field, during two moments of the productive cycle.

Analyze the sequences of the genomic regions obtained to determine the composition of the microbial communities present in the soils studied.

To compare the diversity of the edaphic microbiomes present before the planting of the crop and seventy days after plant germination, of the different management zones of the field.

Evaluate the relationship between the physical-chemical components of the soil and the composition of the edaphic microbiomes.

1.8. References

Andrew, J., & Edwards, A. (2011). *Structure, Variation, and Dynamics of the Root-Associated Microbiota of the Crop Plant Rice*.

Baldrian, P. (2019). The known and the unknown in soil microbial ecology. In *FEMS Microbiology Ecology* (Vol. 95). Oxford University Press. https://doi.org/10.1093/femsec/fiz005

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M. C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., … Schloter, M. (2020). Microbiome definition re-visited: Old concepts and new challenges. In *Microbiome* (Vol. 8). BioMed Central Ltd. https://doi.org/10.1186/s40168-020-00875-0

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Breunig, F. M., Galvão, L. S., Dalagnol, R., Dauve, C. E., Parraga, A., Santi, A. L., Della Flora, D. P., & Chen, S. (2020). Delineation of management zones in agricultural fields using cover–crop biomass estimates from PlanetScope data. *International Journal of Applied Earth Observation and Geoinformation*, *85*, 102004.

https://doi.org/10.1016/j.jag.2019.102004

Carreño, J. del P. (2019). *Evaluación de la diversidad taxonómica y funcional de la comunidad microbiana relacionada con el ciclo del nitrógeno en suelos de cultivo de arroz con diferentes manejos del tamo*. Universidad Nacional de Colombia.

Chandra, R. (2021). Soil Biodiversity and Community Composition for Ecosystem Services. In A. Rakshit, S. K. Singh, P. C. Abhilash, & A. Biswas (Eds.), *Soil Science: Fundamentals to Recent Advances* (pp. 69–84). Springer. https://doi.org/10.1007/978-981-16-0917-6_5

Chang, H. X., Haudenshield, J. S., Bowen, C. R., & Hartman, G. L. (2017). Metagenome-wide association study and machine learning prediction of bulk soil microbiome and crop productivity. *Frontiers in Microbiology*, *8*(APR). https://doi.org/10.3389/fmicb.2017.00519

Chauhan, B. S., Jabran, K., & Mahajan, G. (Eds.). (2017). *Rice Production Worldwide* (1st ed. 2017). Springer International Publishing : Imprint: Springer. https://doi.org/10.1007/978-3-319-47516-5

Chen, S., Du, T., Wang, S., Parsons, D., Wu, D., Guo, X., & Li, D. (2021). Quantifying the effects of spatial-temporal variability of soil properties on crop growth in management zones within an irrigated maize field in Northwest China. *Agricultural Water Management*, *244*, 106535. https://doi.org/10.1016/j.agwat.2020.106535

DANE. (2014). Censo Nacional Agropecuario 2014. In *Departamento Administrativo Nacional de Estadística (DANE)*. https://www.dane.gov.co/files/images/foros/foro-de-entrega-de-resultados-y-cierre-3-censo-nacional-agropecuario/CNATomo2-Resultados.pdf

DANE. (2021). *Encuesta Nacional de Arroz Mecanizado (ENAM) Primer semestre de 2021*. https://fedearroz.s3.amazonaws.com/media/documents/comunicado_ENAM_Isem21_2_XnnAwff.pdf

De Gannes, V., Eudoxie, G., Bekele, I., & Hickey, W. J. (2015). Relations of microbiome characteristics to edaphic properties of tropical soils from Trinidad. *Frontiers in Microbiology*, *6*(SEP), 1045. https://doi.org/10.3389/fmicb.2015.01045

Ding, L.-J., Cui, H., Nie, S., Long, X., Duan, G., & Zhu, Y.-G. (2019). Microbiomes inhabiting rice roots and rhizosphere. *FEMS Microbiology Ecology*. https://doi.org/10.1093/femsec/fiz040

Doerge, T. (2005). Management Zone Concepts. *South Dakota State University*, *1*. http://www.ipni.net/publication/ssmg.nsf/0/C0D052F04A53E0BF852579E500761AE3/$FILE/SSMG-02.pdf

Doni, F., Suhaimi, N. S. M., Mispan, M. S., Fathurrahman, F., Marzuki, B. M., Kusmoro, J., & Uphoff, N. (2022). Microbial Contributions for Rice Production: From Conventional Crop Management to the Use of 'Omics' Technologies. *International Journal of Molecular Sciences*, *23*(2), 737. https://doi.org/10.3390/ijms23020737

Fahad, S., Adnan, M., Noor, M., Arif, M., Alam, M., Khan, I. A., Ullah, H., Wahid, F., Mian, I. A., Jamal, Y., Basir, A., Hassan, S., Saud, S., Amanullah, Riaz, M., Wu, C., Khan, M. A., & Wang, D. (2018). Major constraints for global rice production. In *Advances in Rice Research for Abiotic Stress Tolerance* (pp. 1–22). Elsevier. https://doi.org/10.1016/B978-0-12-814332-2.00001-0

FAO. (2006). Nutrient management guidelines for some major field crops. In *FOOD AND*

AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS.

FAO. (2022). *FAO Cereal Supply and Demand Brief | World Food Situation | Food and Agriculture Organization of the United Nations*. Https://Www.Fao.Org/Worldfoodsituation/Csdb/En/. https://www.fao.org/worldfoodsituation/csdb/en/

FEDEARROZ. (2018). Adopción Masiva De Tecnologia AMTEC AMTEC FEDEARROZ. *Revista Arroz*, 22–34.

FEDEARROZ. (2019). INFORME DE GESTIÓN FONDO NACIONAL DEL ARROZ. *Revista Arroz*.

FEDEARROZ. (2020). Federación Nacional de Arroceros, Estadísticas Arroceras en Colombia. In *Area, Producción y Rendimientos*. http://www.fedearroz.com.co/new/apr_public.php

Garcés-Varón, G., & Medina-Rubio, J. (2018). LA FISIOLOGIA DEL CULTIVO DEL ARROZ EN EL PROGRAMA AMTEC. *Fedearroz*, *1*(2).

Garrido-Cardenas, J. A., & Manzano-Agugliaro, F. (2017). The metagenomics worldwide research. In *Current Genetics* (Vol. 63). Springer Verlag. https://doi.org/10.1007/s00294-017-0693-8

Gebbers, R., & Adamchuk, V. I. (2010). Precision agriculture and food security. In *Science* (Vol. 327). https://doi.org/10.1126/science.1183899

Harwood, C. & Buckley, M. (2008). A golden age for microbial ecology. In *Nature Reviews Microbiology* (Vol. 6). Nature Publishing Group. https://doi.org/10.1038/nrmicro1957

Illumina. (2015). *Technology Spotlight: Illumina ® Sequencing*.

IRRI. (1993). Rice Researcn in a time of change. *International Rice Research Institute*, *1*(1).

Jansson, J. (2013). Encyclopedia of Metagenomics: Soil Metagenomics. In *Encyclopedia of Metagenomics* (Vol. 2). Springer New York. https://doi.org/10.1007/978-1-4614-6418-1

Jing, J., Cong, W.-F., & Bezemer, T. M. (2022). Legacies at work: Plant–soil–microbiome interactions underpinning agricultural sustainability. *Trends in Plant Science*, *27*(8), 781–792. https://doi.org/10.1016/j.tplants.2022.05.007

Justo, C., & Scianca, ; Carlos. (2011). Agricultura Por Ambientes. Estrategias De Manejo De Maiz En Suelos Con Diferentes Aptitud Productiva. *EEA INTA GENERAL VILLEGAS*, *1*.

Kim, H., & Lee, Y.-H. (2020). The Rice Microbiome: A Model Platform for Crop Holobiome. *Phytobiomes Journal • 2020 •*, *4*, 5–18. https://doi.org/10.1094/PBIOMES-07-19-0035-RVW

Kutílek, M., & Nielsen, D. R. (2017). *Soil The Skin of the Planet Earth* (1st ed.). Springer Books.

Lopes, R. (2013). *Towards a sustainable rice culture: The role of microbiota* [PhD Thesis]. Universidade do Porto.

Lukac, M., Grenni, P., & Gamboni, M. (2017). Soil Biological Communities and Ecosystem Resilience. In *Soil Biological Communities and Ecosystem Resilience* (Vol. 1). Springer International Publishing. https://doi.org/10.1007/978-3-319-63336-7

Maguire, V. G., Bordenave, C. D., Nieva, A. S., Llames, M. E., Colavolpe, M. B., Gárriz,

A., & Ruiz, O. A. (2020). Soil bacterial and fungal community structure of a rice monoculture and rice-pasture rotation systems. *Applied Soil Ecology*, *151*, 103535. https://doi.org/10.1016/j.apsoil.2020.103535

Marchesi, J. R., & Ravel, J. (2015). The vocabulary of microbiome research: A proposal. *Microbiome*, *3*(1), 1–3. https://doi.org/10.1186/s40168-015-0094-5

Moharana, P. C., Jena, R. K., Pradhan, U. K., Nogiya, M., Tailor, B. L., Singh, R. S., & Singh, S. K. (2020). Geostatistical and fuzzy clustering approach for delineation of site-specific management zones and yield-limiting factors in irrigated hot arid environment of India. *Precision Agriculture*, *21*(2), 426–448. https://doi.org/10.1007/s11119-019-09671-9

Nannipieri, P., Ascher, J., Ceccherini, M. T., Petramellara, G., Giancarlo, R., & Schloter, M. (2020). Beyond microbial diversity for predicting soil functions: A mini review. *Pedosphere*, *30*(1), 5–17. https://doi.org/10.1016/S1002-0160(19)60824-6

Nielsen, U. N., Wall, D. H., & Six, J. (2015). Soil Biodiversity and the Environment. *Annual Review of Environment and Resources*, *40*(1), 63–90. https://doi.org/10.1146/annurev-environ-102014-021257

Perera, T., & Tirimanne, S. (2021). *Role of Microbial Communities in Sustainable Rice Cultivation* (pp. 189–223). https://doi.org/10.1007/978-981-15-9912-5_7

Prasad, R., Shivay, Y. S., & Kumar, D. (2017). Current status, challenges, and opportunities in rice production. In *Rice Production Worldwide* (pp. 1–32). Springer International Publishing. https://doi.org/10.1007/978-3-319-47516-5_1

Prasad, S., Malav, L. C., Choudhary, J., Kannojiya, S., Kundu, M., Kumar, S., & Yadav, A. N. (2021). Soil Microbiomes for Healthy Nutrient Recycling. In A. N. Yadav, J.

Singh, C. Singh, & N. Yadav (Eds.), *Current Trends in Microbial Biotechnology for Sustainable Agriculture* (pp. 1–21). Springer. https://doi.org/10.1007/978-981-15-6949-4_1

Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., Green, J. L., Green, L. E., Killham, K., Lennon, J. J., Osborn, A. M., Solan, M., van der Gast, C. J., & Young, J. P. W. (2007). The role of ecological theory in microbial ecology. *Nature Reviews Microbiology*, *5*(5), 384–392. https://doi.org/10.1038/nrmicro1643

Reddy, B., Kumar, A., Mehta, S., & Sahu, K. P. (2021). Methods of Assessments of Microbial Diversity and Their Functional Role in Soil Fertility and Crop Productivity. In S. K. Dubey & S. K. Verma (Eds.), *Plant, Soil and Microbes in Tropical Ecosystems* (pp. 293–314). Springer. https://doi.org/10.1007/978-981-16-3364-5_14

Scudiero, E., Teatini, P., Manoli, G., Braga, F., Skaggs, T. H., & Morari, F. (2018). Workflow to establish time-specific zones in precision agriculture by spatiotemporal integration of plant and soil sensing data. *Agronomy*, *8*(11), 1–21. https://doi.org/10.3390/agronomy8110253

Tarafdar, J. C. (2022). Role of Soil Biology on Soil Health for Sustainable Agricultural Production. In B. Giri, R. Kapoor, Q.-S. Wu, & A. Varma (Eds.), *Structure and Functions of Pedosphere* (pp. 67–81). Springer Nature. https://doi.org/10.1007/978-981-16-8770-9_3

Trevors, J. T. (2009). One gram of soil: A microbial biochemical gene library. *Antonie van Leeuwenhoek*, *97*(2), 99. https://doi.org/10.1007/s10482-009-9397-5

Tripathi, R., Nayak, A. K., Shahid, M., Lal, B., Gautam, P., Raja, R., Mohanty, S., Kumar, A., Panda, B. B., & Sahoo, R. N. (2015). Delineation of soil management zones for a rice cultivated area in eastern India using fuzzy clustering. *Catena*, *133*, 128–136. https://doi.org/10.1016/j.catena.2015.05.009

Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., van Elsas, J. D., Bailey, M. J., Nalin, R., & Philippot, L. (2009). TerraGenome: A consortium for the sequencing of a soil metagenome. In *Nature Reviews Microbiology* (Vol. 7). Nature Publishing Group. https://doi.org/10.1038/nrmicro2119

Wang, J., Song, Y., Ma, T., Raza, W., Li, J., Howland, J. G., Huang, Q., & Shen, Q. (2017). Impacts of inorganic and organic fertilization treatments on bacterial and fungal communities in a paddy soil. *Applied Soil Ecology*, *112*, 42–50. https://doi.org/10.1016/j.apsoil.2017.01.005

Wang, T. D., Huang, T. D., Huang, H. S., & Hua, Y. W. (2017). Origin of Secondary Somatic Embryos and Genetic Stability of the Regenerated Plants in Hevea brasiliensis. *Journal of Rubber Research*, *20*(2), 101–116. https://doi.org/10.1007/BF03449145

Whittaker, R. H. (1972). EVOLUTION AND MEASUREMENT OF SPECIES DIVERSITY. *TAXON*, *21*(2–3), 213–251. https://doi.org/10.2307/1218190

Wieneke, X. (2017). *TEMPROAL CHANGE IN SOIL MICROBIOME UNDER INFLUENCE OF EIGHT YEARS OF MANURE AMENDMENT IN CULTIVATED SOILS*. Oklahoma State University.

Williams, A., Kane, D. A., Ewing, P. M., Atwood, L. W., Jilling, A., Li, M., Lou, Y., Davis, A. S., Grandy, A. S., Huerd, S. C., Hunter, M. C., Koide, R. T., Mortensen, D. A.,

Smith, R. G., Snapp, S. S., Spokas, K. A., Yannarell, A. C., & Jordan, N. R. (2016). Soil functional zone management: A vehicle for enhancing production and soil ecosystem services in row-crop agroecosystems. *Frontiers in Plant Science*, *7*(FEB2016), 65. https://doi.org/10.3389/fpls.2016.00065

Wolińska, A. (2018). Metagenomic Achievements in Microbial Diversity Determination in Croplands: A Review. A Review. In *Microbial Diversity in the Genomic Era* (pp. 15–35). Elsevier. https://doi.org/10.1016/B978-0-12-814849-5.00002-2

Yuan, C. L., Zhang, L. M., Wang, J. T., Hu, H. W., Shen, J. P., Cao, P., & He, J. Z. (2019). Distributions and environmental drivers of archaea and bacteria in paddy soils. *Journal of Soils and Sediments*, *19*(1), 23–37. https://doi.org/10.1007/s11368-018-1997-0

Zhong, Y., Hu, J., Xia, Q., Zhang, S., Li, X., Pan, X., Zhao, R., Wang, R., Yan, W., Shangguan, Z., Hu, F., Yang, C., & Wang, W. (2020). Soil microbial mechanisms promoting ultrahigh rice yield. *Soil Biology and Biochemistry*, *143*, 107741. https://doi.org/10.1016/j.soilbio.2020.107741

2. Chapter 2: Amplicon Sequencing with SWIFT AMPLICON™ 16S+ITS PANEL kit

2.1 Introduction

The study of the vast diversity of microorganisms within an environmental sample has been possible thanks to the development of nucleic acid sequencing technologies and bioinformatics tools. The application of Next Generation Sequencing (NGS) for metagenomic analysis has elucidated the fundamental role that microorganisms play in a certain space and time, which contributes directly and indirectly to the maintenance and homeostasis of various ecosystems (Bruno et al., 2015; Gołębiewski & Tretyn, 2020). In particular, the knowledge of the composition of microorganisms associated with soils for agricultural is of interest, to understand their interaction with the biotic and abiotic factors,

and the influence on the productivity and sustainability of agrosystems (Banerjee & van der Heijden, 2022; Dubey et al., 2019; Prasad et al., 2021).

One of the techniques to elucidate the taxonomic composition of the microbial communities present in a given environment is the sequencing of amplicons of marker genes. The most common molecular markers used by that end, are the small subunit of the 16s ribosomal DNA (16S rDNA) and the Internal transcribed spacer (ITS) regions, which are applied for prokaryotes and eukaryotes microorganisms respectively (Francioli et al., 2021; Semenov, 2021). This approach relies on the 16S rDNA and/or ITS marker PCR-directed sequencing, and the further discrimination of the hypervariable regions in order to identify the environment microbial composition. Hypervariable regions are useful to track microbial communities due to this variability Moreover, they are flanked by highly conserved regions that serve as binding sites for PCR primers, giving rise to accurate and robust techniques to differentiate the microbial organisms present in a sample (Bukin et al., 2019; Knight et al., 2018).

Amplicon sequencing has allowed a significant development in the field of meta-taxonomic characterization of communities in certain environmental samples such as soil samples (Alteio et al., 2021). The usage of molecular markers has been considered a gold standard to assess microbial communities due to the decrease in the cost of the technologies, the availability of different methods of extraction of nucleic acids and library preparation kits, and the constant enlargement of taxonomic databases (Starke et al., 2021).

The workflow of amplicon sequencing is initially based on the extraction of DNA from an environmental sample, in which commercial extraction kits are generally used, these grant the lysis of microbial cells within the sample, allowing the isolation of high yields of pure DNA (Knauth et al., 2013). The next step is DNA library preparation, where a wide variety of protocols and commercial kits exist. Its objective is the amplification of marker genes through PCR, then amplicons are fused with sequencing platform-specific adapters (van Dijk et al., 2014). Finally, the sequencing of these amplicons is carried out with next-generation sequencing platforms that allow high-throughput sequencing of DNA fragments.

followed by sequence analysis using various bioinformatics pipelines, such as QIME2 and MOTHUR (Bolyen et al., 2019; Liu et al., 2020; Schloss, 2020).

The 16S rRNA gene is the RNA component of the 30S subunit of a prokaryotic ribosome (SSU rRNA), is approximately 1500 base pairs long, and includes nine regions of varying conservation (V1-V9) (Bukin et al., 2019). Likewise, the study of microbial communities of eukaryotes has used the Internal transcribed spacer (ITS), that is the spacer DNA situated between the small-subunit ribosomal RNA (rRNA) and large-subunit rRNA genes (Nilsson et al., 2019). It has been proposed that the ITS regions provide greater resolution to predict the richness and taxonomic composition of microbial communities compared with markers located within the ribosomal SSU and LSU genes (Li et al., 2020). However, as the entire ITS region and 16s gene were considered too long to be fully sequenced (Francioli et al., 2021), only targeting one of the regions from the 16s and either the ITS1 or the ITS2 region is considered practical for this kind of studies.

In this study the SWIFT AMPLICON 16S+ITS PANEL, library preparation kit was chosen, due the kit enables highly efficient, sensitive, and specific microbial identification by targeting the 16S rRNA (V1-V9) and ITS genes in a single primer pool, along with a simple two-hour workflow (Swift Biosciences, USA). Nevertheless, various authors have explored the use of different regions of the 16s and the ITS, and it has been concluded that these present variations in the resolution of the microbial composition within a sample (Bukin et al., 2019; Pecundo et al., 2021; Rajeev et al., 2020; Yang et al., 2018). Even so, the use of different DNA fragments from these marker genes continues to be debated, since the changes in the level of taxonomic resolution obtained might be related to the biases provided by the DNA extraction techniques, PCR amplification, sequencing, and the use of different bioinformatics pipelines (Liu et al., 2021; Pollock et al., 2018).

With the objective to have greater coverage of the 16s rRNA gene and the ITS genomic regions of microorganisms associated with soil samples from a rice field, an amplicons metataxonomic study was carried out, to understand better how the composition of the

microbiome varies, in three different management zones within a rice crop field of 33 ha, in the eastern plains of Colombia, during two moments of the productive cycle.

2.2 Materials and Methods

2.2.1 Area of study

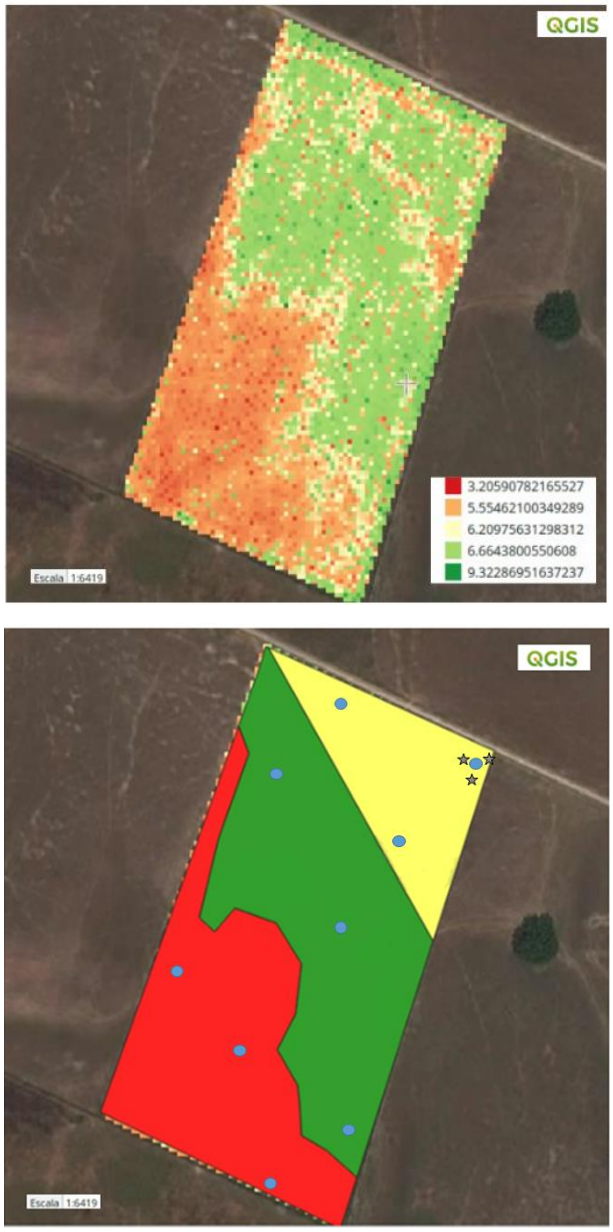
The analysis of the amplicons was carried out from soil samples obtained in a field of rice from the La Arabia farm at coordinates 4° 23' 6.54" North, 72° 57' 56.65" West, from the village of Las Delicias, in the municipality of Cabuyaro in the department of Meta, Colombia (Fig. 2-1), located in the biogeographic zone of Colombian easter plains (llanos Orientales); during a production cycle in a rice crop lot *(Oryza sativa)* under an irrigation system and inorganic fertilization management.

Figure 2-1: Geographic localization of la Arabia farm in Meta, Colombia.

Source: IGAC, 2022

Initially, an identification of three different management zones was carried out within the 33-hectares rice field, these management zones were selected according to the yield in previous production cycles by researchers from FEDEARROZ-FNA in 2021 (Fig. 2-2).

Figure 2-2: Map of the three management zones and the nine georeferenced points, in Red color the management zone of low yield are observed, in green color the management zones of high yield, and in yellow color, the management zone of medium yield is observed, crop yield (ton/ha).



Source: FEDEARROZ-FNA, 2021.

After this, three different management zones were assessed including a low-yield management zone in red color on the map with a historic crop yield of 4.6 ton/ha, a medium

yield management zone (yellow color with a yield of 5.4 ton/ha), and a high yield management zone (green color with a yield of 5.6 ton/h). Following this, within the field, nine different points were randomly georeferenced. The preceding allowed the selection of three points composing biological replicates for each management zone for taking bulk soil samples (Table 2-1), which were the biological resource for DNA extraction.

Table 2-1: Exact coordinates of each point by management zone of the rice field where the bulk soil samples were taken.

Management Zone	Sample	Latitude	Longitude
Medium yield / yellow	S1, S11, R1	4,38512618641	-72,9665052
	S2, S12, R2	4,38715381988	-72,9673261
	S3, S13, R3	4,38636818605	-72,965625
Low yield / red	S4, S14, R4	4,38337848013	-72,9699936
	S5, S15, R5	4,38221363317	-72,9688437
	S6, S16, R6	4,38051847106	-72,9684401
High yield / green	S7, S17, R7	4,38618515582	-72,9683755
	S8, S18, R8	4,38119154140	-72,9675532
	S9, S19, R9	4,38374735126	-72,9671808

2.2.2 Sampling experimental design

Two different samplings were carried out during a productive cycle of a rice crop field (Figure 2-3). The first bulk soil sampling occurred in August 2021 and was carried out 20 days before planting the crop with the FL-Itagua variety. The second sampling was carried out in November 2021, 70 days after planting the crop and after the last chemical fertilization. A total of 27 bulk soil samples were obtained (3 management zones x 3 composite samples x 3 replicates). The nine georeferenced points were taken with a 5 cm diameter soil auger (Figure 2-4). Each composite sample consisted of three different soil

cores collected at a depth of 0 to 20 cm, which were mixed in a plastic bag to finally obtain soil replicates at each of the points, then they were packaged in three different 50 ml falcon tubes previously labeled with the information of each point.

Figure 2-3: Pictures of the sampling field at the “La Arabia” farm at 2 different moments of the sampling.

Saavedra, 2021

The obtained samples were immediately preserved in a container with dry ice to prevent DNA degradation during transport to the laboratory of molecular markers characterization in the “Instituto de Biotecnología de la Universidad Nacional de Colombia (IBUN)” in Bogotá, Colombia. Once the samples arrived at the laboratory they were stored in an ultra-freezer at a temperature of -80°C to continue with the further DNA extraction.

Figure 2-4: Bulk soil sampling method pictures.

Saavedra, 2021

2.2.3 DNA extraction

Total bulk soil DNA was extracted from 0.25 grams of soil from each replicate of the composite samples from each of the nine sites considered in this study. For this purpose, the DNA isolation kit DNeasy PowerSoil Pro Kit ® (Qiagen, Germany) was implemented modifying the manufacturer's instructions in the cell lysis step, since a high DNA degradation was observed. The DNA extraction modifications consisted of reducing the

agitation speed of the soil samples from 2500 to 500 rpm. The modified agitation was applied for ten minutes in the first step. In order to improve the DNA integrity parameter, we maintain the cold chain throughout the whole extraction process.

Thereafter, the genomic DNA extraction, the concentration, and purity parameters were evaluated through 1% agarose gel electrophoresis, spectrophotometry with Nanodrop ND-2000 UV-VIS (NanoDrop Technologies, Wilmington, DE, USA) and fluorometry with Qubit Fluorometric Quantitation (Thermo Fisher Scientific) respectively. The DNA that accomplished the sequencing quality measures was stored at -80 °C before library preparation.

2.2.4 Library preparation and sequencing

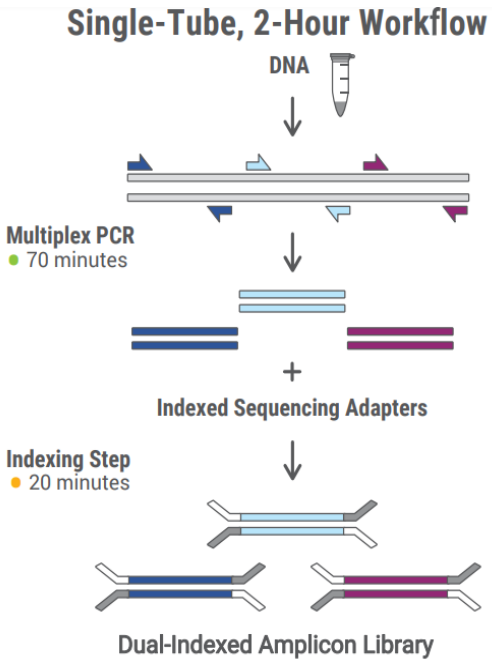
The libraries for sequencing were prepared in two steps: the DNA pool preparation and the library preparation with the commercial kit. The DNA pool preparation was crucial for minimizing the heterogeneity of the samples of each point in order to obtain a representative sample, and also to achieve the minimum input of DNA for further sequencing (Mayday et al., 2019). We started from a pool of DNA belonging to the replicas taken at each point of the three environments, for this the (3) replicas of each sample were unified forming a pool of 15 µl for each of the nine points per sampling, which was the input material for DNA sequencing.

In addition, a ZymoBIOMICS ® Microbial Community Standards mock community was used. The mock is composed of predetermined ratios of DNA from a mixture of bacterial and fungi species (Highlander, 2013), since the expected composition of the mock microbial community is known, it can be used to identify and understand problems and biases in the experimental methods, such as the biases provided in the DNA library preparation, amplification, sequencing and bioinformatic analysis (Karstens et al., 2018). In this case,

the mock community was used as a control in the assessment of the microbial soil communities from the three management zones.

Sequencing libraries were prepared from the soil DNA pools belonging to each georeferenced point as well as from the mock community. To that end, all regions (V1-V9) of the 16s rRNA gene for prokaryotes, and ITS1 and ITS2 for eukaryotes, were amplified using the SWIFT AMPLICON 16S+ITS PANEL kit (Swift Biosciences, USA) following the manufacturer's instructions (Fig. 2-5).

Figure 2-5: Workflow in a brief from “The Swift Amplicon 16S+ITS panel” library preparation kit.



Source: https://www.idtdna.com/

The input material was initially prepared by taking 10 microliters of each DNA pool at a concentration of 20 ng. Initially, a first multiplex PCR round was carried out to amplify each region of the marker genes were generated. The multiplex PCR protocol consisted of 30 seconds at 98°C for the initial DNA denaturation, followed by four cycles at 98ºC for 10 seconds, 63ºC for 5 minutes, 65ºC for 1 minute, and finally 98ºC for 10 minutes, then there were 22 cycles at 64ºC for 1 minute and finally the final extension at 65ºC for 1 minute.

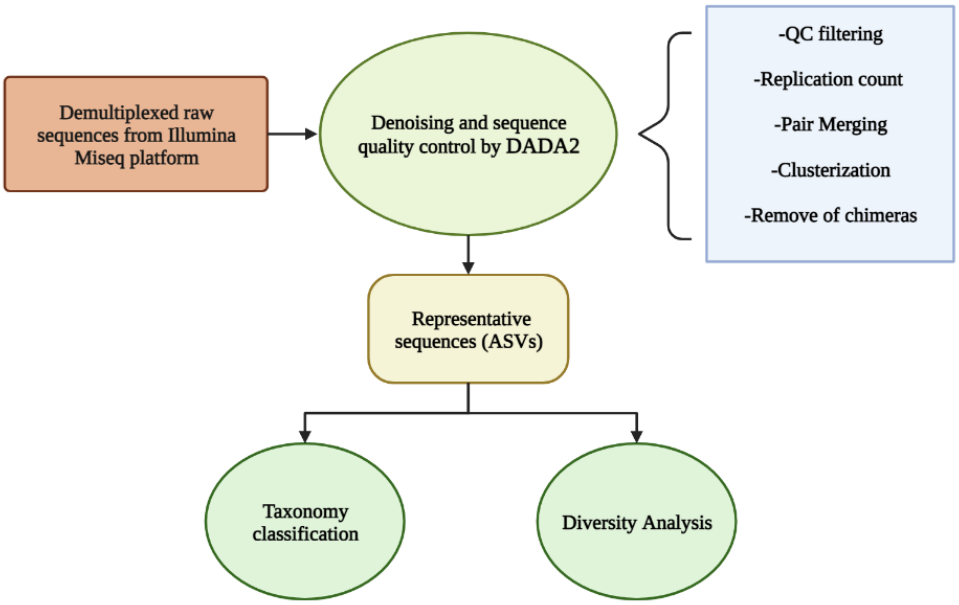
Following obtaining the amplicons of every marker region, a selection of the size, and cleaning process were carried out for all the samples. The cleaning process was done using, magnetic beads. Following this, the PCR round was applied with the aim to index each sample for their subsequent differentiation after sequencing. The second PCR round consisted of a 20-minute cycle at 37° then sample size selection and a cleaning step made to select amplicons. The quantification of the libraries was performed using Qubit Fluorometric Quantitation (Thermo Fisher Scientific). Finally, a new pool with the amplicon libraries was prepared by mixing all the libraries together which was at a final concentration of 4 nanomolar, followed by denaturing the libraries before loading them on the Illumina MiSeq platform for their respective sequencing.

Consequently, the sequencing of the libraries was performed by the MiSeq platform (Illumina, San Diego CA) using the Illumina MiSeq reagent Kit V2 (300 cycles), at the IBUN laboratories, obtaining paired reads (pair-end) of an insert size of 2x150 bp, that were subsequently demultiplexed for further bioinformatic analysis.

2.2.5 DNA sequences analysis

In order to obtain the taxonomic profile of the edaphic communities of the rice field, the sequences obtained with the Illumina MiSeq technology were analyzed using QIIME2 version 2021.8 (Bolyen et al., 2019) (Fig. 2-6). Initially, denoising, dereplication, clusterization, and filtering of chimeras from the raw pair-end sequences demultiplexed using the DADA2 (Divisive Amplicon Denoising Algorithm 2) plugin (Callahan et al., 2016).

Figure 2-6: Qiime2 pipeline overview for analyzing the sequences from the bulk soil samples for both samplings.



DADA2 uses a parametric model to infer true biological sequences from reads. The model relies on input read abundances and distances (less abundant reads only a few base differences away from a more abundant sequence are likely error-derived). In DADA2, reads are quality-filtered and also retain a summary of the quality scores associated with

each unique sequence; the end product is an amplicon sequence variant (ASV). Also, unlike all the other pipelines, DADA2 denoises the forward and the reverse reads independently. ASVs from the forward and reverse reads are only merged at the end of the workflow prior to the removal of chimeric ASVs. Chimeric sequences are identified if they can be exactly reconstructed by combining left and right segments from two more abundant "parent" sequences. (Hall & Beiko, 2018; Prodan et al., 2020).

Following the denoising step, the ASVs (Amplicon Sequence Variants) data found were used for the rarefaction, diversity, and taxonomic analysis. The last one with the reference database SILVA SSURef (version 138)(Callahan et al., 2016; Quast et al., 2013), for this, the classifier was initially trained in QIIME2 to perform the taxonomic determination of the ASVs, this was used for the taxonomic analysis of the ASVs to the classify-consensus-blast method (Bokulich et al., 2018).

2.3 Results

2.3.1 Bulk soil DNA extraction

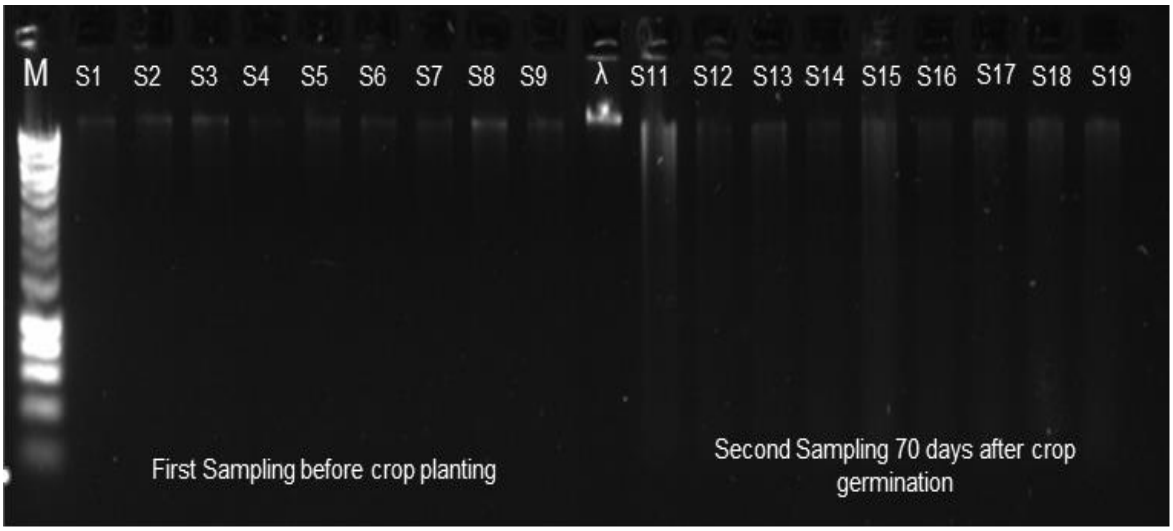
The DNA extraction measurements from the soil belonging to the nine points distributed in the three management zones for each sampling, revealed the following results as shown in table 2-2 and figure 2-7. The spectrophotometric relationships were measured in the absorbance ratio at 260 nm and 280 nm. In general, the A260/A280 ratio is used as an indicator of DNA purity since nucleic acids (DNA and RNA) absorb the most at a wavelength of 260 nm, while protein is best absorbed at 280 nm. The results obtained for this relationship were in a range of 1.62 to 2.18 for all samples for A260/A280nm, with a mean of 1.93 which was an indicator of low contamination caused by proteins on these samples.

Table 2-2: Bulk Soil DNA concentration (µg/µl) from both samplings in the nine georeferenced points.

Sampling	Management Zone	Sample	DNA Concentration (µg/µl)	A260/A280 Ratio
Bulk Soil First	Medium Yield / Yellow	S1	3,76	1,88
		S2	4,38	1,96
		S3	6,26	2,03
	Low Yield / Red	S4	2,66	1,97
		S5	5,41	1,79
		S6	8,64	2
	High Yield / Green	S7	3,43	2,04
		S8	6,41	2,1
		S9	5,74	1,87
Bulk Soil Second	Medium Yield / Yellow	S11	24,05	1,75
		S12	13,11	1,94
		S13	10,75	1,78
	Low Yield / Red	S14	10,77	1,62
		S15	26,2	1,99
		S16	8,15	2,18
	High Yield / Green	S17	13,55	1,99
		S18	14,05	2,16
		S19	15,85	1,71

An average DNA yield concentration of 10.17 ng/µl was obtained for all samples. The DNA yields varied between samplings, generally, the DNA concentration in the samples from the first sampling before the crop planting ranged from 2.66 to 8.64 ng/µl of DNA with an average of 5.18 ng/µl of genomic DNA. On the other hand, DNA concentration obtained in the sampling performed 70 days after the germination of the crop was higher, with a range of 8.15 to 26.2 ng/µl of DNA and an average of 15.16 ng/µl of DNA between all samples.

Figure 2-7: Metagenomic DNA extracted from Bulk soil during the 2 samplings. M = Hyperladder IV molecular weight marker, λ = phage Lambda genomic DNA.



2.3.2 Sequencing data quality

The cluster density on the flow cell obtained at the end of the sequencing was 600 K/mm2, this metric showed that there was an underclustering of the samples, which has an impact on sequencing performance in terms of data quality and total data output. In reference to the quality of the raw reads. An average of 88% of bases were found distributed in the quality score Q30 (sequencing error rate less than 0.1%).

For the first sampling, a total of 1.473.233 paired-end reads have obtained the mean of the number of reads per sample was 163.692. The number of reads in the second sampling was higher consisting of 1.683.372 paired-end reads which presented a mean of 153.033 reads per sample. Likewise, the total number of reads obtained in the three management zones presented differences between the first and second sampling, these are observed in table 2-3.

The differences found in the number of reads could be explained according to the DNA extraction method, library preparation, sequencing, and the changes by which the soil samples from the rice field underwent; different values were observed between samplings, which were taken ten days before crop planting and seventy days later of the crop germination, could be related to the effect of the growth of the rice plants, in addition to external factors such as the management of the crop in terms of chemical fertilization. Furthermore, it is important to remark, that even though the results showed good quality, the underclusterization of the sequencing affected the output of the data in the samples, since a low number of reads was found overall.

Table 2-3: Sequencing data (number of reads) of bulk soil samples by management zone for both samplings.

Management Zone	First Sampling		Second Sampling	
	Total reads	Reads mean per sample	Total reads	Reads mean per sample
Low yield	423.286	141.095	329.555	109.851
Medium yield	600.462	200.154	399.452	133.150
High yield	449.485	149.828	612.710	182.924

In reference to the filtering process of the reads, the denoising was carried out using the dada2 plugin, where initially forward and reverse reads were truncated at position 130 bp, and singletons were discarded as part of the data processing within DADA2. After quality filtering, a total of 1,457,847 and 1,328,999 reads were obtained in the first and second samplings respectively. This represents only 1% of the total reads were filtered by quality in the bulk samples from the sampling before the crop planting, compared to the second

sampling (70 days after the crop germination) where 21% of the total sequences were filtered.

Despite the fact that the filtered reads presented acceptable quality parameters, the denoising and chimeras removal process excluded an average of 88% of the dataset. Only 217,285 reads of the first sampling passed this pipeline step, this value represents a loss of 85.25% of the reads obtained in the 9 points from this sampling. Likewise, only 122,479 reads of the sampling two (bulk soil samples) passed the removal of chimera's process, which represents a loss of 92.72% of this dataset. The differences at the management zone level between the two samplings can be seen in table 2-4 after denoising them.

The high percentage of chimeras obtained might be related to biases specifically provided in the library preparation step, since other studies have shown how various biases in this step, especially in PCR amplification, directly affect the formation of chimeras. Chimeric reads can significantly confound downstream analyses, leading to overestimating the actual community diversity (Brooks et al., 2015; Haas et al., 2011; Sze & Schloss, 2019). Undetected chimeric sequences, caused by the hybridization of DNA fragments from different species, also reduce the reliability of 16S rRNA and ITS sequence-based phylogenetic composition of microbial communities (Auer et al., 2017), thus it is crucial not to include these reads to obtain a more accurate analysis of the composition of the microbial community in the samples.

Table 2-4: Number of reads after chimeras removal performed by DADA2 pipeline in the three management zones of the rice field for both samplings.

Management Zone	First Sampling		Second Sampling	
	Total reads	Reads mean per sample	Total reads	Reads mean per sample
Low yield	41.965	13.988	37.310	12.437
Medium yield	81.863	27.288	31.228	10.409
High yield	93.457	31.152	53.941	17.980

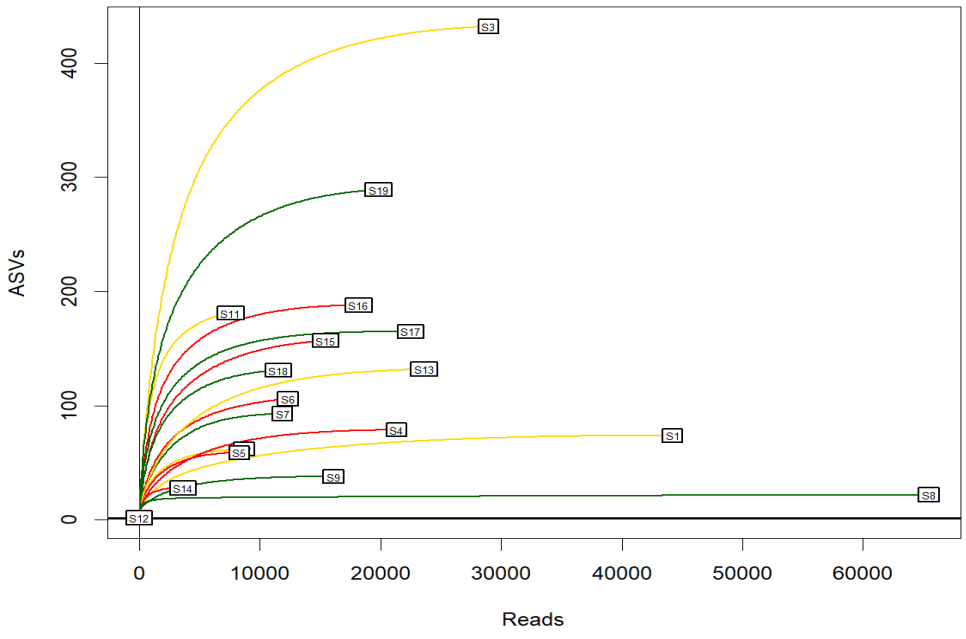
Regarding the results obtained with the mock community used as a control, a total of 170.827 paired-end reads were generated. Quality filtering was applied, resulting in a total of 170.065 reads which represents 99.56% of reads with high quality, these results could be related to the Q-scores above 88% obtained on average. Subsequently, denoising, dereplication, and chimera filtering only 7.598 reads remain, which represents a loss of 95.54% of the total reads.

Nonetheless, a trend was observed after filtering the reads associated with chimeras, since all of the results showed a loss of most of the data generated in the sequencing after this step. Furthermore, it is important to highlight the effect of the underclusterization found after the sequencing of the libraries. Even so, due to the massive loss of data, it was decided to continue with the analysis of the reads that passed the DADA2 workflow, to track any glimpse of the composition of the communities associated with the bulk soil from the three management zones, during the two samplings.

2.3.3 Soil microbial community diversity and composition

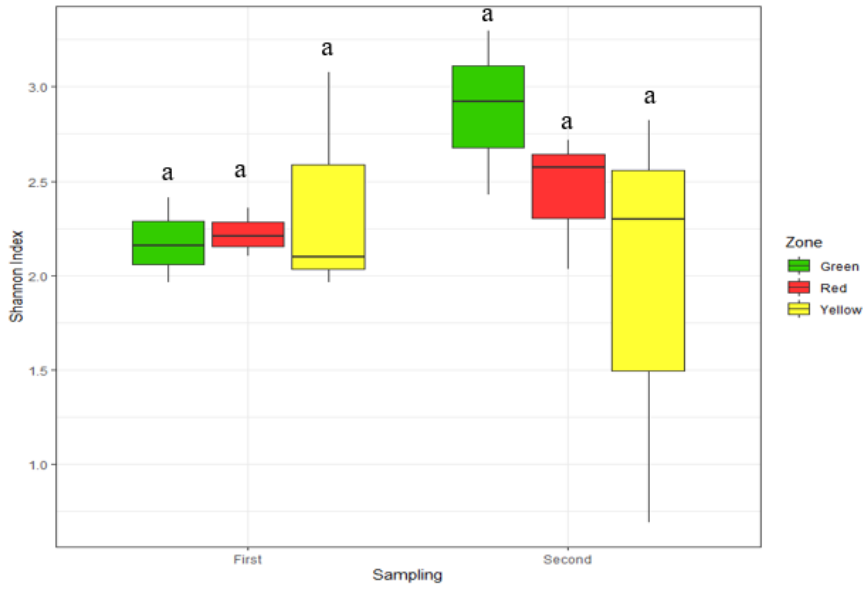
Once the reads with optimal quality were obtained, a total of 1,714 different amplicon sequence variants (ASVs) were identified throughout all samples. As we expected, The rarefaction analysis showed that the sequencing effort was not acceptable due to the loss of data specifically caused by the content of chimeras. In addition, it was observed that the samples were not comparable, besides, was not possible to reach the asymptote of the curve in most of the samples (Fig 2-8), indicating that the actual diversity of the bulk soil from these samples is far greater and a better sampling effort is required to be able to capture the diversity without any biases derived from the library preparation and sequencing methods applied.

Figure 2-8: Rarefaction curve for each sample for both bulk soil samplings, the colors of the curves mean the management zone: red for low yield management zone, yellow for medium yield management zone, and finally green for high yield management zone. Samples from the first sampling before the crop planting(S1-S9) and samples from the second sampling 70 days after the crop germination (S11-S19).



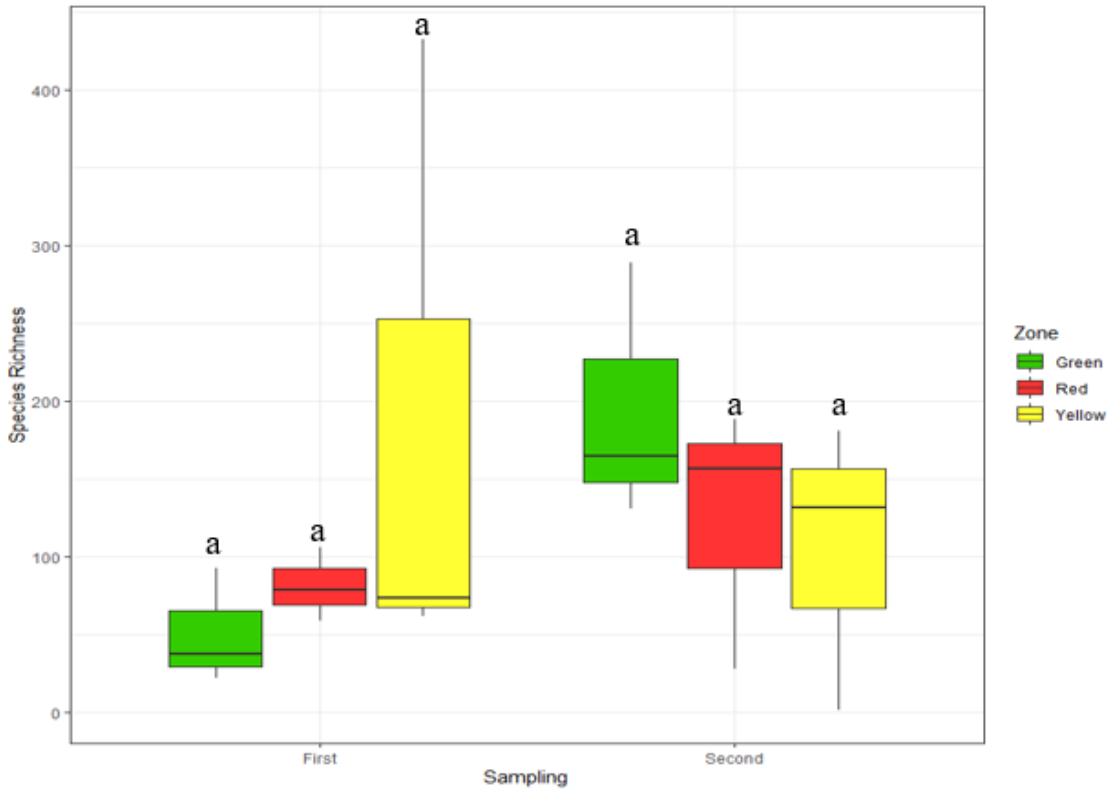
Although the results obtained after denoising was not reliable, alpha diversity analyzes were performed to have a small insight into the structure of the community that was captured. For this, according to the Shannon diversity index, the α-diversity was low in both samplings in all the samples analyzed, with an average value of 2,34 which compared to other similar studies in soils associated with rice crops the index is considerably low. Shannon diversity index obtained did not show significant differences according to the non-parametric Kruskal Wallis test among the three management zones. In the first sampling (p-value=0.8), likewise the second sampling (p-value= 0.339), and also between samplings (p-value = 0.145). However, a lower diversity was obtained in the samples belonging to the first sampling compared to the second sampling (Fig. 2-9).

Figure 2-9: Boxplot of the Shannon indexes of the different management zones during both samplings. Different letters indicate significant differences among samplings and management zones by one-way ANOVAs (Tukey, P < 0.05).



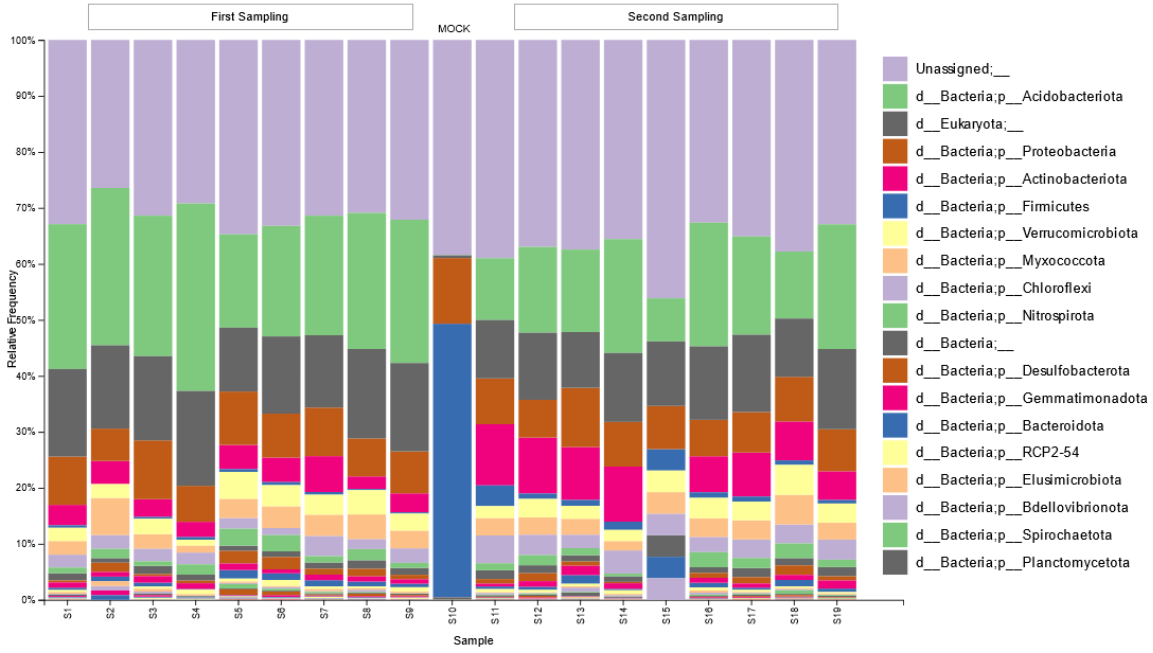
In the same way, a similar behavior was observed in the observed richness of ASVs found, due to during the first sampling a lower richness with an average of 107 ASVs was found compared to the second sampling with an average of 141 ASVs (Fig. 2-10), about the observed richness there were no significant differences between the zones in the first sampling (p-value = 0.4298) measured with Kruskal Wallis test, likewise for the second sampling the sampling (p-value= 0.458) measured with ANOVA, also there were no significant differences between both samplings in the richness (p-value = 0.1451). It is also important to highlight that these results obtained do not represent the real microbial community of the bulk soil samples.

Figure 2-10: Boxplot of the richness of ASVs of the different management zones during both samplings. Different letters indicate significant differences among samplings and management zones by one-way ANOVAs (Tukey, P < 0.05).



Finally, according to the taxonomy results, the phylum Acidobacteria was the most abundant in all samples which accounted for an average of 19.33% of the total sequences, followed by Proteobacteria, Actinobacteria, Myxococcota, and Chloroflexi, which accounted for 8.16%, 6.41%, 4.63%, and 3.78% respectively (Fig. 2-11).

Figure 2-11: Relative abundance at the phylum level of the samples from the three different management zones during both samplings.

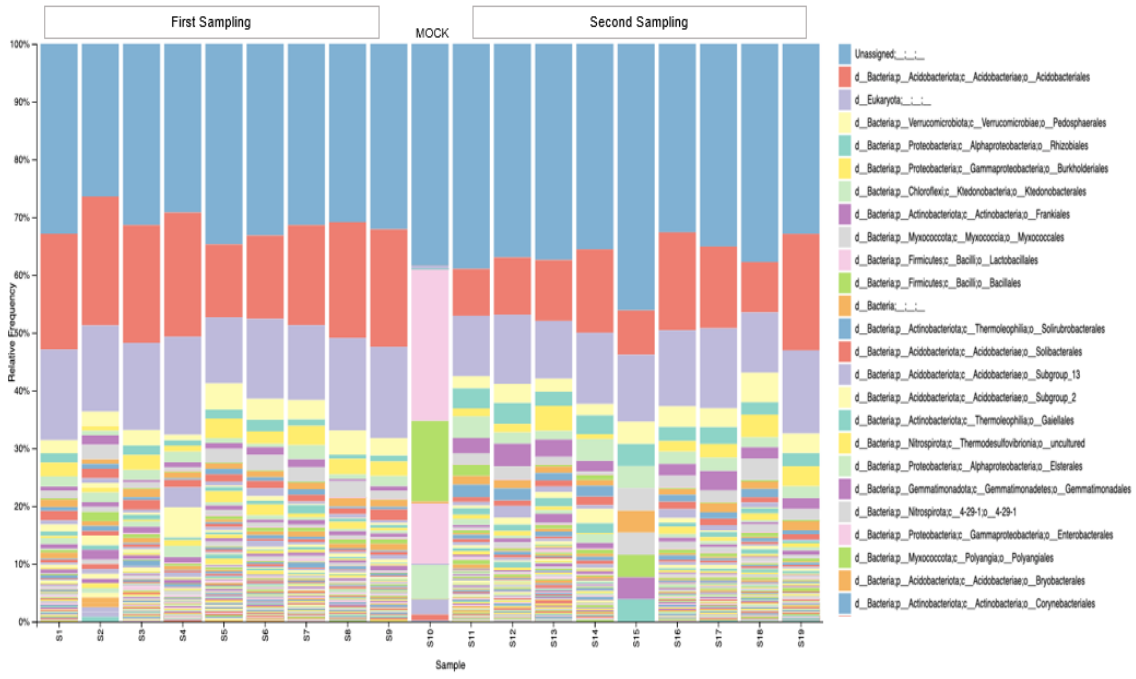


Moreover, it was observed that up to 40% of the relative abundance of ASVs in all the samples could not be classified in any taxon at the order level, which would be directly related to the loss of most of the reads from the sequences during the denoising due to the removal of chimeras. However, at the order level, the most dominant taxa in all the samples were Acidobacteriales with a mean relative abundance of 14.82%, followed by

Pedosphaerales with 3.24% relative abundance, and finally the order Rhizobiales with a mean relative abundance of 2.78% (Fig. 2-12).

Even so, it is important to remark that the taxonomic results for the mock community can capture a part of the composition of this community. Thereafter it was observed that the expected taxa were obtained but their relative abundances were not the same among themselves as expected, considering this community is characterized by having similar relative abundances in the 8 species that the community contains from the orders Pseudomonales, Enterobacteriales, Lactobacillales, Bacillales, and Saccharomycetales.

Figure 2-12: Relative abundance at order level of the samples from the three different management zones during both samplings, and the mock community from ZymoBIOMICS ® Microbial Community Standards.



2.4 Conclusions

In conclusion, the results obtained are not reliable, therefore it is not worth discussing and delving into these results, since the bias contributed by a large number of chimeras affected the result of the composition and structure of the microbial communities from these bulk soil samples, therefore the low number of reads, that were achieved at the end of the denoising performed with DADA2 affected the results negatively. However, this methodological practice was helpful for the Institute of Biotechnology of the National University of Colombia and for the bioprocess research group, since it contributed to the establishment of areas of opportunity to improve the development of the applied methodologies.

According to the results obtained, the decision was made to continue with a different strategy, which consisted of sequencing the complete metagenome (shotgun metagenomics), of all the samples in order to have reliable results and meet the previously proposed objectives. This new methodological approach will be discussed in the next chapter.

2.5 References

Alteio, L. V., Séneca, J., Canarini, A., Angel, R., Jansa, J., Guseva, K., Kaiser, C., Richter, A., & Schmidt, H. (2021). A critical perspective on interpreting amplicon sequencing data in soil ecological research. *Soil Biology and Biochemistry*, *160*, 108357. https://doi.org/10.1016/j.soilbio.2021.108357

Auer, L., Mariadassou, M., O'Donohue, M., Klopp, C., & Hernandez-Raquet, G. (2017). Analysis of large 16S rRNA Illumina data sets: Impact of singleton read filtering on microbial community description. Molecular Ecology Resources, 17(6), e122–e132. https://doi.org/10.1111/1755-0998.12700

Banerjee, S., & van der Heijden, M. G. A. (2022). Soil microbiomes and one health. Nature Reviews Microbiology, 1–15. https://doi.org/10.1038/s41579-022-00779-w

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome, 6(1), 90. https://doi.org/10.1186/s40168-018-0470-z

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive,

scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Brooks, J. P., Edwards, D. J., Harwich, M. D., Rivera, M. C., Fettweis, J. M., Serrano, M. G., Reris, R. A., Sheth, N. U., Huang, B., Girerd, P., Strauss, J. F., Jefferson, K. K., Buck, G. A., & Vaginal Microbiome Consortium (additional members). (2015). The truth about metagenomics: Quantifying and counteracting bias in 16S rRNA studies. BMC Microbiology, 15(1), 66. https://doi.org/10.1186/s12866-015-0351-6

Bruno, F., Marinella, M., & Santamaria, M. (2015). e-DNA Meta-Barcoding: From NGS Raw Data to Taxonomic Profiling. In E. Picardi (Ed.), RNA Bioinformatics (pp. 257–278). Springer. https://doi.org/10.1007/978-1-4939-2291-8_16

Bukin, Y. S., Galachyants, Y. P., Morozov, I. V., Bukin, S. V., Zakharenko, A. S., & Zemskaya, T. I. (2019). The effect of 16S rRNA region choice on bacterial community metabarcoding results. Scientific Data, 6(1), Article 1. https://doi.org/10.1038/sdata.2019.7

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High resolution sample inference from Illumina amplicon data. Nature Methods, 13(7), 581–583. https://doi.org/10.1038/nmeth.3869

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(8), 2405–2429. https://doi.org/10.1007/s10531-019-01760-5

Francioli, D., Lentendu, G., Lewin, S., & Kolb, S. (2021). DNA Metabarcoding for the

Characterization of Terrestrial Microbiota—Pitfalls and Solutions. Microorganisms, 9(2), Article 2. https://doi.org/10.3390/microorganisms9020361

Gołębiewski, M., & Tretyn, A. (2020). Generating amplicon reads for microbial community assessment with next-generation sequencing. Journal of Applied Microbiology, 128(2), 330–354. https://doi.org/10.1111/jam.14380

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methé, B., DeSantis, T. Z., Human Microbiome Consortium, Petrosino, J. F., Knight, R., & Birren, B. W. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research, 21(3), 494–504. https://doi.org/10.1101/gr.112730.110

Hall, M., & Beiko, R. G. (2018). 16S rRNA Gene Analysis with QIIME2. In R. G. Beiko, W. Hsiao, & J. Parkinson (Eds.), Microbiome Analysis: Methods and Protocols (pp. 113–129). Springer. https://doi.org/10.1007/978-1-4939-8728-3_8

Highlander, S. (2013). Mock Community Analysis. In K. E. Nelson (Ed.), Encyclopedia of Metagenomics (pp. 1–7). Springer. https://doi.org/10.1007/978-1-4614-6418-1_54-1

Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W., Wolfe, A., Braun, J., & Mcweeney, S. (2018). Controlling for contaminants in low biomass 16S rRNA gene sequencing experiments. https://doi.org/10.1101/329854

Knauth, S., Schmidt, H., & Tippkötter, R. (2013). Comparison of commercial kits for the extraction of DNA from paddy soils. Letters in Applied Microbiology, 56(3), 222–228. https://doi.org/10.1111/lam.12038

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.-I., McDonald, D., Melnik, A. V., Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D., Thompson, L. R., Tripathi, A., Xu, Z. Z., … Dorrestein, P. C. (2018). Best practices for analysing microbiomes. Nature Reviews Microbiology, 16(7), Article 7. https://doi.org/10.1038/s41579-018-0029-9

Li, S., Deng, Y., Wang, Z., Zhang, Z., Kong, X., Zhou, W., Yi, Y., & Qu, Y. (2020). Exploring the accuracy of amplicon-based internal transcribed spacer markers for a fungal community. Molecular Ecology Resources, 20(1), 170–184. https://doi.org/10.1111/1755-0998.13097

Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. Ecological Entomology, 45(3), 373–385. https://doi.org/10.1111/een.12831

Liu, Qin, Y., Chen, T., Lu, M., Qian, X., Guo, X., & Bai, Y. (2021). A practical guide to amplicon and metagenomic analysis of microbiome data. Protein & Cell, 12(5), 315–330. https://doi.org/10.1007/s13238-020-00724-8

Mayday, M., Khan, L., Chow, E. D., Zinter, M. S., & DeRisi, J. L. (2019). High-Throughput Library Pooling for NGS. https://www.protocols.io/view/high-throughput-library-pooling-for-ngs-tcdeis6

Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., & Tedersoo, L. (2019). Mycobiome diversity: High-throughput sequencing and identification of fungi. Nature Reviews Microbiology, 17(2), Article 2. https://doi.org/10.1038/s41579-018-0116-y

Pecundo, M. H., Chang, A. C. G., Chen, T., dela Cruz, T. E. E., Ren, H., & Li, N. (2021). Full-Length 16S rRNA and ITS Gene Sequencing Revealed Rich Microbial Flora in Roots of Cycas spp. In China. Evolutionary Bioinformatics, 17, 1176934321989713. https://doi.org/10.1177/1176934321989713

Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The Madness of Microbiome: Attempting To Find Consensus “Best Practice” for 16S Microbiome Studies. Applied and Environmental Microbiology, 84(7), e02627-17. https://doi.org/10.1128/AEM.02627-17

Prasad, S., Malav, L. C., Choudhary, J., Kannojiya, S., Kundu, M., Kumar, S., & Yadav, A. N. (2021). Soil Microbiomes for Healthy Nutrient Recycling. In A. N. Yadav, J. Singh, C. Singh, & N. Yadav (Eds.), Current Trends in Microbial Biotechnology for Sustainable Agriculture (pp. 1–21). Springer. https://doi.org/10.1007/978-981-15-6949-4_1

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A. H., Nieuwdorp, M., & Levin, E. (2020). Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLOS ONE, 15(1), e0227434. https://doi.org/10.1371/journal.pone.0227434

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Research, 41(Database issue), D590-596. https://doi.org/10.1093/nar/gks1219

Rajeev, M., Sushmitha, T. J., Toleti, S. R., & Pandian, S. K. (2020). Sediment-associated bacterial community and predictive functionalities are influenced by

choice of 16S ribosomal RNA hypervariable region(s): An amplicon-based diversity study. Genomics, 112(6), 4968–4979. https://doi.org/10.1016/j.ygeno.2020.09.006

Schloss, P. D. (2020). Reintroducing mothur: 10 Years Later. Applied and Environmental Microbiology, 86(2), e02343-19. https://doi.org/10.1128/AEM.02343-19

Semenov, M. V. (2021). Metabarcoding and Metagenomics in Soil Ecology Research: Achievements, Challenges, and Prospects. Biology Bulletin Reviews, 11(1), 40–53. https://doi.org/10.1134/S2079086421010084

Starke, R., Pylro, V. S., & Morais, D. K. (2021). 16S rRNA Gene Copy Number Normalization Does Not Provide More Reliable Conclusions in Metataxonomic Surveys. Microbial Ecology, 81(2), 535–539. https://doi.org/10.1007/s00248-020-01586-7

Sze, M. A., & Schloss, P. D. (2019). The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data. MSphere, 4(3), e00163-19. https://doi.org/10.1128/mSphere.00163-19

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends in Genetics, 30(9), 418–426. https://doi.org/10.1016/j.tig.2014.07.001

Yang, R.-H., Su, J.-H., Shang, J.-J., Wu, Y.-Y., Li, Y., Bao, D.-P., & Yao, Y.-J. (2018). Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. PLOS ONE, 13(10), 1–17. https://doi.org/10.1371/journal.pone.0206428

Chapter 3: Rice soil microbiomes using Shotgun Metagenomics

3.1 Introduction

Soil microbiomes related to rice crops are reported to be an essential factor in maintaining soil integrity and quality. Rice soil microbiomes influence growth, nutrient uptake, and pathogen resistance, improving the yields of the crops. (Ding et al., 2019; Dong et al., 2022; Smith & Gallaher, 2019; Zhong et al., 2020). Moreover, it has been established that microbial diversity influences soil fertility, thanks to the various functions and ecosystem services, such as supporting services, in which soil microbial communities are involved in soil formation processes, nutrient cycling, water cycling, and primary production (Hartmann & Six, 2022; Kendzior et al., 2022; Saccá et al., 2017). Consequently, soil fertility is the key factor for sustaining crop production. However, in modern agriculture, it is maintained by external fertilizer inputs, which has a variety of detrimental effects on the environment and substantially raise production costs (Mustafa et al., 2023; Walling & Vaneeckhaute, 2020; X. Wang et al., 2020). It is important to promote agronomic practices adapted to the needs of each agricultural zone, and focused on the soil microbial biodiversity maintained These practices could have a positive impact on rice crop production, reducing costs, and mitigating the negative impact on the environment (Amanullah & Khalid, 2020; Caulfield et al., 2020; Farooq et al., 2023; FEDEARROZ, 2018). Understanding the connection between the rice crop soil microbial diversity, and their interaction with the biotic and abiotic factors associated with the soils is crucial to generate strategies that allow the achievement of new agronomic management strategies.

Multi-omics methods have been used to disentangle the contribution of the microbial composition in edaphic environments associated with crucial crops (Jansson, 2013; Rolf, 2005; Sabale et al., 2019). Instead of focusing on a single marker gene, shotgun metagenomic sequencing offers a powerful and broad option to examine a microbial community. This method briefly consists in obtaining all genomic DNA contained in an environmental sample followed by randomly shearing it into smaller fragments, which are sequenced using second or third-generation sequencing platforms to generate a signal of the composition of the community associated with the sample (Garlapati et al., 2019; Yadav, 2021). Metagenomic sequencing has facilitated the ability to predict taxonomic and functional gene compositions in soil samples, providing new access to the microbial world knowledge (Biswas & Sarkar, 2018; Quince et al., 2017; Rolf, 2005; Sabale et al., 2019).

The analysis of metagenomics shot-gun data can employ a variety of bioinformatics analysis pipelines, which commonly could have two perspectives: i) reads assembly and ii) assembly-free approaches (Pérez-Cobas et al., 2020). The assembly depended approach includes the quality control of the raw reads, followed by metagenomes assembly through concatenated reads into longer contiguous sequences (contigs), and contigs binning, which entails a grouping of contigs into bins that represent taxonomic groups, and finally, the binned contigs are designated as metagenome-assembled genomes (MAGs) (Breitwieser et al., 2017; Chandrasiri et al., 2022; Sharpton, 2014; Yang et al., 2021) and are the definitive resource for the taxonomical classification and annotation of the representative function associated to the sample. Nevertheless, this kind of approach has challenges, such as low sequencing coverage and depth brought on by variations in species abundance, and the characteristics of metagenomic samples. As a result, fragmented and error-prone metagenomic assemblies tend to be generated (Lapidus & Korobeynikov, 2021). Similarly, these fractionated metagenomes are a result of the limited ability of sequencing to capture complex repeat genomic regions accurately, and the difficulties encountered in reconstructing the sequence of phylogenetically close organisms (Haryono et al., 2022).

The free assembly strategies utilize the information contained into the raw or preprocessed reads to predict both microbial composition and functions associated to an environmental sample. The prediction of the community traits is performed by comparing them with reference genomes available in public databases. The assignment of the high-quality reads could be made using marker genes or signature-based profiling, genome alignments, *k-mer* mapping or protein alignment (Quince et al., 2017). The approach can speed up computation and mitigate assembly problems like fragmented genomes, along with profiling of low-abundance organisms that cannot be assembled de novo (Quince et al., 2017). The online server MG-RAST was chosen for this investigation, due is a platform that offers reliable results and robust analysis in taxonomic classification and annotation. MG-RAST's advantage is handling raw reads directly from a sequencing service providing consistent analyses as deep and unbiased as possible at an affordable computational cost. (Meyer et al., 2019).

In MG-RAST, initially, a pre-processing of the sequences occurs to trim low-quality regions from FASTQ data using SolexaQA (Cox et al., 2010). Then the pipeline includes a dereplication step, where a simple k-mer approach is used, this step is required in order to remove Artificial Duplicate Reads. Thereafter, the rRNA-similar reads are then clustered at 97% identity, and the longest sequence is picked as the cluster representative, the pipeline presents reconstructions of the species content of the sample based on the similarity results. The species composition of the sample is reconstructed by looking at the phylogenetic origin of the database sequences hit by the similarity searches (Meyer et al., 2019; Wilke et al., 2016).

Lastly, the annotation was made using the BLAST-like alignment tool (BLAT algorithms) against M5NR database which offers a non-redundant combination of various databases: GenBank,(Benson et al., 2013), SEED (Overbeek et al., 2014), IMG (Markowitz et al. 2008), UniProt (Magrane & UniProt Consortium, 2011), KEGG (Kanehisa & Goto, 2000), and eggNOG (Jensen et al., 2008), in order to assign genes to a clustering based subsystems

as functional groups. MG-RAST builds clusters of proteins at the 90% identity level, these clusters greatly reduce the computational burden of comparing all pairs of short reads, while clustering at 90% identity preserves sufficient biological signals. (Keegan et al., 2016). MG-RAST has become a widely used platform for compiling information, on 2019 had processed >295 000 data sets from 23 000 researchers around the world (Meyer et al., 2019), leading to an increase in the collection of metagenome data, in this context related to agricultural soils.

In light of the aforementioned, it is critical to examine the microbiomes associated with rice crops. Since these types of studies are novel in the context of rice production in Colombia. Therefore, this study hypothesized that the composition of the edaphic microbiomes is different within three different zones of a rice field in Colombia, evaluated in two different moments of a productive cycle. Here we investigate the edaphic microbiomes of a commercial rice field and its relationship with the physical-chemical properties of the soil, using a shotgun metagenomic approach.

3.2 Materials and Methods

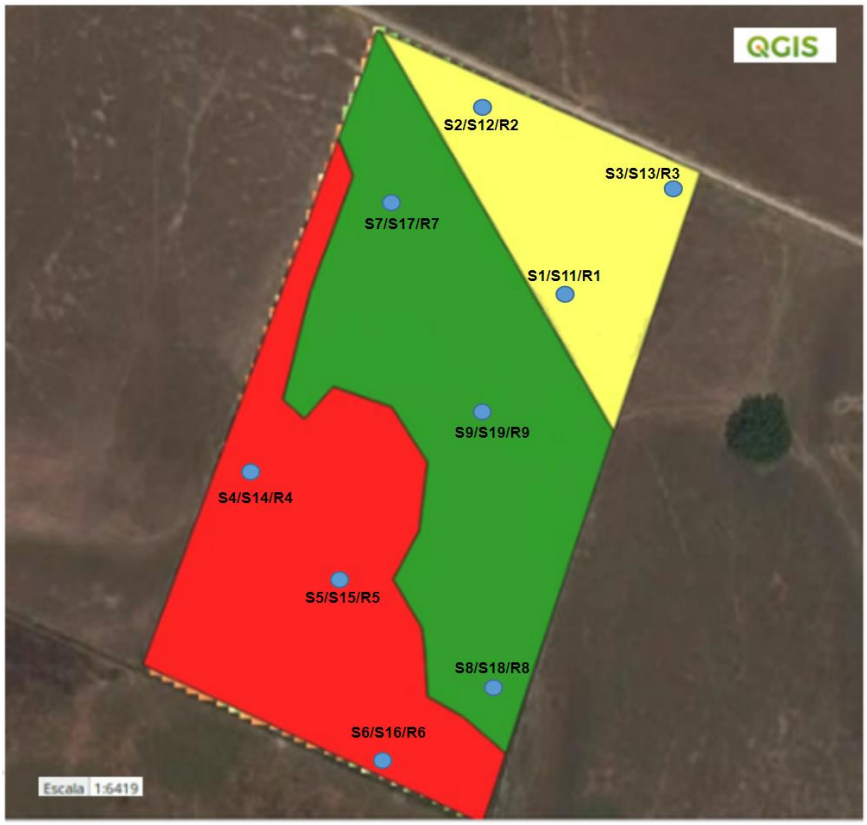
3.2.1 Area of study

The analysis of the bulk and rhizosphere soil samples was carried out in a rice field of 33 hectares, during one production cycle in 2021, in the department of Meta, Colombia. In this study, three different management zones were previously defined according to the historic yield data of rice crops (Fig. 3-1). As mentioned in the last chapter, two different samplings over time were carried out, the first one was before the planting of the crop, and the second seventy days after rice plants germination. In each management zone bulk and rhizosphere soil samples were taken to further DNA extraction and microbiome analysis.

It is important to emphasize that the crop was subjected to conventional agronomic management practices that were determined by FEDEARROZ agronomists, according to previous information from soil chemical characterization of previous cycles in this 33-hectare field in Meta, Colombia. These management was performed along with a non-differential fractional inorganic fertilization approach across three distinct zones. The adopted approach encompassed the precise application of fertilizers comprising essential macronutrients and micronutrients. The fertilization process consisted of three distinct applications, strategically timed between the crop's germination stage and 60 days thereafter, aligning with the reproductive phenological phase. These applications were designed to incorporate dosage levels tailored to fulfill the crop's nutritional requirements at each stage of development. By synchronizing nutrient provision with the crop's specific demands, this optimizes nutrient availability while concurrently mitigating losses due to leaching or runoff. Moreover, this method enhances the efficiency of nutrient uptake and utilization, enabling the crop to extract nutrients more effectively when they are most needed. Consequently, this integrated approach ensures the crop receives essential nutrients in a manner that maximizes its growth potential, minimizing environmental impact and contributing to sustainable agricultural practices (FEDEARROZ-FNA, 2021).

Moreover, the field had an irrigation system consisting of intermittent flooding, during different moments of the crop, unlike continuous flooding, where the fields are kept submerged throughout the growing season, intermittent flooding involves alternating periods of flooded and non-flooded conditions. The water level is periodically raised and lowered, allowing the rice plants to experience both submerged and aerobic conditions. This technique helps in reducing water usage, controlling weeds, and improving nutrient availability to the crop. In this case contrary to the first sampling, the field was under flooded conditions at 70 days following the germination of the plants (second sampling).

Figure 3-1: Map of the three management zones and the nine georeferenced points for each sampling, (S1-S9 and S11-S19 for the bulk soil and R1-R9 for rhizosphere soil samples), in red color the management zones of low productivity observed, in green color the management zones of high productivity are observed, and in yellow color, the management zones of medium productivity are observed.



Source:FEDEARROZ-FNA, 2021.

3.2.2 Sampling experimental design

A total of 27 bulk soil samples (3 management zones x 3 composite samples x 3 replicates) were obtained for each sampling, and 27 samples were obtained for the rhizosphere soil sampling from plants (3 management zones x 3 composite samples x 3 replicates). In the

three georeferenced points of each environment, samples composed of rhizosphere and bulk soil were taken under a completely randomized experimental design. Bulk soil samples were taken as described in the previous chapter.

Rhizosphere samples were collected by selecting three randomly fresh rice plants at the points of the three management zones. Each plant was shaken to remove all soil particles not considered rhizosphere soil, then the soil associated with the roots of the three plants, belonging to each point, were pooled and stored in plastic bags for later transport to the laboratory under cold chain conditions.

The soil remaining on the roots was collected using the method proposed and adapted by various authors (Barillot et al., 2013; Chang et al., 2021; He et al., 2017; Otero-Jiménez et al., 2021). Where 10 g of fresh roots from the pooled sample of each point with rhizosphere soil were placed into 50 ml sterile falcon, then 30 ml of 1x phosphate buffer was added with Tween 30 2 % (v/v). Each tube was shaken for 10 minutes in a vortex at maximum speed. The samples were centrifuged for 5 min at 7250 g and the pellets were conserved for a second washing with the phosphate buffer followed by another centrifugation. Lastly, the remaining rhizosphere soil was collected and stored at -80°C to continue with the DNA extraction.

3.2.2 Physicochemical properties

Soils samples were characterized for soil organic matter (SOM), soil organic carbon (SOC), the content of macronutrients (P, K), and micronutrients (Cu, Zn, Mn, Mg, Al), pH, electrical conductivity (EC), and the effective cation exchange capacity (ECEC) for both samplings in each georeferenced point, the soil samples were sent to be characterized at the facility of AGROSAVIA (“Corporación Colombiana de Investigación Agropecuaria Agrosavia”) in Bogota, Colombia.

3.2.3 DNA extraction

The methodology related to DNA extraction and the analysis of quality and integrity of the extraction of the nucleic acids, are described in detail in the methods of chapter 2.

3.2.4 Library preparation and sequencing

The genomic DNA obtained in triplicate from each point of the three management zones, were pooled in equal concentration (10 ng/ µL) for each sample and sent to Novogene Corporation Inc. facility in Sacramento, CA (EEUU), for further library preparation and metagenome sequencing in the Illumina NovaSeq™ 6000 NGS platform.

The library preparation consisted of randomly shearing into smaller fragments from the pooled DNA of the bulk and rhizosphere samples. The obtained fragments were end-repaired, A-tailed, and ligated with Illumina adapters. The fragments with the adapters were PCR amplified followed by a size selection and purification. Finally, the DNA fragments were sequenced on NovaSeq™ 6000 Sequencing System (Illumina, Inc, 2022)

3.2.5 DNA sequences analysis

Raw reads were analyzed using the MG-RAST server (Metagenomics Rapid Annotation using Subsystem Technology) an open-source service built for taxonomic assignment and function analyses of metagenomes, the server compares DNA sequences against a large group of protein and nucleotide databases for the assignment of metagenomic sequences to their respective taxonomic and functional groups (Keegan et al., 2016; Meyer et al., 2008; Wilke et al., 2016).

Initially, in the research group, a comparative methodology was carried out to define the following workflow with the sequences obtained in MG-RAST. For this, a subset of the total sequenced data was taken, and three different strategies were applied for its analysis: (i) raw reads were subjected to previous quality control and pre-processing, (ii) raw reads of the subset were assembled, and (iii) raw reads were uploaded to MG-RAST, with no prior processing. It was observed that the results obtained with the three different approaches, did not present significant differences (results not shown). Therefore, it was decided to continue the analysis with the reads without pre-processing. An average of 6.75 Gb of high-quality sequence data per sample was uploaded to the server for further analysis.

The quality control of the uploaded sequences was made through a series of automated pipeline steps to remove artificial sequences, ambiguous bases, and followed by a read length filtering (Keegan et al., 2016). Additionally, to suppress the influence of the experimental error, the data normalization option was selected on the server, after quality control analysis, the annotation of the sequences was carried out, for taxonomic assignment the MG-RAST server identifies candidate RNA genes by comparing the sequence against rRNA databases.

And for the taxonomic and assignment and function annotation, the match threshold was set at an expected value of $< 1 \times 10^{-5,}$ e-value cut off, and 60% minimum sequence similarity. In order to explore the general functions associated with the metagenomes rice field, a functional analysis was carried out based on the subsystem classification provided by SEED protein database (Overbeek et al., 2014).

3.2.6 Statistical Analysis

The statistical analysis of soil properties, macronutrients, and micronutrients was performed with Analysis of Variance (ANOVA) and Kruskal-Wallis test in order to compare the effects of these soil properties among the management zones and between samplings. Tukey (HSD) for ANOVA and Wilcoxon rank-sum test for Kruskal-Wallis, were used to assess these differences, with a defined p-value ≤ 0.05, using R ® 4.2.2 language (R Core Team, 2013).

To compare the structure of soil microbial communities, the MG-RAST dataset was extracted. For each sample alpha diversity and richness were estimated through Shannon, Inv Simpson, Chao-1, and ACE indexes using the *vegan* package (Oksanen et al., 2019). ANOVA analysis was performed to determine the differences among the α-diversity results. Principal coordinate analysis (PCoA) was performed to determine the microbial community beta diversity, which was based on Bray-curtis dissimilarity matrix. Differences in significance between samplings and management zones were tested with permutational multivariate analysis of variance (PERMANOVA) using the adonis (999 permutations) function of the vegan package.

A Distance-based Redundancy Analysis (db-RDA) was carried out to identify the relationships among the rhizosphere and bulk soil microbiomes related to the soil properties and nutrients availability, the significance of db-RDA model was tested using a PERMANOVA. Moreover, parametric Pearson's correlation analysis was performed to determine the relationships between the soil characteristics and alpha diversity and richness for each sampling, the correlation plots were generated using the R package “corrplot” version 0.77 (Wei & Simko, 2021); the significance of Pearson's correlations were determined at p-value of 0.05. Consequently, a regression analysis were done using the linear model function (lm), between the soil properties that had a significative correlation with the microbial diversity (Shannon index).

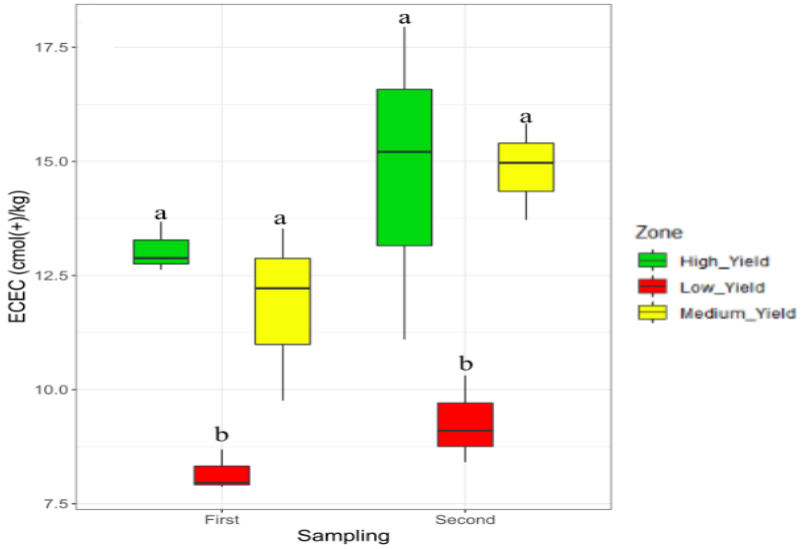
Furthermore, A heat map based on the z-score with the relative abundances of the reads assigned to each subsystem identified with the SEED database, was made for both bulk soil samplings and rhizosphere soil sampling, using the R package “pheatmap” version 1.0.12 (Kolde, 2019). Additionally, the non-parametric Kruskal-Wallis rank sum tests were applied to identify significant differences in their abundances.

3.3 Results

3.3.1 Soil physicochemical properties

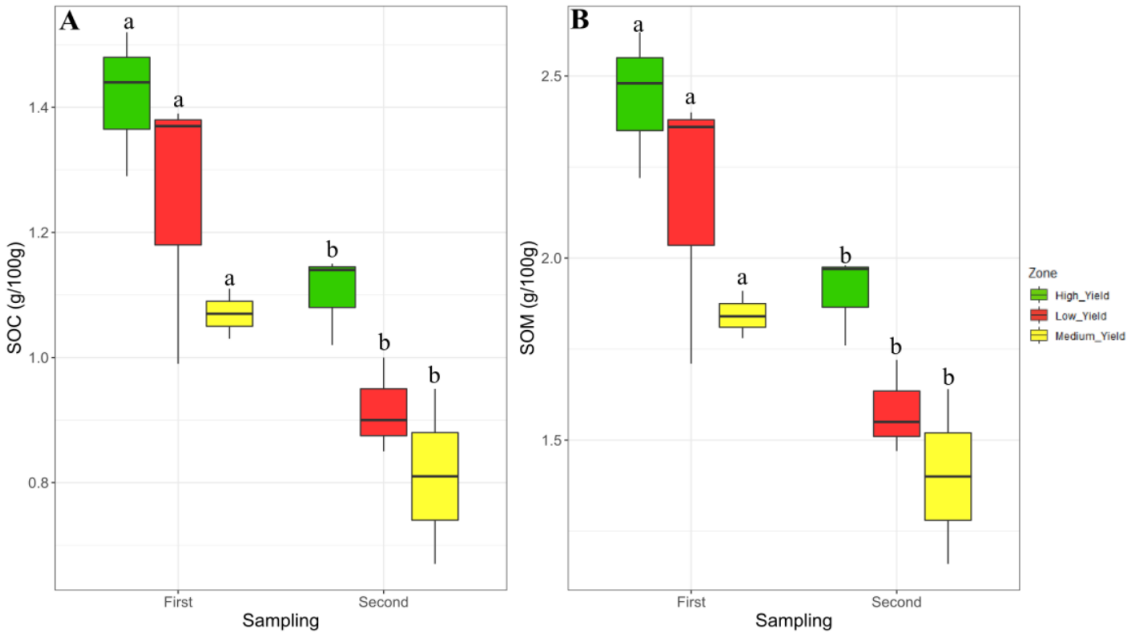
The physicochemical properties did not present differences at the management zone level in general, however, it was observed that some of the measured properties, presented significant differences over time at the sampling level. The parameters that were significantly different among the two samplings are: i) the effective cation exchange capacity (ECEC) with differences between low and high yield management zones (p-value=0.005) and low and medium management zone (p-value = 0.020) in the first sampling. For the second sampling, differences were found between the low and high yield management zone (p-value= 0.04) and the low and medium yield management zone (p-value = 0.044) (Fig. 3-2).

Figure 3-2: Boxplot of ECEC for both samplings by Management Zone. The letters above the boxes indicate significant differences ($p < 0.05$).



ii) Soil Organic Matter (SOM) (p-value = 0.002), c) Soil Organic Carbon (SOC) (p-value = 0.002) (Fig. 3-3). The results of SOC and SOM had a significant decrease in the samples obtained 70 days after the germination of the plants, moreover, it was evidenced that the high-yield management area obtained the highest values in both variables.

Figure 3-3: Boxplot of (A) soil organic carbon, and (B) soil organic matter for both samplings by Management Zone The letters above the boxes indicate significant differences ($p < 0.05$).



Elements such as potassium, calcium, boron, and magnesium differed significantly among samplings. The potassium and Calcium concentrations were significantly lower in the second sampling (p-value < 0.05). In addition, Calcium concentration was lowest in the

low-yield management area of the same sampling, in relation to the other management areas (p-value < 0.05) (Fig. 3-4).

On the other side, Boron concentration (Fig. 3-5A) appeared to be higher in the second sampling compared to the first one in all management zones. (p-value = 0.0012). Additionally, the micronutrient magnesium obtained (p-value = 0.022) in the first sampling and (p-value = 0.028) in the second sampling between the low and high yield management zone (Fig. 3-5B).

Figure 3-4: Soil physicochemical variables that were significantly different among samplings, (A) boxplot of K by samplings. (B) boxplot of Ca by samplings. The letters indicate significant differences (p < 0.05)

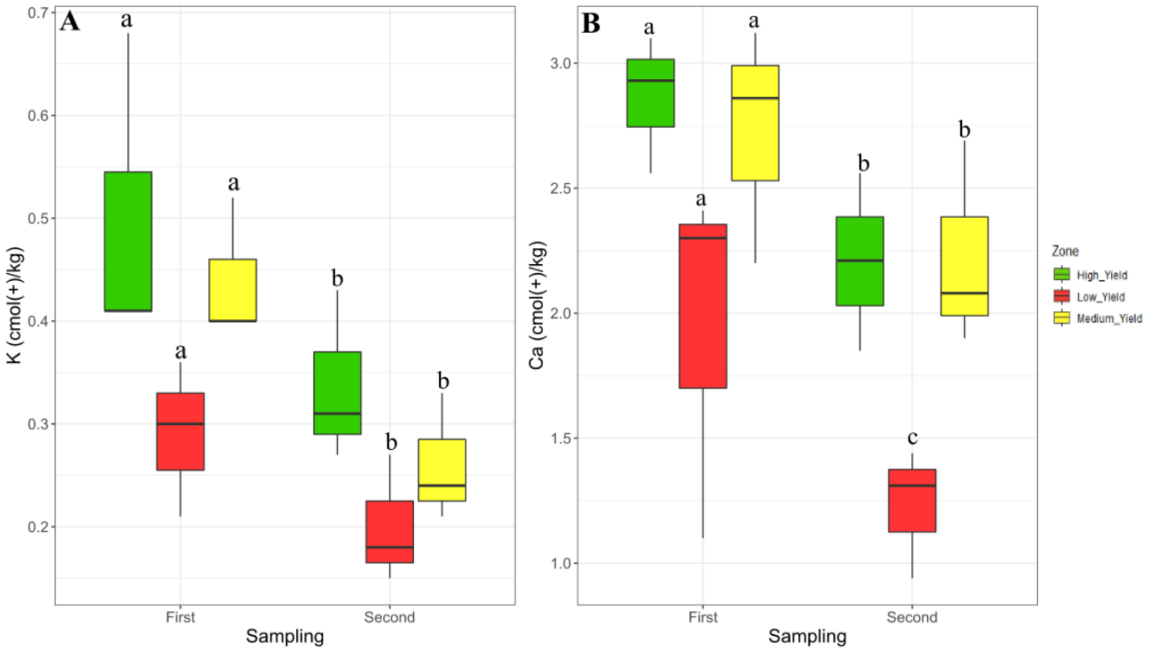
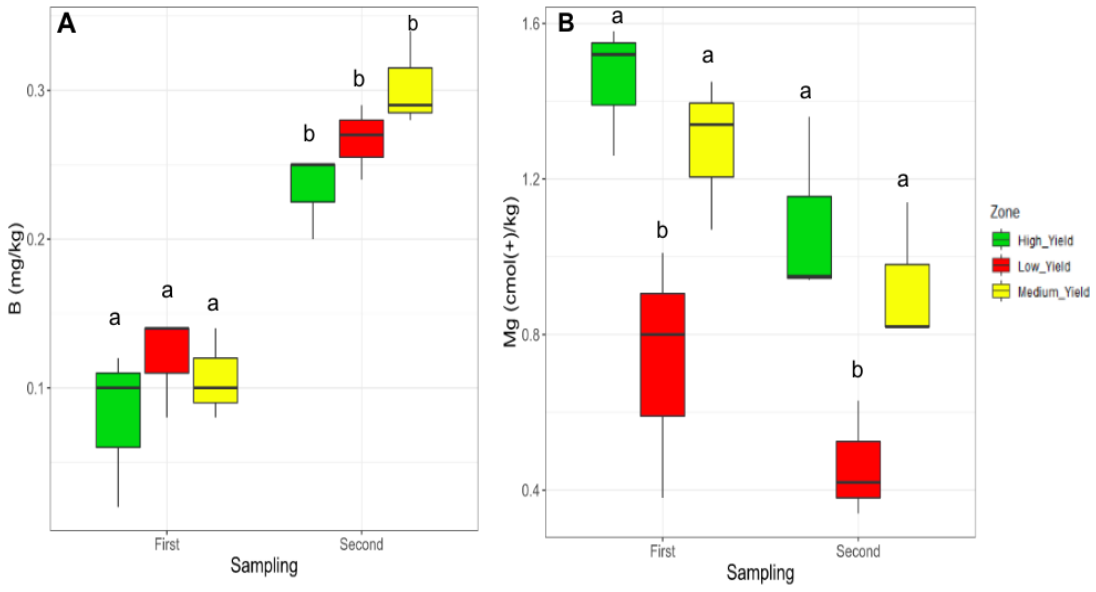


Figure 3-5: Soil physicochemical variables that were significantly different among samplings, (A) boxplot of Boron by samplings. (B) boxplot of Magnesium by samplings. The letters indicate significant differences ($p < 0.05$).



Regarding the results of other elements and properties like pH, there were no significant differences between management zones and sampling, however, it is possible to characterize the soils as acid with an average value of 4.90 ±0.12. About elements such as N, P, Zn, EC, Al, Mn, S, Fe, and Cu the means, were no significantly different among the three management zones nor between both samplings (p-value > 0.05), the means of each variable are shown in Table 3-1.

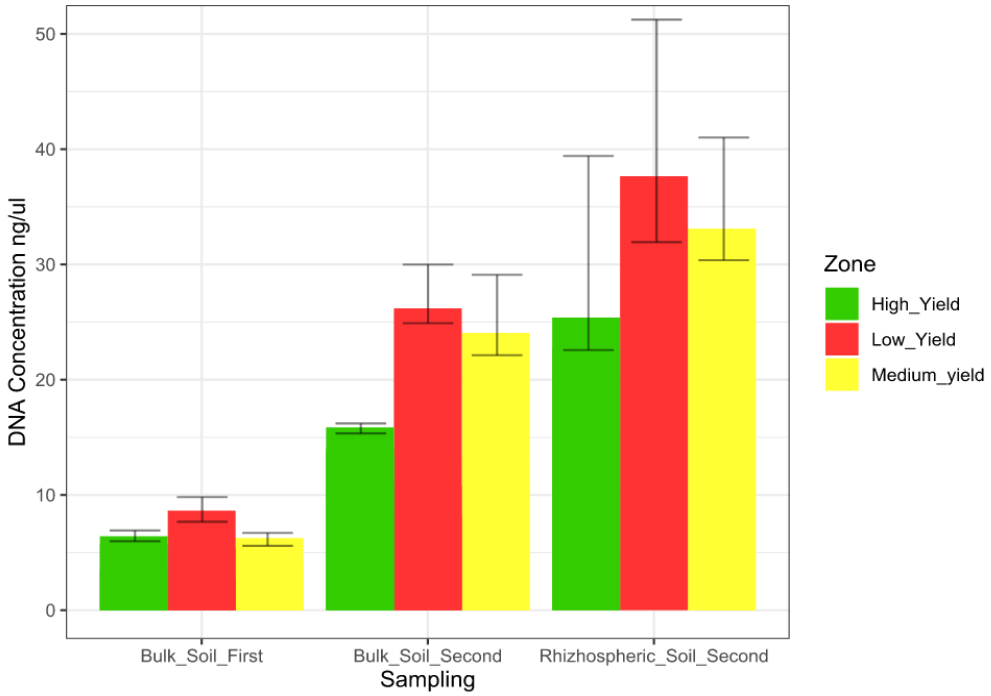
Table 3.1: Soil physicochemical factors among different management zones and samplings, means with different letter differ significantly P, statistical test ANOVA - Tukey´s HSD and Kruskal Wallis - Wilcoxon test.

		High Yield Management Zone		Low Yield Management Zone		Medium Yield Management Zone	
		1st Sampling	2nd Sampling	1st Sampling	2nd Sampling	1st Sampling	2nd Sampling
Soil properties	**pH**	4,94 a ±0.03	4,97 a ±0.01	4,98 a ±0.14	4,87 a ±0.20	4,82 a ±0.12	4,84 a ±0.07
	Soil Organic Matter (SOM) (g/100g)	2,44 a ±0.20	1,9 b ±0.12	2,16 a ±0.38	1,58 b ±0.12	1,84 a ±0.06	1,4 b ±0.24
	Soil Organic Carbon (SOC) (g/100g)	1,42 a ±0.11	1,1 a ±0.07	1,25 a ±0.22	0,92 b ±0.07	1,07 b ±0.04	0,81 b ±0.14
	Effective cation exchange capacity (ECEC) (cmol/kg)	8,12 a ±3.4	7,82 a ±0.54	5,32 b ±0.96	4,38 b ±0.44	6,67 a ±1.06	6,81 a ±1.91
	Soil Electric Conductivy (EC) (dS/m)	0,186 a ±0.04	0,146 a ±0.01	0,203 a ±0.02	0,136 a ±0.04	0,206 a ±0.06	0,213 a ±0.08
Soil macronutrients	**N (mg/kg)**	0,12 a ±0.01	0,07 a ±0.006	0,11 a ±0.01	0,08 a ±0.006	0,09 a ±0.003	0,10 a ±0.01
	P (mg/kg)	20,96 a ±8.73	19,78 a ±5.43	39,67 a ±10.39	40,01 a ±10.27	38,27 a ±9.05	31,41 a ±10.04
	K (mg/kg)	0,5 a ±0.15	0,34 b ±0.08	0,29 a ±0.07	0,2 b ±0.06	0,44 a ±0.06	0,26 b ±0.06
	S (mg/kg)	12,27 a ±0.25	13,88 a ±3.8	10,98 a ±2.67	10,79 a ±0.66	11,05 a ±4.23	15,95 a ±6.89
	Ca (cmol/kg)	2,86 a ±0.27	2,21 b ±0.35	1,94 a ±0.72	1,23 c ±0.25	2,73 a ±0.47	2,22 b ±0.41
	Mg (cmol/kg)	1,45 a ±0.17	1,08 b ±0.23	0,73 a ±0.32	0,46 b ±0.14	1,29 a ±0.19	0,93 b ±0.18
Soil micronutrients	**B (mg/kg)**	0,08 a ±0.05	0,23 b ±0.02	0,12 a ±0.03	0,27 b ±0.02	0,11 a ±0.03	0,3 b ±0.03
	Al (mg/kg)	2,5 a ±0.64	3,32 a ±0.69	1,86 a ±0.28	1,98 a ±0.47	1,75 a ±0.30	2,77 a ±1.04
	Na (cmol/kg)	<0,14 a	<0,14 a	<0,14 a	<0,14 a	<0,14 a	<0,14 a
	Fe (mg/kg)	632,53 a ±82.16	609,57 a ±115.5	589,24 a ±131.4	425,83 a ±65.9	513,62 a ±118.7	372,59 a ±98.3
	Cu (mg/kg)	5,33 a ±1.40	5,09 a ±0.63	2,41 a ±0.88	2,45 a ±1.44	3,51 a ±0.85	3,75 a ±0.62
	Mn (mg/kg)	57,1 a ±41.45	71,22 a ±41.56	20,45 a ±11.32	16,57 a ±8.62	27,43 a ±19.98	39,35 a ±11.47
	Zn (mg/kg)	3,44 a ±0.66	2,95 a ±0.80	2,5 a ±0.67	1,53 a ±0.24	2,93 a ±1.01	2,42 a ±0.07

3.3.2 DNA extraction

An average DNA yield concentration of 38.55 ng/µl was obtained for all the samples belonging to the rhizosphere, seventy days after plant germination. The DNA yield in these samples ranged from 22.6 to 50.9 ng/µl. The DNA from the rhizosphere system could be distinguished from bulk soil DNA with considerably more nucleic acid extracted, especially compared to the samples taken before the crop germination (first sampling) (Fig. 3-6).

Figure 3-6: Soil DNA mean yields from bulk and rice rhizosphere soil samples taken before crop planting (bulk soil first sampling) and seventy days after the crop germination for (bulk and rhizosphere second soil sampling) for each management zone. Error bars represent standard deviation.



3.3.3 Metagenome sequencing and quality control

A total of twenty-seven soil samples (18 bulk soil samples and 9 rhizosphere soil samples) were sequenced with an output of 111,953,828 paired-end raw reads, obtained using Illumina NovaSeq™ 6000 with an average size of 250 bp for all samples. 47,035,536 reads were obtained from rhizosphere soil compared with 32,492,934 and 32,425,358 total reads for the first and second sampling of the bulk soil samples respectively. An average of 4,146,438 individual sequence raw reads was obtained per sample.

Following the quality control and filtering carried out with MG-RAST, a total of 88,715,520 sequenced reads were obtained, which represents that a 20.75% of the total reads, were filtered by quality. For the first bulk soil sampling, there were 81.13% of reads passed quality control, followed, in the second sampling bulk soil there was 79.69%, and finally in the rhizosphere soil samples 77.62% of reads after that step (Table 3-2).

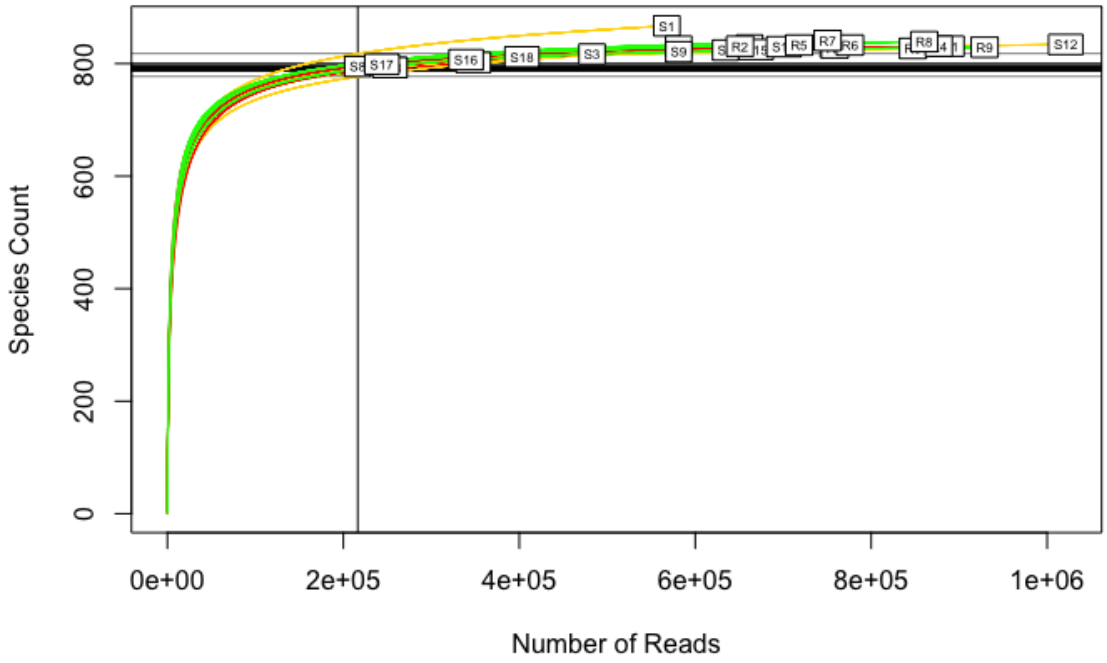
Table 3-2: Sequencing data (number of reads and total reads after quality control) of bulk and rhizosphere soil samples.

Sampling	Total number of reads	Total reads after quality control	Percentage of high-quality reads
Bulk Soil First Sampling	32,492,934	26,364,133	81.13%
Bulk Soil Second Sampling	32,425,358	25,841,478	79.69%
Rhizospheric Soil Second Sampling	47,035,536	36,509,909	77.62%

3.3.4 Soil microbial diversity and structure

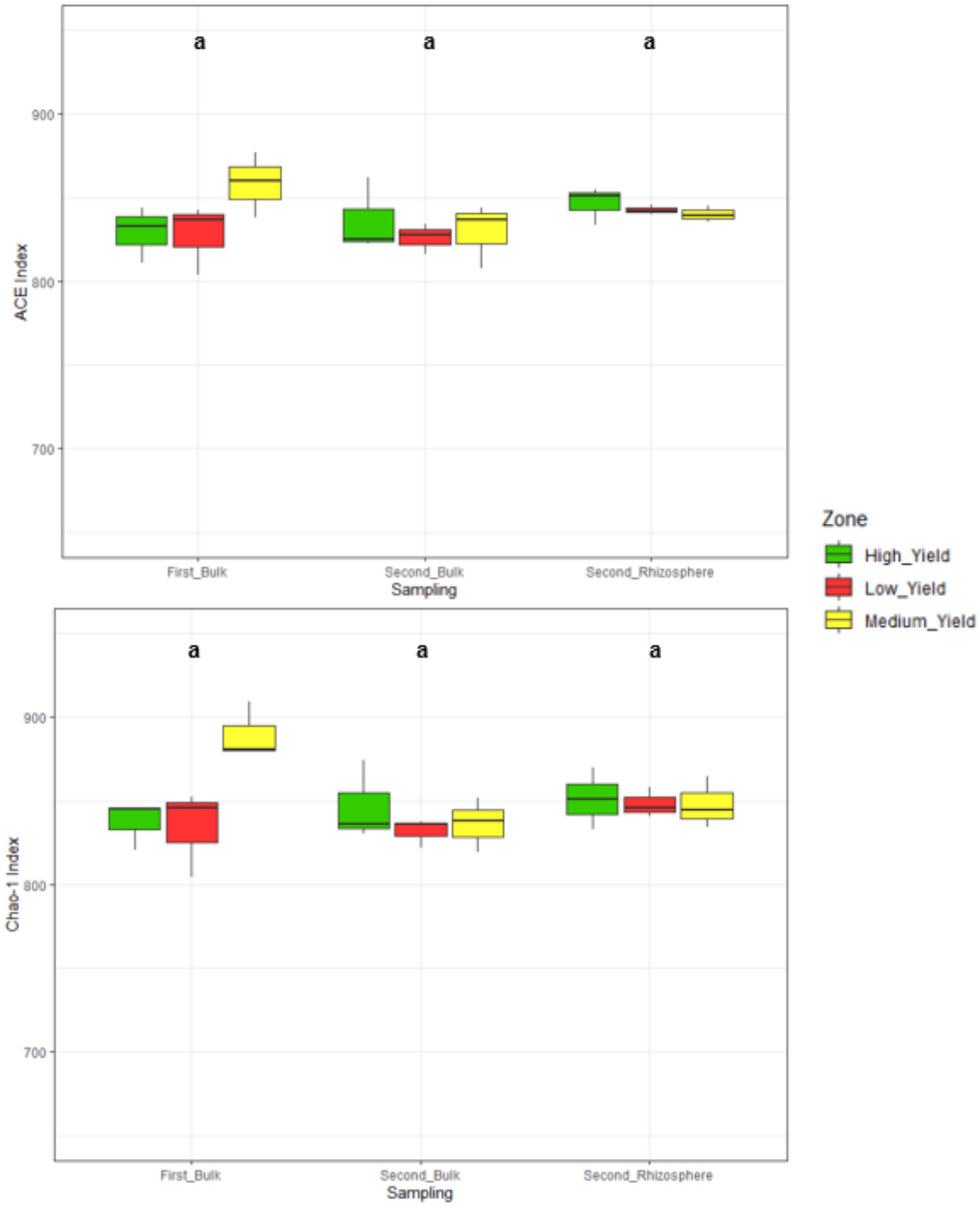
Rarefaction analysis showed that the sequencing effort was adequate to capture a greater part of the microbial diversity in the samples, since increasing the number of reads had minimal impact on the number of species that were found using MG-RAST (Fig. 3-7).

Figure 3-7: Rarefaction curves for each sample comparing the number of reads with the species count from bulk soil first sampling (S1-S10), Second bulk soil sampling (S11-S19) and Rhizosphere soil sampling (R1-R9).



The Chao-1 and ACE richness indices that were used to estimate the species richness showed that there were no significant differences in richness between the three samplings and neither was it between the management zones. Although, it was observed that the richness in the medium yield management zone from the bulk soil taken during the first sampling was slightly higher with respect to the other two management zones (Fig. 3-8).

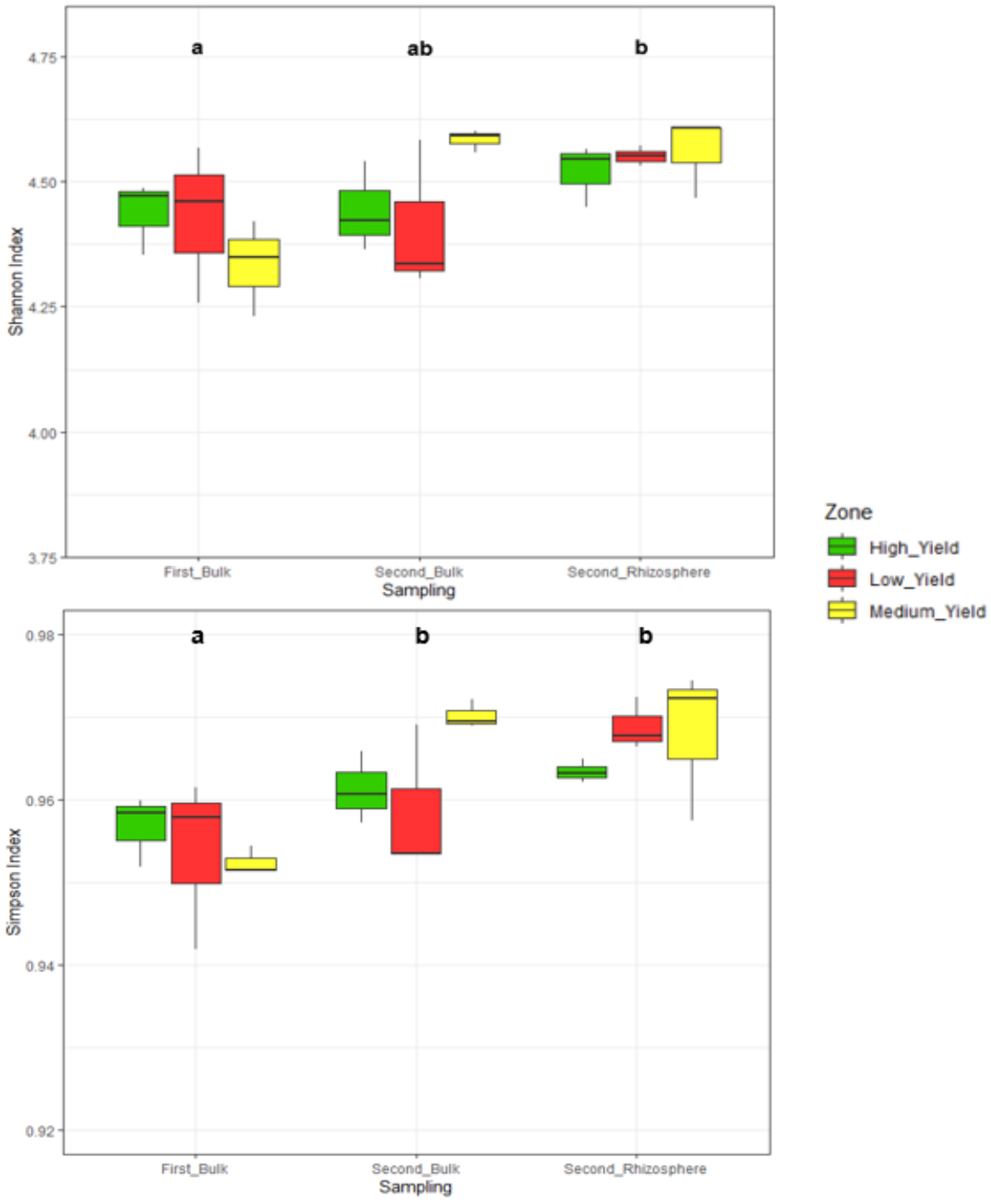
Figure 3-8: Chao-1 and ACE α-diversity indexes among the three samplings for each management zone. Different letters indicate significant differences among samplings and management zones by one-way ANOVAs (Tukey, $P < 0.05$).


a
a
a
900
800
700
ACE Index
First_Bulk
Second_Bulk
Second_Rhizosphere
Sampling
Zone
High_Yield
Low_Yield
Medium_Yield
a
a
a
900
800
700
Chao-1 Index
First_Bulk
Second_Bulk
Second_Rhizosphere
Sampling

Shannon and Inverse Simpson indexes were analyzed to evaluate the α-diversity from each management zone in the three different samplings. The Shannon diversity index was significantly higher in the rhizosphere soil compared to the bulk soil taken before the planting of the crop ($p = 0.0139$) (Fig. 3-9). Similar results were observed for the Inv-Simpson index, showing significant differences between the diversity of rhizosphere soil samples and bulk soil samples compared to the diversity from the first sampling, since a lower diversity was obtained ($p < 0.05$) (Fig. 3-9). However, the overall diversity was not significantly different between the three management zones in each sampling.

.

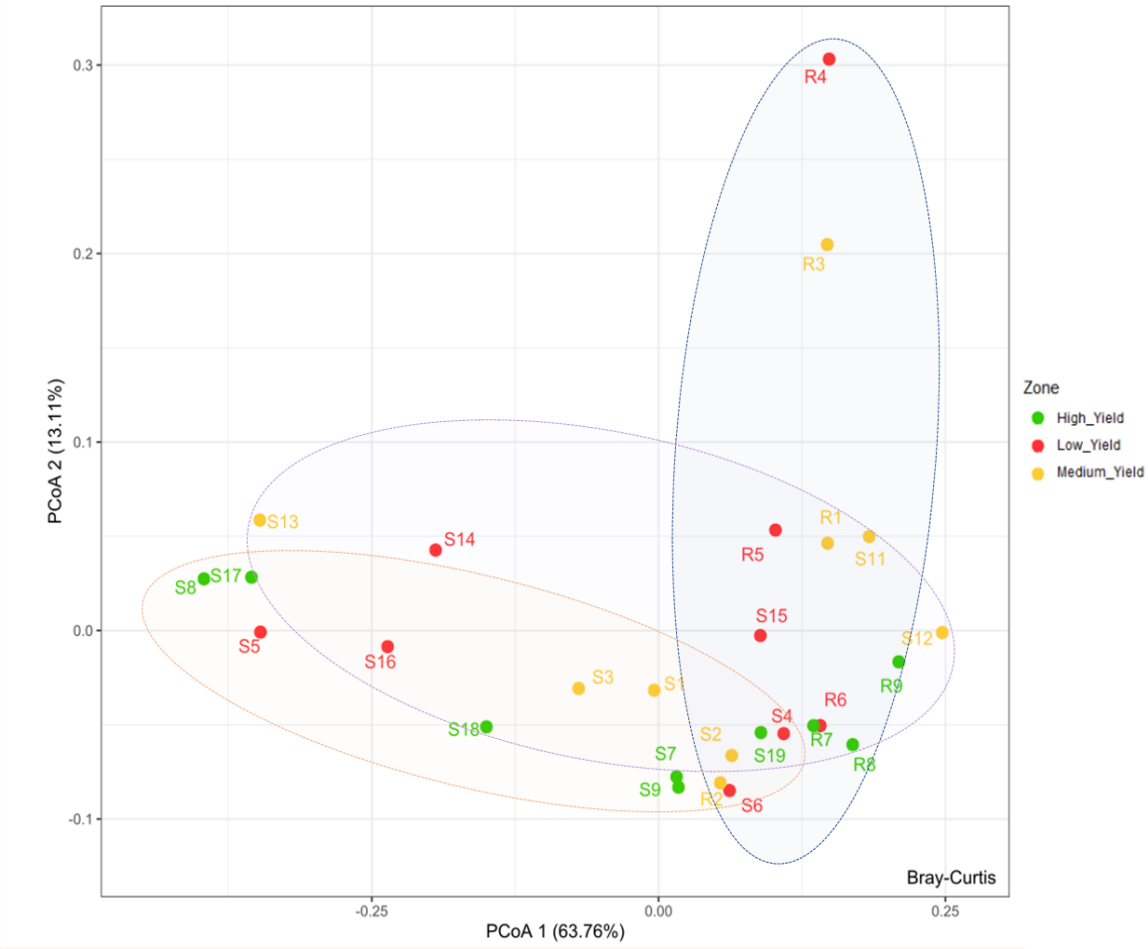
Figure 3-9: Shannon and InvSimpson α-diversity indexes among the three samplings for each management zone. Different letters indicate significant differences among samplings by one-way ANOVAs (Tukey, $P < 0.05$).


a
ab
b
4.75
4.50
4.25
4.00
3.75
Shannon Index
First_Bulk
Second_Bulk
Second_Rhizosphere
Sampling
a
b
b
0.98
0.96
0.94
0.92
Simpson Index
First_Bulk
Second_Bulk
Second_Rhizosphere
Sampling
Zone
High_Yield
Low_Yield
Medium_Yield

Beta diversity were measured with PCoA and permutational ANOVA based on Bray-curtis dissimilarity. There were not significant differences in the community β-diversity between management zones in each sampling (PERMANOVA R^2 =0.042; F = 0.595, *P* =0.668). Conversely, when the diversity was compared among the bulk and soil rhizosphere microbiomes, they revealed significant differences (PERMANOVA R^2 =0.257; F = 0.595, *P* =0.005).

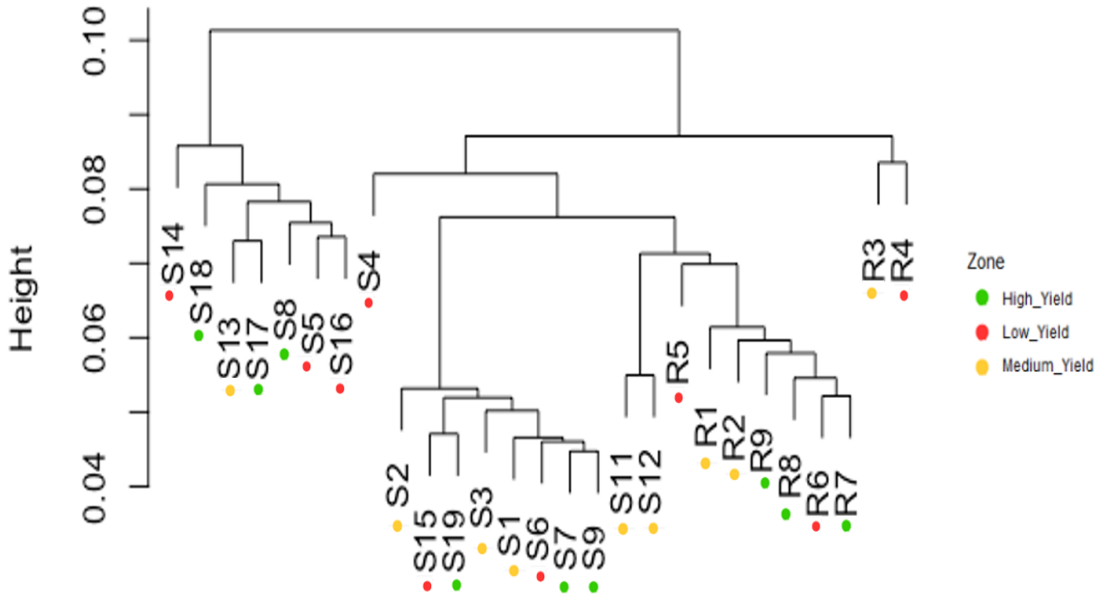
The PCoA percentages of variation explained by the axis 1 and axis 2 were 63,76% and 13,11% correspondingly, as shown in figure 3-10. Even though, there was no clear clusterization in the microbiomes of the bulk and rhizosphere soil and among management zones; the rhizosphere soil microbiomes exhibited a comparable composition among them, as well as the bulk soil samples between them. Despite the fact that there was no evident differentiation in the microbiomes from the rhizosphere and bulk soils, as well as between management zones; the rhizosphere and bulk soil microbiomes had comparable compositions among them.

Figure 3-10: Principal coordinate analysis (PCoA) showing the bulk soil rice and rhizosphere microbiota shifts, according to Bray-Curtis dissimilarity matrix. Bulk soil samples from first sampling (S1-S9) second sampling (S11-S19) and rhizosphere sampling (R1-R9).



Furthermore, a hierarchical cluster analysis of the microbial communities based on Bray Curtis dissimilarity showed that the soil microbiomes from bulk and rhizosphere soil differed. In particular, a cluster of rhizosphere soil microbiomes is observed, showing that they are similar in their microbial composition. Moreover, there are two more clusters that were seen in the analysis, where bulk soil samples from the first and second sampling were grouped together. However, no clusters were identified between the field's management zones (Fig. 3-11).

Figure 3-11: Hierarchical clustering based on Bray-Curtis dissimilarity matrix. Bulk soil samples from first sampling (S1-S9), second sampling (S11-S19) and rhizosphere sampling (R1-R9).



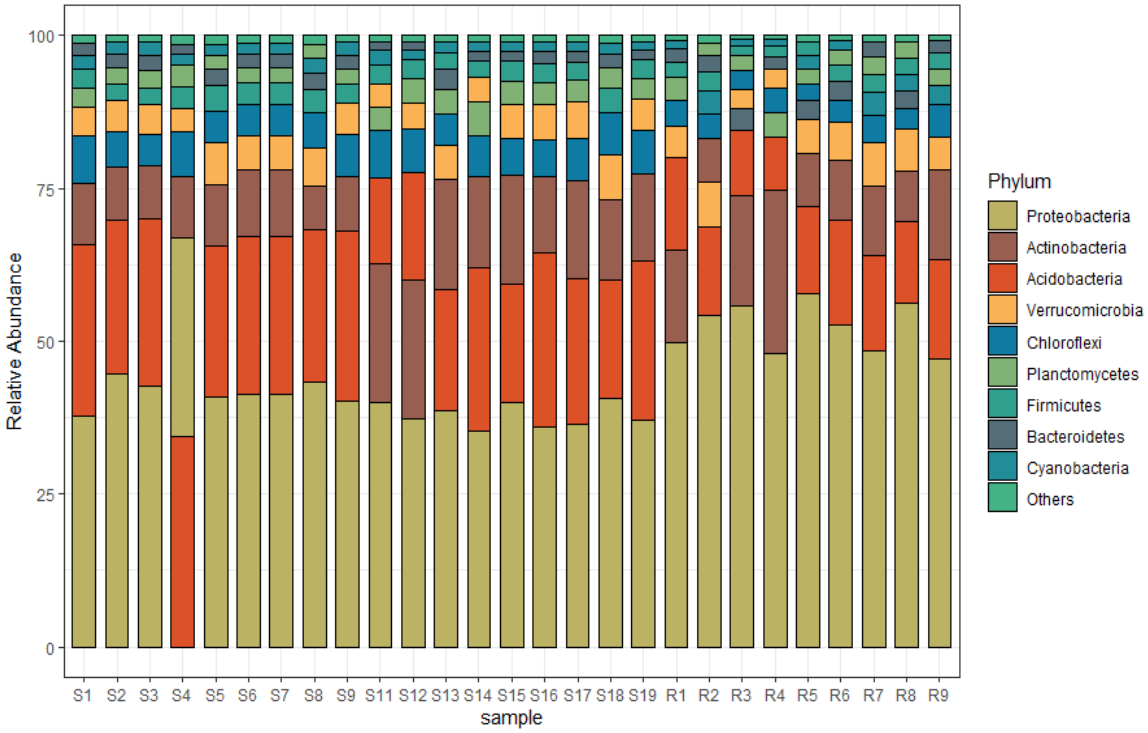
3.3.2 Bulk and rhizosphere soil microbial composition

Among the 27 seven soil samples MG-RAST identified 97.67% of rRNA features were Prokaryotes, followed by 1.51% were Archaea, 0.66% were Eukaryotes and < 0.005% were other sequences. From these reads a total of 47 phyla were found; the most abundant identified taxa at the phylum level were *Proteobacteria* with an average relative abundance of 43.5%, followed by *Acidobacteria* (20.9%), *Actinobacteria* (13.2%), *Chloroflexi* (5.51%*) and Verrucomicrobia* (5.33%). There were no significant differences in the microbiome composition between the samplings (bulk and rhizosphere soil) and between the evaluated management zones. Nevertheless, it was found that in the bulk soil sample S4, Acidobacteria was higher, accounting for 34.36% of the relative abundance of that sample, followed by Proteobacteria with 32.5%.

In the analysis of the Archaea and Eukaryotic communities in the sampled rice soils, several phyla exhibited notable relative abundance. Among these phyla, for archaea the samples were dominated by the phyla *Euryarchaeota, Thaumarchaeota* and *Crenarchaeota.* About *e*ukaryotes it was found that Ascomycota, Chlorophyta, Basidiomycota, and Nematoda emerged as the most dominant taxa across all samples.

The microbiota structure from bulk soils was consistent over time, however, when compared to the microbiota from the rhizosphere soil, it was observed that the abundance of the dominant phyla presented changes, having a greater dominance of Proteobacteria with an average of 52.2% of the total sequences of these samples. (Fig. 3-12).

Figure 3-12: Relative abundances of the microbial phyla from the bulk soil samples taken 10 days before crop planting (S1-S9), bulk soil (S11- S19) and rhizosphere soil samples (R1-R9) taken 70 days after crop germination.

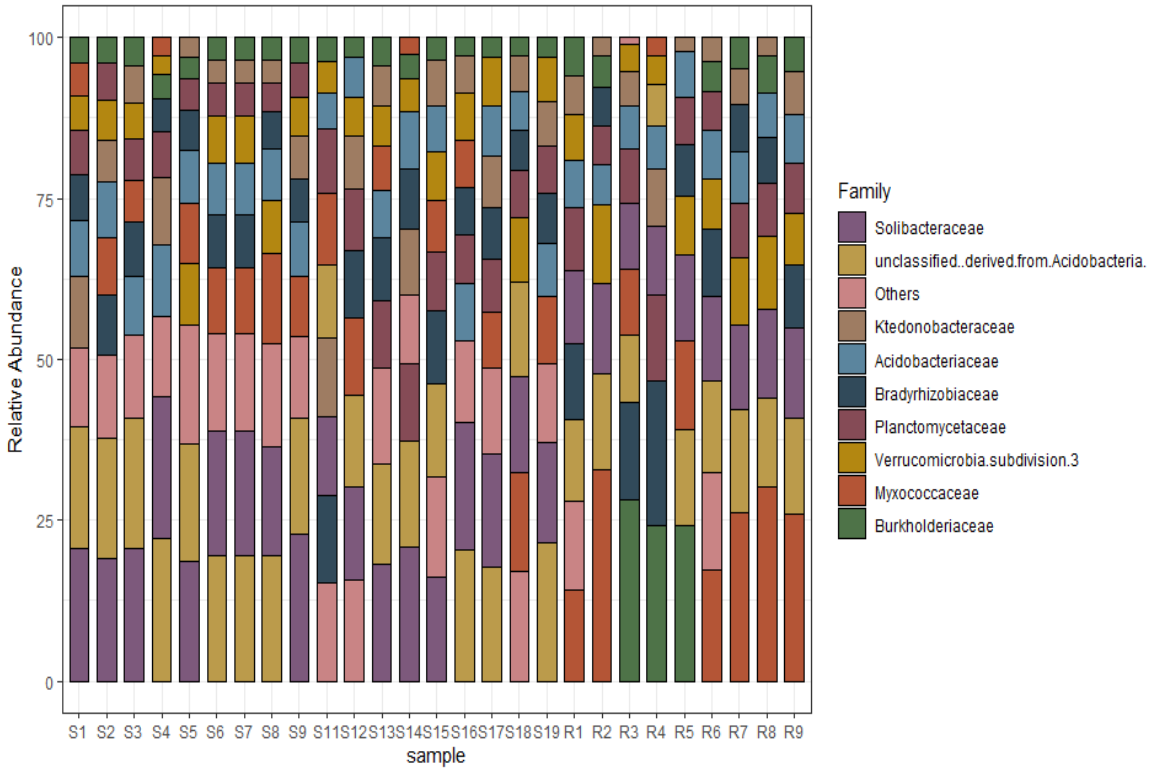


The most abundant families for the bulk soil samples were *Solibacteraceae* with (18.2%) of reads assigned, followed by *unclassified family derived from acidobacteria* with (17.8%), *Myxoccocaceae* (8.8%), *Bradyrhizobiacea*e (8.2%) and *Acidobacteriaceae* (7.8%). It was determined that the composition of the samples at the family level was not similar over time in the bulk soil samples, since the abundances of the most dominant families are variable between both samples (Fig. 3-13).

On the other hand, differences were observed in the composition at family level of the rhizosphere soil samples, since the most dominant family was *Myxococcaceae* with a relative abundance of 19.2%, followed by the *family unclassified derived from acidobacteria* with (13.1%)*, Solibacteraceae (*12.6%), *Burkholderiaceae* (11.9%) and *Bradyrhizobiaceae* (10.9%). It is notable that the *Burkholderiaceae* family in samples R3 from the medium yield management zone, R4 and R5 from the low yield management zone (rhizosphere soil), was

the most dominant family with 28.2%, 24% and 24.1% respectively of the relative abundance of these samples.

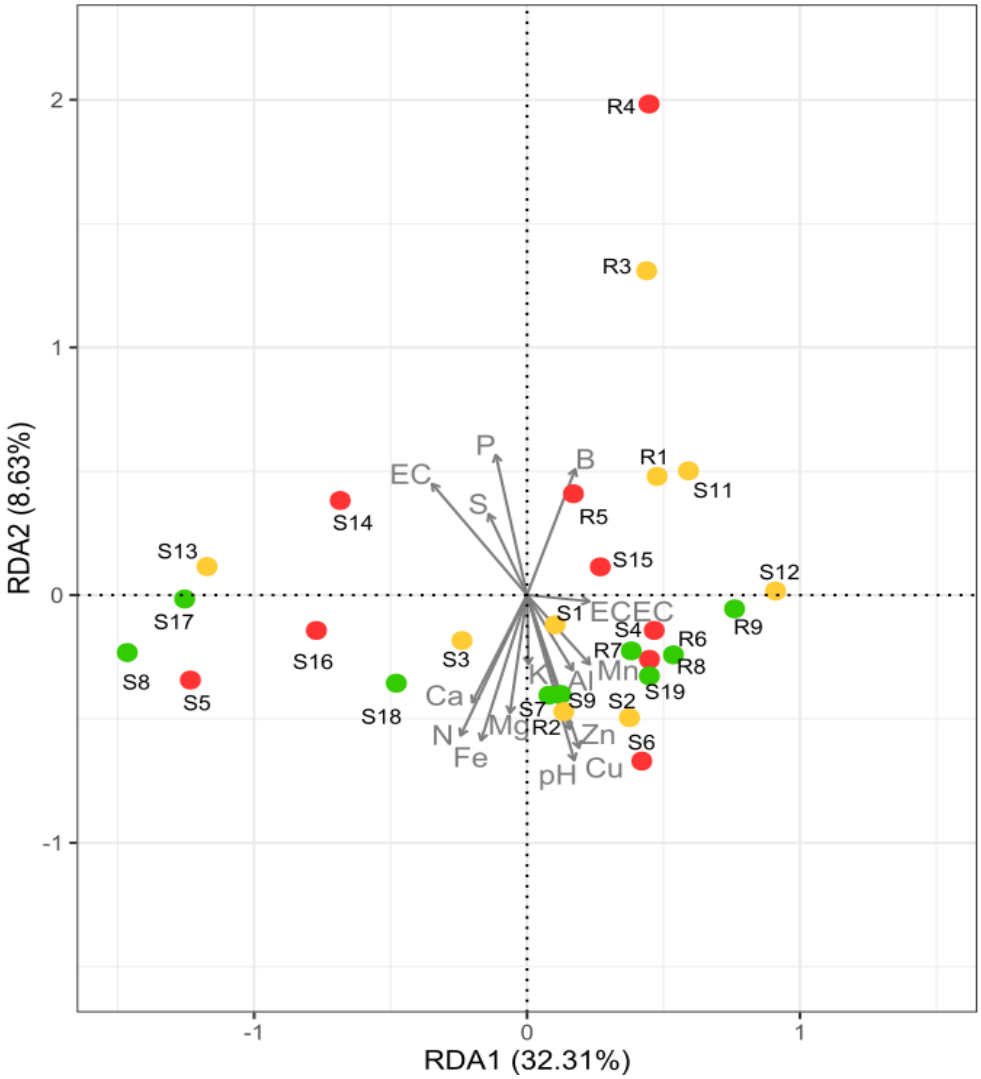
Figure 3-13: Relative abundances of the microbial families from the bulk soil samples taken 10 days before crop planting (S1-S9), bulk soil (S11- S19) and rhizosphere soil samples (R1-R9) taken 70 days after crop germination.



3.3.3 Correlations with the physico-chemical properties of the soil

A Bray-curtis distance-based redundancy analysis (RDA) was performed to identify the soil characteristics that could modulate soil microbial community structure (Fig. 3-14). The soil properties and nutrients together explained 40.94% of the total variance, but there were no significant differences in the db-RDA and the physico-chemical variables (p-value >0.05).

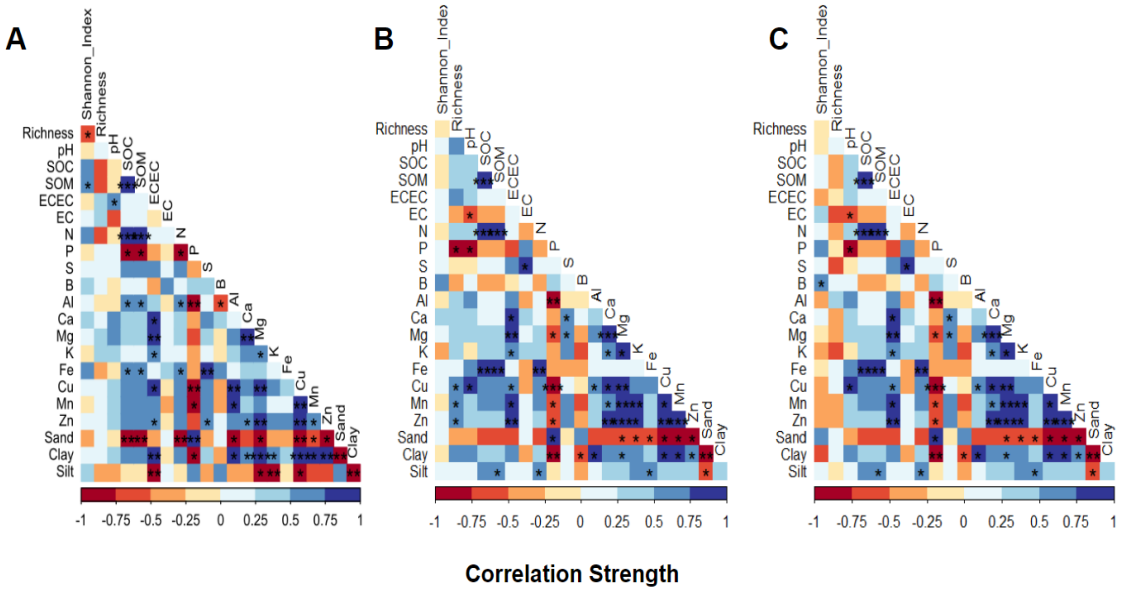
Figure 3-14: Distance-based Redundancy analysis (RDA) of microbial community structure from bulk (S1-S9 and S11-S19), rhizosphere soil samples (R1-R9), and the soil properties measured in both samplings.



The Shannon diversity of the bulk soil microbiome from the first sampling was negatively correlated with richness (rho = –0.7, P < 0.05). On the contrary was positively correlated with the SOM (*rho = 0.6, P < 0.05)* (Fig. 3-15A). Regarding the bulk soil samples taken 70 days after the germination of the crop, it was observed that the Shannon index of the bulk soil did not have a significant correlation with the characteristics and nutrients of the soil

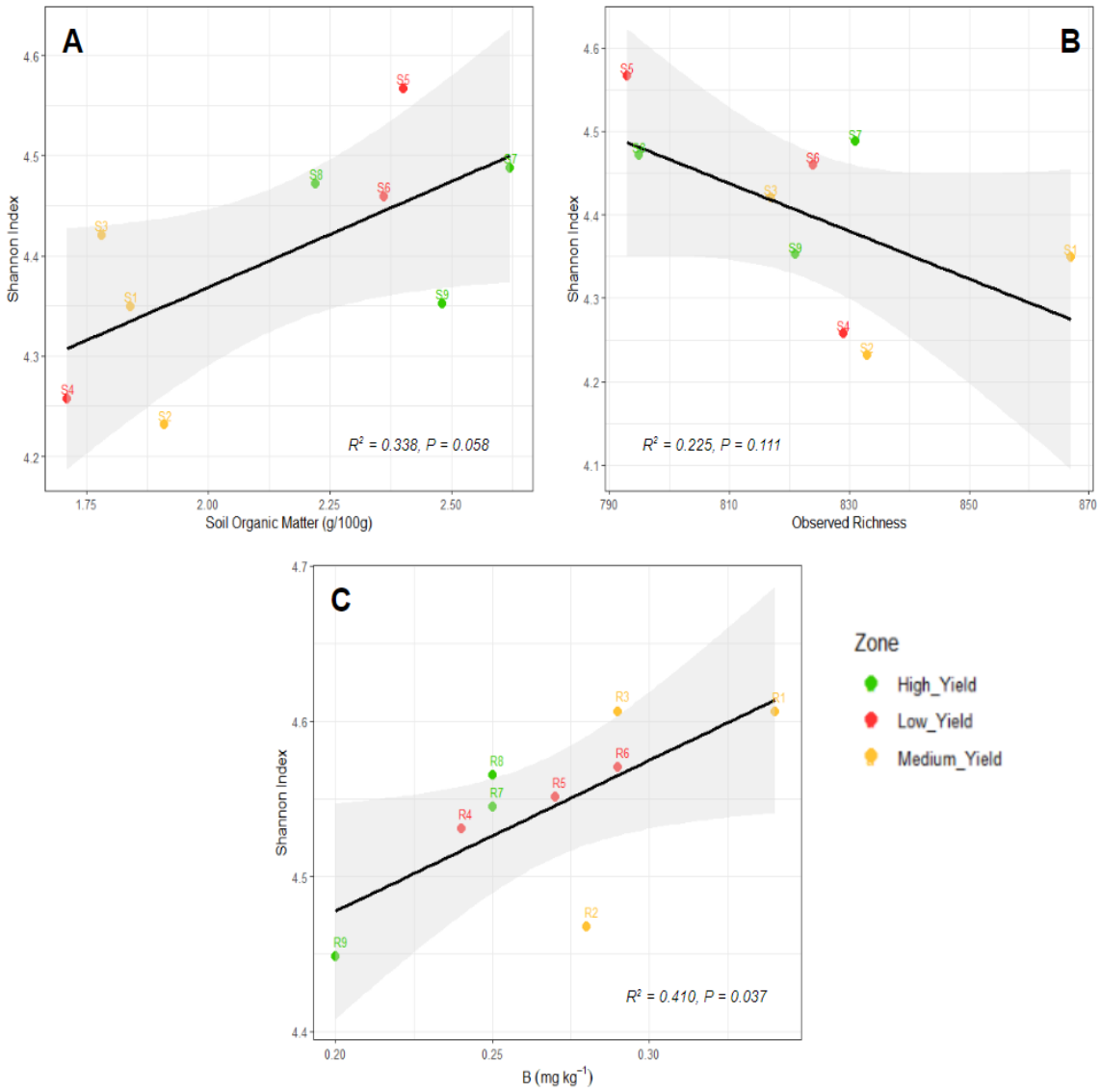
(Fig. 3-15B). However, the Shannon index in the rhizosphere soil samples showed a significant positive correlation with Boron *(rho= 0.7, P < 0.05*) (Fig. 3-15C).

Figure 3-15: Pairwise Pearson's correlation analysis. Blue and red color indicate positive and negative correlations, respectively. The color density on the squares reflects the strength of correlation and *, **, *** indicates a significant correlation at P < 0.05. (A) Correlations in bulk soil from the first sampling. (B) Correlations in bulk soil from the second sampling. (C) Correlations in rhizosphere soil from the second sampling.



The content of copper *(rho= 0.7, P < 0.05)*, zinc *(rho= 0.7, P < 0.05)*, and manganese *(rho= 0.6, P < 0.05)* in the bulk soil had a significant positive correlation with the observed richness during the second sampling. nonetheless, it was found that there was a strong negative correlation with phosphorus in the same sampling *(rho= -0.8, P < 0.05)* (Fig. 3-16).

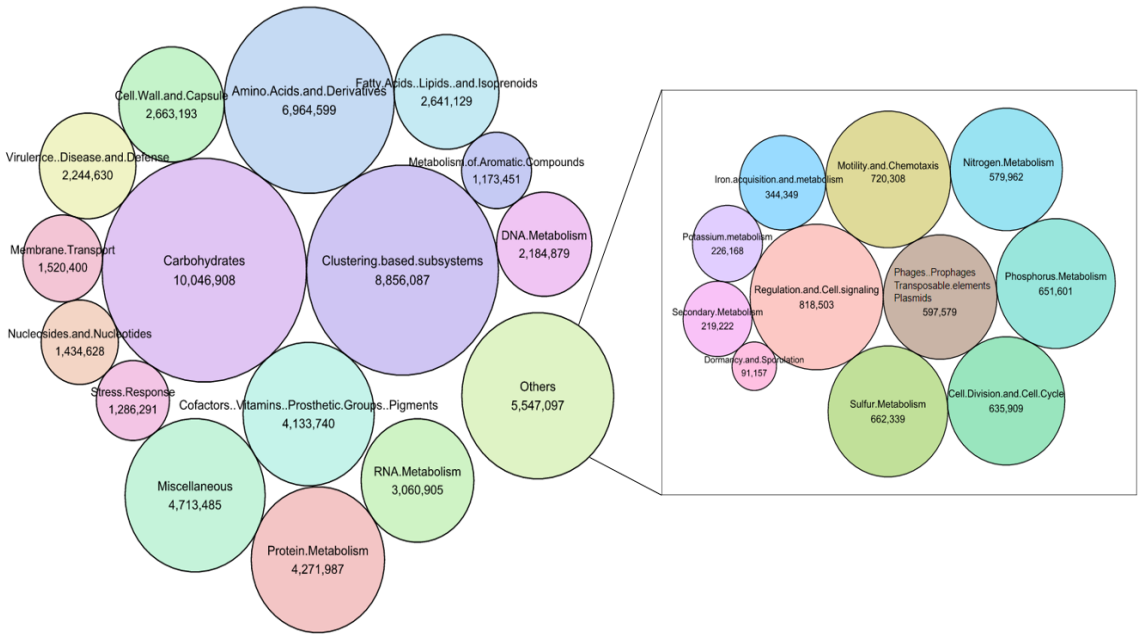
Figure 3-16: Linear regression analysis of the relationships of soil alpha-diversity (Shannon index) and: (A) Soil Organic Matter, (B) Observed Richness in the first sampling (Bulk soil samples S1-S9). (C) and Boron in the second sampling (Rhizosphere soil samples R1-R9)



3.3.2 Soil microbial functional prediction

The SEED subsystem data revealed 28 important functional categories attributable to the rhizosphere and bulk soil microbiomes from the rice field. A total of 65,004,782 hits were obtained in the SEED data-base, which was distributed in 18,582,166 hits for the first sampling and 18,428,148 for the bulk soil in the second sampling. As for the rhizosphere soil, it was obtained that were 27,994,438 hits assigned. Carbohydrates metabolism, clustering-based subsystems, aminoacids and derivatives, and cofactor, vitamins, prosthetic groups, and pigments had the largest quantity of annotated reads assigned in the three samplings (Fig. 3-17).

Figure 3-17: Subsystems general classification and assigned reads from the three samplings (Bulk and Rhizosphere soil).

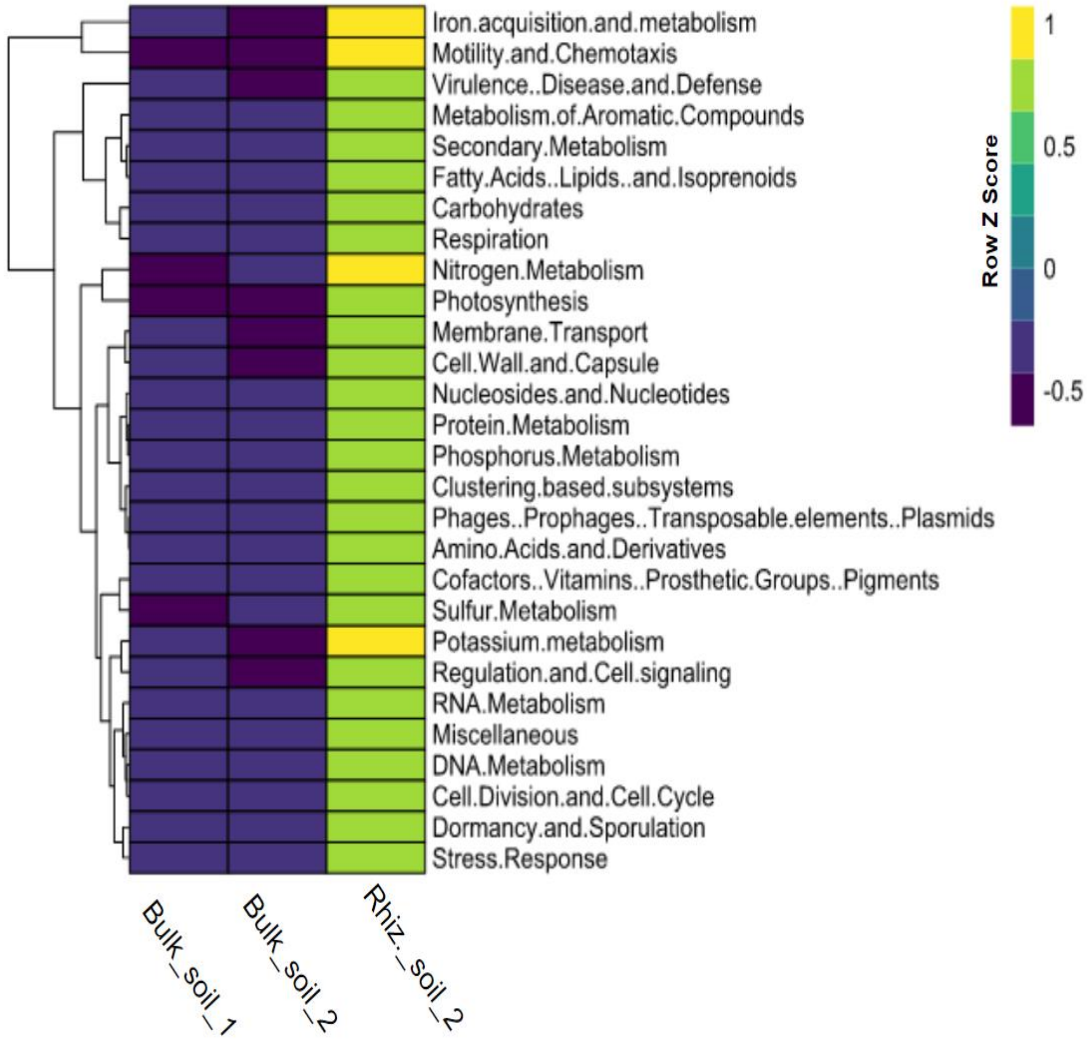


It was observed that in the rhizosphere soil samples taken 70 days after germination, there was a significantly higher relative abundance in the number of sequences ($p < 0.05$) assigned to the 28 key categories. Specifically, greater differences were observed in the categories iron acquisition and metabolism, motility chemotaxis, nitrogen metabolism, and potassium metabolism ($p < 0.01$), having a greater representation in the rhizosphere soil (Fig. 3-18A).

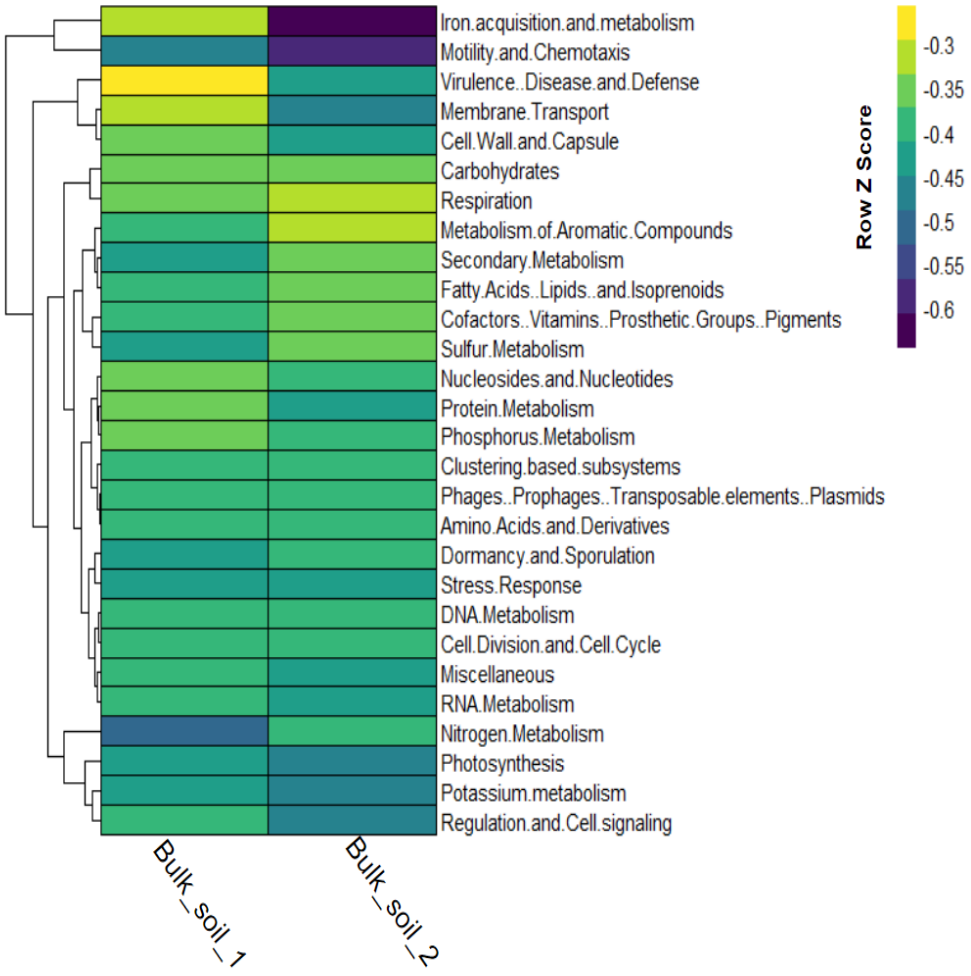
On contrary, among the bulk soil samples, a greater relative abundance in the categories of virulence, disease, defense, and iron acquisition and metabolism in the samples taken 10 days before the planting of the crop. Regarding the bulk soil samples taken 70 days after the plant's germination, it was observed that there was an increase in the relative abundance of respiration metabolism of nitrogen and aromatic compounds (Fig. 3-18B).

Figure 3-18: Heatmap showing the abundance of sequences similar to key metabolisms (A) in bulk and rhizosphere soil samples and (B) in the first and second sampling of the bulk soil. The relative abundance is indicated with different colors as represented with the scale bar with z-score.

A


Iron.acquisition.and.metabolism
Motility.and.Chemotaxis
Virulence..Disease.and.Defense
Metabolism.of.Aromatic.Compounds
Secondary.Metabolism
Fatty.Acids..Lipids..and.Isoprenoids
Carbohydrates
Respiration
Nitrogen.Metabolism
Photosynthesis
Membrane.Transport
Cell.Wall.and.Capsule
Nucleosides.and.Nucleotides
Protein.Metabolism
Phosphorus.Metabolism
Clustering.based.subsystems
Phages..Prophages..Transposable.elements..Plasmids
Amino.Acids.and.Derivatives
Cofactors..Vitamins..Prosthetic.Groups..Pigments
Sulfur.Metabolism
Potassium.metabolism
Regulation.and.Cell.signaling
RNA.Metabolism
Miscellaneous
DNA.Metabolism
Cell.Division.and.Cell.Cycle
Dormancy.and.Sporulation
Stress.Response
Bulk_soil_1
Bulk_soil_2
Rhiz._soil_2
Row Z Score
1
0.5
0
-0.5

B


Iron.acquisition.and.metabolism
Motility.and.Chemotaxis
Virulence..Disease.and.Defense
Membrane.Transport
Cell.Wall.and.Capsule
Carbohydrates
Respiration
Metabolism.of.Aromatic.Compounds
Secondary.Metabolism
Fatty.Acids..Lipids..and.Isoprenoids
Cofactors..Vitamins..Prosthetic.Groups..Pigments
Sulfur.Metabolism
Nucleosides.and.Nucleotides
Protein.Metabolism
Phosphorus.Metabolism
Clustering.based.subsystems
Phages..Prophages..Transposable.elements..Plasmids
Amino.Acids.and.Derivatives
Dormancy.and.Sporulation
Stress.Response
DNA.Metabolism
Cell.Division.and.Cell.Cycle
Miscellaneous
RNA.Metabolism
Nitrogen.Metabolism
Photosynthesis
Potassium.metabolism
Regulation.and.Cell.signaling
Bulk_soil_1
Bulk_soil_2
Row Z Score
-0.3
-0.35
-0.4
-0.45
-0.5
-0.55
-0.6

3.4 Discussion

3.4.1 Soil physicochemical properties

The soil's average pH was 4.90 ±0.12, classifying it as a strongly acid soil with an average composition of 44% clay, 32% sand, and 24% silt which is suitable for agriculture use since high contents of clay are related with advantages in terms of water holding capacity, nutrient retention, and cation exchange capacity (Dou et al., 2016). However, when the samples from the low-yield management zone were compared with the other zones, it was observed that this was characterized by a higher content of sand and a low content of clay. This composition aligns with the characteristic attributes of sandy soils, primarily influenced by the prevalence of larger sand particles. Sandy soils are known for their excellent drainage properties, owing to the larger pore spaces created by the coarse texture. Although, sandy soils exhibit a reduced capacity to retain nutrients due to the lower surface area provided by the larger sand particles. Consequently, nutrient availability in sandy soils is often limited, and there is an increased risk of nutrient leaching (Osman, 2013).

In addition, the low electric conductivity values, enable to characterize the field as non-saline soil; also, it was noted that the organic matter in the soil was low, similarly, macronutrient and micronutrient contents were mostly low, apart from the iron content, which was found to be high in comparison to other studies done in soils used for rice crops (Azadi et al., 2021; Chen et al., 2022; Mahender et al, 201). The previously determined three management zones and the change over time generate significant differences in some evaluated soil properties. In the high-yield management zone in both samplings; ECEC, SOM, and SOC increased significantly. Likewise, nutrients available in the soil such as potassium, magnesium, and calcium were significantly increased in this zone of the field. The low-yield management zone, on the other hand, exhibits a significant decrease in these soil attributes in both samplings.

The ECEC represents the total amount of exchangeable cations (Na^{+}, K^{+}, Mg^{2+}, Ca^{2+}, Mn^{2+}, Al^{3+}, Fe^{2+}) that a soil can adsorb, which is later released for crop uptake. These cations are held by the negatively charged clay and organic matter particles in the soil through electrostatic forces (Gliński et al., 2011). Soils with a low ECEC are more likely to develop deficiencies in potassium (K^{+}), magnesium (Mg^{2+}), and other cations. In contrast, soils with high ECEC are less susceptible to the leaching of these cations (Ketterings et al., 2007). Due to its direct impact on soil fertility, this metric is crucial and is related to the amounts of nutrients found in the three management zones.

SOM and SOC content are key factors that determine the health and quality of soil. SOC is the dominant component of soil organic matter, which is an important source of carbon for microbial soil processes and a sink for carbon sequestration (Arunrat et al., 2017). SOM also affects many properties of soils, including their ability to retain water and nutrients, to provide structure promoting efficient drainage and aeration, and to minimize loss of topsoil via erosion (Oldfield et al., 2019). Climatic factors have a large influence on SOC content and turnover, resulting in a strong feedback of the terrestrial C cycle (Luo et al., 2017). Furthermore, plant and microbial community biotic activity, as well as soil physicochemical parameters, influence the fate of SOC dynamics (Jungkunst et al., 2022).

Considering that the values for the previously described properties were higher in the high-yield management zone (Fig. 3-2), is related to the content of K, Ca, and Mg nutrients found in this area in both samplings due to its direct correlation; it was also possible to determine that in the low-yield management zone, these nutrients presented the lowest values in their contents.

Additionally, it was observed that the SOC, SOM, K, Ca, and Mg values in the samples obtained 70 days following the plant's germination, were lower than those in the soils from the first sampling, in the three management zones. This may be explained in part, variations and effect of the other physical-chemical characteristics in the soils, such as the pH, EC

and ECEC; as well as the impact of the fertilization scheme and the water inputs caused by the weather and the irrigation system. Those factors contributes to the soil nutrient loss by leaching, that is the downward movement of dissolved nutrients in the soil profile with percolating water. (Nguyen et al., 2020; Phongchanmixay et al., 2019; Schroth & Sinclair, 2003)

Furthermore, considering that differences in plant nutrient consumption are typically linked to the phenological stage of the crop, plant nutrient uptake may also be a factor. At this case, in the second sampling, the plants were in the phenological stage of the rice reproductive phase, were most of the nutrients have been actively used by the plant. On the other hand, different results were observed for boron, having a significant increase in the second sampling, it has been demonstrated that this element has a greater uptake at the end of the vegetative stage (Shahid et al., 2018; J. L. Wang et al., 2021), which may be related to the concentration found in these soils, given that in the second sampling the plants had already passed the vegetative stage, besides, the plants in the second sampling had already been fertilized a week before, this might affect changes in this micronutrient.

3.4.2 Soil microbial diversity and structure

The DNA concentration obtained and the number of total reads in the rhizosphere soil was higher, which could be associated with differences in the DNA extraction method for rhizosphere soils and the rhizosphere effect on microorganisms, where the biomass related with the population density and their activity are amplified by exudates from plant roots (Carreño, 2019; Hussain et al., 2012).

The results of the chao-1 and ACE richness indices indicated that there were no significant differences between the three management zones and samplings, even so, it was observed that on average there was a higher richness in the rhizosphere soil compared to the bulk soil microbiomes. Due to the field spatial scale of the study area, no significant variations in the three management zones could have been detected in the richness of the microbiomes. Similar findings in soils related to cereal croplands and tropical and temperate forests, demonstrated that the spatial scale where the study is carried out, is a decisive factor in the variation of the richness of soil microbial communities. Since the soil environment is so complex and heterogeneous, a large portion of this variation can be attributable to random sampling effects. As a result, samples taken from the same location can produce widely different results, being almost impossible to re-sample exactly the same community in a natural environment (De Gruyter et al., 2020; O'Brien et al., 2016).

Regarding diversity, assessed using the Shannon and Inv-Simpson indices in the samples, we found that alpha diversity of the studied rice microbiomes, was generally low when its compared with other investigations of these microbiomes in the world and also in Colombia (Barro et al., 2022; Carreño, 2019; H.-Y. Li et al., 2021; Sun et al., 2017; J. Wang et al., 2017). There were no significant differences across the management zones. Instead, between the rhizosphere and the bulk soil sample (first sampling), an increase in the diversity of the rhizosphere microbiomes was observed. The variations in the diversity

among the samplings can be explained by the differences in bacterial and fungi physiology, which is mainly related to the shifts in the physical-chemical conditions found in the bulk and rhizosphere soil (Naveed et al., 2016).

The beta diversity analysis revealed that were significant differences in the composition of the microbiomes at the sampling level since the microbiota related to the root of the plant, varied and clustered together, concerning the samples taken in the bulk soil. Conversely, no significant differences were obtained between the management zones. The hierarchical cluster analysis, however, revealed that different soil microbiomes associated with the high-yield and medium-yield zone management were clustered together, despite that no differences were found among the management zones, suggesting that the composition of their microbiomes could be similar.

It is well known that the rhizosphere and bulk soils exhibit niche differentiation in the formation of a spatially diverse environment and subsequently differentiate archaea, bacterial, and eukaryote communities. This is mainly because plants deposit a significant proportion of their photosynthates in the soil as root biomass or exudates (Nuccio et al., 2020; Pausch & Kuzyakov, 2018).

According to research, the phenology of many plants causes changes in the soil microbial communities that they harbor, since specific phytochemicals in the root exudates are differentially produced at distinct stages of plant growth; to promote an efficient variety of functions, such as act as substrates, chemotactic or signaling molecules to orchestrate changes in microbial composition, to facilitate relationships with microorganisms and enhance plant development and resilience (Chaparro et al., 2014; Lopes et al., 2022; Lyu & Smith, 2022; W. Wang et al., 2019; Wu et al., 2018).

The crop-flooded conditions during the second soil sampling could induce changes on the soil, like increasing of the anaerobic conditions on the bulk soil (Liesack et al., 2000). This factor could be linked to changes in the composition and structure of the soil microorganisms. Additionally, it is known that oxygen (O_2) is secreted through the

aerenchyma of rice roots, creating oxic zones surrounding the roots (such as the rhizosphere) that are surrounded by anoxic bulk soil. This oxic-anoxic interface is colonized by either aerobic, anaerobic, or facultative anaerobic microbes, which could result in a higher diversity in the rhizosphere (Bridgemohan et al., 2020; Ding et al., 2019).

3.4.3 Bulk and rhizosphere soil microbial composition

In the present study, the results showed that the composition at the phylum level was similar in the bulk soil over time and across the three management zones, and this composition is consistent with the results from other studies of rice crop soil microbiomes (Chialva et al., 2020; Majumdar et al., 2022; Sun et al., 2017; J. Wang et al., 2017; Zhong et al., 2020). Besides, *Proteobacteria* was the most abundant phylum followed by *Actinobacteria* in all samples. Similarly, it was found that *Proteobacteria* were enriched in soil samples from the rhizosphere microbiome. In contrast, sample S4 from the low-yield management zone of the first bulk soil sampling showed that the most dominant taxon was *Actinobacteria.* The abundances of *Proteobacteria* and *Acidobacteria* are related to the nutrient status of soils, and high ratios of abundances of *Proteobacteria/Acidobacteria* are indicative of copiotrophic soils (Huang et al., 2020).

The phylum *Proteobacteria*, or certain classes within this phylum, has been proposed as a copiotrophic lineage of Gram-negative bacteria, which indicates that its members are more abundant and have high growth rates under nutrient-rich conditions (Arunrat et al., 2022). The phylum has a high prevalence in agricultural soil environments and is linked to a wide range of activities involved in the cycling of carbon, nitrogen, and sulfur (Mhete et al., 2020), the presence of these bacterial groups in soils is also related to their metabolic plasticity to decompose an assortment of organic matter (S. Li et al., 2023; Pittol et al., 2018). Colonization by proteobacteria could reflect different nutrient conditions in the rhizosphere, and they might contribute to important processes of plant metabolism with their versatility in nutrient cycling (Santoyo et al., 2021).

According to research by Banerjee et al. (2018), the phylum *Acidobacteria* is one of the keystone bacterial taxa in soil involved with the breakdown of soil organic matter (SOM), indicating their importance in carbon turnover, which contributes to the breakdown of

complex biopolymers and plant residues. Members of the *Acidobacteria* phylum have important ecological roles in plant-soil ecosystems, such as regulating biogeochemical cycles and affecting plant growth (Kalam et al., 2020).

Actinobacteria, Chloroflexi, and *Verrucomicrobia* also showed a relative abundance that was higher compared to the other taxa. These soil bacteria associated with rice crops, particularly in agricultural soils, exhibit a wide distribution of these phyla (Doni et al., 2022). The phylum of *Actinobacteria* is associated with a number of significant processes, including the breakdown and decomposition of various organic materials, including cellulose, polysaccharides, protein lipids, organic acids, and sugars (Anandan et al., 2016). Additionally, it has been demonstrated that soil fertilization affects actinobacteria abundance (Ahn et al., 2012), as seen in the bulk soil samples from the second sampling where this taxon was enriched (Fig. 3-12).

Chloroflexi is a facultative anaerobic phylum including autotrophic, heterotrophic, and mixotrophic taxa particularly common in soils across the globe (Speirs et al., 2019), Finally, *Verrucomicrobia* are involved in plant growth promotion through the release of siderophores and chelators, degradation of cellulose, other are related with the oxidation of methane and biological nitrogen fixation (Nayak & Mishra, 2020), the phylum appear to be dominant in many soil bacterial communities across the globe (Fuerst, 2019).

At the family level, it was possible to observe changes in the general composition of the microbiome of rhizosphere soil samples, since an increase in the abundance of the *Myxococcaceae* family was evidenced. Members of this family are generally aerobic Gram-negative bacteria and are widely distributed in soil and the rhizosphere (Garcia & Müller, 2014). The family has species that feed on other microorganisms or macromolecular organic matter that contribute to nutrient cycling and carbon turnover in the soil, via the two nutritional behaviors that group them into predators or saprophytes(Adaikpoh, 2021; Wu et al., 2005).

Similarly, a higher dominance of the family Bradyrhizobiaceae was seen in the rhizosphere samples, their different species may or may not be able to perform fixation and/or other pathways of nitrogen assimilation, carbon assimilation, photosynthesis, and aerobic or anaerobic respiration. The majority of the family's genera are crucial players in the soil biogeochemical nitrogen cycle due to their capacity to utilize a variety of nitrogen sources in their metabolism (Marcondes de Souza et al., 2014).

Additionally, it was revealed that the *Burkholderiaceae* family showed the most dominance in rhizosphere soil samples R3 (medium yield zone management), R4, and R5 (low yield management). The family is ubiquitous in diverse soil environments and associated with different plant species. Their species are metabolically versatile, and some of them are capable to fixing nitrogen, nitrate reduction, and denitrification; Some species produce secondary metabolites, such as VOCs, that restrict fungal growth, while other species are major plant pathogens. (Carrión et al., 2018; Coenye, 2013; Xiong et al., 2021).

Regarding to the results from the archaea composition, Euryarchaeota had the higher abundance, this diverse phylum, exhibits widespread distribution and thrives in various environmental conditions. While their specific roles in rice soils remain elusive, certain members of Euryarchaeota are capable of methanogenesis, a process contributing to methane emissions in paddy fields and influencing greenhouse gas dynamics (Hu, et al, 2013; Zhang, et al, 2019). Thaumarchaeota was another prominent phylum, particularly the subgroup known as ammonia-oxidizing archaea (AOA), plays a pivotal role in nitrogen cycling. AOA oxidize ammonia, a crucial step in the nitrogen cycle, potentially impacting nitrogen availability and interactions with other soil microorganisms, including plants (Wang, et al, 2021). Finally, Crenarchaeota, although less abundant, encompass a diverse array of archaea, such as thermophiles and acidophiles. Their precise functions in rice soils remain less understood, yet they likely contribute to carbon and nitrogen cycling, as well as other biogeochemical processes (Yuan, et al, 2019).

Lastly, the eukaryotic communities in the analyzed rice soils exhibited distinct patterns of relative abundance among various phyla. Notably, Ascomycota, Chlorophyta, Basidiomycota, and Nematoda were the dominant phyla. Being Ascomycota the most abundant, their presence suggests their significant involvement in essential ecological processes such as decomposition, nutrient cycling, and symbiotic interactions with plants (Guo et al, 2022; Xu,. Et al, 2022). Following by Chlorophyta, encompassing green algae, could contribute to primary production and nutrient cycling, potentially influencing microbial diversity and carbon fixation processes within rice soils (Devi, et al, 2022). Then Basidiomycota fungi, known for wood decomposition and mycorrhizal associations, likely played a crucial role in nutrient cycling and interactions with rice plants (Li, et al, 2023). And Nematoda, a diverse group of microscopic worms, which their presence implying important contributions to plant health as decomposers, predators, or parasites (Mathesius & Costa, 2021). The prevalence of these eukaryotic phyla underscores their ecological significance and potential impact on nutrient cycling, organic matter decomposition, and plant-microbe interactions in rice soil ecosystems (Wang, et al 2021).

3.4.4 Correlations with the physico-chemical properties of the soil

Although soil pH is often recognized as the primary predictor of soil microorganisms in numerous research (De Gruyter et al., 2020; C. Wang et al., 2019) no significant associations between soil microbiota and pH were found in the current study. This is mostly attributable to the slight pH variation in the field, which had a pH range of 4.6 to 5.2, however, these soils have an acid pH, which is an important factor that determines the soil microbiome across the world (Table 3-1).

There was a significant positive correlation between the soil organic matter and Shannon diversity in the first sampling on the bulk soil. This may be due to the fact that the amount of organic matter present, directly influences soil microbes and their microbial activity. The soil organic matter transformation is primarily carried out by soil microbes, via processes including decomposition, polymerization, and immobilization (Ng et al., 2014). Other studies have demonstrated how the loss of this soil quality impacts the diversification of microbiomes. Additionally, it has an impact since microbes are crucial for soil carbon cycle, which affects agricultural productivity, greenhouse gas emissions, and carbon balance (Kalbitz et al., 2000; H.-Y. Li et al., 2018; Luo et al., 2017).

Likewise, the amount of boron in the soil and the Shannon diversity index based on samples from the rhizosphere were shown to be positively correlated. Boron is transformed into an organic form by microorganisms using the organic matter for their nutrient uptake, contributing to the soil's fertility load. It is implicit to mention that boron is a crucial component for plant development and is reported that is an essential element for cyanobacteria and diatoms. Nevertheless, the research on the influence of B on soil biological characteristics is very limited (Bhupenchandra et al., 2022; Bilen et al., 2011).

The observed richness and the Shannon index values had an inverse relationship in this same sampling. Low uniformity in the relative abundances of the discovered taxa may be the cause of this, which results in an increased number of species but reduces the diversity

of the community due to the effect of the dominance of particular taxa (Delgado-Baquerizo et al., 2017).

Moreover, significant positive correlations were obtained between the richness observed in the bulk soil samples taken in the second sampling with the concentration of Copper, Zinc, and Manganese in the soil. It has been shown how the application of these elements during the typical fertilization of the crop has an effect on the structure and composition of the soil microbial communities (Singh et al., 2022; Xiao et al., 2022), in this case, it is possible that the last chemical fertilization of the crop has an effect on the changes in the microbial richness from the bulk soil.

3.4.5 Soil microbial functional prediction

All of the specific functional groups were more representative in the rhizosphere than in the bulk soil samples. The data imply that plants can select a particular microbial community in the rhizosphere based on functional features favorable to their own performance (Ding et al., 2019; Jones et al., 2019; Mendes et al., 2014; Minz et al., 2013). There was a significant increase in iron acquisition and metabolism, motility chemotaxis, nitrogen metabolism, and potassium metabolism genes in the rhizosphere soil samples.

The genes related to the metabolism of iron, nitrogen, and phosphorus suggest that there may be a greater metabolic activity of microorganisms related to specific biogeochemical cycles in the rhizosphere of plants, which could be involved to the fact, that this soil compartment is considered a hot-spot for various processes and microbial interactions (Kuzyakov & Blagodatskaya, 2015; Pathan et al., 2020; Pramanik et al., 2020). The results suggest that the microbial communities in the rhizosphere samples may contribute to the

provision of major nutrients for plant growth and development. Besides, microorganisms motility and chemotaxis play an important role during the rhizosphere colonization and their composition definition, chemotaxis in the rhizosphere can indeed enhance both symbiotic and pathogenic interactions of microorganisms with their plant hosts (Colin et al., 2021; Matilla & Krell, 2018).

For bulk soil, the virulence, disease, defense, and iron acquisition and metabolism genes had an increase in the samples taken 10 days before the planting of the crop, compared to the bulk soil from the second sampling (Fig. 3-18). Iron is a micronutrient that is essential in the activation of metabolic pathways and a component of the prosthetic group in living organisms (Jin et al., 2014) One mechanism of defense of many microorganisms is to tightly chelate iron to several proteins to limit its extracellular availability (Chukwuneme et al., 2021). The results also suggest that the rise in this virulence, disease, and defense genes may be linked to competition for resources and other ecological interactions among microbes.

The relative abundance of genes assigned to respiration, metabolism of aromatic compounds, and metabolism of nitrogen and sulfur subsystems, were higher in the bulk soil samples collected 70 days following the plant's germination and the crop's last fertilization (Fig. B). Given that standard inorganic fertilization of crops alters the composition and structure of soil microbial communities, the changes in this genes abundance could be associated with the differences in their functional diversity (J. Wang et al., 2017; Z.-B. Zhao et al., 2020). The higher abundance of genes related to respiration, which is an essential part of microbial activity, is related to soil nutrients and carbon availabilities, which may play greater roles in regulating soil microbial respiration (Cleveland et al., 2007; C. Zhao et al., 2016). The increase in soil respiration rates has also been shown to be influenced by fertilization with inorganic nitrogen, mainly due to an increase in soil microbial biomass and root biomass or activity after the nitrogen addition (Chen et al., 2019; Yan et al., 2021).

In addition, the abundance of genes related to the metabolism of aromatic compounds was increased in this support soil sampling, may be related to the application of pesticides, many

of which have aromatic compounds, that could stimulate the abundance of these genes in the soil, since various microorganisms, use this compounds as part of their catabolism(Han et al., 2022; Seo et al., 2009). Finally, the changes in these relative abundances in the functional genes of bulk soils, may also be related with the intermittent irrigation conditions, which involve repeatedly flooding the field with water throughout the various rice development stages. Flooded soils are usually anaerobic, where oxidation processes are predominantly executed by mandatory and facultative anaerobic microorganisms, which reduce both organic (fermentation) and inorganic (anaerobic respiration) oxidized compounds (Colmer & Greenway, 2005; Denardin et al., 2020; X. Wang et al., 2020).

¿

3.5 Conclusions

In general, in this study the microbiome of the rice crop soil at the La Arabia farm in Meta, Colombia, are characterized by their low microbial diversity, which may be related to the intensive agronomic use of the soil for agronomic purposes (more than 10 years) and its acid soils conditions. The results showed that there were no significant differences in the composition of the soil microbiomes at the level of the three management zones in bulk and rhizosphere soil. It was discovered that there are variations in the composition of the soil microbiomes at the sampling level, which means that changes in the microbial community were evidenced over time.

The composition of the microbiomes in the management zones did not exhibit a clear trend, but it is possible to infer that there were differences between the three randomly selected points that composed each management zone, where soil samples were collected for the metagenomic analysis. This could be explained by the high variation of microorganisms when they are studied in complex ecosystems like soil, where we can find greater differences on a small spatial scale. Moreover, the rhizosphere microbiomes showed compositional and structural differences, when it was compared to the results from the bulk soil. This confirms that rice plants have a significant impact on the selection and modulation of their root microbiomes.

Additionally, is concluded that the soil microbiomes are influenced by the changes in physico-chemical properties; however, the effect of the soil characteristics, such as organic matter, or the availability of macronutrients and micronutrients, is not consistent over time, as well as its correlations with the diversity and richness. On the other hand, the functional prediction of microbiomes allowed us to determine that a higher relative abundance of

genes related to motility and chemotaxis, and the metabolism of nitrogen, iron, and potassium were found in the rhizosphere soil.

Finally, the results of this study deepen our understanding of the structure, composition, and prediction of the function of edaphic microbiomes associated with bulk and rhizosphere soils of a rice field, characterized by having highly acidic soils in Colombia, allowing us to identify that the agricultural practices such as management and fertilization alter microbial communities.

3.6 Recommendations

For future studies, it is advisable to initially take into account climatic variables such as precipitation and soil moisture, as well as to collect more physical-chemical data, since it has been described as a determining factor on soil microbiome shaping.

Likewise, it is advisable to carry out a more in-depth analysis of the yield obtained in the georeferenced points that were randomly selected to take the soil samples, since there was no evidence of representativeness of the performance of these with the proposed management zones since within the same point the variations in performance they were very high on a special very small scale.

3.6 References

Amanullah, D., & Khalid, S. (2020). Agronomy: Climate Change (Vol. 1). IntechOpen. https://doi.org/10.5772/intechopen.78102

Arunrat, N., Pumijumnong, N., & Hatano, R. (2017). Practices sustaining soil organic matter and rice yield in a tropical monsoon region. Soil Science and Plant Nutrition, 1-14. https://doi.org/10.1080/00380768.2017.1323546

Arunrat, N., Sansupa, C., Kongsurakan, P., Sereenonchai, S., & Hatano, R. (2022). Soil Microbial Diversity and Community Composition in Rice-Fish Co-Culture and Rice Monoculture Farming System. Biology, 11(8), 1242. https://doi.org/10.3390/biology11081242

Atique-ur-Rehman, Farooq, M., Rashid, A., Nadeem, F., Stuerz, S., Asch, F., Bell, R. W., & Siddique, K. H. M. (2018). Boron nutrition of rice in different production systems. A review. Agronomy for Sustainable Development, 38(3), 25. https://doi.org/10.1007/s13593-018-0504-8

Averill, C., Anthony, M. A., Baldrian, P., Finkbeiner, F., van den Hoogen, J., Kiers, T., Kohout, P., Hirt, E., Smith, G. R., & Crowther, T. W. (2022). Defending Earth's terrestrial microbiome. Nature Microbiology, 1-9. https://doi.org/10.1038/s41564-

022-01228-3

Azadi, A., Baghernejad, M., Gholami, A., & Shakeri, S. (2021). Forms and distribution pattern of soil Fe (Iron) and Mn (Manganese) oxides due to long-term rice cultivation in fars Province Southern Iran. Communications in Soil Science and Plant Analysis, 52(16), 1894-1911. https://doi.org/10.1080/00103624.2021.1900226

Barillot, C. D. C., Sarde, C.-O., Bert, V., Tarnaud, E., & Cochet, N. (2013). A standardized method for the sampling of rhizosphere and rhizoplan soil bacteria associated to a herbaceous root system. Annals of Microbiology, 63(2), 471-476. https://doi.org/10.1007/s13213-012-0491-y

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. Nucleic Acids Research, 41(Database issue), D36-D42. https://doi.org/10.1093/nar/gks1195

Biswas, R., & Sarkar, A. (2018). 'Omics' Tools in Soil Microbiology: The State of the Art. En T. K. Adhya, B. Lal, B. Mohapatra, D. Paul, & S. Das (Eds.), Advances in Soil Microbiology: Recent Trends and Future Prospects: Volume 1: Soil-Microbe Interaction (pp. 35-64). Springer. https://doi.org/10.1007/978-981-10-6178-3_3

Caulfield, M. E., Fonte, S. J., Groot, J. C. J., Vanek, S. J., Sherwood, S., Oyarzun, P., Borja, R. M., Dumble, S., & Tittonell, P. (2020). Agroecosystem patterns and land management co-develop through environment, management, and land-use interactions. Ecosphere, 11(4), e03113. https://doi.org/10.1002/ecs2.3113

Chang, J., Sun, Y., Tian, L., Ji, L., Luo, S., Nasir, F., Kuramae, E. E., & Tian, C. (2021). The Structure of Rhizosphere Fungal Communities of Wild and Domesticated Rice:

Changes in Diversity and Co-occurrence Patterns. Frontiers in Microbiology, 12. https://doi.org/10.3389/fmicb.2021.610823

Chaparro, J. M., Badri, D. V., & Vivanco, J. M. (2014). Rhizosphere microbiome assemblage is affected by plant development. The ISME Journal, 8(4), 790-803. https://doi.org/10.1038/ismej.2013.196

Chen, L., Zhao, D., Han, G., Yang, F., Gong, Z., Song, X., Li, D., & Zhang, G. (2022). Iron loss of paddy soil in China and its environmental implications. Science China Earth Sciences, 65(7), 1277-1291. https://doi.org/10.1007/s11430-021-9936-6

Chialva, M., Ghignone, S., Cozzi, P., Lazzari, B., Bonfante, P., Abbruscato, P., & Lumini, E. (2020). Water management and phenology influence the root-associated rice field microbiota. FEMS microbiology ecology, 96. https://doi.org/10.1093/femsec/fiaa146

Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics, 11(1), 485. https://doi.org/10.1186/1471-2105-11-485

De Gruyter, J., Weedon, J. T., Bazot, S., Dauwe, S., Fernandez-Garberí, P.-R., Geisen, S., De La Motte, L. G., Heinesch, B., Janssens, I. A., Leblans, N., Manise, T., Ogaya, R., Löfvenius, M. O., Peñuelas, J., Sigurdsson, B. D., Vincent, G., & Verbruggen, E. (2020). Patterns of local, intercontinental and interseasonal variation of soil bacterial and eukaryotic microbial communities. FEMS Microbiology Ecology, 96(3), fiaa018. https://doi.org/10.1093/femsec/fiaa018

Delgado-Baquerizo, M., Trivedi, P., Trivedi, C., Eldridge, D. J., Reich, P. B., Jeffries, T. C., & Singh, B. K. (2017). Microbial richness and composition independently drive soil

multifunctionality. Functional Ecology, 31(12), 2330-2343. https://doi.org/10.1111/1365-2435.12924

Devi, R., Kaur, T., Kour, D., Yadav, A., Yadav, A. N., Suman, A., ... & Saxena, A. K. (2022). Minerals solubilizing and mobilizing microbiomes: A sustainable approach for managing minerals' deficiency in agricultural soil. Journal of Applied Microbiology, 133(3), 1245-1272.

Ding, L.-J., Cui, H., Nie, S., Long, X., Duan, G., & Zhu, Y.-G. (2019). Microbiomes inhabiting rice roots and rhizosphere. FEMS microbiology ecology. https://doi.org/10.1093/femsec/fiz040

Dong, H., Sun, H., Jiang, L., Ma, D., & Fan, S. (2022). Characteristics of root-associated bacterial community and nitrogen biochemical properties of two Japonica rice cultivars with different yields. Food and Energy Security, 11(1), e357. https://doi.org/10.1002/fes3.357

Doni, F., Suhaimi, N. S. M., Mispan, M. S., Fathurrahman, F., Marzuki, B. M., Kusmoro, J., & Uphoff, N. (2022). Microbial Contributions for Rice Production: From Conventional Crop Management to the Use of 'Omics' Technologies. International Journal of Molecular Sciences, 23(2), Art. 2. https://doi.org/10.3390/ijms23020737

Dou, F., Soriano, J., Tabien, R. E., & Chen, K. (2016). Soil texture and cultivar effects on rice (Oryza sativa, L.) grain yield, yield components and water productivity in three water regimes. PLOS ONE, 11(3), e0150549. https://doi.org/10.1371/journal.pone.0150549

Farooq, A., Farooq, N., Akbar, H., Hassan, Z. U., & Gheewala, S. H. (2023). A Critical

Review of Climate Change Impact at a Global Scale on Cereal Crop Production. Agronomy, 13(1), Art. 1. https://doi.org/10.3390/agronomy13010162

FEDEARROZ. (2018). Adopción Masiva De Tecnologia AMTEC AMTEC FEDEARROZ. Revista Arroz, 22-34.

Garlapati, D., Charankumar, B., Ramu, K., Madeswaran, P., & Ramana Murthy, M. V. (2019). A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. Reviews in Environmental Science and Bio/Technology, 18(3), 389-411. https://doi.org/10.1007/s11157-019-09501-4

Gliński, J., Horabik, J., & Lipiec, J. (Eds.). (2011). Cation Exchange Capacity. En Encyclopedia of Agrophysics (pp. 110-110). Springer Netherlands. https://doi.org/10.1007/978-90-481-3585-1_550

Guo, X., Liu, J., Xu, L., Sun, F., Ma, Y., Yin, D., ... & Lv, Y. (2022). Combined organic and inorganic fertilization can enhance dry direct-seeded rice yield by improving soil fungal community and structure. Agronomy, 12(5), 1213.

Hartmann, M., & Six, J. (2022). Soil structure and microbiome functions in agroecosystems. Nature Reviews Earth & Environment, 1-15. https://doi.org/10.1038/s43017-022-00366-w

He, H., Li, W., Yu, R., & Ye, Z. (2017). Illumina-Based Analysis of Bulk and Rhizosphere Soil Bacterial Communities in Paddy Fields Under Mixed Heavy Metal Contamination. Pedosphere, 27(3), 569-578. https://doi.org/10.1016/S1002-0160(17)60352-7

Huang, M., Tian, A., Chen, J., Cao, F., Chen, Y., & Liu, L. (2020). Soil bacterial communities in three rice-based cropping systems differing in productivity. Scientific Reports, 10,

9867. https://doi.org/10.1038/s41598-020-66924-8

Hu, H. W., Zhang, L. M., Yuan, C. L., & He, J. Z. (2013). Contrasting Euryarchaeota communities between upland and paddy soils exhibited similar pH-impacted biogeographic patterns. Soil Biology and Biochemistry, 64, 18-27.

Jensen, L. J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., & Bork, P. (2008). eggNOG: Automated construction and annotation of orthologous groups of genes. Nucleic Acids Research, 36(Database issue), D250-254. https://doi.org/10.1093/nar/gkm796

Kalam, S., Basu, A., Ahmad, I., Sayyed, R. Z., El-Enshasy, H. A., Dailin, D. J., & Suriani, N. L. (2020). Recent Understanding of Soil Acidobacteria and Their Ecological Significance: A Critical Review. Frontiers in Microbiology, 11. https://www.frontiersin.org/articles/10.3389/fmicb.2020.580024

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research, 28(1), 27-30. https://doi.org/10.1093/nar/28.1.27

Keegan, K. P., Glass, E. M., & Meyer, F. (2016). MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. En F. Martin & S. Uroz (Eds.), Microbial Environmental Genomics (MEG) (pp. 207-233). Springer. https://doi.org/10.1007/978-1-4939-3369-3_13

Kendzior, J., Warren raffa, D., & Bogdanski, A. (2022). The soil microbiome: A game changer for food and agriculture : Executive summary for policymakers and researchers. FAO. https://doi.org/10.4060/cc0717en

Ketterings, Q., Reid, S., & Rao, R. (2007). Cation Exchange Capacity (CEC) [Agronomy

Fact Sheet Series]. Cornell University, Department of crop ans soil sciences. http://nmsp.cals.cornell.edu/publications/factsheets/factsheet22.pdf

Li, S., Li, G., Huang, X., Chen, Y., Lv, C., Bai, L., Zhang, K., He, H., & Dai, J. (2023). Cultivar-specific response of rhizosphere bacterial community to uptake of cadmium and mineral elements in rice (Oryza sativa L.). Ecotoxicology and Environmental Safety, 249, 114403. https://doi.org/10.1016/j.ecoenv.2022.114403

Lopes, L. D., Wang, P., Futrell, S. L., & Schachtman, D. P. (2022). Sugars and Jasmonic Acid Concentration in Root Exudates Affect Maize Rhizosphere Bacterial Communities. Applied and Environmental Microbiology, 88(18), e0097122. https://doi.org/10.1128/aem.00971-22

Lyu, D., & Smith, D. L. (2022). The root signals in rhizosphere inter-organismal communications. Frontiers in Plant Science, 13. https://www.frontiersin.org/articles/10.3389/fpls.2022.1064058

Mahender, A., Swamy, B. P. M., Anandan, A., & Ali, J. (2019). Tolerance of iron-deficient and -toxic soil conditions in rice. Plants, 8(2), 31. https://doi.org/10.3390/plants8020031

Magrane, M. & UniProt Consortium. (2011). UniProt Knowledgebase: A hub of integrated protein data. Database: The Journal of Biological Databases and Curation, 2011, bar009. https://doi.org/10.1093/database/bar009

Mathesius, U., & Costa, S. R. (2021). Plant signals differentially affect rhizosphere nematode populations. Journal of Experimental Botany, 72(10), 3496-3499. https://doi.org/10.1093/jxb/erab149

Meyer, F., Bagchi, S., Chaterji, S., Gerlach, W., Grama, A., Harrison, T., Paczian, T.,

Trimble, W. L., & Wilke, A. (2019). MG-RAST version 4—Lessons learned from a decade of low-budget ultra-high-throughput metagenome analysis. Briefings in Bioinformatics, 20(4), 1151-1159. https://doi.org/10.1093/bib/bbx105

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., & Edwards, R. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics, 9(1), 386. https://doi.org/10.1186/1471-2105-9-386

Mhete, M., Eze, P. N., Rahube, T. O., & Akinyemi, F. O. (2020). Soil properties influence bacterial abundance and diversity under different land-use regimes in semi-arid environments. Scientific African, 7, e00246. https://doi.org/10.1016/j.sciaf.2019.e00246

Mustafa, G., Hayat, N., & Alotaibi, B. A. (2023). Chapter fifteen—How and why to prevent over fertilization to get sustainable crop production. En T. Aftab & K. R. Hakeem (Eds.), Sustainable Plant Nutrition (pp. 339-354). Academic Press. https://doi.org/10.1016/B978-0-443-18675-2.00019-5

Naveed, M., Herath, L., Moldrup, P., Arthur, E., Nicolaisen, M., Norgaard, T., Ferré, T. P. A., & de Jonge, L. W. (2016). Spatial variability of microbial richness and diversity and relationships with soil organic carbon, texture and structure across an agricultural field. Applied Soil Ecology, 103, 44-55. https://doi.org/10.1016/j.apsoil.2016.03.004

Nayak, S. K., & Mishra, B. B. (2020). Frontiers in Soil and Environmental Microbiology (S.

K. Nayak & B. B. Mishra, Eds.; 1.a ed.). CRC Press. https://doi.org/10.1201/9780429485794

Nguyen, B. T., Phan, B. T., Nguyen, T. X., Nguyen, V. N., Van Tran, T., & Bach, Q.-V. (2020). Contrastive nutrient leaching from two differently textured paddy soils as influenced by biochar addition. Journal of Soils and Sediments, 20(1), 297-307. https://doi.org/10.1007/s11368-019-02366-8

Nuccio, E. E., Starr, E., Karaoz, U., Brodie, E. L., Zhou, J., Tringe, S. G., Malmstrom, R. R., Woyke, T., Banfield, J. F., Firestone, M. K., & Pett-Ridge, J. (2020). Niche differentiation is spatially and temporally regulated in the rhizosphere. The ISME Journal, 14(4), 999-1014. https://doi.org/10.1038/s41396-019-0582-x

Li, W., Li, Z., Liu, Y., Nie, X., Zheng, H., Zhang, G., ... & Ma, Y. (2023). Soil nutrients shape the composition and function of fungal communities in abandoned ancient rice terraces. Journal of Environmental Management, 329, 117064.

O'Brien, S., Gibbons, S., Owens, S., Hampton-Marcell, J., Johnston, E., Jastrow, J., Jack, G., Meyer, F., & Antonopoulos, D. (2016). Spatial scale drives patterns in soil bacterial diversity. Environmental microbiology, 18. https://doi.org/10.1111/1462-2920.13231

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). Vegan: Community ecology package. http://CRAN.R-project.org/package=vegan

Oldfield, E. E., Bradford, M. A., & Wood, S. A. (2019). Global meta-analysis of the relationship between soil organic matter and crop yields. SOIL, 5(1), 15-32.

https://doi.org/10.5194/soil-5-15-2019

Osman, K. T. (2013). Plant Nutrients and Soil Fertility Management. In K. T. Osman (Ed.), Soils: Principles, Properties and Management (pp. 129–159). Springer Netherlands. https://doi.org/10.1007/978-94-007-5663-2_10

Otero-Jiménez, V., Carreño-Carreño, J. del P., Barreto-Hernandez, E., van Elsas, J. D., & Uribe-Vélez, D. (2021). Impact of rice straw management strategies on rice rhizosphere microbiomes. Applied Soil Ecology, 167, 104036. https://doi.org/10.1016/j.apsoil.2021.104036

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Research, 42(Database issue), D206-214. https://doi.org/10.1093/nar/gkt1226

Pausch, J., & Kuzyakov, Y. (2018). Carbon input by roots into the soil: Quantification of rhizodeposition from root to ecosystem scale. Global Change Biology, 24(1), 1-12. https://doi.org/10.1111/gcb.13850

Phongchanmixay, S., Bounyavong, B., Khanthavong, P., Khanthavong, T., Ikeura, H., Matsumoto, N., & Kawamura, K. (2019). Rice plant growth and nutrient leaching under different patterns of split chemical fertilization on sandy soil using a pot. Paddy and Water Environment, 17(2), 91-99. https://doi.org/10.1007/s10333-019-00701-w

Pittol, M., Scully, E., Miller, D., Durso, L., Mariana Fiuza, L., & Valiati, V. H. (2018). Bacterial

Community of the Rice Floodwater Using Cultivation-Independent Approaches. International Journal of Microbiology, 2018, 6280484. https://doi.org/10.1155/2018/6280484

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., & Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. Nature Biotechnology, 35(9), Art. 9. https://doi.org/10.1038/nbt.3935

R Core Team. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www. R-project. org/.

Schroth, G., & Sinclair, F. L. (2003). Trees, Crops, and Soil Fertility: Concepts and Research Methods. CABI.

Smith, S., & Gallaher, C. M. (2019). Soil and Agriculture. En D. M. Kaplan (Ed.), Encyclopedia of Food and Agricultural Ethics (pp. 2211-2217). Springer Netherlands. https://doi.org/10.1007/978-94-024-1179-9_539

Speirs, L. B. M., Rice, D. T. F., Petrovski, S., & Seviour, R. J. (2019). The Phylogeny, Biodiversity, and Ecology of the Chloroflexi in Activated Sludge. Frontiers in Microbiology, 10. https://www.frontiersin.org/articles/10.3389/fmicb.2019.02015

Sun, W., Xiao, E., Pu, Z., Krumins, V., Dong, Y., Li, B., & Hu, M. (2017). Paddy soil microbial communities driven by environment-and microbe-microbe interactions: A case study of elevation-resolved microbial communities in a rice terrace. https://doi.org/10.1016/j.scitotenv.2017.08.275

Walling, E., & Vaneeckhaute, C. (2020). Greenhouse gas emissions from inorganic and organic fertilizer production and use: A review of emission factors and their variability. Journal of Environmental Management, 276, 111211.

https://doi.org/10.1016/j.jenvman.2020.111211

Wang, Q., Liang, A., Chen, X., Zhang, S., Zhang, Y., McLaughlin, N. B., ... & Jia, S. (2021). The impact of cropping system, tillage and season on shaping soil fungal community in a long-term field trial. European Journal of Soil Biology, 102, 103253.

Wang, W., Luo, X., Chen, Y., Ye, X., Wang, H., Cao, Z., Ran, W., & Cui, Z. (2019). Succession of Composition and Function of Soil Bacterial Communities During Key Rice Growth Stages. Frontiers in Microbiology, 10, 421. https://doi.org/10.3389/fmicb.2019.00421

Wang, X., He, T., Gen, S., Zhang, X.-Q., Wang, X., Jiang, D., Li, C., Li, C., Wang, J., Zhang, W., & Li, C. (2020). Soil properties and agricultural practices shape microbial communities in flooded and rainfed croplands. Applied Soil Ecology, 147, 103449. https://doi.org/10.1016/j.apsoil.2019.103449

Wang, L., & Huang, D. (2021). Soil ammonia-oxidizing archaea in a paddy field with different irrigation and fertilization managements. Scientific Reports, 11(1), 1-11.

Wei, T., & Simko, V. (2021). R package «corrplot»: Visualization of a Correlation Matrix. (Version 0.92). (Version 0.92). https://cran.r-project.org/web/packages/corrplot/citation.html

Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P., Paczian, T., Trimble, W. L., Bagchi, S., Grama, A., Chaterji, S., & Meyer, F. (2016). The MG-RAST metagenomics database and portal in 2015. Nucleic Acids Research, 44(D1), D590-D594. https://doi.org/10.1093/nar/gkv1322

Wu, Z., Liu, Q., Li, Z., Cheng, W., Sun, J., Guo, Z., Li, Y., Zhou, J., Meng, D., Li, H., Lei, P.,

& Yin, H. (2018). Environmental factors shaping the diversity of bacterial communities that promote rice production. BMC Microbiology, 18(1), 51. https://doi.org/10.1186/s12866-018-1174-z

Yadav, A. N. (2021). Soil Microbiomes for Sustainable Agriculture: Functional Annotation. Springer Nature.

Yuan, C. L., Zhang, L. M., Wang, J. T., Hu, H. W., Shen, J. P., Cao, P., & He, J. Z. (2019). Distributions and environmental drivers of archaea and bacteria in paddy soils. Journal of Soils and Sediments, 19, 23-37.

Xu, C., Li, Y., Hu, X., Zang, Q., Zhuang, H., & Huang, L. (2022). The influence of organic and conventional cultivation patterns on physicochemical property, enzyme activity and microbial community characteristics of paddy soil. Agriculture, 12(1), 121.

Zhang, Q., Li, Y., Xing, J., Brookes, P. C., & Xu, J. (2019). Soil available phosphorus content drives the spatial distribution of archaeal communities along elevation in acidic terrace paddy soils. Science of the Total Environment, 658, 723-731.

Zhong, Y., Hu, J., Xia, Q., Zhang, S., Li, X., Pan, X., Zhao, R., Wang, R., Yan, W., Shangguan, Z., Hu, F., Yang, C., & Wang, W. (2020). Soil microbial mechanisms promoting ultrahigh rice yield. Soil Biology and Biochemistry, 143, 107741. https://doi.org/10.1016/j.soilbio.2020.107741