University of São Paulo
"Luiz de Queiroz" College of Agriculture

From communities to genomes: a multifaceted approach to depict bacterial life in soils

Lucas Dantas Lopes

Thesis presented to obtain the degree of Doctor in Science. Area: Soil and Plant Nutrition

Piracicaba
2017

Lucas Dantas Lopes
Agronomist

From communities to genomes: a multifaceted approach to depict bacterial life in soils

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **FERNANDO DINI ANDREOTE**

Thesis presented to obtain the degree of Doctor in Science. Area: Soil and Plant Nutrition

Piracicaba
2017

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Lopes, Lucas Dantas

From communities to genomes: a multifacetd approach to depict bacterial life in soils / Lucas Dantas Lopes. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2017.

126 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Diversidade bacteriana 2. Microbioma radicular 3. *Pseudomonas* spp. associadas a plantas 4. Evolução genômica 5. Ecologia evolutiva I. Título

ACKNOWLEDGEMENTS

I would like to thank every person that helped me in this process, both personally and professionally.

Firstly, to my family for the best support and love that one could have: my bride Camila de S. Varize; my parents Kátia Maria F. Dantas and Paulo Cezar B. Lopes; my brother Luís Paulo D. Lopes and sister Luciana D. Lopes; my nephew David D. B. Ribeiro; my grandparents Nilma A. Fontes, Francisco José C. Dantas, Maria Lúcia D. Farra, Cleonice B. Lopes, and especially José Ives M. Lopes, who I lost during this journey. I also thank my aunts, uncles, cousins, other relatives and friends;

To my advisor Dr. Fernando D. Andreote, for the great support in everything I needed during my PhD course, the important teachings, scientific conversations and compatibility of ideas;

To Dr. Jeff Chang and Dr. Joyce Loper, for kindly accepting me in their groups last year (2016), for helping me to improve my research, and for everything that I learned in their labs;

To Edward Davis and Alexandra Weisberg for the great help in the computational and genomic analyses; Michele C. P. Silva for the help in Biolog and bioinformatics analyses; Qing Yan for making the mutant strains used in the last article of this work; Chih-Feng and Camila S. Varize for helping in the TECAN analyses; Dr. Tal Pupko for the help in evolutionary inferences based on phylogenetics; and everybody in the Soil Microbiology lab (ESALQ/USP), in Dr. Loper lab (ARS/USDA-USA) and in Dr. Chang lab (OSU-USA) for the friendly coexistence;

To my past advisors, Dr. Rodrigo Mendes and Dr. Marcelo F. Fernandes, who also highly contributed to my formation as a researcher;

To Soil and Plant Nutrition program of the University of São Paulo, where I did my PhD course; Oregon State University where I passed one year during my internship abroad; and CAPES for the scholarship granted during my PhD course and my internship abroad.

"All things flow, nothing abides. No man ever steps in the same river twice, for it's not the same river and he's not the same man. Nothing is permanent except change."

Heraclitus of Ephesus

CONTENTS

RESUMO

De comunidades a genomas: uma análise multifacetada para descrever a vida bacteriana nos solos

Desvendar a ecologia microbiana do solo é essencial para aumentar a produtividade agrícola sustentável. Estudos baseados em comunidades revolucionaram esse campo nas últimas décadas, mas ainda há muito a ser revelado. Esta tese propôs uma abordagem para aumentar a resolução desses estudos, combinando sequenciamento em larga escala de rDNA 16S e genômica populacional, com o objetivo de explorar mais a fundo as diferenças apontadas por análises de comunidades, assim como superar as limitações do uso de unidades taxonômicas operacionais (UTOs) como entidades ecológicas e introduzir o pensamento evolutivo na ecologia microbiana. Nossa principal meta foi entender as características que tornam as bactérias hábeis em colonizar a rizosfera de cana-de-açúcar ou viver no solo saprofiticamente. Rizosfera e solo são hábitats contrastantes para a vida microbiana, já que são altamente distintos em suas características físicas, químicas e, consequentemente, biológicas. Nossos resultados indicaram que a cana-de-açúcar modifica o microbioma da rizosfera e o metabolismo do ácido D-galacturônico é uma função chave para colonizar este nicho. Dentre os táxons que prevalecem na rizosfera, o gênero *Pseudomonas* foi escolhido para um estudo mais detalhado, considerando os seus atributos de promoção de crescimento de plantas. Setenta e seis *Pseudomonas* spp. fluorescentes foram isoladas e submetidas ao sequenciamento do genoma. Uma análise de genômica comparativa foi realizada entre as populações obtidas do solo e rizosfera. As análises filogenéticas classificaram os isolados nos grupos *P. fluorescens* (57) ou *P. putida* (19). Doze prováveis novas espécies e dois novos subgrupos propostos de *P. fluorescens* foram encontrados no solo tropical prospectado. A genômica comparativa revelou que genes de fosfatases e de uso de xilose foram significativamente enriquecidos nas populações da rizosfera e solo do grupo *P. fluorescens*, respectivamente. O catabolismo do ácido D-galactônico foi maior na população da rizosfera do grupo *P. putida*, baseado tanto em resultados genotípicos quanto fenotípicos. O crescimento em D-xilose foi mais explorado usando linhagens geneticamente modificadas e confirmou que este açúcar é mais utilizado por membros da população do solo do que da rizosfera no grupo *P. fluorescens*, um padrão também observado no microbioma do solo. Em resumo, essas descobertas constituem um passo adiante no entendimento da ecologia bacteriana do solo e rizosfera, por superar algumas limitações de análises de comunidades e mostrar diferenças genômicas entre populações bacterianas destes hábitats.

Palavras-chave: Diversidade bacteriana; Microbioma radicular; *Pseudomonas* spp. associadas a plantas; Evolução genômica; Ecologia evolutiva

ABSTRACT

From communities to genomes: a multifaceted approach to depict bacterial life in soils

Unraveling soil microbial ecology is essential for improving sustainable agricultural productivity. Community-based studies revolutionized this field in the last decades, but much is yet to be disclosed. This thesis proposed an approach to increase the resolution of such studies by combining 16S rDNA high-throughput sequencing and population genomics, aiming to further explore the differences pointed by community analyses, as well as to overcome the limitations of using operational taxonomic units (OTUs) as ecological entities, and to introduce the evolutionary thinking in microbial ecology. Our main goal was to understand the features that make bacteria able to colonize sugarcane rhizosphere or live saprophytically in bulk soil. Rhizosphere and bulk soil are contrasting habitats for microbial life as they are highly distinct in its physical, chemical and consequently biological characteristics. Our results indicated that sugarcane shapes the rhizosphere microbiome and metabolism of D-galacturonic acid is a key function for colonizing this niche. Among the taxa prevailing in the rhizosphere, *Pseudomonas* genus was targeted for a more detailed study considering its known attributes for plant growth promotion. Seventy-six fluorescent *Pseudomonas* spp. were isolated and submitted to whole genome sequencing (WGS). A comparative genomic analysis was performed between populations from rhizosphere and bulk soil. Phylogenetic analyses classified the isolates in the *P. fluorescens* (57) or *P. putida* (19) groups. Twelve putative new species and two new proposed *P. fluorescens* subgroups were found in the prospected tropical soil. Comparative genomics revealed that phosphatases or xylose-utilization genes were significantly enriched in the rhizosphere and bulk soil populations of the *P. fluorescens* group, respectively. D-galactonate catabolism was higher in the rhizosphere population of the *P. putida* group based on both genotypic and phenotypic results. Growth in D-xylose was further explored using genetic modified strains and confirmed that this sugar is more used by members of the bulk soil than the rhizosphere population of the *P. fluorescens* group, a pattern also observed in the bulk soil microbiome. In summary, these findings constitute a step forward in understanding the ecology of rhizosphere and bulk soil bacteria, by overcoming some limitations of community-based analyses and showing genomic differences between bacterial populations of these habitats.

Keywords: Bacterial diversity; Root microbiome; Plant-associated *Pseudomonas* spp.; Genome evolution; Evolutionary ecology

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Soil is the basis for macroscopic life, supporting the growth and health of animals and plants. The physical and chemical characteristics of soils promoting high crop production are widely known, while the biological features related to this issue are less known, although very important. Thus, understanding soil microbial ecology is pivotal for improving crops productivity. Additionally, microbial utilization has the potential to replace expensive or polluting substances, reducing environmental impact and promoting sustainability. In this context, it is necessary to investigate the different spots in soils available for microbial activity (*i.e.* niches), in order to unravel how microbes live and act in these soil fractions aiming to manage soil and cropping systems favoring microbial beneficial activities.

The most distinct microbial habitats in the soil are the rhizosphere, *i.e.* the nutritionally rich layer of soil influenced by rhizodeposition; and the bulk soil, *i.e.* the layer of soil without presence or influence of roots, known as poor in nutrients availability. There are increasing evidences that bulk soil and rhizosphere host different microbial communities, selecting for taxa able to live and spread under the specific conditions of each habitat. The rhizosphere microbiome is characterized by the presence of mutualist microbes able to help plant growth and development.

Sugarcane is a crop adapted to tropical countries like Brazil, the largest producer in the world, where the harvested material is mainly destined to sugar and bioethanol production. Targeting the microbial ecology of bulk soil and rhizosphere communities of sugarcane is thus important as any advance in this subject would have a great impact in the production of this important crop, here used as a model plant. In order to achieve innovative information, the particularities of bulk soil and rhizosphere need to be investigated regarding the genome content of populations, which knowledge lacks in the literature either for sugarcane or any other plant. Assessing the existence of genomic shifts in phylogenetically closely related bacteria inhabiting contrasting habitats allows inferences in the differential evolution of populations, besides in community structuring. Following this reasoning, the same group of bacteria could be able to perform distinct activities in different habitats. Thus, the population genomics approach allows inferences on the ecology of bacteria in a higher level of resolution.

Pseudomonas is a bacterial genus ideal for this study, as it is widely studied, it is composed by a high number of species, has a great availability of genome sequences in databases, hosts divergent species in accordance to ecological and functional diversity; it is

easy to culture; it is characterized by a high genome mobility; and it has several plant growth promoting rhizobacteria (PGPR) species (mainly the fluorescent *Pseudomonas* spp.). In addition, prospecting *Pseudomonas* spp. in tropical soils increases the possibility of finding new species, as most part of the strains available in the databases came from countries of temperate regions.

The general objective of this study was to understand the ecological differences between sugarcane rhizosphere and bulk soil regarding microbial communities and bacterial populations, more specifically of *Pseudomonas* spp. For that, we split this thesis in 4 sections, each one specifying a particular issue. The section 2 is a literature review, showing concepts and previous studies, clarifying why choosing fluorescent *Pseudomonas* as a model for our study, and describing the lacks in the knowledge and future prospects in this research field. The section 3 is an article that assessed the bacterial communities of sugarcane rhizosphere and bulk soil, aiming to identify the taxa enriched in each habitat, and to understand the bacterial features needed for rhizosphere colonization, analyzing the microbiome as a whole. The section 4 is an article with a parallel goal. The aim was to explore the diversity of fluorescent *Pseudomonas* spp. in the same Brazilian tropical soil under sugarcane cultivation, using genomic and phylogenetic approaches in order to find new species and genetic diversity in this genus. Finally, the section 5 is an article that combined the outcomes of the two previous articles. It identified differences in genome content of the fluorescent *Pseudomonas* spp. between populations from bulk soil and rhizosphere, and those differences were further analyzed to strengthen and confirm the ecological distinctions suggested in this study. The final part of this section was based on phenotypic analyses using genetically modified strains, revealing that the prevalence of specific populations in the bulk soil or rhizosphere could be determined by few genes.

2 LITERATURE REVIEW

2.1 Soil Microbial Ecology

Soils are the results of rock weathering, accumulation of organic material and living organisms. Edaphology is the science that studies soils in the context of plant development, the primary source of carbon and energy for an intricate food chain that sustain the terrestrial macroscopic life (Weil and Brady, 2002). In addition to solids, soil is also composed of air and water, and the last contains all the solubilized nutrients needed for root absorption and plant growth, named the soil solution. Soil is a dynamic system where nutrients and organic material undergoes innumerous transformations, highly mediated by microbial activity. Furthermore, microorganisms can associate to plants promoting growth and health. Therefore, microbial life is an essential part of soil in the edaphological context, and it is important to be studied, providing innovation for improving agriculture (Weil and Brady, 2002; Paul, 2014).

Soils are not uniform but highly heterogeneous, changing on its physical and chemical characteristics along space and time. Consistently, the distribution of organic substrates used for microbial metabolism is randomly dispersed, creating the microbial hotspots – the sites of high microbial activity (Kuzyakov and Blagodatskaya, 2015). The soil hotspots comprise particular ecological niches colonized by specific microbial communities. Several hotspots are available for microbial colonization in soil, e.g. the organic matter deposited by dead organisms or its residues (detritusphere); the surface of aggregates which are structured by binding organic molecules; the body surfaces of meso and macroorganisms (biopores); and the soil surrounding plant roots (rhizosphere), an environment rich in carbon derived from plant rhizodeposition (Kuzyakov and Blagodatskaya, 2015).

Microbial ecology studies the relation between microorganisms and their natural environments, nominating microbial traits or environmental characteristics supporting the presence of specific communities or populations in each occupied environment. Microbes are being studied for centuries by culturing in growth media and much progress was achieved in understanding its biology. However, although the knowledge in microbial ecology has slowly progressed in the past, the last decades were marked by a great number of studies on this subject (Snyder et al., 2009), mainly supported by the upgrade in the analytical methodologies, which allowed the assessment of uncultured microbes in soils. Sequencing technologies got more resolution power, more accuracy and decreased its costs, becoming accessible for the scientific community (Shokralla et al., 2012). In parallel, data analyses are

becoming more robust and faster, allowing deeper and more detailed studies on microbial ecology (Pallen, 2016).

The advance in omics and culture independent approaches also generated new concepts in soil microbial ecology. It improved our understanding of the plant microbiome, and pointed for distinctions among each plant compartment (phyllosphere, endosphere, rhizosphere) regarding the associated microbial communities (Turner et al., 2013; Andreote et al., 2014). Also, these approaches showed how soil microbial communities change when cropping different plant species (Kuske et al., 2002), when changing the soil characteristics (Rousk et al., 2010), or when using distinct soil managements (Hartmann et al., 2015). Culture independent studies on soil microorganisms revealed biogeography patterns for microbes (Fierer et al., 2006; Fierer et al., 2012; Gumiere et al., 2016), and supported the development/application of theories and models to exclusively describe the specific ecology of microorganisms (Fierer et al., 2007; Faust and Raes, 2012). Nevertheless, much yet remains to be deciphered in microbial ecology: the limitations of the current methodologies are starting to slow down the advances, and new methods/approaches have to be developed to overcome these problems, avoiding the plateau in knowledge, and supporting the continuous improving in the knowledge of soil microbial ecology.

2.2 Microbial communities in rhizosphere and bulk soil

From several microbial habitats of the soil environment, rhizosphere and bulk soil are the most distinct ones. Rhizosphere is defined as the root surfaces and the soil surrounding it, where a constant and rich supply of carbon is deposited by plants (Prashar et al., 2014). The carbon sources are mainly formed by low molecular weight compounds, which can be easily used for microbial metabolism. Simple sugars, amino acids and organic acids are components of root exudates, and this carbon abundance contributes to the enrichment of microbes in this hotspot (Dennis et al., 2010). On the other hand, bulk soil is the layer of soil not reached by the roots or rhizodeposition. Therefore, it is nutrionally poorer than rhizosphere, predominantly composed of complex and recalcitrant organic molecules, resulting from the humification process of soil organic matter (Dennis et al., 2010; Paul, 2014). The bulk soil is considered an oligotrophic environment, where k-strategists seem to be favored (Fierer et al., 2007).

The shaping on rhizosphere and bulk soil microbial communities is determined by many factors related to microbial and environmental characteristics. Rhizosphere communities are more influenced by the plant hosts (Aira et al., 2010; Philippot et al., 2013),

where the microbiome is enriched in microbial groups able to support plant growth, development and health (Mendes et al., 2013). These microbes associate to the plants, both symbiotically – in the case of nitrogen fixing rhizobia and phosphate transporter mycorrhizal fungi – or mutualistically, in the case of plant growth promoting rhizobacteria (PGPR) (Philippot et al., 2013; Mendes et al., 2013). PGPR are able to perform several direct or indirect beneficial activities for plants, e.g. produce phytohormones, solubilize inorganic phosphate, fix atmospheric nitrogen, promote abiotic stress tolerances, defend against pathogens, among others (Berendsen et al., 2012; Mendes et al., 2013). However, the rhizosphere microbiome in not only composed by beneficial organisms, but it has neutral, commensal, plant pathogen and even opportunistic human pathogen microbes (Berendsen et al., 2012; Mendes et al., 2013).

The rhizosphere microbial communities of several plants are different from bulk soil, indicating that the taxonomic and/or functional selection is divergent in the two habitats, but also specific for each plant species. Several studies using distinct plant species suggested these shifts and disclosed the specificities for many plants in different regions (Berg and Smalla, 2009; Uroz et al., 2010; Mendes et al., 2014). However, more specific information about microbes and populations found in this habitat are needed. For example, the knowledge about the set of bacterial features required for rhizosphere colonization is still incomplete; it was not shown whether intra-species variations result in different rhizosphere competences; and the core or minimal microbiome essential for a satisfactory and healthy development was not disclosed for any plant. Correlation between sequencing data and microbial phenotypes is mandatory to better explore the ecological differences between microbial communities. The community level physiological profile (CLPP) is a method widely used in the last decades, and can be applied to assess the phenotypic differences between bulk soil and rhizosphere communities. In combination to metagenomics, this method can support the identification of microbial functions important for the rhizosphere colonization (Garland, 1997).

Studies based on community ecology have methodological limitations. Sequencing of molecular biomarkers have no taxonomic resolution for the species level; metagenomic or even metatranscriptomic data cannot provide accurate information about the phenotype of microbes in each distinct habitat; and culture independent techniques have no accuracy to understand the influence of evolutionary changes on the ecological differences. Therefore, there is a demand for new approaches that investigate deeply, more accurately and with higher resolution the ecological differences of microorganisms in the bulk soil and in the rhizosphere. The continuous decline in sequencing costs is making possible the analysis of

several genome sequences in a single study, supporting inferences about the ecology of microbial populations in the environment (Shapiro et al., 2012; Cordero and Polz, 2014).

In this study, we selected sugarcane as the model plant to investigate this issue, aiming to assess the occurrence of the rhizosphere shaping in the bacterial communities and specific populations associated to this plant species, and to provide more detailed information about the occupation of this habitat. The selection of this crop was based on the fact that Brazil is its largest world producer, and it is one of the most cultivated crops in this country (Rudorff et al., 2010). Sugarcane is used for many purposes, mainly for sugar and bioethanol production, supplying world demands for both food and renewable energy (Loarie et al., 2011). Studying microbial ecology of economically relevant crops is important for generating knowledge that can have both theoretical and practical applications. In combination to the selection of sugarcane as a model, we also selected a bacterial genus as a model for the development of this thesis, aiming to compare the genomic and functional profiles between isolates of the contrasting soil habitats discussed.

2.3 The genus *Pseudomonas*: diversity and agricultural importance

The genus *Pseudomonas* is allocated in the class *Gammaproteobacteria*, composed of gram-negative bacteria mostly characterized by a heterotrophic and aerobic metabolism, and a copiotrophic ecological strategy (r-strategists), quickly responding to nutrient availability (Garbeva et al., 2004). This genus is considered ubiquitous, found in a multitude of terrestrial and aquatic environments due to the existence of a huge species diversity (Palleroni, 1992; Silby et al., 2011). Consistently, *Pseudomonas* also has a high genetic, ecological, and physiological diversity (Spiers et al., 2000). Among the species hosted in this genus, it is possible to find soil saprophytes; xenobiotic degraders; mutualists or pathogens associated to plants, animals, humans, arthropods; and species used for biotechnological application in the industry, agriculture and bioremediation of contaminated environments (Silby et al., 2011; Loper et al., 2012). The great versatility of the genus is also related to its high genome mobility and horizontal gene transfer events (HGT), allowing its fast adaptation and diversification (Silby et al., 2011)

The high genetic diversity of this genus makes its taxonomical classification challenging and constantly revisited. In the last decades, the generation of new analytical methods allowed the description of many species previously classified as *Pseudomonas* spp. in new genera, families and even classes, like *Sphingomonas* spp., *Xanthomonas* spp., *Burkholderia* spp., *Ralstonia* spp., and *Comamonas* spp. (Swings et al., 1983; Yabuuchi et al.,

1990; Kersters et al., 1996). However, even after that, the genus *Pseudomonas* still contains multiple and diverse species misclassified in the past, making its systematics often problematic. To overcome these problems, the *Pseudomonas* genus is currently split in 10 phylogenetic groups, each containing many related species, which were named as *P. fluorescens*, *P. aeruginosa*, *P. putida*, *P. syringae*, *P. anguilliseptica*, *P. oryzihabitans*, *P. stutzeri*, *P. oleovorans*, *P. straminae*, and *P. pertucinogena* groups (Gomila et al., 2015; Garrido-Sanz et al., 2016).

In agricultural systems, the fluorescent *Pseudomonas* spp. are the main targets of studies and biotechnological utilization, commonly related to soil processes of nutrients cycling and pollutants degradation, or associated to plants as PGPR (Silby et al., 2011). Among the beneficial activities that fluorescent *Pseudomonas* spp. perform to plants, we can highlight phytohormones synthesis (Patten and Glick, 2002), production of iron-chelating siderophores (Trapet et al., 2016), protection against abiotic stresses (Cho et al., 2015), phosphate solubilization (de Werra et al., 2009), and mainly biocontrol against insects or microbial pathogens (Hofte and Altier, 2010; Loper et al., 2012; Hol et al., 2013). The beneficial fluorescent species of *Pseudomonas* are classified in the *P. putida* and mostly in the *P. fluorescens* groups. Nevertheless, the diversity of species in the *P. fluorescens* group is also huge and several strains were classified with incorrect species names. Thus, the species of this group are now classified inside 10 subgroups in order to organize its systematics, i.e. *P. fluorescens*, *P. koreensis*, *P. jessenii*, *P. protegens*, *P. corrugata*, *P. fragi*, *P. asplenii*, *P. gessardii*, *P. chlororaphis*, and *P. mandelii*, based on their degree of phylogenetic and genomic relatedness (Gomila et al., 2015; Garrido-Sanz et al., 2016).

It has been shown that *Pseudomonas* spp. are distinctly selected by plant rhizosphere in comparison to bulk soil (Lemanceau, 1994; Costa et al., 2007; Garcia-Salamanca et al., 2012), and the same could be true for sugarcane. If it is the case, this genus constitutes a great candidate to perform a deeper analysis comparing the communities and populations inhabiting bulk soil and rhizosphere. A comparative genomic analysis aiming to identify differences in gene frequencies between populations of the two contrasting habitats could reveal key features for understanding their ecology. These bacteria are easy to isolate and culture (the first requirement of this approach); they have a high species, genetic, ecological, and functional diversity; they are present in the soil and plant environments; they are agriculturally relevant since several species are PGPR; they are differentially selected by rhizosphere and bulk soil in many plants; there are several genomic sequences available in

public databases; and at last, they have a high genome mobility, which is important for the analysis of genome content and short-term evolution.

In addition to the study of comparative genomics between bulk soil and rhizosphere populations, the prospection of fluorescent *Pseudomonas* spp. from tropical soils is important for other reasons. Cho and Tiedje (2000) suggested a degree of endemicity of fluorescent *Pseudomonas* spp. in soils caused by biogeographical patterns. Hence, yet *Pseudomonas* genus is ubiquitous, its species are probably unequally distributed in the regions of the planet. As few studies targeted this genus in the tropical compared to the temperate regions, there is a high probability of finding new fluorescent *Pseudomonas* species in Brazilian soils. Moreover, the climate, plant diversity, geology and consequently the soils vary more in the tropics than in the temperate and arctic regions, which amplifies the potential of finding a new fluorescent *Pseudomonas* spp. diversity in these soils (Kalpage, 1974). Accessing the genome sequences of tropical soil isolates have the potential to accurately describe and disclose not only new species, but also their functional potential and perhaps discover genetic-based novelties for application in agriculture.

2.4 Evolutionary ecology of bacterial populations

Genetic variability is common among individuals of the same species, either for prokaryotes or eukaryotes. The difference between them is the source for genetic variation, which comes from sexual reproduction in eukaryotes, and gene losses or gains in prokaryotes (Snel et al., 2002; Jain et al., 2003; Bolotin and Hershberg, 2015). The genetic variability found among different genotypes of a given species compose the pan-genome, which is defined as the full set of genes that compose a prokaryotic species (Medini et al., 2005; Tetelin et al., 2008). It is split in core genome, *i.e.* the set of genes shared by all individuals in a species; and the dispensable or accessory genome, *i.e.* the set of genes not present in all individuals of a species. The core genome is considered essential for the primary survival of any individual of a given species, while the accessory genome is responsible for the intra-species diversity resulting from the genetic variation among individuals, and possibly related to the niche occupation of different genotypes (Medini et al., 2005; Tetelin et al., 2008).

The general feeling when considering microbial populations is that they are composed by genetically identical individuals in natural environments, what is commonly found in pure cultures of microbes under laboratorial conditions. However, bacteria can live in genetically diverse populations in the environment (Cordero and Polz, 2014), and the better

comprehension of the microbial pan-genome is mandatory to properly infer on the ecology of bacterial (or archaeal) populations.

Population is defined as the set of coexisting individual genotypes of the same species (Cordero and Polz, 2014). However, it becomes complicated when considering that the concept of bacterial species is not fully accepted and applicable due to the lack of sexual reproduction in prokaryotes, which makes it more difficult to separate bacterial populations (Gevers et al., 2005; Fraser et al., 2009). Thus, the taxonomic classification of bacteria is more operational than biological, differing from the classification of species in sexual eukaryotes.

The primary sources of genetic variation for both eukaryotic and prokaryotic organisms are the single nucleotide polymorphisms (SNPs), caused by point mutations and homologous recombination (Cordero and Polz, 2014). However, the process from which the genetic variations spread in the populations is distinct between sexual eukaryotes and prokaryotes. In most eukaryotes, sexual reproduction spreads the new alleles in the populations generating different genotypes, while in prokaryotes gene losses or gains are the main processes to accelerate the intra-species genotypic diversity (Jain et al., 2003; Cordero and Polz, 2014). A gene can be lost in a bacterial chromosome by many ways: *e.g*. accumulating mutations that obliterate the function of the gene, or physical gene loss by heterologous recombination. Genes without function tend to be also physically lost during bacterial cell division, by neutral or selection processes, since pseudogenes were shown to be an exception in these organisms (Bolotin and Hershberg, 2015). On the other hand, gain of new genes is dependent on HGT processes, where bacteria can receive new alleles either from free DNA in the environment or from other donor species, promoted by heterologous or homologous recombination. The main described HGT methods in literature are conjugation (transfer of plasmids by mating), transformation (acquisition of free DNA from the environment), and transduction (acquisition of genes by the infection of phages) (Jain et al., 2003). In both prokaryotes and eukaryotes, the phenotype fitness of the new allele can determine if its frequencies will increase or decrease in populations. Thus, genes are suggested to be the units of selection, since they perpetuate or disappear in the gene pool of populations based on the phenotype advantage that they confer to the host organisms, a feature observed in all forms of life (Dawkins, 1989). Inversely, genes without phenotype change its frequencies in populations based on neutral processes (Cordero and Polz, 2014).

Having these concepts in mind, it is possible to correlate the evolution and ecology of bacteria. Ecology and evolution are highly connected phenomena that take place together,

determining the niche occupation of individuals in a given species. One could separate the ecological selection, usually mentioned in microbiome studies as the differential shaping in community structure between treatments, hosts or environments; from the evolutionary selection, also named natural selection, the Darwinian phenomena regarding the changes along time in species. Actually, they can be considered the same process occurring in different time scales (Pianka, 2011). While microbial community studies have no resolution to show evolutionary differences, the studies of population genomics are able to infer on both evolutionary and ecological effects at the same time. The selection that favors some members of a community also favors some genotypes inside those taxa, causing changes in gene frequencies between populations occupying different habitats. In counterpart, the further diversification changes after niche partitioning increase the specificity of each subpopulation, becoming more adapted to the new niche (Falush, 2009; Cordero and Polz, 2014).

Evolutionary ecology studies the genetic changes in populations related to niche adaptation, focusing on the ecological factors affecting evolution and vice-versa, constituting an intersection of these two fields of biology (Pianka, 2011). Evolutionary ecology of bacteria is on its infancy, as only in the last decade scientists had the opportunity to study several genome sequences at the same time. Some studies addressed this issue and successfully showed eco-evolutionary differences in bacterial populations, *e.g. Prochlorococcus* spp. (Johnson et al., 2006), *Vibrio* spp. (Hunt et al., 2008), and *Bacillus* spp. (Connor et al., 2010). Analysis of 19 fluorescent *Pseudomonas* spp. isolates also showed genomic and phenotypic differences between populations isolated from endosphere or rhizosphere compartments of *Populus deltoides* (Timm et al., 2015). Hundreds (698) of fluorescent *Pseudomonas* spp. isolates were phenotypically analyzed between bulk soil and rhizosphere, but any function was found significantly enriched in the populations of the compared habitats (Vacheron et al., 2016). Nevertheless, there is no genome comparison between populations of different soil habitats (nor for *Pseudomonas* spp. or any other microbial genus or species). Therefore, there is an open avenue to unraveling key eco-evolutionary differences between microbial populations inhabiting soil environment. In this thesis, it is expected to achieve a better understanding of bacterial traits needed for the occupation of the bulk soil or rhizosphere habitats. The following sections of this thesis assessed it, starting with a community ecology study and culminating in a population genomics comparison of fluorescent *Pseudomonas* spp. inhabiting rhizosphere or bulk soil.

References

Aira M, Gómez-Brandón M, Lazcano C, Baath E, Domínguez J (2010) Plant genotype strongly modifies the structure and growth of maize rhizosphere microbial communities. *Soil Biol Biochem*, 42: 2276-2281

Andreote FD, Gumiere T, Durrer A (2014) Exploring interactions of plant microbiomes. *Scientia Agricola*, 71: 528–539. 10.1590/0103-9016-2014-0195

Berendsen RL, Pieterse CMJ, Bakker PAHM (2012) The rhizosphere microbiome and plant health. *Trends Plant Sci*, 17(8): 478-486

Berg G, Smalla K (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol*, 68: 1-13

Bolotin E, Hershberg R (2015) Gene loss dominates as a source of genetic variation within clonal pathogenic bacterial species. *Genome Biol Evol*, 7(8): 2173-2187

Cho JC, Tiedje JM (2000) Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl Env Microbiol*, 66(12): 5448-5456.

Cho ST, Chang HH, Egamberdieva D, Kamilova F, Lugtenberhg B, Kuo CH (2015) Genome analysis of *Pseudomonas fluorescens* PCL1751: a rhizobacterium that controls root diseases and alleviates salt stress for its plant host. *PLoS ONE*, 10(10): e0140231. doi:10.1371/journal.pone.0140231.

Connor N, Sikorski J, Rooney AP, Kopac S, Koeppel AF, Burger A, Cole SG, Perry EB, krizank D, Field NC, Slaton M, Cohan FM (2010) Ecology of speciation in the genus *Bacillus*. *Appl Environ Microbiol*, 76(5): 1349-1358

Cordero OX, Polz MF (2014) Explaining microbial genomic diversity in light of evolutionary ecology. *Nat Rev Microbiol*, 12: 263-272

Costa R, Gomes NCM, Krogerrecklenfort E, Opelt K, Berg G, Smalla K (2007) *Pseudomonas* community structure and antagonistic potential in the rhizosphere: insights gained by combining phylogenetic and functional gene-based analyses. *Environ Microbiol*, 9(9): 2260-2273

Dawkins R (1989) *The selfish gene*. Oxford, Oxford University Press

de Werra P, Peche-Tarr M, Keel C, Maurhofer M (2009) Role of gluconic acid production in the regulation of biocontrol traits of *Pseudomonas fluorescens* CHA0. *Appl Env Microbiol*, 75(12): 4162-4174.

Dennis PG, Miller AJ, Hirsch PR (2010) Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiol Ecol*, 72: 313-327

Falush D (2009) Toward the use of genomics to study microevolutionary change in bacteria. *PLoS Genetics*, 5(10): e1000627. doi:10.1371/journal.pgen.1000627

Faust K, Raes J (2012) Microbial interactions: from networks to models. *Nat Rev*, 10: 538-550

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A*, 103(3): 626-631

Fierer N, Bradford MA, Jackson RB (2007) Toward an ecological classification of soil bacteria. *Ecology*, 88(6): 1354-1364

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG. Cross-biome metagenomics analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci U S A*, 109 (52): 21390-21395

Fraser C, Alm EJ, Polz MF, Spratt BG, Hanage WP (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science*, 323: 741-746

Garcia-Salamanca A, Molina-Henares MA, Dillewijn P, Solano J, Pizarro-Tobias P, Roca A, Duque E, Ramos JL (2012) Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil. *Microb Biotechnol*, 6(1): 36-44

Garland JL (1997) Analysis and interpretation of community-level physiological profiles in microbial ecology. *FEMS Microbiol Rev*, 24: 289-300

Garbeva P, van Veen JA, van Elsas JD (2004) Assessment of the diversity, and antagonism towards *Rhizoctonia solani* AG3, of *Pseudomonas* species in soil from different agricultural regimes. *FEMS Microbiol Ecol*, 47: 51-64

Garrido-Sanz D, Meier-Kolthoff JP, Göker M, Martín M, Rivilla R, Redondo-Nieto M (2016) Genomic and genetic diversity within the *Pseudomonas fluorescens* complex. *PLoS One*, doi:10.1371/journal.pone.0150183.

Gevers D, Cohan FM, Lawrence JG, Spratt BG, Coenye T, Feil EJ, Stackebrandt E, Van de Peer Y, Vandamme P, Thompson FL, Swings J (2005) Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol*, 3(9): 733-739

Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E (2015) Phylogenomics and systematics in *Pseudomonas*. *Front Microbiol* 6(214), doi:10.3389/fmicb.2015.00214

Gumiere T, Durrer A, Bohannan B, Andreote FD (2016) Biogeographical patterns in fungal communities from soils cultivated with sugarcane. *J Biogeogr*, 43: 2016-2026

Hartmann M, Frey B, Mayer J, Mader P, Widmer F (2015) Distinct soil microbial diversity under long-term organic and conventional farming. *ISME J*, 9: 1177-1194

Hofte M, Altier N (2010) Fluorescent pseudomonads as biocontrol agents for sustainable agricultural systems. *Res Microbiol*, 161(6): 464-471.

Hol WHG, Beemer TM, Biere A (2013) Getting the ecology into interactions between plants and the plant growth-promoting bacterium *Pseudomonas fluorescens*. *Front Plant Sci*, 4(81): doi:10.3389/fpls.2013.00081.

Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science*, 320(5879): 1081-1085

Jain R, Rivera MC, Moore JE, Lake JA (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol*, 20: 1598–1602. 10.1093/molbev/msg154

Johnson ZI, Zinser ER, Coe A, McNulty NP, Woodward EM, Chisholm SW (2006) Niche partitioning among *Prochlorococcus* ecotypes along ocean-scale environmental gradients. *Science*, 311(5768): 1737-1740

Kalpage FSCP (1974) *Tropical soils: classification, fertility and management*. St. Martin's Press, New York, USA, 294 pp.

Kersters K, Ludwig W, Vancanneyt M, DeVos P, Gillis M, Schleifer KH (1996) Recent changes in the classification of the pseudomonads: an overview. *Syst Appl Microbiol*, 19: 465-477

Kuske CR, Ticknor LO, Miller ME, Dunbar JM, Davis JA, Barns SM, Belnap J (2002) Comparison of soil bacterial communities in rhizosphere of three plant species and the interspaces in an arid grassland. *Appl Environ Microbiol*, 68(4): 1854-1863

Kuzyakov Y, Blagodatskaya E (2015) Microbial hotspots and hot moments in soil: concept & review. *Soil Biol Biochem*, 83: 184-199

Lemanceau P, Corberand T, Gardan L, Latour X, Daguerre G, Boeufgras JM, Alabouvette C (1995) Effect of two plant species, flax (*Linum usitatissinum* L.) and tomato (*Lycopersicon esculentum* Mill.), on the diversity of soil borne populations of fluorescent Pseudomonads. *Appl Environ Microbiol*, 61(3): 1004-1012

Loarie SR, Lobell DB, Asner GP, Mu Q, Field CB (2011) Direct impacts on local climate of sugar-cane expansion in Brazil. *Nature Climate Change*, 1: 105-109

Loper JE, Hassan KA, Mavrodi DV, Davis EW, Lim CK, Shaffer BT, Elbourne LDH, Stockwell VO, Hartney SL, Breaklwell K, Henkels MD, Tetu SG, Rangel LI, Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE, Paulsen IT (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet*, 8(7): e1002784.

Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. *Curr Opin Genet Dev*, 15: 589-594

Mendes LW, Kuramae EE, Navarrete AA, van Veen JA, Tsai SM (2014) Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J*, 8: 1577–1587. 10.1038/ismej.2014.17

Mendes R, Garbeva P, Raaijmakers JM (2013) The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Rev*, 37: 634-663

Pallen MJ (2016) Microbial bioinformatics 2020. *Microb Biotechnol*, 9(5): 681-686

Palleroni NJ (1992) Introduction to the Pseudomonadaceae. *The prokaryotes, a handbook on the biology of bacteria, ecophysiology, isolation, identification and applications*, vol. III, 2nd ed (Balows A, Truper HG, Dworkin M, Harder W, Schlepper KH, eds), 3071-3085. Springer, New York.

Paul EA (2014) *Soil microbiology, ecology and biochemistry*. Academic press.

Patten CL, Glick BR (2002) Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system. *Appl Env Microbiol*, 68(8): 3795-3801.

Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH (2013) Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol*, 11: 789-799

Pianka ER (2011) *Evolutionary ecology*. Eric R. Pianka.

Prashar P., Kapoor N., Sachdeva S. (2014). Rhizosphere: its structure, bacterial diversity and significance. *Rev Environ Sci Biotechnol* 13 63–77. 10.1007/s11157-013-9317-z

Rousk J, Baath E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J*, 4: 1340-1351

Rudorff BFT, Aguiar DA, da Silva WF, Sugawara LM, Adami M, Moreira MA (2010) Studies on the rapid expansion of sugarcane for ethanol production in São Paulo state (Brazil) using landsat data. *Remote Sens*, 2(4): 1057-1076

Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, Polz MF, Alm EJ (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science*, 336: 48-51

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Mol Ecol*, 21: 1794-1805

Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev*, 35(4): 652-680.

Snel B, Bork P, Huynen MA (2002) Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res*, 12:17-25

Snyder LAS, Loman N, Pallen MJ, Penn CWP (2009) Next-generation sequencing – the promise and perils of charting the great microbial unknown. *Microb Ecol*, 57:1-3

Spiers AJ, Buckling A, Rainey PB (2000) The causes of *Pseudomonas* diversity. *Microbiol*, 146: 2345-2350

Swings J, de Vos P, van den Mooter M, de Ley J (1983) Tranfer of *Pseudomonas maltophilia* Hugh 1981 to the genus *Xanthomonas* as *Xanthomonas maltophilia* (Hugh 1981) comb. nov. *Int J Syst Bacteriol*, 33: 409-413

Timm CM, Campbell AG, Utturkar SM, Jun SR, Parales RE, Tan WA, Robeson MS, Lu TYS, Jawdy S, Brown SD, Ussery DW, Schadt CW, Tuskan GA, Doktycz MJ, Weston DJ, Pelletier DA (2015) Metabolic functions of Pseudomonas fluorescens strains from *Populus deltoides* depend on rhizosphere or endosphere isolation compartment. *Front Microbiol*, 6: 1118, doi: 10.3389/fmicb.2015.01118

Tetelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol*, 12: 472-477

Turner TR, James EK, Poole PS (2013) The plant microbiome. *Genome Biol*, 14:209 10.1186/gb-2013-14-6-209

Trapet P, Avoscan L, Klinguer A, Pateyron S, Citerne S, Chervin C, Mazurier S, Lemanceau P, Wendehenne D, Besson-Bard A (2016) The *Pseudomonas fluorescens* siderophore pyoverdine weakens *Arabidopsis thaliana* defense in favour of growth in iron-deficient conditions. *Plant Physiol*, pp. 15.01537, doi:http://dx.doi.org/10.1104

Uroz S, Buée M, Murat C, Frey-Klett P, Martin F (2010) Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ Microbiol Rep*, 2: 281–288. 10.1111/j.1758-2229.2009.00117.x

Vacheron J, Moënne-Loccoz Y, Dubost A, Gonçalves Martins M, Muller D, Prigent-Combaret C (2016) Fluorescent *Pseudomonas* strains with only few plant-beneficial properties are favored in the maize rhizosphere. *Front Plant Scie*, 7: 1212, doi: 10.3389/fpls.2016.01212

Weil RR, Brady N (2002) *Elements of the nature and properties of soils*. Upper Saddle River, NJ, USA: Prentice Hall.

Yabuuchi E, Yano I, Oyazu H, Hashimoto Y, Ezaki T, Yamamoto H (1990) Proposals of *Sphingomonas paucimobilis* gen. nov. and comb. nov., *Sphingomonas parapaucimobilis* sp. nov., *Sphingomonas yanoikuyae* sp. nov., *Sphingomonas adhaesiva* sp. nov., *Sphingomonas* capsulate comb. nov., and two genospecies of the genus *Sphingomonas*. *Microbiol Immunol*, 34(2): 99-119

3 BACTERIAL ABILITIES AND ADAPTATION TOWARD THE RHIZOSPHERE COLONIZATION

The article of this section is already published in Frontiers in Microbiology Journal, section Terrestrial Microbiology:

Lopes LD, Pereira e Silva MC, Andreote FD (2016) Bacterial abilities and adaptation toward the rhizosphere colonization. *Front Microbiol*, 7: 1341, doi: 10.3389/fmicb.2016.01341

Abstract

The rhizosphere harbors one of the most complex, diverse, and active plant-associated microbial communities. This community can be recruited by the plant host to either supply it with nutrients or to help in the survival under stressful conditions. Although selection for the rhizosphere community is evident, the specific bacterial traits that make them able to colonize this environment are still poorly understood. Thus, here we used a combination of community level physiological profile (CLPP) analysis and 16S rRNA gene quantification and sequencing (coupled with *in silico* analysis and metagenome prediction), to get insights on bacterial features and processes involved in rhizosphere colonization of sugarcane. CLPP revealed a higher metabolic activity in the rhizosphere compared to bulk soil, and suggested that D-galacturonic acid plays a role in bacterial selection by the plant roots (supported by results of metagenome prediction). Quantification of the 16S rRNA gene confirmed the higher abundance of bacteria in the rhizosphere. Sequence analysis showed that of the 252 classified families sampled, 24 were significantly more abundant in the bulk soil and 29 were more abundant in the rhizosphere. Furthermore, metagenomes predicted from the 16S rRNA gene sequences revealed a significant higher abundance of predicted genes associated with biofilm formation and with horizontal gene transfer (HGT) processes. In sum, this study identified major bacterial groups and their potential abilities to occupy the sugarcane rhizosphere, and indicated that polygalacturonase activity and HGT events may be important features for rhizosphere colonization.

Keywords: microbiome, sugarcane, metagenome prediction, plant selection, horizontal gene transfer

3.1 Introduction

Of the various niches that comprise the plant microbiome – i.e., phyllosphere, spermosphere, rhizosphere, internal tissues (Turner et al., 2013; Andreote et al., 2014)- the most complex, diverse, and active microbial community is located in the rhizosphere, the soil layer immediately influenced by the rhizodeposition (Prashar et al., 2014). In the rhizosphere, there is a selection effect imposed by the physicochemical changes created in the roots' surroundings that shape the microbial composition. This selection can happen both indirectly, favoring the growth of opportunistic microbes adapted to specific chemical conditions, or by the active recruitment of microbes that will support plant development, support plant growth/nutrition, and/or enhancing the plant's ability to resist biotic and abiotic stressors (Bulgarelli et al., 2013; Philippot et al., 2013; Mendes R. et al., 2014). These microbial partners are understood to receive labile forms of carbon and a continuing supply of nutrients from the plants in exchange to their services (Bais et al., 2006).

Bulk soil, i.e., the zone of soil not under the influence of rhizodeposition, houses a huge microbial diversity and is considered the ultimate reservoir of microbes available for plant colonization, from which microbes are selected to the rhizosphere microbiome (Uroz et al., 2010; Peiffer et al., 2013; Mendes L.W. et al., 2014). However, the set of features required for bacteria to efficiently colonize the rhizosphere is yet to be properly described. For instance, studies have shown the importance of motility (de Weger et al., 1987) and lipopolysaccharide (LPS) production (de Weger et al., 1989) for the colonization of potato roots by a *Pseudomonas fluorescens* strain. The capacity to form biofilm was shown to be related to the rhizosphere colonization in *Bacillus amyloliquefaciens* and rhizobia species (Rinaudi and Giordano, 2010; Tan et al., 2013), but not in *P. fluorescens* (Barahona et al., 2010). Fast growth rate was suggested to be important for rhizosphere colonization in *Pseudomonas* spp. and *B. amyloliquefaciens* (Simons et al., 1996; Tan et al., 2013). In addition, the chemotaxis exerted by some organic acids was significant for the colonization of tomato roots by *B. amyloliquefaciens* (Tan et al., 2013). Another less explored ecological aspect possibly involved in the rhizosphere colonization by soil bacteria is the suitability of bacteria for horizontal gene transfer (HGT) processes, which could hypothetically support the rapid adaptation of bacteria in the face of environmental shifts. There are increasing indications that the rhizosphere could be a hot spot of HGT events, for example, by the increase of the transference of conjugative plasmids between rhizosphere inhabitants (Lilley et al., 1994; Pukall et al., 1996; Van Elsas et al., 1998). Although some studies have identified

traits related to the colonization of specific bacterial species, there is still a gap in the knowledge of key features involved in rhizosphere colonization by the soil microbiome.

Thus, the aim of the present study was to identify key bacterial traits for rhizosphere colonization using sugarcane as a model system. We studied the bacterial communities and assessed shifts in the taxonomic and functional profiles in bulk soil in comparison to the rhizosphere. Analyzes were performed using a combination of bacterial quantification, metabolic capacity to degrade carbon sources, high-throughput sequencing and metagenome prediction. Using this approach we gained a better understanding about the taxonomic and ecological relationships that microorganisms establish in the sugarcane rhizosphere, as well as of some of the characteristics needed for bacterial communities to colonize this soil habitat, including polygalacturonase activity and the possible importance of HGT in the rhizosphere.

3.2 Materials and Methods

3.2.1 Bulk Soil and Rhizosphere Sampling

Bulk soil and rhizosphere were sampled in a sugarcane cultivation field (cultivar SP-3250) located at ESALQ/USP (Piracicaba, Brazil). The sugarcane crop is being used in this field for 10 years under a green harvest management. Plants sampled in this study were at 9-months of cultivation (average height of 2.0 m) and did not show evidence for pest attack, disease, or nutritional deficiency. Bulk soil samples were made of soils free of roots, collected in the interline area of planting, at the layer of 0–10 cm. Rhizosphere samples were obtained by separating soil from plants roots (similar soil depth), focusing on a soil layer not thicker than 2 mm from the roots surface. Six biological replicates were used, each comprising of a single plant from which bulk soil and rhizosphere samples were collected, generating a total of 12 samples.

3.2.2 DNA Extraction

Total DNA was extracted from each sample using a MoBio Power Soil DNA Isolation Kit (Mobio, USA) according to the manufacturer’s instructions. The resulting DNA was checked for integrity by electrophoresis in a 1% agarose gel, stained with ethidium bromide and visualized under UV light, and stored at -20°C.

3.2.3 Quantification of Bacterial Community

The 16S rRNA gene copy numbers in bulk soil and rhizosphere were assessed to investigate whether these two habitats harbor significantly different bacterial abundances. Real-time PCR amplification was performed using P1/P2 primers (Muyzer et al., 1993) in a reaction with 12.5 μL of Sybr Green (1X), 10 mM of each primer, and 1 μL of template DNA for a total volume of 25 μL. The amplification was conducted using a StepOne Real-Time System (Applied Biosystems) under the following conditions: 1 cycle of 95°C for 3 min, 35 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 30 s, followed by a melting curve analysis. For the standard curve, soil-derived amplicons of the 16S rRNA gene were diluted from 10^{-1} to 10^{-8} and quantified. These dilutions were submitted to amplifications under the same conditions described above. In addition, the efficiency of the reaction was calculated and was 101.5%, between the acceptable values of 90 and 110%, indicating the absence of PCR inhibitors.

3.2.4 Sequencing of the 16S rRNA Gene

We performed a sequence-based analysis using the Illumina MiSeq platform and the Nextera XT index kit for library preparation (Illumina, USA), targeting the V3-V4 region of the 16S rRNA gene. A nested protocol was performed. In the first reaction the whole gene was amplified using the universal primers 27F (Lane, 1991) and 1387R (Marchesi et al., 1998). This first reaction contained 1 μL of template, 2 μL of dNTPs (2.5 mM), 3.75 μL of $MgCl_2$ (25 mM), 2.5 μL of Taq buffer (10X), 0.1 μL of each primer (100 mM) and 0.3 μL of Taq polymerase (5U/μL) for a total volume of 25 μL. The thermal cycles consisted of 1 cycle of 94°C for 4 min, 25 cycles of 94°C (30 s), 63°C (1 min) and 72°C (1 min), ending with 1 cycle of 72°C for 10 min. In the second reaction, primers S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21 (Klindworth et al., 2013) were used (coupled with Illumina adapters), which cover the hypervariable regions V3-V4 of the 16S rRNA gene. This reaction was conducted using 1 μL of the amplicons produced in the first reaction, 2 μL of dNTPs (2.5 mM), 3 μL of MgCl2 (25 mM), 3 μL of Taq Buffer (10X), 0.1 μL of each primer and 0.3 μL of Taq DNA polymerase for a total volume of 25 μL. The amplification was conducted following the thermal cycles: an initial denaturation cycle of 95° for 3 min, 30 cycles of 95°C for 45 s (denaturation), 57°C for 1 min and 45 s (annealing) and 72°C for 1 min (extension), followed by a final extension cycle of 72°C for 4 min. The primers sequences are found in Supplementary Table 3.1. Nextera indices and sequencing adapters were ligated to amplicons from the second reaction, and submitted for paired-end sequencing on the Illumina MiSeq

platform. The reads were deposited at the SRA database (NCBI), BioProject PRJNA319762, and the BioSample accession numbers are: SAMN04904299 (B1); SAMN04904300 (B2); SAMN04904301 (B3); SAMN04904302 (B4); SAMN04904303 (B5); SAMN04904304 (B6); SAMN04904305 (R1); SAMN04904306 (R2); SAMN04904307 (R3); SAMN04904308 (R4); SAMN04904309 (R5); and SAMN04904310 (R6) (R = Rhizosphere and B = Bulk Soil samples).

3.2.5 Next Generation Sequence Analysis

The sequenced paired-end reads were separated by sample and analyzed using the QIIME software pipeline (Caporaso et al., 2010). Quality control analyses were performed to eliminate low quality reads, short reads, chimeric sequences, and to trim the low quality 3′ region of individual reads in order to achieve a minimum quality of Q28 (Phred scale). The UCLUST algorithm was used to cluster the reads in operational taxonomic units (OTUs) with a 97% cutoff, and to assign the taxonomy using the Greengenes database (version gg_13_8_otus) with a minimum fraction of 0.51. The reads were then aligned using the PyNAST algorithm and filtered. An OTU table was generated, the singletons were excluded and the OTU table was rarified (170,000 sequences) to avoid bias related to different number of reads in the samples. Richness and diversity indices were calculated, and the OTU table was exported to STAMP (Parks et al., 2014) for statistical analyzes.

In addition to the amplicon-based analysis, we also used a metagenome prediction approach to infer probable functions performed by the bacterial communities, which is more informative than a purely taxonomic community structure approach. For this, we used the PICRUSt software package, which uses 16S rRNA libraries to make a predicted reconstruction of the metagenome (Langille et al., 2013). The PICRUSt software uses an evolutionary approach to handle the OTUs that match with unavailable genomes by using their sequenced relatives as a reference for the prediction. This means that there is some uncertainty in the genome prediction, as all microbiome sequencing produces some OTUs that match with unavailable genomes and different strains of the same species/OTUs have some distinct gene content. The nearest sequenced taxon index (NSTI) value show the level of uncertainty of the metagenome prediction, increasing the accuracy of the prediction the smaller are its values. First, a new OTU table was created using a closed-reference picking OTU protocol against the Greengenes database (version gg_13_5_otus) at 97% identity, which was then normalized by dividing the abundances of each OTU by known or predicted 16S rRNA gene copy number abundances. The normalized data was then submitted to

metagenome prediction and categorized by function using KEGG level 2 Gene Ontology (GO) terms for classification. The NSTI was used to quantify the availability of nearby genome representatives in the samples. The prediction tables (the raw gene prediction table and the classified table with the GO terms) were then exported to STAMP software (Parks et al., 2014) for statistical analysis.

3.2.6 Community Level Physiological Profile (CLPP) Analysis

To better characterize the functional profile of the microbial communities in our samples, we performed a CLPP analysis using BIOLOG Ecoplates (Biolog Inc., USA), which contains 31 different carbon sources. For this analysis, 10 g of either soil or rhizosphere samples were weighed, and a soil suspension was obtained by shaking the samples in 90 mL of saline solution (0.1 M NaCl) for 30 min. This suspension was then centrifuged for 30 min at 250 rpm, the supernatant was diluted to 10^{-3} and 150 µl was used to inoculate the Ecoplates. Plates were incubated at 25°C and read at the 590 nM wavelength in 24 h intervals for a total of 168 h. Absorbance readings were corrected using blank samples as controls.

3.2.7 Statistical Analysis

Results of qPCR were analyzed using the PAST software (Hammer et al., 2001), where ANOVA and Tukey pairwise comparison tests were performed to assess differences in rhizosphere and bulk soil samples.

The OTU table of the 16S rRNA gene sequencing was exported and analyzed in PRIMER-6 software (Clarke and Gorley, 2006) where a Non-metric Multidimensional Scaling (NMDS) as well as an Analysis of Similarity (ANOSIM) were performed using the Bray–Curtis dissimilarity matrix with the aim of detecting differences in bacterial community structure between the 2 habitats studied (rhizosphere × bulk soil). These identified differences were further explored using the STAMP software (Parks et al., 2014), where bacterial groups with significant differences were identified using the Welch's t-test (Parks et al., 2014). The same approach was used to identify significant differences in the predicted metagenomes of rhizosphere and bulk soil samples. The Benjamini–Hochberg *P*-value correction was utilized to avoid type 1 and 2 errors. In addition, a Bonferroni *P*-value correction was applied to the raw predicted genes table to account for potential bias related to PICRUSt predictions and statistical analysis, as this correction is even more conservative in detecting significant differences.

The PAST software was used to perform an ANOVA and Tukey test for each variable (C-sources) of the CLPP results. Alternatively, the results were exported to the CANOCO 4.5 software, where a biplot Principal Component Analysis (PCA) was performed to check the differences in the metabolic profiles of the microbial communities of rhizosphere and bulk soil, as well as to detect any correlation with the C-sources (ter Braak and Smilauer, 2002). An ANOSIM was also performed to validate the significance of these differences using the PRIMER-6 software package (Clarke and Gorley, 2006). The niche of a given community, bulk soil or rhizosphere, was calculated according to Salles et al. (2009). This analysis is based on the performance of the total community in each of the carbon sources, calculated as the sum of the best performances on each source present on the environment where that community is functioning.

3.3 Results

3.3.1 Distinctions of Bacterial Abundance and Community Structure between Rhizosphere and Bulk Soil

Quantification of total bacterial communities in bulk soil and rhizosphere samples revealed significant differences in bacterial abundance ($P < 0.05$). The average bacterial abundance in rhizosphere samples was 4.96×10^9 rRNA gene copies per gram of soil, while bulk soil samples contained an average of 1.67×10^9 (Figure 3.1A). This distinction was complemented by the differential structuring of bacterial communities in each niche, as indicated by the results from partial sequencing of the 16S rRNA gene. After trimming sequences of low quality and rarification of sequences per sample, a total of 2,040,000 sequences – 1,020,000 from each rhizosphere and bulk soil samples – were used in the analysis.

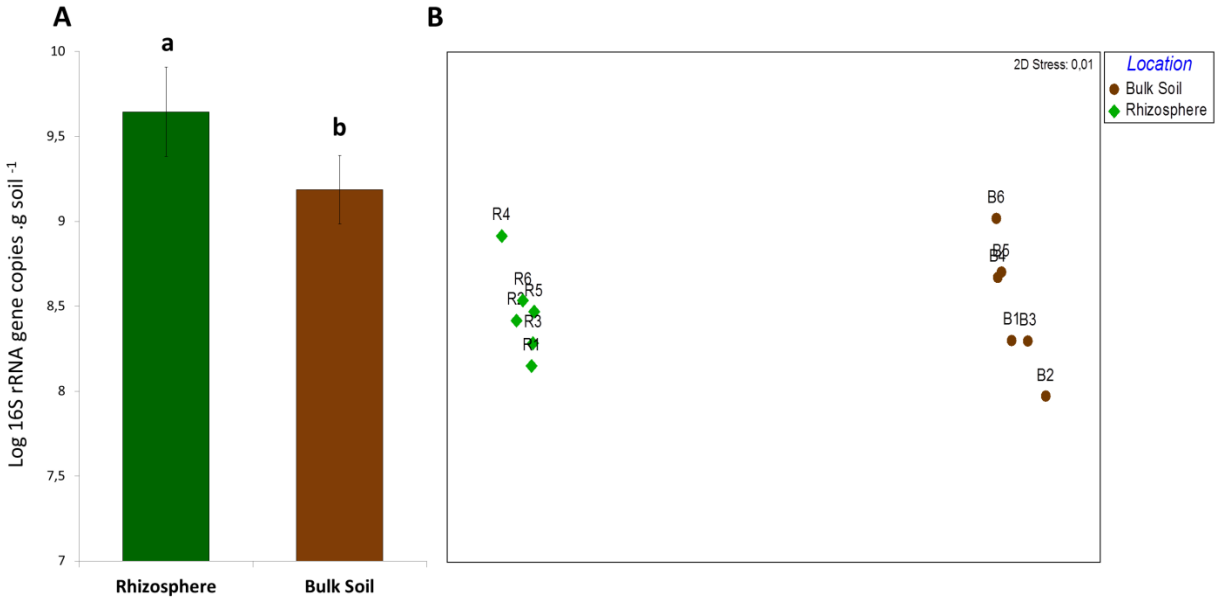


Figure 3.1 (A) Quantification of 16S rRNA gene in samples of bulk soil and rhizosphere. Different letters means significant differences according to the pairwise comparison test of Tukey (*P* < 0.05). (B) Non-Metric Multidiomensional Scaling (NMDS) comparing the structure of bacterial community in bulk soil and rhizosphere on the basis of the OTU table

Non-metric Multidimensional Scaling results indicated the distinct structure of bacterial communities in rhizosphere and bulk soil samples – confirmed by ANOSIM (*P* = 0.008/*R* = 0.82). The two groups of samples were separated in the first axis, and replicates grouped together, clearly indicating the differential selection exerted by each of these niches upon the bacterial community (Figure 3.1B). Although there was a clear difference in β-diversity, no differences concerning the α-diversity estimators (*P* < 0.05) were observed between the niches (Table 3.1).

Table 3.1 The output of sequences analysis after quality filtering per sample, before and after the OTU table rarefaction

		Before rarefaction	Post rarefaction on OTU table		
Location	Samples	N° of Reads	N° of Reads	Shannon index	Chao 1 index
Rhizosphere	R1	230,123	170,000	11.9991	93,359
	R2	271,867	170,000	12.0379	91,204
	R3	216,554	170,000	12.1031	93,485
	R4	243,438	170,000	12.2418	92,860
	R5	211,310	170,000	11.9151	85,191
	R6	172,230	170,000	11.8505	85,779
	Total	1,345,522	1,020,000	$\mu = 12.0246$ $SD = 0.1389$	$\mu = 89,804$ $SD = 4,050.56$
Bulk Soil	B1	213,884	170,000	11.9015	90,660
	B2	277,195	170,000	12.1156	95,930
	B3	271,546	170,000	11.9811	91,201
	B4	304,783	170,000	12.0649	92,421
	B5	201,595	170,000	11.9613	90,310
	B6	280,574	170,000	11.9765	96,428
	Total	1,549,577	1,020,000	$\mu = 12.0002$ $SD = 0.0770$	$\mu = 93,350$ $SD = 2,734.53$

Calculation of Shannon diversity index, Chao 1 species richness index and their average (μ) and standard deviation (SD).

Sequences were classified into 24 known and 31 candidate bacterial phyla, with prevalence of sequences affiliated to Acidobacteria and Proteobacteria in both communities (Figure 3.2A). Divergences in the relative abundance of phyla Actinobacteria, Bacteroidetes, Cyanobacteria and Gemmatimonadetes were observed, which were more abundant in the bulk soil ($P < 0.01$). On the other hand, Verrucomicrobia, Nitrospirae and Tenericutes were more abundant in the rhizosphere ($P < 0.01$) (Figure 3.2B). Higher divergences were observed at family level (Figure 3.3), from which 24 were significantly more abundant in the bulk soil, and 29 were more abundant in the rhizosphere ($P < 0.05$).

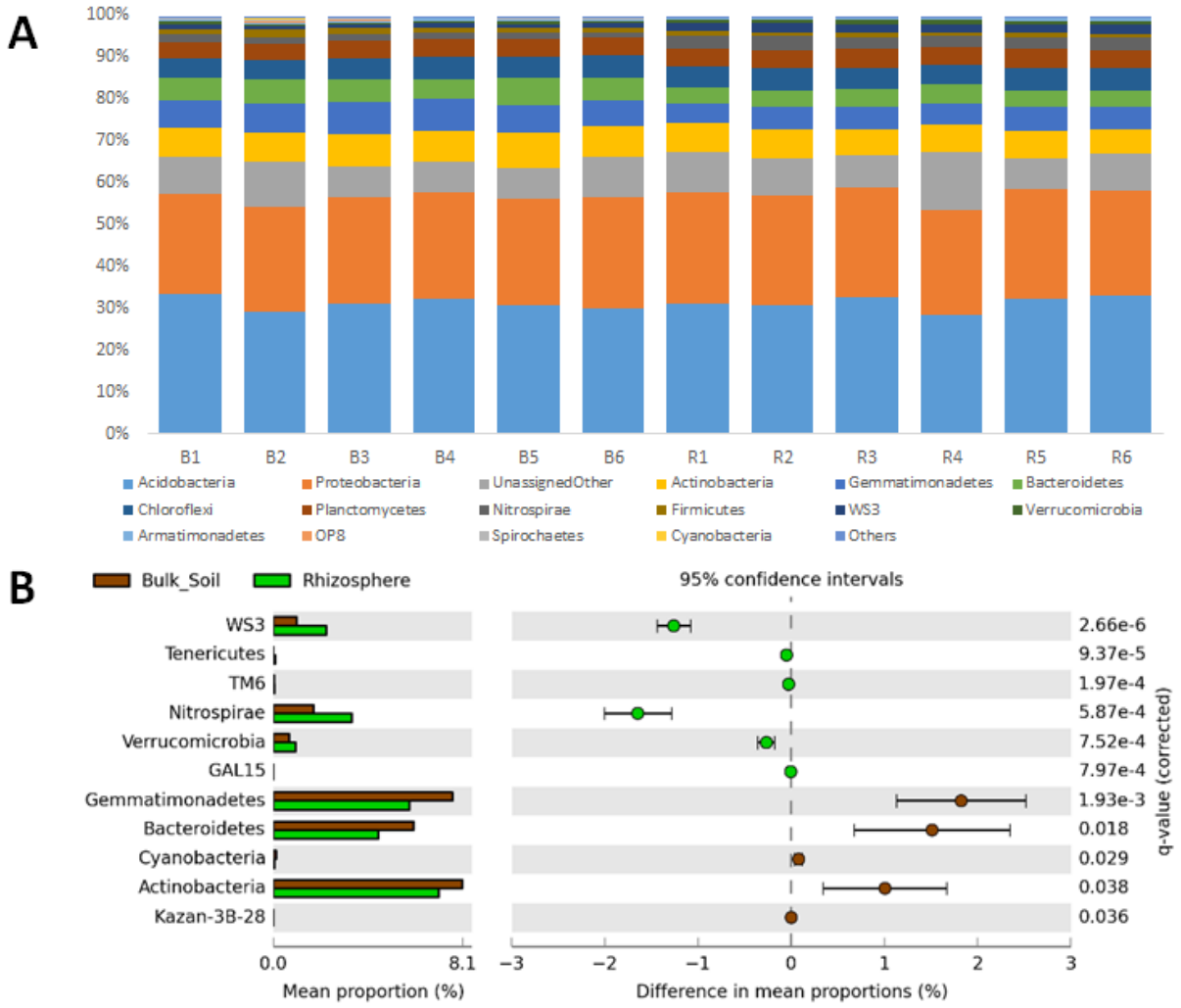


Figure 3.2 (A) Taxonomic classification of OTUs (at level of phyla) based on Greengenes database using the QIIME software. (B) Statistical comparison (Welch's *t*-test) between the phyla abundance on rhizosphere and bulk soil using the Benjamini–Hochberg *P*-value correction ($P < 0.05$)

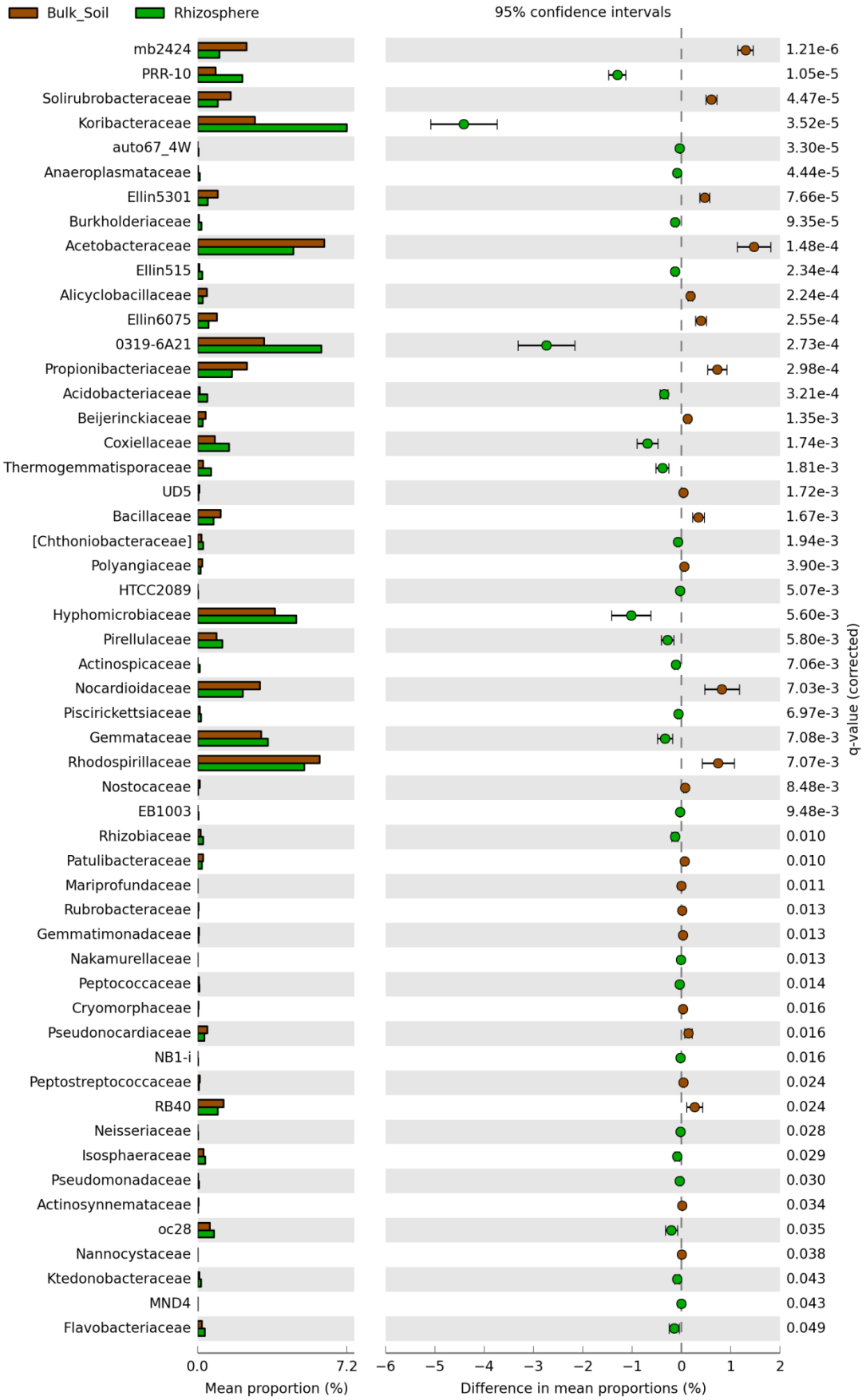


Figure 3.3 Statistical comparison (Welch's *t*-test) between the families abundance on rhizosphere and bulk soil using the Benjamini-Hochberg *P*-value correction ($P < 0.05$)

3.3.2 Metabolic Profiles of Rhizosphere and Bulk Soil Microbial Communities

CLPP analysis indicated differential metabolic patterns of the microbial communities found in bulk soil and rhizosphere. A broad view of such patterns was obtained with a PCA, which revealed clear distinctions between the carbon degradation profiles of the communities from each niche, supported by ANOSIM ($P = 0.002$; $r^2 = 0.51$) (Figure 3.4). From the 31 C-sources, 11 were significantly more utilized by the rhizosphere community while 4 were more utilized by the bulk soil community ($P < 0.05$). The performances of the communities on individual carbon sources were used to calculate the community niche of bulk soil and rhizosphere samples, and were significantly higher ($P < 0.05$) in the rhizosphere (29.03 ± 2.90) compared to bulk soil (23.57 ± 0.30). The most catabolized C-sources by the microbial community from rhizosphere ($P < 0.05$) were D-galacturonic acid, D-galactonic acid, D-glucosaminic acid, L-asparagine, 4-hydroxybenzoic acid, D,L,a-glycerolphosphate, L-phenylalanine, L-threonine, L-serine, tween-40 and putrescine. On the other hand, A,D-lactose, D-xylose, *i*-erythritol and *y*-hydroxybutyric acid were the most consumed C-sources by the microbial community from the bulk soil ($P < 0.05$) (Figure 3.4).

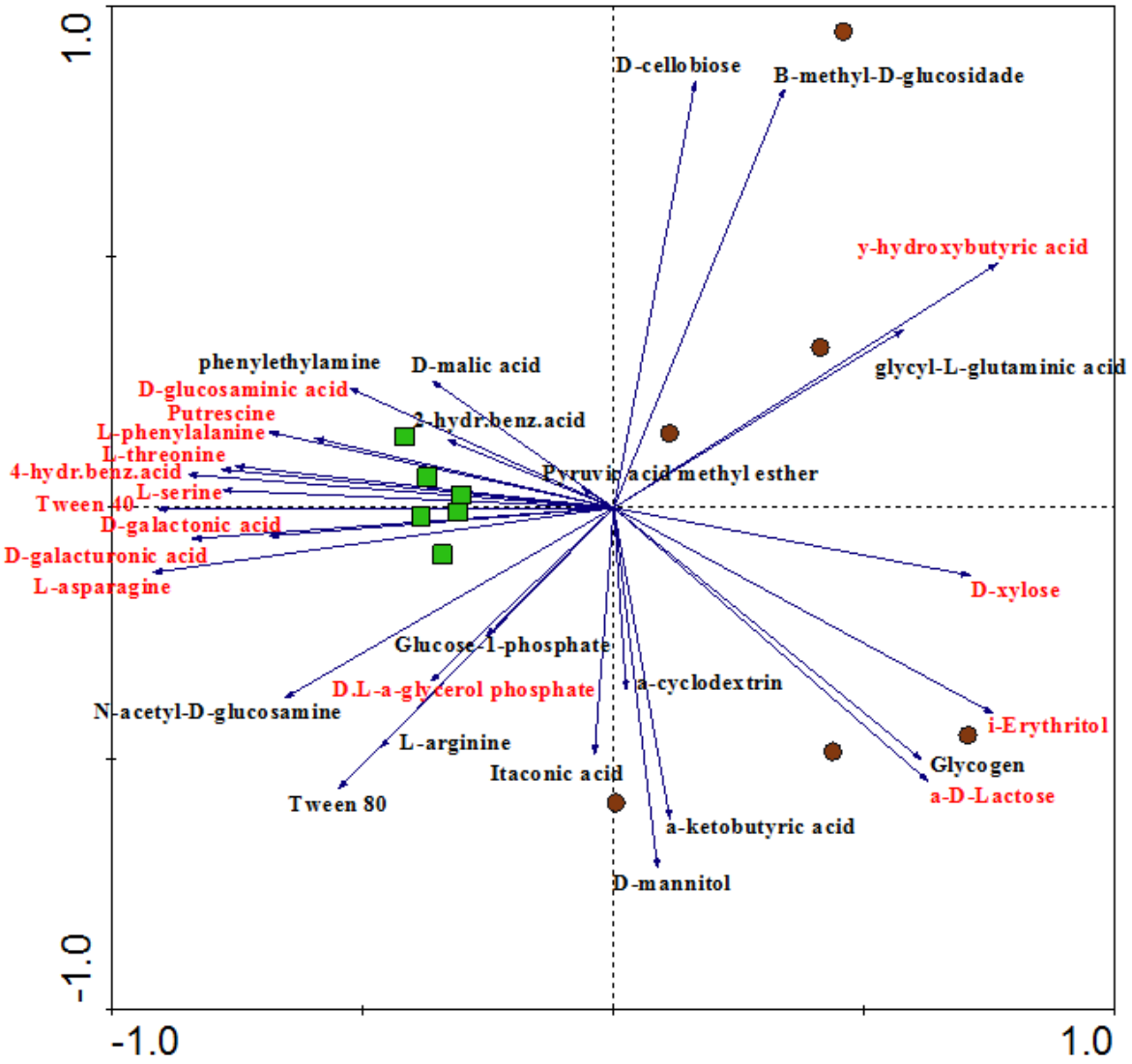


Figure 3.4 Principal component analysis (PCA) showing the ordination of rhizosphere and bulk soil samples according to the oxidation of C-sources in the CLPP analysis. Variables (C-sources) colored in red means significantly different between rhizosphere and bulk soil in the Tukey test ($P < 0.05$). Samples represented by green squares are from rhizosphere and samples represented by brown circles are from bulk soil (PC1 = 32.5%, PC2 = 21.3%)

3.3.3 Metagenome Prediction of Bulk Soil and Rhizosphere Bacterial Communities

Metagenomes were predicted from the 16S rRNA gene sequences with an accuracy based on a NSTI average value of 0.23 ± 0.02, which is typical for soil samples analyzed in other studies (Zarraonaindia et al., 2015; Chen et al., 2016). Despite such a high NSTI value, Zarraonaindia et al. (2015) found a significant similarity between the sequenced and predicted metagenome. Metagenome prediction resulted in more than 6,900 protein-coding genes, of which 153 were differentially predicted in accordance to their frequencies in bulk soil or

rhizosphere ($P < 0.05$); 80 were significantly more abundant in the rhizosphere and 73 were significantly higher in bulk soil (Supplementary Figure S3.1). The predicted protein-coding genes were categorized by function using the KEGG level 2 GO, revealing the prevalence (~50%) of genes related to amino acid metabolism, carbohydrate metabolism, membrane transport, replication and repair and energy metabolism, in both bulk soil and rhizosphere samples (Figure 3.5A). The other half of the predicted functions were affiliated with more than 25 other functional categories. A Welch's t-test indicated that rhizosphere and bulk soil were enriched in 6 and 7 functional categories (KEGG level 2), respectively ($P < 0.05$). More specifically, predicted genes more abundant in the rhizosphere were mainly associated with Cellular Processes and Signaling, Cell Growth and Death, Carbohydrate Metabolism, Metabolism, Glycan Biosynthesis and Metabolism and Transcription (Figure 3.5B). On the other hand, genes related to Metabolism of Amino Acids, Lipids, Nucleotide, Terpenoids and Polyketides, and Cell Communication and Environmental Adaptation were more abundant in the bulk soil (Figure 3.5B).

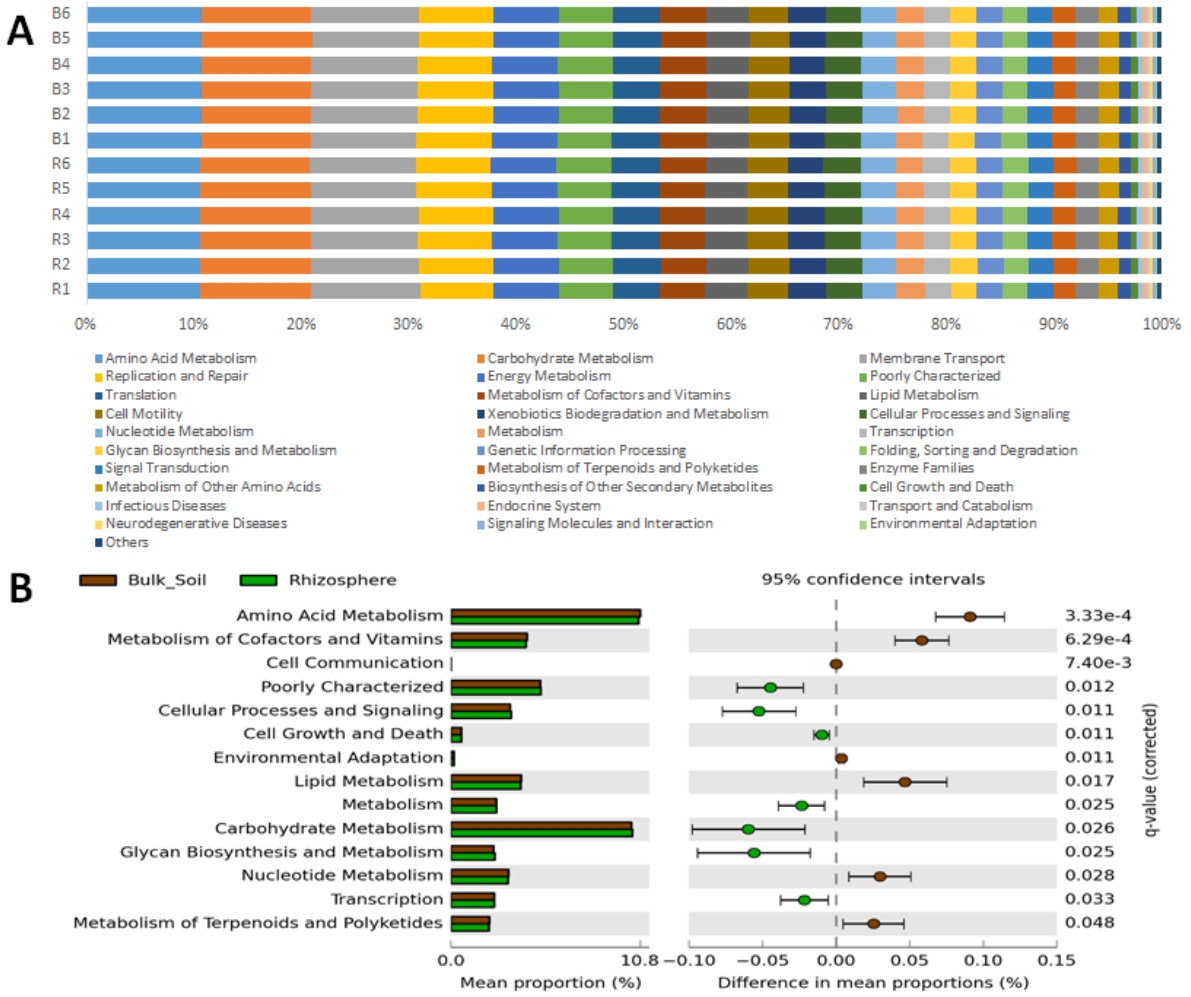


Figure 3.5 (A) Metagenome predicted functions classified using KEGG level 2 database in PICRUSt software, showing the most abundant functions throughout the 12 samples. (B) Statistical comparison (Welch's *t*-test) between the predicted functions abundance on rhizosphere and bulk soil using the Benjamini–Hochberg *P*-value correction ($P < 0.05$)

We observed a significantly higher abundance of predicted genes associated with HGT in the rhizosphere (Figure 3.6B) than in bulk soil. We found a higher prevalence of genes associated with conjugation, such as the pilus assembly protein CpaE and type IV pilus assembly protein PilV (Cabezón et al., 2015); as well as genes involved in transformation/conjugation, such as the type IV secretion system proteins VirB4, VirB5, VirB6, and VirB9 (Figure 3.6B) (Christie et al., 2005; Cabezón et al., 2015; Juhas, 2015). We also observed a greater number of predicted genes indirectly related to transduction, like DNA polymerase bacteriophage-type. Another interesting predicted gene, also more abundant in the rhizosphere, was related to colanic acid biosynthesis protein WcaH, which is associated with biofilm formation. To validate the results from the PICRUSt analysis, we looked for predicted genes coding for enzymes involved in the metabolism of the carbon sources present in the Biolog plate assays. We found a predicted gene associated with polygalacturonase, which was

more abundant in the rhizosphere (Figure 3.6A). Likewise, the rhizosphere samples had higher degradation levels of D-galacturonic acid in the Biolog assays (Figure 3.4).

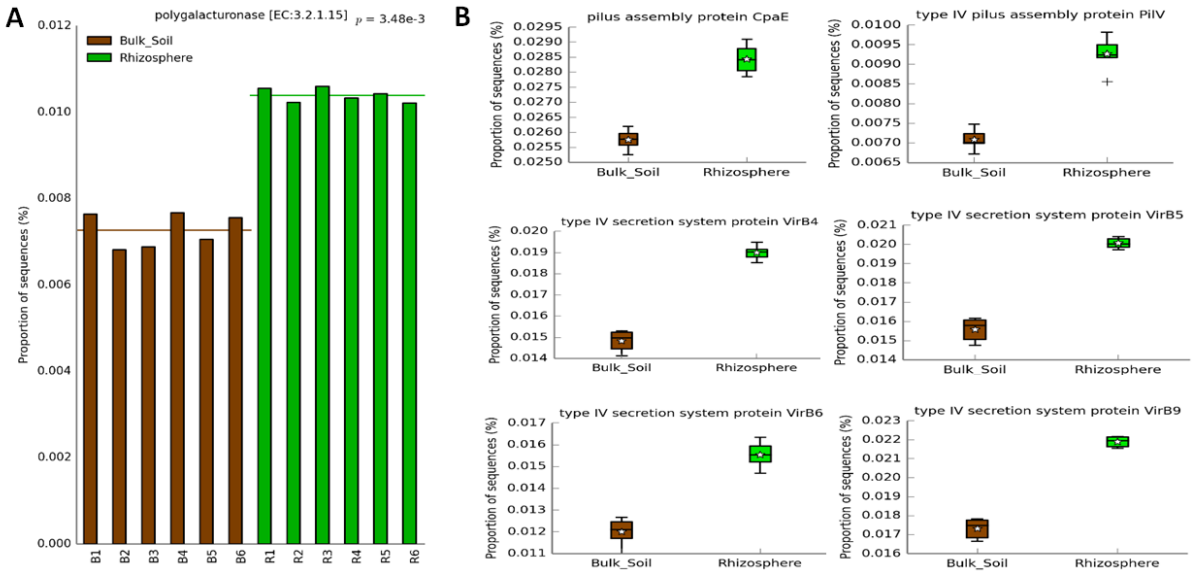


Figure 3.6 Predicted genes on PICRUSt related to (A) CLPP analysis and (B) HGT, which showed significantly higher abundance in the rhizosphere (Welch's *t*-test, $P < 0.05$) using the Bonferroni *P*-value correction

3.4 Discussion

This study accessed the bacterial communities present in the sugarcane rhizosphere and bulk soil using real-time PCR and high-throughput sequencing in order to identify differences in bacterial abundance, diversity and taxonomic structure. Furthermore, the metabolic profile of the microbial communities was also assessed using CLPP and metagenome prediction with the aim of identifying potential traits important for the rhizosphere colonization.

Results of bacterial quantification (based on the quantity of 16S rRNA gene copies) provided the first indication of growth stimulus in the roots surroundings in our study, corroborating other studies (Wessén et al., 2010; Rachid et al., 2013). Differences in the bacterial community structure and composition between rhizosphere and bulk soil support the hypothesis that the rhizosphere selects a bacterial community from a pool in the bulk soil, which changes the abundance of specific groups according to their ability to adapt to the rhizosphere conditions. In this line, the similar values of richness and α-diversity found in these two environments provide supporting evidences that the rhizosphere shapes the bacterial community by modulating its composition (Mendes L.W. et al., 2014). Such distinction is better observed in the community structure based on the analysis of OTUs, where samples

from each niche are separately plotted in the NMDS indicating a high β-diversity between rhizosphere and bulk soil.

Despite the fact that the bacterial communities were very different considering the analysis of OTUs, small changes were observed at the phyla level, where the taxonomic profile is quite similar (Figure 3.1A), suggesting that the shifts in abundances of specific groups are not phylogenetically dramatic. Three phyla, Acidobacteria, Proteobacteria and Actinobacteria were the most abundant in our study, similarly to those observed by Rachid et al. (2013), who analyzed bacterial communities of sugarcane fields under two different managements. In our study, from these 3 phyla only Actinobacteria was present at relatively higher abundance in bulk soil. Members of this phylum are known for their resistance to stress conditions (particularly drought stress) and/or ability to degrade complex carbon sources, explaining their higher abundance in the oligotrophic bulk soil environment (DeBruyn et al., 2011; Roes-Hill et al., 2011; Thomas et al., 2011; Kavamura et al., 2013). On the other hand, less abundant phyla, such as Nitrospirae, Verrucomicrobia, and Tenericutes, were enriched in the rhizosphere. Nunes da Rocha et al. (2013) showed that groups affiliated to Verrucomicrobia might be important to the rhizosphere of several plants, such as grasses, leek, and potatoes. Similarly, Verrucomicrobia and Nitrospirae were also detected with significant higher abundance in the rhizosphere of oats (DeAngelis et al., 2009) and soybean (Mendes L.W. et al., 2014). This evidence suggests that although understudied, these phyla can be commonly selected by the roots of several different plants and might play a critical role for their development. Proteobacteria is another bacterial phylum known for colonizing the rhizosphere (Uroz et al., 2010; Peiffer et al., 2013). Here we did not find differences when compared the abundance of members of this phylum in each niche. However, we found relatively higher abundances of families such as Burkholderiaceae, Pseudomonadaceae and Rhizobiaceae, which belong to Proteobacteria and are widely described as plant growth promoters (García-Fraile et al., 2012; Suárez-Moreno et al., 2012; Redondo-Nieto et al., 2013), in the rhizophere (Figure 3.3). These observations indicate the need to compare these communities in more detailed taxonomy levels than phyla, where they seem more similar.

Community level physiological profile results provide the first evidence of how the plant roots select microbes for its rhizosphere microbiome. PCA based on C-sources oxidation showed that the rhizosphere samples were closer to each other, while the bulk soil samples were more dispersed in the chart, suggesting that the microbial communities from the rhizosphere share a very similar functional community structure (Figure 3.4). This suggests

that the rhizosphere selection effect can be possibly more effective for functional attributes than for taxonomy – as observed in the ordination of samples based on OTUs (Figure 3.1B).

We could conclude from both the number of carbon sources consumed in each habitat and the community niche analysis that the rhizosphere microbial community is more metabolically active. It might be that cells in an active state preferentially colonize the rhizosphere microbiome, as the roots are constantly supplying nutrients to this niche. In counterpart, the bulk soil community is more prone to undergo long periods without labile carbon, which might decrease the metabolic rate of most microbes. Our results also indicate that the labile C-sources released in the rhizosphere might be the main plant compounds acting in the microbiome selection. This is supported by the preferential use of the D-galacturonic acid, and its reduced form D-galactonic acid, in the rhizosphere.D-galacturonic acid is the most abundant component of pectin (Zhang et al., 2011), a major constituent of plant cell walls, which is consequently released in the rhizosphere. D-galacturonic acid has also been found as a component of root exudates (Tawaraya et al., 2015). It is important to note that CLPP is culture dependent, meaning that the measurements of C-sources oxidation reflect the activity of only part of the community in the sample, and is therefore limited in its ability to provide the real functional capacity of a microbial community *in vivo*.

The only common plant-produced C-source that was more utilized by the bulk soil bacterial community was D-xylose, a component of hemicellulose. Nevertheless, hemicellulose is one of the polysaccharide constituents of lignocellulose, the most recalcitrant component of the plant cell wall, which is only degraded by specialized enzymes (Mansour et al., 2016). In sugarcane field soils, the presence of hemicellulose is even higher, as leaves are constantly falling and accumulating on the soil surface (Chandel et al., 2012).

Although evidence for selection in the rhizosphere is clear, the bacterial abilities that enable rhizosphere colonization are still poorly described. To gain more insights on these abilities, PICRUSt was used to predict the metagenome of each of the samples. PICRUSt is dependent on available sequenced genomes and some bias can be produced if references for the species inhabiting the sampled environment are missing. However, it is a useful tool to provide insights about the community functional potential in the absence of shotgun metagenomic data. The combination of metagenome prediction and CLPP provided insights into community functional structure and reduced the bias related to each approach. For both communities, rhizosphere and bulk soil, the most prevalent predicted functions were Amino Acid Metabolism and Carbohydrate Metabolism, corroborating a previous study of sugarcane rhizosphere communities by metaproteomics (Lin et al., 2013). Some significant differences

between bulk soil and rhizosphere were found, but the overall functional profiles classified by KEGG (level 2) were very similar, meaning that all higher functions are probably executed in both environments. Functions related to basic metabolism were found in higher abundance in the rhizosphere compared to the bulk soil (Cell Growth and Death, Metabolism, Carbohydrate Metabolism, Glycan Biosynthesis and Transcription), what can be linked to the higher metabolic activity in the rhizosphere communities detected on the CLPP analysis. These convergent results also suggest that the fast growth rate observed in previous studies of single species can be a widespread feature for rhizosphere colonization (Simons et al., 1996; Tan et al., 2013). In contrast, the bulk soil community showed a higher ratio of functions related to secondary metabolism, including the degradation of complex compounds and environmental resistance (Metabolism of Terpenoids and Polyketides, and Environmental Adaptation).

A deeper investigation of the metagenome prediction showed a greater abundance of genes related to Colanic Acid Biosynthesis Protein WcaH in the rhizosphere. Colanic acid is associated with biofilm formation in *Escherichia coli* (Danese et al., 2000), which is in agreement with previous data that showed the importance of biofilm formation for colonization of the roots surface (Ramey et al., 2004; Rinaudi and Giordano, 2010; Tan et al., 2013). Another interesting result obtained from the metagenome prediction was the higher abundance of genes related to HGT in the rhizosphere. We found genes related to bacterial transformation and conjugation, such as those associated with the Type IV Secretion System, enriched in the rhizosphere (Christie et al., 2005; Cabezón et al., 2015; Juhas, 2015). We also found a higher abundance of genes related to conjugation, as indicated by the occurrence of genes codifying the pilus assembly protein CpaE and type IV pilus assembly protein PilV (Cabezón et al., 2015); and genes possibly related to transduction, such as genes related to bacteriophages. These results suggest that HGT could be an important mechanism for bacterial adaptation in the rhizosphere. Given that the accumulation of beneficial mutations is a slow process to generate genetic variability in the short time scale (Gogarten et al., 2002; Jain et al., 2003), the horizontal transfer of entire genes may be a more efficient way for bacteria to rapidly adapt to the rhizosphere niche. This process is widely understood as a powerful tool for the rapid evolution of antibiotic resistance in bacteria inhabiting hospital environments (Palmer and Kishony, 2013), and may happen in an analogous way for rhizosphere colonization.

Although there is some bias in using 16S rRNA gene libraries to predict metagenomes, other studies based on real metagenome sequencing found results consistent with and supporting our data. Mendes L.W. et al. (2014) found an enrichment of Type IV Secretion

System genes in the rhizosphere of soybean. Similarly, Alzubaidy et al. (2016) found a significant higher number of genes linked to Phages, Prophages, Transposable Elements, and Plasmids in the rhizosphere of the mangrove plant *Avicennia marina*. Together with our results, this evidence suggests that HGT may be a widespread mechanism that facilitates rhizosphere colonization in plants.

Moreover, to support the results of the functional analyzes, we checked for matches between the C-sources degradation in the CLPP analysis and the enzyme-coding genes predicted by PICRUSt. Even considering the bias in each approach, a match was observed for genes related to polygalacturonase, more abundant in the rhizosphere samples, and the preferential degradation of D-galacturonic acid by the rhizospheric microbial community. This meaningful finding indicates that the capacity to metabolize the galacturonic acid may be an important trait for rhizosphere colonization. Other genes related to the degradation of specific CLPP C-sources were found in the metagenome prediction, but no significant differences were observed between rhizosphere and bulk soil. The lack of other matches might be related to the fact that CLPP measures differences in gene content as well as differential expression of genes, while metagenome prediction is based exclusively on the prediction of genotypes. It might also be that the low frequency of groups hosting these genes limited its prediction in the PICRUSt approach, or even that this approach may not have the accuracy to properly detect small changes in gene abundance of functional groups. The high NSTI values found in our samples and the lack of other approach to confirm the prediction also tell that the finding of HGT-related genes higher in the rhizosphere is an interesting suggestion that this activity can be enriched in this environment, but we highlight the need to experimentally confirm this issue in next studies.

Taken together, our results show important findings concerning the taxonomic and functional selection exerted by the sugarcane rhizosphere on the bacterial community, as well as identified potential traits that allow bacteria to colonize the rhizosphere environment. We firstly highlight the ability to degrade the D-galacturonic acid, a compound potentially used by the plant in a biochemical selection; and secondly the potential ability to perform HGT, which can spread important genes related to rhizosphere colonization, providing the bacteria an evolutionary and ecological advantage.

3.5 Supplementary Material

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.01341

Supplementary Table 3.1 Primers used in the second reaction for Illumina sequencing

16SV4FPCR1_1 F	5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAYTGGGYDTAAAGNG
16SV4FPCR1_2 F	5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNAYTGGGYDTAAAGNG
16SV4FPCR1_3 F	5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNAYTGGGYDTAAAGNG
16SV4FPCR1_4 F	5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNAYTGGGYDTAAAGNG
16SV4RPCR1_1 R	5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCCGTCAATTCMTTTRAGT
16SV4RPCR1_2 R	5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNCCGTCAATTCMTTTRAGT
16SV4RPCR1_3 R	5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNCCGTCAATTCMTTTRAGT
16SV4RPCR1_4 R	5' GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNCCGTCAATTCMTTTRAGT

References

Alzubaidy H, Essack M, Malas TB, Bokhari A, Motwalli O, Kamanu FK, et al. (2016) Rhizosphere microbiome metagenomics of gray mangroves (*Avicennia marina*) in the Red Sea. *Gene,* 576(2 Pt. 1): 626–636. 10.1016/j.gene.2015.10.032

Andreote FD, Gumiere T, Durrer A (2014) Exploring interactions of plant microbiomes. *Sci Agric*, 71: 528–539. 10.1590/0103-9016-2014-0195

Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM (2006) The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu Rev Plant Biol*, 57: 233–266. 10.1146/annurev.arplant.57.032905.105159

Barahona E, Navazo A, Yousef-Coronado F, Aquirre de Cárcer D, Martínez-Granero F, Espinosa-Urgel M, et al. (2010) Efficient rhizosphere colonization by *Pseudomonas fluorescens* f113 mutants unable to form biofilms on abiotic surfaces. *Environ Microbiol,* 12: 3185–3195. 10.1111/j.1462-2920.2010.02291.x

Bulgarelli D, Spaepen SS, Themaat EVL, Shulze-Lefert P (2013) Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol*, 64: 807–838. 10.1146/annurev-arplant-050312-120106

Cabezón E, Ripoll-Rozada J, Peña A, de la Cruz F, Arechaga I (2015) Towards an integrated model of bacterial conjugation. *FEMS Microbiol Rev*, 39: 81–95. 10.1111/1574-6976.12085

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*, 7: 335–336. 10.1038/nmeth.f.303

Chandel AK, Silva SS, Carvalho W, Singh OM (2012). Sugarcane bagasse and leaves: foreseeable biomass of biofuel and bio-products. *J Chem Technol Biotechnol*, 87: 11–20. 10.1002/jctb.2742

Chen L, Luo Y, Xu J, Yu Z, Zhang K, Brookes PC (2016) Assessment of bacterial communities and predictive functional profiling in soils subjected to short-term fumigation-incubation. *Microb Ecol*, 72: 240–251. 10.1007/s00248-016-0766-0

Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S, Cascales E (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol*, 59: 451–485. 10.1146/annurev.micro.58.030603.123630

Clarke KR, Gorley RN (2006) *PRIMER v6: User Manual/Tutorial*. Plymouth: PRIMER-E.

Danese PN, Pratt LA, Kolter R (2000) Exopolysaccharide production is required for development of Escherichia coli K-12 biofilm architecture. *J Bacteriol*, 182: 3593–3596. 10.1128/JB.182.12.3593-3596.2000

de Weger LA, Bakker PAHM, Schippers B, van Loosdrecht MCM, Lugtenberg BJJ (1989). "*Pseudomonas* spp. with mutational changes in the O-antigenic side chain of their lipopolysaccharide are affected in their ability to colonize potato roots," in *Signal Molecules in Plants and Plant-Microbe Interactions* ed. Lugtenberg B. J. J., editor. (Berlin: Springer Verlag;) 197–202.

de Weger LA, van der Vlugt CIM, Wijfjes AHM, Bakker PAHM, Schippers B, Lugtenberg BJJ (1987) Flagella of a plant growth stimulating *Pseudomonas fluorescens* strain are required for colonization of potato roots. *J Bacteriol*, 169: 2769–2773

DeAngelis KM, Brodie EL, DeSantiz TZ, Andersen GL, Lindow SE, Firestone MK (2009) Selective progressive response of soil microbial community to wild oat roots. *ISME J*, 3: 168–178. 10.1038/ismej.2008.103

DeBruyn JM, Nixon LT, Fawaz MN, Johnson AM, Radosevich M (2011) Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. *Appl Environ Microbiol*, 77: 6295–6300. 10.1128/AEM.05005-11

García-Fraile P, Carro L, Robledo M, Ramírez-Bahena MH, Flores-Félix JD, Fernández MT, et al. (2012) *Rhizobium* promotes non-legumes growth and quality in several production steps: towards a biofertilization of edible raw vegetables healthy for humans. *PLoS ONE*, 7:e38122 10.1371/journal.pone.0038122

Gogarten J P, Doolittle WF, Lawrence JG (2002) Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* 19: 2226–2238. 10.1093/oxfordjournals.molbev.a004046

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. *Paleontol Electron* 4: 1–9.

Jain R, Rivera MC, Moore JE, Lake JA (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol*, 20: 1598–1602. 10.1093/molbev/msg154

Juhas M (2015) Type IV secretion systems and genomic islands-mediated horizontal gene transfer in *Pseudomonas* and *Haemophylus*. *Microbiol Res*, 170: 10–17. 10.1016/j.micres.2014.06.007

Kavamura VN, Taketani RG, Lançoni MD, Andreote FD, Mendes R, Melo IS (2013) Water regime influences bulk soil and rhizosphere of *Cereus jamacaru* bacterial communities in the brazillian caatinga biome. *PLoS ONE*, 8:e73606 10.1371/journal.pone.0073606

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res*, 41: e1 10.1093/nar/gks808

Lane DJ (1991) "16S/23S rRNA sequencing," in *Nucleic acid techniques in bacterial systematics* eds Stackebrandt E., Goodfellow M., editors. (Chichester: John Wiley and Sons;).

Langille MG, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*, 31: 814–821. 10.1038/nbt.2676

Lilley AK, Fry JC, Day MJ, Bailey MJ (1994) *In situ* transfer of an exogenously isolated plasmid between indigenous donor and recipient *Pseudomonas* spp. in sugar beet rhizosphere. *Microbiology*, 140: 27–33. 10.1099/13500872-140-1-27

Lin W, Wu L, Lin S, Zhang A, Zhou M, Lin R, et al. (2013) Metaproteomic analysis of ratoon sugarcane rhizospheric soil. *BMC Microbiol*, 13: 135 10.1186/1471-2180-13-135

Mansour AA, Da Costa A, Arnaud T, Lu-Chau TA, Fdz-Polanco M, Moreira MT, et al. (2016) Review of lignocellulolytic enzyme activity analyses and scale-down to microplate-based assays. *Talanta*, 150: 629–637. 10.1016/j.talanta.2015.12.073

Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, et al. (1998) Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl Environ Microbiol*, 64: 795–799

Mendes LW, Kuramae EE, Navarrete AA, van Veen JA, Tsai SM (2014) Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J*, 8: 1577–1587. 10.1038/ismej.2014.17

Mendes R, Garbeva P, Raaijmakers JM (2014) The rhizosphere microbiome: significance of plant beneficial, plant pathogenic and human pathogenic microorganisms. *FEMS Microbiol Rev*, 37: 634–63. 10.1111/1574-6976.1202

Muyzer G, De Wall EC, Uitterlinden AG (1993) Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol*, 59: 695–700

Nunes da Rocha U, Plugge CM, George I, van Elsas JD, van Overbeek LS (2013) The rhizosphere selects for particular groups of *Acidobacteria* and *Verrucomicrobia*. *PLoS ONE*, 8: e82443 10.1371/journal.pone.0082443

Palmer AC, Kishony R (2013) Understanding, predicting and manipulating the genotypic evolution of antibiotic resistance. *Nat Rev Genet*, 14: 243–248. 10.1038/nrg3351

Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*, 30: 3123–3124. 10.1093/bioinformatics/btu494

Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, et al. (2013) Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc Natl Acad Sci U.S.A.*, 110: 6548–6553. 10.1073/pnas.1302837110

Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH (2013) Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol*, 11: 789–799. 10.1038/nrmicro3109

Prashar P, Kapoor N, Sachdeva S (2014) Rhizosphere: its structure, bacterial diversity and significance. *Rev Environ Sci Biotechnol*, 13: 63–77. 10.1007/s11157-013-9317-z

Pukall R, Tschäpe H, Smalla K (1996) Monitoring the spread of broad host and narrow host range plasmids in soil microcosms. *FEMS Microbiol Ecol*, 20: 53–66. 10.1111/j.1574-6941.1996.tb00304.x

Rachid CT, Santos AL, Piccolo MC, Balieiro FC, Coutinho HLC, Peixoto RS, et al. (2013) Effect of sugarcane burning or green harvest methods on the Brazilian cerrado soil bacterial community structure. *PLoS ONE*, 8: e59342 10.1371/journal.pone.0059342

Ramey BE, Koutsoudis M, von Bodman SB, Fuqua C (2004) Biofilm formation in plant-microbe associations. *Curr Opin Microbiol*, 7: 602–609. 10.1016/j.mib.2004.10.014

Redondo-Nieto M, Barret M, Morrissey J, Germaine K, Martínez-Granero F, Barahona E, et al. (2013) Genome sequence reveals that Pseudomonas fluorescens F113 possesses a large and diverse array of systems for rhizosphere function and host interaction. *BMC Genomics*, 14: 54 10.1186/1471-2164-14-54

Rinaudi LV, Giordano W (2010) An integrated view of biofilm formation in rhizobia. *FEMS Microbiol Lett*, 304: 1–11. 10.1111/j.1574-6968.2009.01840.x

Roes-Hill M, Khan N, Burton SG (2011) Actinobacterial peroxidases: an unexplored resource for biocatalysis. *Appl Biochem Biotechnol*, 164: 681–713. 10.1007/s12010-011-9167-5

Salles JF, Poly F, Schmid B, Roux XL (2009) Community niche predicts the functioning of denitrifying bacterial assemblages. *Ecology*, 90: 3324–3332. 10.1890/09-0188.1

Simons M, van der Bij AJ, Brand I, de Weger LA, Wijffelman CA, Lugtenberg BJJ (1996) Gnotobiotic system for studying rhizosphere colonization by plant-growth promoting *Pseudomonas* bacteria. *Mol Plant Microbe Interact*, 9: 600–607. 10.1094/MPMI-9-0600

Suárez-Moreno ZR, Caballero-Mellado J, Coutinho BG, Mendonça-Previato L, James EK, Venturi V (2012) Common features of environmental and potentially beneficial plant-associated *Burkholderia*. *Microb Ecol*, 63: 249–266. 10.1007/s00248-011-9929-1

Tan S, Yang C, Mei X, Shen S, Raza W, Shen Q, et al. (2013) The effect of organic acids from tomato root exudates on rhizosphere colonization of *Bacillus amyloliquefaciens* T-5. *Appl Soil Ecol*, 64: 15–22. 10.1016/j.apsoil.2012.10.011

Tawaraya K, Horie R, Saito A, Shinano T, Wagatsuma T, Saito K, et al. (2015) Metabolite profiling of shoot extracts, root extracts, and root exudates of rice plant under phosphorus deficiency. *J Plant Nutr*, 36: 1138–1159. 10.1080/01904167.2013.780613

ter Braak CJF, Smilauer P (2002) *CANOCO reference manual and CanoDraw for windows user's guide: software for canonical community ordination (version 4.5)*. Ithaca, NY: Microcomputer Power.

Thomas F, Hehemann JH, Rebuffet E, Czjzek M, Michel G (2011) Environmental and gut Bacteroidetes: the food connection. *Front Microbiol*, 2: 93 10.3389/fmicb.2011.00093

Turner TR, James EK, Poole PS (2013) The plant microbiome. *Genome Biol*, 14: 209 10.1186/gb-2013-14-6-209

Uroz S, Buée M, Murat C, Frey-Klett P, Martin F(2010) Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ Microbiol Rep*, 2: 281–288. 10.1111/j.1758-2229.2009.00117.x

Van Elsas JD, McSpadden Gardener BB, Wolters AC, Smit E (1998) Isolation, characterization, and transfer of cryptic gene-mobilizing plasmids in the wheat rhizosphere. *Appl Environ Microbiol*, 64: 880–889.

Wessén E, Hallin S, Philippot L (2010) Differential responses of bacterial and archaeal groups at high taxonomical ranks to soil management. *Soil Biol Biochem*, 42: 1759–1765. 10.1016/j.soilbio.2010.06.013

Zarraonaindia I, Owens SM, Weisenhorn P, West K, Hampton-Marcell J, Lax S, et al. (2015) The soil microbiome influences grapevine-associated microbiota. *mBio*, 6: e2527-14 10.1128/mBio.02527-14

Zhang L, Thiewes H, van Kan JAL (2011) The D-galacturonic acid catabolic pathway in Botrytis cinerea. *Fungal Genet Biol*, 48: 990–997. 10.1016/j.fgb.2011.06.002

4 TROPICAL SOILS ARE A RESERVOIR FOR FLUORESCENT *Pseudomonas* spp. BIODIVERSITY: PROPOSITION FOR NEW *P. fluorescens* SUBGROUPS

Abstract

Fluorescent *Pseudomonas* spp. are an exceptionally diverse group of bacteria containing several species able to perform beneficial activities to plants. To explore the genetic diversity of *Pseudomonas* spp. in tropical regions, we collected 76 isolates from a Brazilian soil. The genomes were sequenced and various phylogenetic and genomic analyses were used to assess and compare them to known *Pseudomonas* strains, which were mostly collected from temperate regions. Multi-Locus Sequence Analysis (MLSA) classified the 76 tropical isolates in the *P. fluorescens* (57) and *P. putida* (19) groups. Among the isolates in the *P. fluorescens* group, most (37) were classified in the *P. koreensis* subgroup and two were classified in the *P. jessenii* subgroup. The remaining 18 isolates fell into two phylogenetic clades that are distinct from currently recognized *P. fluorescens* subgroups. Accordingly, we propose their classification in two new *P. fluorescens* subgroups. Consistent with their phylogenetic distance from previously described subgroups, the genome sequences of strains in the new subgroups are asyntenous to genomes of members of their neighbor subgroups. In addition, members of the new subgroups share exclusive genes not observed in the other subgroups, and point to the potential for novel functions. The isolates seem to host several functional genes also present in the genomes of known fluorescent *Pseudomonas* spp. strains. We also identified 12 potential new species among the strains classified in the known and in the new subgroups. The high degree of *Pseudomonas* phylogenetic and species novelty in this tropical soil is possibly related to endemism and biogeographical patterns.

Keywords: Soil bioprospecting, bacterial diversity, tropical Pseudomonas spp., fluorescent Pseudomonads, P. fluorescens classification

4.1 Introduction

The bacterial genus *Pseudomonas* contains multiple species of bacteria (Palleroni, 1992; Silby et al., 2011; Loper et al., 2012). These species are genetically, ecologically and functionally diverse, and can be found in many terrestrial and aquatic habitats (Spiers et al., 2000). *Pseudomonas* species can be soil saprophytes; degrade pollutants; be mutualists or pathogens associated to plants, insects, animals and humans; and produce several compounds of industrial interest (Silby et al., 2011; Loper et al., 2012). Currently, the *Pseudomonas*

species are classified in 10 groups, *P. fluorescens*, *P. aeruginosa*, *P. putida*, *P. syringae*, *P. anguilliseptica*, *P. oryzihabitans*, *P. stutzeri*, *P. oleovorans*, *P. straminae*, and *P. pertucinogena* (Gomila et al., 2015; Garrido-Sanz et al., 2016). The *P. fluorescens* group is one of the most diverse, and is further sub-classified into ten subgroups (Gomila et al., 2015; Garrido-Sanz et al., 2016). Nonetheless, as a consequence of its high diversity, the classification of species of *Pseudomonas* is challenging and is constantly being amended as new species are described and new methods are developed (Gomila et al., 2015; Garrido-Sanz et al., 2016).

In agricultural systems, there is particular interest in fluorescent *Pseudomonas* spp. that live in the soil/plant environments because of their potential to improve plant productivity. Strains of fluorescent *Pseudomonas* characterized as plant growth promoting rhizobacteria (PGPR) are most commonly within the *P. putida* and *P. fluorescens* groups (Silby et al., 2011). PGPR strains produce iron-chelating siderophores (Trapet et al., 2016), synthesize phytohormones (Patten and Glick, 2002), solubilize phosphorus (de Werra et al., 2009), protect against abiotic stresses (Cho et al., 2015), or control against insects or microbial pathogens (Hofte and Altier, 2010; Loper et al., 2012; Hol et al., 2013).

The tropics occupy approximately one-third of Earth's land surface. The climate, vegetation, geomorphology, lithology and consequently the soils within these regions vary more than those of temperate and arctic regions (Kalpage, 1974). Such variation provides expanded habitats and niches highly contributing to the evolution and diversification of organisms. Therefore, tropical ecosystems are major reservoirs of biological diversity on our planet. The diversity of macroscopic species decreases from the tropics to the poles but the existence of a latitudinal gradient in microbial diversity is controversial. A latitudinal gradient has been described for *Streptomyces* (Andam et al., 2016) and geographical distance was shown to be the main factor for the structuring of soil fungal communities (Gumiere et al., 2016). On the other hand, edaphic variables were suggested to be the main factors controlling biogeography of soil bacterial communities (Fierer and Jackson, 2006). Fluorescent *Pseudomonas* spp. seem to have some degree of endemicity, i.e. specificity of some species/strains to different geographical locations, and were suggested to be unequally distributed around the planet following biogeographical patterns (Cho and Tiedje, 2000).

Understanding the diversity of microorganisms in the tropics is key to resolving these fundamental concepts of global biogeography. Although *Pseudomonas* spp. are important in agricultural systems, knowledge of these bacteria in tropical soils has lagged behind those from temperate soils. The few reported examples of tropical strains include the

PGPR strain *P. fluorescens* Ps006, isolated from banana roots in Colombia (Gamez et al., 2016); the biocontrol and PGPR strain *P. fluorescens* UM270 isolated from the rhizosphere of wild *Medicago* sp. in Mexico (Hernandez-Salmeron et al., 2016); strains *P. fluorescens* BRIP34879 and *P. fluorescens* SRM1 obtained from cereal crops and spoiled raw milk, respectively in Queensland, Australia, (Gardiner et al., 2013; Lo et al., 2015); and strain *P. fluorescens* RP47 was isolated from the maize rhizosphere in Brazil (Araujo et al., 1993).

Most studies in microbial biogeography have used molecular biomarkers, which are unable to accurately analyze the intra-species diversity. The goal of this study was to expand knowledge on the diversity of fluorescent *Pseudomonas* spp. in a tropical soil using phylogenetic and genomic approaches. We discovered that tropical soils harbor novel members of the *P. fluorescens* group that fall outside of previously-defined subgroups and with gene inventories that are distinct from known strains in this group of bacteria. The genomic approach allowed the intra-species comparison of fluorescent *Pseudomonas* spp., and the high degree of phylogenetic and species novelty in the tropical soil argue for an unequal distribution of soil fluorescent *Pseudomonas* throughout the world and support the possibility of endemism.

4.2 Material and Methods

4.2.1 Isolation of fluorescent *Pseudomonas* spp. from soil samples

Isolates were obtained in 2014 from a soil (Oxisol) in Piracicaba-SP, Brazil, managed for the past 10 years using green-harvest system for sugarcane cultivation. Samples were collected from six different points of the field and from two soil compartments: bulk soil, the root-free soil located between the crop rows and at a depth of 0-10 cm; and rhizosphere, the 1-2 mm layer of soil adhering to plant roots. All bulk soil or rhizosphere samples were homogenized, and serially diluted. Aliquots were plated on the selective media Pseudomonas Agar Base (Oxoid, Basingstoke, Hampshire, UK) supplemented with the antibiotics cetrimide, fucidin and cephalosporin (10, 10 and 50 mg L^{-1}, respectively). Plates were incubated at 28ºC and colonies of *Pseudomonas* were periodically visualized under UV light over the span of two days. A total of 76 single fluorescent colonies were obtained in this process and stored in a 35% glycerol solution at -80ºC.

4.2.2 DNA isolation and genome sequencing

Genomic DNA from the 76 isolates was extracted using the Wizard Kit (Promega, Madison, Wisconsin, USA) after growing cultures overnight with the same (broth) medium and conditions described above. Nextera libraries were constructed and sequenced by the Center for Genome Research and Biocomputing (CGRB) at Oregon State University. The 76 barcoded libraries were sequenced using the 150mer paired end sequencing kit, in a single lane on an Illumina HiSeq 3000.

4.2.3 Genome assemblies and annotations

FastQC (Andrews, 2010) was used to assess the quality of the sequencing reads and the Cutadapt 1.9.1 software (Martin, 2011) was used to trim off adapter sequences. Paired-end sequencing reads were *de novo* assembled, using SPAdes 3.7.0 (Bankevich et al., 2012), with a phred-offset of 33, and multiple kmers (21, 33, 55, 77, 99). Small contigs (<500 bp) and palindromic sequences were removed from the assemblies. The scaffold genomes were annotated using the Prokka software (Seeman, 2014), which also generated the .gbk files. This Whole Genome Shotgun project (BioProject PRJNA383709) has been deposited at DDBJ/ENA/GenBank databases under the accession numbers NEHD00000000-NEKA00000000. The versions described here are NEHD01000000-NEKA01000000.

4.2.4 Phylogenetic and genomic analyses

Using sequences of the housekeeping genes *dnaE*, *guaA*, *mutL*, *ppsA*, *pyrC*, *recA*, and *rpoB* of *P. aeruginosa* PA01 as queries and autoMLSA v. 1.0 (Davis II et al., 2016), homologous sequences from only *Pseudomonas* genome sequences having all seven housekeeping genes were retrieved from the Genome database of NCBI (Loper et al., 2012). Sequences of the housekeeping genes were also extracted from the genome assemblies of each isolate sequenced herein. The gene sequences corresponding to each genome were aligned (MAFFT v. 7.271, L-insi-i algorithm), concatenated, and RAxML v. 8.2.8 was used to generate a maximum likelihood multi-locus sequence analysis (MLSA) phylogenetic tree with bootstrap support (1,000 bootstrap replicates) (Stakamakis, 2014; Katoh et al., 2013). The ItoL platform was used to visualize results and produce the phylogenetic trees (Letunic & Bork, 2016). Analysis was performed on 06/2016.

A second MLSA tree, based on the analysis of four housekeeping genes (16S rRNA, *rpoB*, *rpoD* and *gyrB*), was constructed to infer subgroup classifications. All strains with sequences available in the NCBI Genome database and classified in the *P. fluorescens* or *P.*

putida groups, as well as the type strains of all *P. fluorescens* subgroups and the *P. putida* group, were included in this analysis. Analysis was performed on 03/2017.

The software autoANI v. 1.0 was used to calculate all possible pairwise average nucleotide identity values (ANI; Davis et al., 2016). Each genome sequence was chunked into 1,020 nt fragments and used as queries in all possible reciprocal pairwise BLAST searches. BLAST hits above the default 30% identity and 70% coverage were averaged to calculate the pairwise ANI values. The R-based software program ggplot2 was used to generate a heatmap of the ANI matrix (Wickham, 2009). The Past software (Hammer et al., 2001) was used to normalize the ANI data and a non-metric multidimensional scaling (NMDS) was performed using the Bray-Curtis index.

The orthoMCL algorithm in the Get Homologues software (Contreras-Moreira and Vinuesa, 2013) was used to identify clusters of homolog genes from the coding sequences (CDSs). The CDSs from 18 reference strains (*P. fluorescens* Pf0-1, *Pseudomonas* sp. MS586, *P. aeruginosa* DK2, *P. alkylphenolia* KL28, *P. chlororaphis* PCL1606, *P. fluorescens* FW300-N2E3, *P. fluorescens* NCIMB 11764, *P. mandelii* JR-1, *P. monteilii* SB 3078, *P. mosselii* SJ10, *P. protegens* Pf-5, *P. putida* H8234, *P. putida* HB3267, *P. putida* S12, *P. putida* W619, *Pseudomonas* sp. FGI182, *Pseudomonas* sp. UW4, and *P. syringae* CC1557) - representing phylogenetically related and distant strains - were used for comparisons. Mauve 2.4.0 software, using the progressiveMauve algorithm, was used to align multiple genome sequences for synteny analyses (Darling et al., 2010).

A database of translated sequences that exemplify key functions of fluorescent *Pseudomonas* spp. was developed and used as queries in BLAST searches (tBLASTn) against the genome assemblies of the 76 isolates from Brazilian soil. The amino acid sequences were extracted from the .faa files corresponding to: *P. fluorescens* Q8r1-96 (ACC deaminase; T3SS-1; T3SS-2; T6SS HSI-I), *P. fluorescens* Pf0-1 (AprA; chitinase; GABA catabolism; T2SS Hxc; T6SS HSI-II; T6SS HSI-III; Tcc4; Tcc5; carocin-like; S-type + cytotoxic domain proteins; prototypic S-type), *P. protegens* Pf-5 (DAPG; HCN; pyoluteorin; pyrrolnitrin; rhizoxin), *P. chlororaphis* O-6 (IAA biosynthesis), *P. fluorescens* SBW25 (T2SS Hxc-2), and *P. fluorescens* Q2-87 (T3SS-3).

4.3 Results and Discussion

4.3.1 Genome Sequencing and Assembly

A total of 560 million reads were generated from the multiplexed 76 libraries. The reads were processed, and independently *de novo* assembled for each genome sample. The genome assemblies had an average of 43 contigs (from 15 to 94 contigs) (Supplementary Table 4.1). The total within-assembly sizes for each genome were comparable to sizes of finished genome sequences of related *Pseudomonas* spp. (Supplementary Table 4.1). We thus, estimate the average coverage was approximately 70X for each sequenced isolate. The assemblies were annotated and yielded an average of 5,559 coding sequences (CDSs) per genome sequence, consistent with the expected density of most bacterial genomes.

4.3.2 Most of the Tropical Isolates are Members of the *P. fluorescens* and *P. putida* Groups

To determine the identity and phylogeny of the isolates, a phylogenetic tree based on sequences of seven housekeeping genes was constructed (Loper et al., 2012). Seventy-four of the 76 isolates clustered with strains that belong to one of two recognized groups of *Pseudomonas*, *P. fluorescens* and *P. putida* (Figure 4.1). The 55 isolates that clustered to the *P. fluorescens* group formed 18 nodes. The 19 isolates that clustered to *P. putida* group formed nine nodes. None of the included reference strains were positioned in the 27 nodes formed by the isolates collected and analyzed in this study. However, the reference strains *P. fluorescens* Pf0-1 and *Pseudomonas* sp. MS586 were closely related to the isolates in the *P. fluorescens* group, while *Pseudomonas sp.* FGI182, *P. putida* HB3267 and *P. putida* W619 were closely related to isolates in the *P. putida* group (Figure 4.1).

The remaining two isolates (B6 and R26), though belonging to *Pseudomonas*, did not associate to either of the aforementioned groups, or any other established group of *Pseudomonas* (Figure 4.1). Despite Figure 4.1 highlights *P. fluorescens* and *P. putida* groups, strains of all other *Pseudomonas* groups were analyzed in the MLSA, and these two isolates did not cluster to any of those groups. The two formed their own node, which shared a common ancestor to the *P. putida* group based on this phylogenetic tree (Figure 4.1).

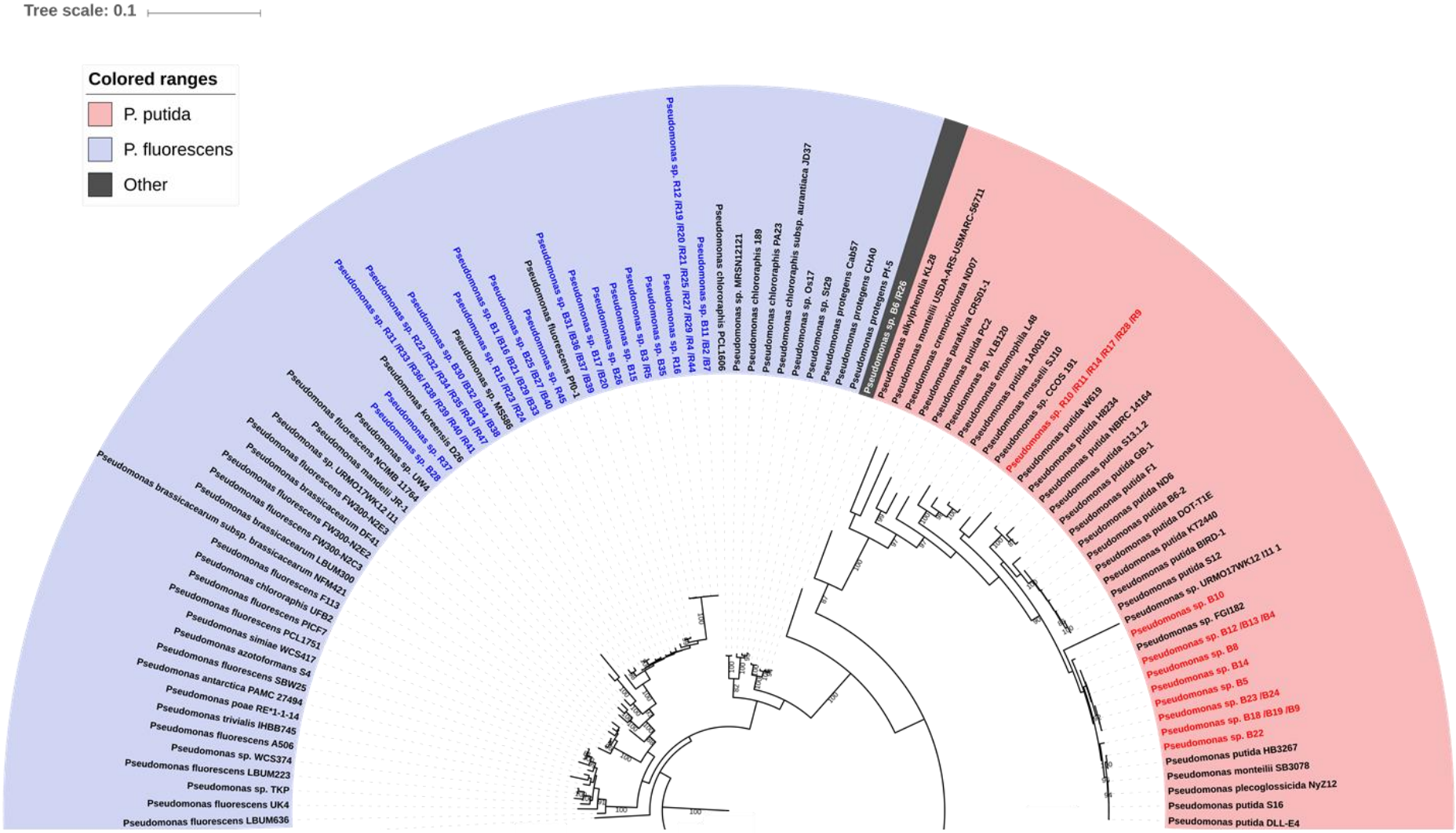


Figure 4.1 MLSA Phylogenetic tree. All *Pseudomonas* genome sequences available in the NCBI Genome database and with the housekeeping genes: *dnaE*, *guaA*, *mutL*, *ppsA*, *pyrC*, *recA*, and *rpoB* were used. Light blue region indicates the *P. fluorescens* group, while light red region indicates the *P. putida* group. The dark grey region comprises the isolates B6 and R26 that were not classified in the known subgroups in this tree (Other). Labels in black are the reference strains, blue and red labels are the ones isolated in this study. Branch lengths are supported by bootstrap values (numbers below each clade). Analysis performed on 06/2016

4.3.3 Tropical Soils Harbor Isolates Belonging to New *P. fluorescens* Subgroups

To test for subgroup associations, we constructed a second phylogenetic tree using just four housekeeping genes (16S rRNA, *rpoB, rpoD* and *gyrB*), which allowed for inclusion of a greater diversity and quantity of type strains and for comparison with previous studies that focused on *P. fluorescens* subgrouping classification (Gomila et al., 2015; Garrido-Sanz et al., 2016). This tree was consistent with the previously described tree in classifying the 55 isolates to the *P. fluorescens* group and 19 isolates to the *P. putida* group, and in placing isolates B6 and R26 in a separate clade (Figure 4.2). Of the 55 isolates in the *P. fluorescens* group, 37 associated to the *P. koreensis* subgroup. Isolates R37 and B28 were classified in the *P. jessenii* subgroup (Figure 4.2). Those from tropical soils that associated to the *P. putida* group were intermingled with reference strains, but represent new haplotypes. This second phylogenetic tree agreed with the previous one in suggesting that the newly sequenced isolates are new haplotypes and may also represent new species. The number of haplotypes was slightly higher using this approach, increasing to 10 and 24 in the *P. putida* and *P. fluorescens* group, respectively (Figure 4.2).

One clade containing 16 isolates within the *P. fluorescens* group did not cluster to any previously recognized subgroup. The separation of this clade from its closest subgroup, *P. koreensis*, is well supported (bootstrap value = 100). These results suggest that some of the isolates collected from the tropical soil are sufficiently distinct from strains available in the databases, which were mostly identified from temperate regions. Additionally, B6 and R26 did not associate to any currently recognized subgroup. However, unlike the previous analysis, these two isolates associated with the *P. fluorescens* group, and were located between the *P. protegens* and *P. chlororaphis* subgroups (Figure 4.2). The distance between these new clades and clades of previously recognized subgroups is high enough to recommend new subgroups of *P. fluorescens*, which will be named after species description.

In order to assess the possibility of other known strains not included in the phylogenies fit in the newly recommended subgroups, a BLASTn approach was used. The genes *rpoD*, *rpoB* and *gyrB* from isolates representing the new subgroups were used as queries. There were no hits with higher identity and similarity than the strains already included in our phylogenetic trees. Thus, we conclude that there are no other strains available in the databases that fit in the new subgroups proposed in this study.

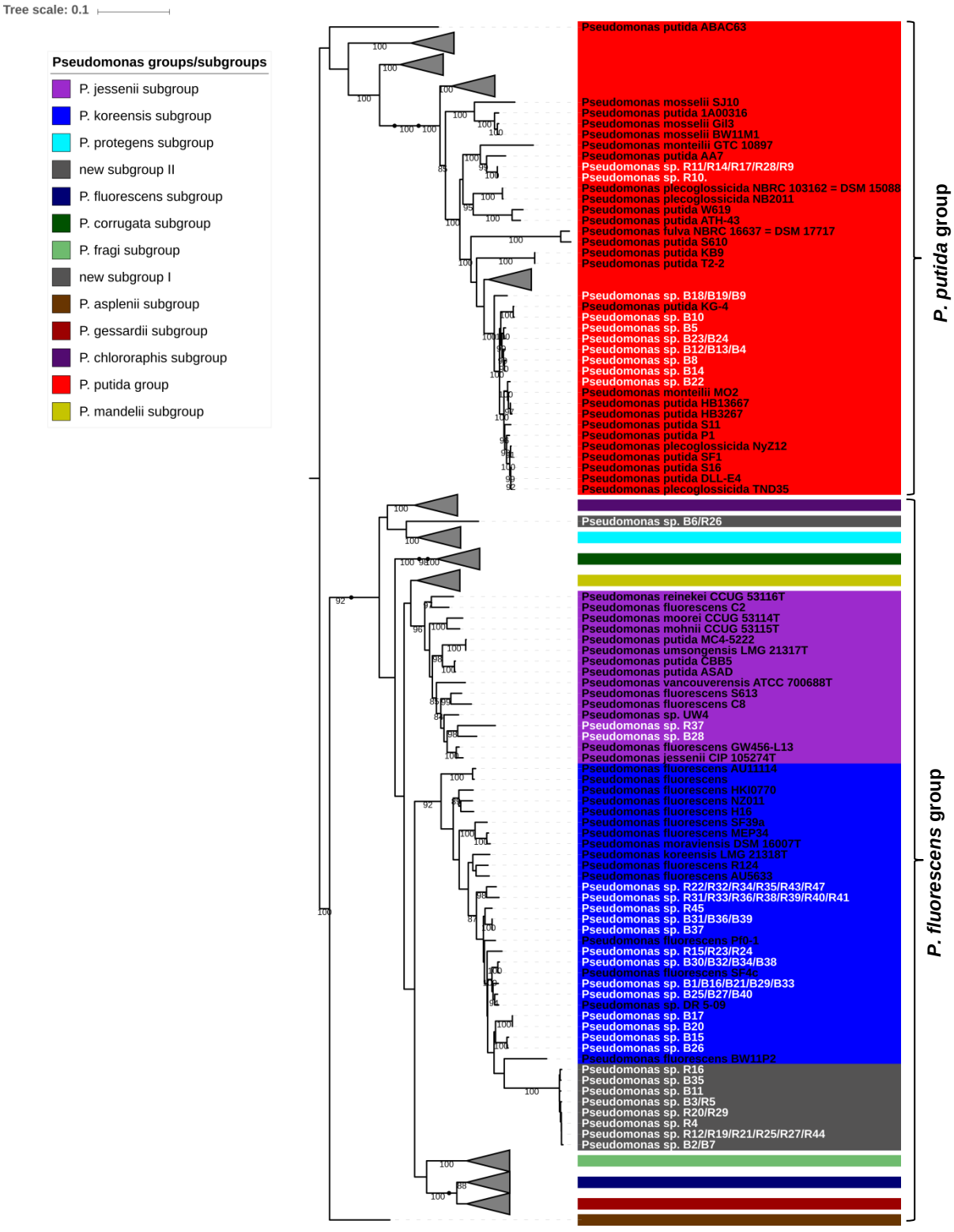


Figure 4.2 Phylogenetic tree inferred by MLSA using the housekeeping genes: *gyrB, rpoD, rpoB* and 16S rRNA. *Pseudomonas* type strains of the *P. putida* group and all *P. fluorescens* subgroups that have those genes were included in the analysis, besides all the reference strains with available genome sequences classified in the *P. putida* and *P. fluorescens* groups. Colors refer to the *Pseudomonas* groups and *P. fluorescens* subgroups. The proposed new subgroups are colored in grey. Labels in black are the reference and type strains, while labels in white are the isolates of this study. Branch lengths are supported by bootstrap values (numbers below each clade). Analysis performed on 03/2017

4.3.4 ANI Supports the Phylogenetic Classification and the Presence of New Subgroups

A whole-genome based approach was used to support conclusions derived from phylogenetic analyses. ANI was calculated for all possible pairwise comparisons between the genome sequences of the 76 isolates and those from key reference strains (Figure 4.3). The pattern based on ANI clustering supported the conclusions drawn from the analyses of phylogenetic trees based on only seven or four marker genes (Figures 4.1, 4.2 and 4.3). Use of ANI clustered isolates into groups that mirrored the clades defined from analyses of phylogenetic trees. More importantly, use of ANI confirmed the existence of new subgroups. Inter-group comparisons between members of the new subgroups and previously defined subgroups had ANI values of less than 85%. Analysis using NMDS of the ANI values was consistent in supporting subgroup assignments and the recommendation for defining new subgroups (Supplementary Figure S4.1).

ANI was also used to infer the number of possible new species found in our study. ANI values of 94-96% correspond to 70% similarity in DNA-DNA hybridization and this threshold can be used to operationally classify bacteria into species (Konstantinidis & Tiedje, 2004; Kampfer, 2012; Oren & Garrity, 2014). When considering different species the ones with <95% ANI, it was found found a total of 18 species among our 76 isolates, 13 classified in the *P. fluorescens* group and five in the *P. putida* group. In the *P. fluorescens* group, eight isolates are potentially the same species of the reference strain *Pseudomonas* sp. DR 5-09; four isolates are the same species of the reference strain *P. fluorescens* SF4c; and three isolates are the same species of the reference strain *Pseudomonas* sp. MS586. The other 42 isolates in this group comprised 10 individual clusters showing <95% ANI to any reference strain and thus are potential new species (Figure 4.3).

In the *P. putida* group, eight isolates are potentially the same species of the reference strain *Pseudomonas* sp. FGI182; one isolate is the same species of the reference strain *P. putida* HB3267; and one isolate is the same species of the reference strain *P. putida* KG-4. The remaining 9 isolates formed two single clusters with <95% ANI to any strain used in this analysis and are also potential new species. Therefore, using the 95% ANI cutoff it was found 10 putative new species in the *P. fluorescens* group and 2 putative new species in the *P. putida* group (Figure 4.3). However, to get the accurate number of new species we need to follow the taxonomy protocol, meaning that other analyses are required and will be performed in future works.

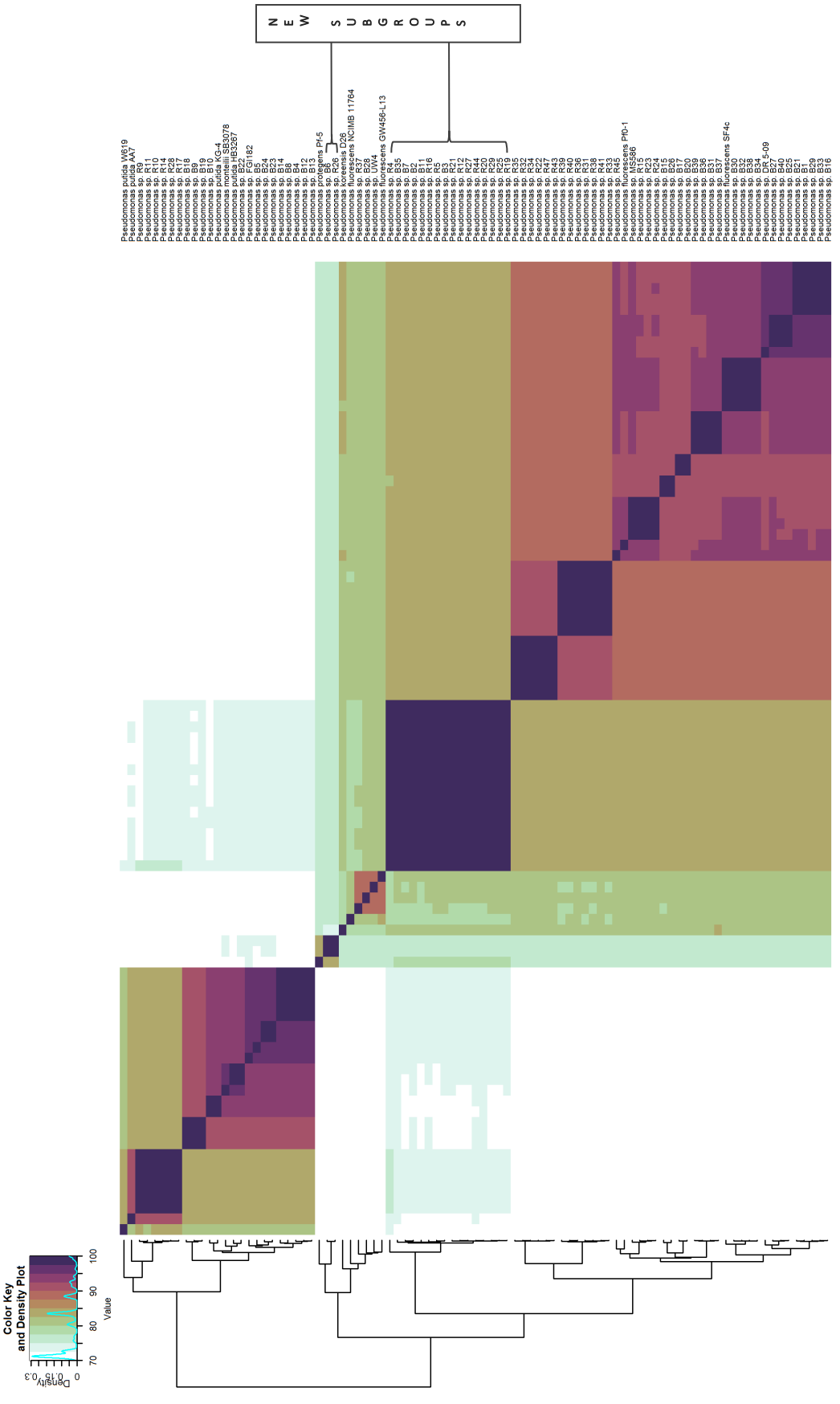


Figure 4.3 Average nucleotide identity (ANI) heatmap showing the 76 isolates obtained in this study and phylogenetically closed reference strains revealed by the phylogenetic inferences. The darker are the colors, the higher are the ANI percentages. The clusters of the new subgroups are highlighted

4.3.5 Specific Gene Clusters and Genome Organization of Isolates in the New Subgroups

With the 76 isolates clustered into groups, subgroups, and potential new species, we next turned our attention to identifying discriminating functions that could be used to distinguish between the strains belonging to the new subgroups from those in the known subgroups. The Get Homologues software was used to identify orthologous clusters. From this, a total of 15,138 gene clusters were found, of which 2,839 clusters were present and had an ortholog in each of the 76 genome sequences (Table 4.1). The remaining 12,299 were polymorphic as presence/absence. Inclusion of 18 reference genome sequences reduced the number of conserved clusters to 2,064 and significantly increased the total number of clusters to 23,561 (Table 4.1).

Table 4.1 Number of homolog genes (gene clusters) among genome sequences analyzed

	All isolates (76)	All isolates + 18 references (94)	Isolates in the *P. fluorescens* group (57)	Isolates in the *P. putida* group (19)
Shared gene clusters	2,839	2,064	3,580	3,816
Accessory gene clusters	12,299	21,497	7,640	4,313
Total gene clusters	15,138	23,561	11,220	8,127

To determine whether the proposed subgroups have distinguishing functions, we focused on clusters that are present in all members of the new subgroups and absent in all other subgroups. For the subgroup next to *P. koreensis*, a total of 105 gene clusters was exclusive to its members. The 105 gene loci are unlinked and distributed throughout each genome, arguing against their acquisition through a limited number of horizontal gene transfer (HGT) events. Sixty-one of the 105 gene clusters were annotated as “hypothetical proteins”. The small number of genes with annotated functions makes it difficult to test for enriched functions that may distinguish this new subgroup from species of other subgroups. Results from synteny analyses of representative genome sequences also supported the formation of a subgroup distinct from *P. koreensis* subgroup, which is consistent with our phylogenetic and ANI analysis (Figure 4.4A). The genome sequence of Brazilian isolate B34 is represented by 15 contigs, and they are co-linear to the genome sequence of strain Pf0-1. In contrast, the genome sequence of isolate R21, which clustered with a new subgroup and is represented by 57 contigs, is slightly different from the B34 or Pf0-1 genomes in organization and has an

inversion event that includes the region of termination. Additionally, Mauve alignment of genome sequences from all isolates that belong to the proposed subgroup next to *P. koreensis*, reveal them to be largely syntenic (Supplementary Figure S4.2A).

Similarly, 527 gene clusters were exclusively identified in the new subgroup next to *P. protegens*. Of these, 315 were annotated as "hypothetical proteins". As observed for the new subgroup next to *P. koreensis*, these exclusive gene clusters are dispersed throughout the genome, which argues against their acquisition through a few HGT events. Genome organization is also quite divergent compared to its neighbor subgroup (*P. protegens*) (Figure 4.4B). As was the case above, the two strains of the new subgroup next to *P. protegens* are co-linear (Supplementary Figure S4.2A and B). Therefore, our results accumulated several evidences supporting the classification of these isolates in new subgroups.

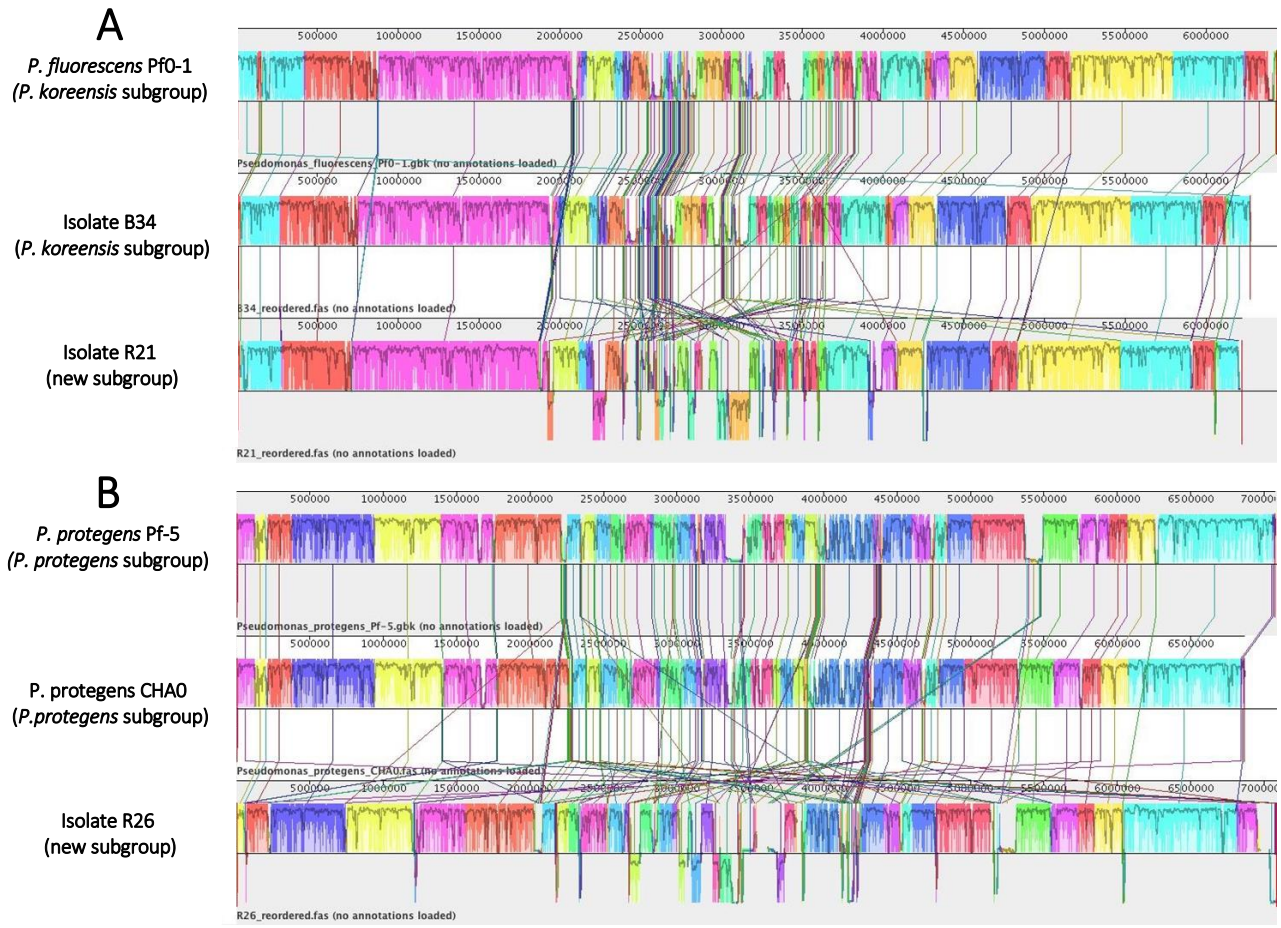


Figure 4.4 Multiple genome alignments. Selected genome sequences were aligned using Mauve software. Different colors in the genome sequences are chunks of high percentage alignment regions. Lines indicate rearranged regions and show their new position in the compared genomes. Origin of replication is in the extreme left of each genome sequence, based on the reference genome sequences used to reorder the contigs of the other genomes analysed, i.e. A*) P. fluorescens* Pf0-1 for the comparison between the *P. koreensis* subgroup with its neighbor new subgroup; and B) *P. protegens* Pf-5 for the comparison between the *P. protegens* subgroup and its neighbor new subgroup

4.3.6 Genetic Content of the Isolates Covers Several Known Functions Performed by Fluorescent *Pseudomonas* spp.

Several functions, such as antibiotic production, protein secretion systems, and mechanisms for interacting with plants have been identified as important for the plant-associated lifestyles of fluorescent *Pseudomonas* spp. (Rezzonico et al., 2005; Kidarsa et al., 2013). We therefore mined the 76 genome sequences to determine whether isolates inhabiting tropical soils could potentially have similar functions and exhibit similar lifestyles as their temperate counterparts (Supplementary Table 4.2). Two sets of genes were found in all 57 genome sequences of the isolates in the *P. fluorescens* group (100%). One set is associated with pyoverdine biosynthesis and the other is associated with Gamma-aminobutyric acid (GABA) catabolism. Almost all the other function-encoding genes analyzed were found in

most, but not all of the genome sequences of the *P. fluorescens* group. Of the functions that were examined, only ACC deaminase was not found in any of the 57 isolates in this group (0%).

Pyoverdine biosynthesis genes are perhaps the most outstanding of the analyzed functions (Supplementary Table 4.2). Pyoverdines are a class of structurally-related siderophores that are produced by fluorescent *Pseudomonas* spp., function in iron chelation and uptake, and are characterized by their green fluorescence (Visca et al., 2006). It is assumed that pyoverdine is a public good and that *Pseudomonas* spp. populations have a balance between pyoverdine producers and pyoverdine defective genotypes (Lujan et al., 2015). In temperate soils, pyoverdine depletes the amount of iron available to pathogens and helps control their populations in the rhizosphere (Visca et al., 2006). However, the tropical soil used in this study was an Oxisol, which is characterized by a high availability of iron due to its chemical characteristics (Brady and Weil, 2002). In principle, the increased availability should reduce the beneficial effect of pyoverdine production for plant protection. Nonetheless, pyoverdine may also be involved in biofilm formation and could be advantageous for survival of bacteria in such soils with lower levels of organic matter (Visca et al., 2006; Ross, 1993; Guimaraes et al., 2013). We therefore speculate that the tropical isolates could synthesize pyoverdine for this latter benefit.

GABA is a non-protein amino acid that functions in stress tolerance of plants as well as in plant-microbe communication (Dagorn et al., 2013). The presence of GABA caused increase accumulation of LPS and affected biofilm maturation in *P. fluorescens* (Dagorn et al., 2013), and could also be related to this process in tropical soils.

Other sets of genes found in >70% of the genome sequences of the *P. fluorescens* group were ones related to hydrogen cyanide (HCN) biosynthesis; the bacteriocins S-type+cytotoxic domain proteins and carocin-like; the exoprotease AprA; the type II secretion systems (T2SS Hxc and T2SS Hxc-2) and type VI secretion system (T6SS HSI-II); and chitinase (Supplementary Table 4.2). In general, these genes function in inter-microbial competition. The extracellular protease AprA is related to both control of protist grazing and entomopathogenic activity (Jousset et al., 2006; Chen et al., 2104). Bacteriocins were shown to play a role in *P. fluorescens* competitiveness with other bacteria (Bruce et al., 2016), and the T6SS delivers toxic effector proteins into competing bacteria, providing a benefit to the produced strain in the complex bacterial communities (Ho et al., 2014). Chitinase degrades chitin, which is present in arthropods and fungal cell walls, and could be related to anti-fungal

activity (Manikandan & Raguchander, 2015). Finally, HCN has both antimicrobial and insecticidal activity (Flury et al., 2017).

Strains B6 and R26, which are classified in the new subgroup next to *P. protegens*, have genes for the biosynthesis of the antibiotics hydrogen cyanide, pyoluteorin, 2,4-diacetylphloroglucinol (DAPG), pyrrolnitrin and rhizoxin (Supplementary Table 4.2). The closest neighbor of this new subgroup is the *P. protegens* subgroup (Garrido-Sanz et al. 2016), which has strains that also contain the genes related to the biosynthesis of these antibiotics (Loper et al., 2012). Although the strains in this new subgroup and in the *P. protegens* subgroup are phylogenetically different, they share common features for interacting in the microbial community. All the functional genes found in the genome sequences of the *P. koreensis* subgroup were also found in the genome sequences of the new subgroup next to it (Supplementary Table 4.2). Both results indicate a high similarity in gene content related to biotic interactions of the new subgroups compared to their most-closely related subgroups, even considering their high phylogenetic distances and the presence of subgroup-specific gene clusters.

The finding of those functional genes in most part of our isolates in the *P. fluorescens* group suggest a considerable degree of functional redundancy between isolates inhabiting tropical soils and previously characterized *P. fluorescens* strains primarily from temperate regions. The ability of fluorescent *Pseudomonas* spp. to produce chemicals for competing in the soil/plant environments is widely known, and several of these functions can help plants to defend against pathogens and/or pests (Chen et al., 2014; Manikandan & Ragunchander, 2015; Flury et al., 2017). It is noteworthy that the rhizosphere isolates of the *P. putida* group have some degree of phylogenetic relatedness to *P. putida* W619 (Figure 4.1 and 4.2), which is a strain isolated from tissues of the plant *Populus trichocarpa* x *deltoide*s cv. 'Hoogvorst, and showed plant growth promotion activity (Taghavi et al., 2009). A close relative of the isolates classified in the *P. jessenii* subgroup (R37, B28) is the reference strain *Pseudomonas* sp. UW4 (Figure 4.1 and 4.2), which was isolated from the rhizosphere of common reeds and able to promote plant growth (Duan et al., 2013). The strains isolated from tropical soils in this study have not yet been tested for their effects on plant growth or health, but the close relationship of some of these tropical isolates to well-characterized PGPR suggests that they have potential to be beneficial to plants.

4.3.7 New Diversity Found May be Related to Geographical Distribution

The results of this study showed that the *Pseudomonas* genus is a plentiful source of new strains and species, reflecting its tremendous diversity (Spiers et al., 2000). Fluorescent *Pseudomonas* spp. are commonly isolated from agroecosystems and are well known for their activities in plant growth promotion and biocontrol (Hofte & Altier, 2010). Even so, our study pointed to the presence of several new haplotypes and 12 possible new species from the 18 (66.7%) identified among our 76 isolates using a 95% ANI cutoff; and recommends the establishment of two new subgroups. Thus, based on phylogenetic analyses and operational criteria based on ANI, the 76 isolates that we collected from tropical soils represent a substantial new diversity.

Despite the attention given to *Pseudomonas* in the past and the ease with which species in this genus are cultured, we isolated several new haplotypes and possible new species. We speculate that this reflects our focus on a relatively unexplored habitat and geographical location. Currently, there are 104 genome sequences of strains named as *P. fluorescens* in the NCBI Genome database. Only eight (*P. fluorescens* Pt14, *P. fluorescens* BRIP34879, *P. fluorescens* SRM1, *P. fluorescens* EGD-AQ6, *P. fluorescens* S12, *P. fluorescens* UM270, *P. fluorescens* et76 and *P. fluorescens* Ps006) were submitted by institutions from tropical countries (https://www.ncbi.nlm.nih.gov/genome/genomes/150?, searched in 02/10/2017). The lack of available data and studies from tropical countries can explain why we found several possible new species of fluorescent *Pseudomonas* spp. in this study.

Microbial biogeography have shown that soil microbial communities change across the world and several selective factors can influence their composition and structure, including pH, carbon sources, plant species, moisture, and temperature (Fierer et al., 2006; Fierer et al., 2012). In addition to shaping communities, geographic distance also influences the evolution and diversification of microbial species, since both genetic drift and natural selection processes take course independently for populations facing geographical separation (Hanson et al., 2012). The biogeographical theory also suggests microbial endemism. Cho and Tiedje (2000) suggested the possibility of endemism on soil fluorescent *Pseudomonas* spp., arguing against the omnipresent distribution of these species in the world. Our results strengthen this idea by finding several putative new species not described in any other place. The existence of novel species and subgroups in the tropical soil analyzed could represent a case of endemism, until proven otherwise. Even considering the high dispersion and migration potential of prokaryotes – amplified by human interference – the Earth might be

big enough to prevent exogenous bacteria from rapidly conquering new ecosystems; while environmental filtering can limit the easy establishment of new isolates elsewhere. In addition, as environment differs with distance, it is possible to suggest a rapid and distinct bacterial evolution in populations inhabiting very distant regions (Hanson et al., 2012). Hence, the biogeographical patterns added to the lack of available genome sequences and strains from tropical soils possibly resulted in our findings, a description of not only new haplotypes and possible species, but also new *P. fluorescens* subgroups.

The similarities found in gene content indicates that the biogeographical effects may have affected the genome arrangement and phylogenetic diversification on the tropical soil isolates, but not their functionality, at least with respect to activities associated with plant interation of *P. fluorescens*. Notwithstanding, the set of specific gene clusters found in the new subgroups suggests that its members may perform some different activities compared to strains in the neighboring subgroups (*P. koreensis and P. protegens*). Biochemical and phenotypic analyses are required to assess this issue more deeply.

4.4 Conclusions

The tropical soil analyzed in our study harbors new haplotypes and 12 possible new species of the *P. fluorescens* and *P. putida* groups. Two clades of isolates have phylogenetic and genomic distinctions compared to the known subgroups of *P. fluorescens,* and therefore we propose the creation of two new subgroups to classify these strains. Biogeography may be the reason for our finding of such a high quantity of putative new species and even subgroups, since most studies and available data on fluorescent *Pseudomonas* spp. are from temperate regions. Although phylogenetically distinct, genes conferring several known functions of fluorescent *Pseudomonas* spp. are present in the strains isolated from the tropical soil of our study, including strains in the new subgroups. Nevertheless, 105 and 527 specific gene clusters were found in the new subgroups next to *P. koreensis* and *P. protegens*, respectively, and further analyses are needed to elucidate the functions conferred by the genes exclusive to these subgroups.

4.5 Supplementary Material

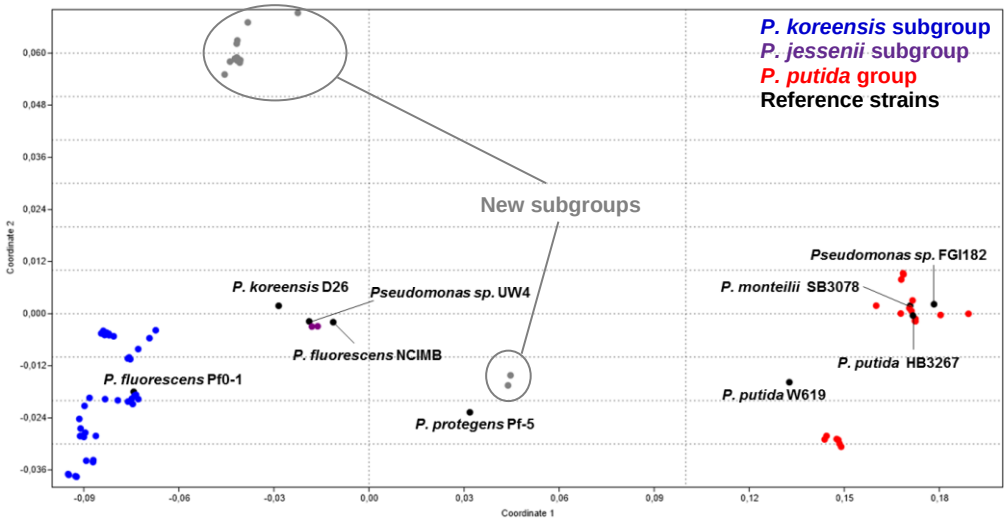
Supplementary Table 4.1 Output of assemblies' main quality parameters

Isolate	No of Contigs	N50 (Kbp)	Longest Contig (Kbp)	Contigs > 1Kbp	Predicted genome size (Mbp)	Isolate	No of Contigs	N50 (Kbp)	Longest Contig (Kbp)	Contigs > 1Kbp	Predicted genome size (Mbp)
B10	51	210.8	526.1	50	5.65	R10	67	280.0	563.4	57	5.98
B11	60	227.5	718.3	54	6.22	R11	47	323.4	668.9	42	5.99
B12	66	217.3	700.0	58	5.95	R12	57	299.8	560.9	52	6.21
B13	61	217.3	700.0	52	5.95	R14	51	280.0	594.4	46	5.99
B14	64	198.5	416.6	57	5.93	R15	34	390.3	777.9	32	6.30
B15	41	261.7	608.6	39	6.40	R16	55	284.7	879.4	53	6.31
B16	37	376.1	587.5	33	6.16	R17	47	320.5	593.8	42	5.99
B17	35	246.5	609.5	34	6.13	R19	63	299.8	561.7	55	6.21
B18	58	242.8	638.8	45	5.56	R20	61	271.1	558.0	57	6.21
B19	57	242.8	638.8	48	5.56	R21	55	299.8	561.4	52	6.21
B1	36	331.2	564.4	35	6.16	R22	23	416.7	680.0	19	6.41
B20	33	270.2	609.5	32	6.13	R23	37	390.3	777.9	34	6.30
B21	38	314.1	652.8	33	6.16	R24	37	390.3	777.9	34	6.31
B22	43	460.4	826.8	34	5.61	R25	58	270.7	558.1	50	6.21
B23	52	285.3	728.6	47	5.93	R26	94	299.3	814.0	73	7.08
B24	64	210.4	476.4	54	5.93	R27	58	299.8	557.9	53	6.21
B25	19	652.9	809.6	18	6.16	R28	44	320.1	593.7	40	5.99
B26	37	287.6	673.9	33	6.23	R29	65	271.1	558.4	57	6.21
B27	21	513.4	808.8	21	6.16	R31	23	636.4	794.6	22	6.26
B28	31	608.8	1,017.0	28	5.93	R32	30	416.7	694.4	27	6.41
B29	34	375.6	803.6	30	6.16	R33	25	597.5	793.1	23	6.26
B2	62	244.4	772.9	55	6.25	R34	27	434.8	836.4	25	6.41
B30	21	572.1	796.6	20	6.27	R35	32	353.7	677.7	30	6.41
B31	32	397.0	1,033.2	30	6.39	R36	30	498.0	793.1	25	6.26
B32	15	706.3	1,189.3	14	6.27	R37	52	305.9	641.6	47	6.03
B33	35	375.3	652.6	30	6.16	R38	24	636.4	744.6	21	6.26
B34	15	717.8	1,189.3	14	6.27	R39	21	597.5	882.8	20	6.26
B35	54	245.1	795.4	49	6.23	R40	24	597.6	882.8	23	6.27
B36	32	382.3	1,033.2	31	6.39	R41	25	636.4	794.6	22	6.27
B37	28	433.1	1,429.8	27	6.39	R43	19	655.1	1,002.2	17	6.41
B38	16	721.6	1,189.3	15	6.27	R44	58	271.0	560.8	53	6.22
B39	31	343.9	1,119.1	30	6.39	R45	29	502.1	953.2	25	6.32
B3	55	280.2	595.1	50	6.20	R47	21	435.1	1,002.1	20	6.41
B40	15	727.7	809.6	14	6.16	R4	93	215.9	561.0	86	6.77
B4	64	216.9	700.0	57	5.92	R5	55	305.3	666.2	52	6.20
B5	50	321.2	603.3	39	5.88	R9	45	319.4	668.4	40	5.99
B6	81	310.2	668.8	67	7.08						
B7	51	323.9	772.6	46	6.25						
B8	56	244.9	753.0	51	5.89						
B9	62	209.9	638.8	51	5.56						

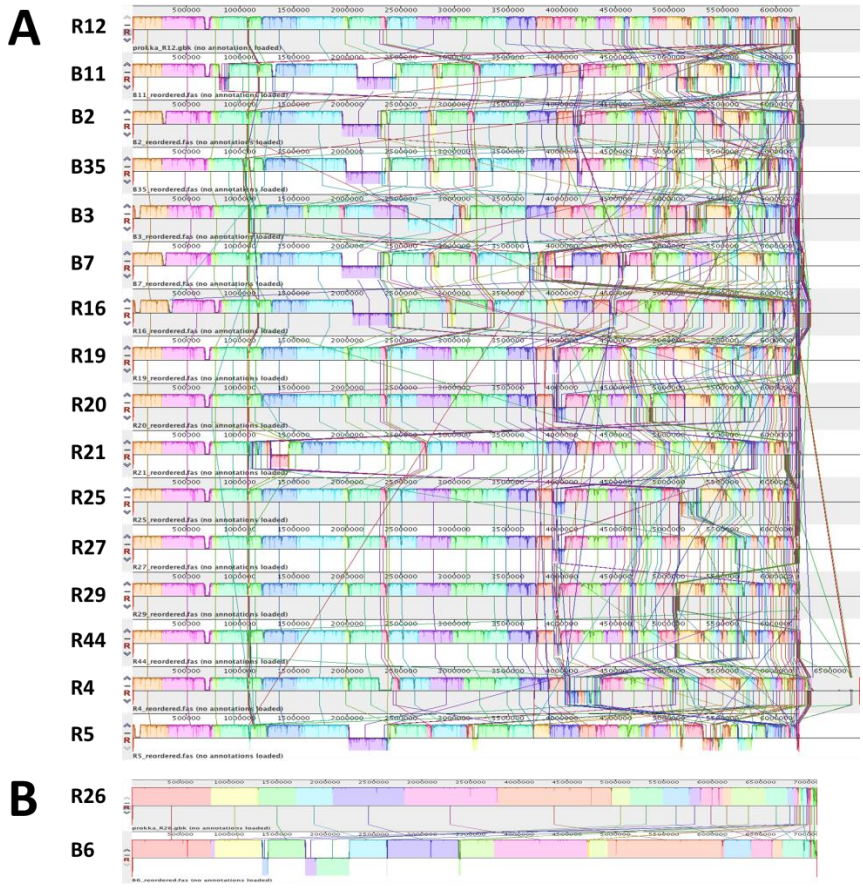
Supplementary Table 4.2 BLAST searches for the set of genes related to important functions performed by known *P. fluorescens* strains. Functions were considered present in a genome only if the complete set of genes were found with high similarity and identity hits in the tBLASTn

		% of genome sequences containing the genes among the 57 isolates of the *P. fluorescens* group
Antibiotics	DAPG	3.5
	HCN	96.5
	Pyrrolnitrin	3.5
	Rhizoxin	3.5
	Pyoluteorin	3.5
Bacteriocin	S-type+cytotoxic domain proteins	77.2
	carocin-like	71.9
	prototypic S-type	21.1
Plant Interaction	GABA catabolism	100.0
	ACC deaminase	0.0
	IAA biosynthesis	47.4
	Pyoverdine biosynthesis	100.0
Insect toxins	Tcc4	98.2
	Tcc5	47.4
	AprA	40.4
	Chitinase	96.5
Secretion Systems	T2SS Hxc	75.4
	T2SS Hxc-2	96.5
	T3SS-1	28.1
	T3SS-2	28.1
	T3SS-3	1.8
	T6SS HSI-I	3.5
	T6SS HSI-II	91.2
	T6SS HSI-III	64.9

Abbreviations: DAPG, 2,4-diacetylphloroglucinol; GABA, Gamma-aminobutyric acid; IAA, indole acetic acid (iaaMH); Tcc4 and Tcc5, toxin complex clusters; AprA, exoprotease; T2SS, type 2 secretion system; T3SS, type 3 secretion system; T6SS, type VI secretion system. Toxin complex clusters and secretion systems are designated according to Loper et al., 2012



Supplementary Figure S4.1 NMDS using the Bray-Curtis similarity index for ordination of samples (isolates and references) according to the ANI results



Supplementary Figure S4.2 Multiple genome alignments (sinteny analysis) inside the new subgroups. Origin of replication is in the middle of each genome sequence, based on the reference genomes used to reorder the contigs of the other assemblies, i.e. A) isolate R12 (~2,962 Mbp) for the comparison inside the new subgroup neighbor to *P. koreensis*; and B) isolate R26 (~3,148 Mbp) for the comparison inside the new subgroup neighbor to *P. protegens*

References

Andam CP, Doroghazi JR, Campbell AN, Kelly PJ, Choudoir MJ, Buckley DH (2016) A latitudinal diversity gradient in terrestrial bacteria of the genus *Streptomyces*. *mBio*, 7(2): e02200-15

Andrews S (2010) *FastQC: a quality control tool for high throughput sequence data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Araujo MAV, Mendonca-Hagler LC, Hagler AN, van Elsas JD (1994) Survival of genetically modified *Pseudomonas fluorescens* introduced into subtropical soil microcosms. *FEMS Microbiol Ecol*, 13: 205-216.

Brady NC, Weil RR (2002) *The nature and properties of soils*. Upper Saddle River, N.J. Prentice Hall

Davis II EW, Weisberg AJ, Tabima JF, Grunwald NJ, Chang JH (2016) Gall-ID: tools for genotyping gall-causing phytopathogenic bacteria. *PeerJ*, 4: e2222 https://doi.org/10.7717/peerj.2222

de Werra P, Peche-Tarr M, Keel C, Maurhofer M (2009) Role of gluconic acid production in the regulation of biocontrol traits of *Pseudomonas fluorescens* CHA0. *Appl Env Microbiol*, 75(12): 4162-4174.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pavel A (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*, 19(5): 455-477.

Bruce JB, West SA, Griffin AS (2017) Bacteriocins and the assembly of natural *Pseudomonas fluorescens* populations. *J Evol Biol*, 30: 352-360.

Chen WJ, Hsieh FC, Hsu FC, Tasy YF, Liu JR, Shih MC (2014) Characterization of an insecticidal toxin and pathogenicity of *Pseudomonas taiwanensis* against insects. *PLoS Pathog* 10(8): e1004288. doi:10.1371/journal.ppat.1004288

Cho JC, Tiedje JM (2000) Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. *Appl Env Microbiol*, 66(12): 5448-5456.

Cho ST, Chang HH, Egamberdieva D, Kamilova F, Lugtenberhg B, Kuo CH (2015) Genome analysis of *Pseudomonas fluorescens* PCL1751: a rhizobacterium that controls root diseases and alleviates salt stress for its plant host. *PLoS ONE*, 10(10): e0140231. doi:10.1371/journal.pone.0140231.

Dagorn A, Chapalain A, Mijouin L, Hillion M, Duclairoir-Poc C, Chevalier S, Taupin L, Orange N, Feuilloley MGJ (2013) Effect of GABA, a bacterial metabolite, on *Pseudomonas fluorescens* surface properties and cytotoxicity. *Int J Mol Sci*, 14:12186-12204, doi:10.3390/ijms140612186.

Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignments with gene gain, loss and rearrangement. *PLoS ONE*, 5(6): e11147. doi:10.1371/journal.pone.0011147

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A*, 103(3): 626-631.

Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST, Lauber CL, Owens S, Gilbert JA, Wall DH, Caporaso JG (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci U S A*, 109(52): 21390-21395.

Flury P, Vesga P, Pechy-Tarr M, Aellen N, Dennert F, Hofer N, Kupferschmied, Metla Z, Ma Z, Siegfried S, de Weert S, Bloemberg G, Hofte M, Keel CJ, Maurhofer M (2017) Antimicrobial and insecticidal: cyclic lipopeptides and hydrogen cyanide produced by plant-beneficial *Pseudomonas* strains CHA0, CMR12a, and PCL1391 contribute to insect killing. *Front Microbiol*, 8: 100. doi:10.3389/fmicb.2017.00100

Gamez RM, Rodriguez F, Ramirez S, Gomez Y, Agarwala R, Landsman D, Marino-Ramirez L (2016) Genome sequence of the banana plant growth-promoting rhizobacterium *Pseudomonas fluorescens* PS006. *Gen Announc*, 4(3): pii: e00329-16. doi: 10.1128/genomeA.00329-16

Gardiner DM, Stiller J, Covarelli L, Lindeberg M, Shivas R, Manners JM (2013) Genome sequences of *Pseudomonas* spp. isolated from cereal crops. *Genome Announc,* 1(3): e00209-13 doi:10.1128/genomeA.00209-13.

Garrido-Sanz D, Meier-Kolthoff JP, Göker M, Martín M, Rivilla R, Redondo-Nieto M (2016) Genomic and genetic diversity within the *Pseudomonas fluorescens* complex. *PLoS One*, doi:10.1371/journal.pone.0150183.

Gomila M, Peña A, Mulet M, Lalucat J, García-Valdés E (2015) Phylogenomics and systematics in *Pseudomonas*. *Front Microbiol* 6(214), doi:10.3389/fmicb.2015.00214

Guimaraes DV, Gonzaga MIS, Silva TO, Silva TL, Dias NS, Matias MIS (2013) Soil organic matter pools and carbon fractions in soil under different land uses. *Soil Till Res*, 126: 177-182.

Gumiere T, Durrer A, Bohannan B, Andreote FD (2016) Biogeographical patterns in fungal communities from soils cultivated with sugarcane. *J Biogeography*, 43: 2016-2026.

Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. *Paleontol Electron*, 4: 1-9.

Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JB (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol*, 10(7): 497-506.

Hernandez-Salmeron JE, Hernandez-Leon R, Orozco-Mosqueda MDC, Valencia-Cantero E, Moreno-Hagelsieb G, Santoyo G (2016) Genome sequence of the biocontrol and plant growth-promoting rhizobacteria *Pseudomonas fluorescens* strain UM270. *Stand Genomic Sci*, 11(5): doi: 10.1186/s40793-015-0123-9.

Ho BT, Dong TG, Mekalanos JJ (2014) A view to a kill: the bacterial type VI secretion system. *Cell Host Microb*, 15: 9-21.

Hol WHG, Beemer TM, Biere A (2013) Getting the ecology into interactions between plants and the plant growth-promoting bacterium *Pseudomonas fluorescens*. *Front Plant Sci*, 4(81): doi:10.3389/fpls.2013.00081.

Hofte M, Altier N (2010) Fluorescent pseudomonads as biocontrol agents for sustainable agricultural systems. *Res Microbiol*, 161(6): 464-471.

Jousset A, Lara E, Wall LG, Valverde C (2006) Secondary metabolites help biocontrol strain *Pseudomonas fluorescens* CHA0 to escape protozoan grazing. *Appl Env Microbiol*, 72(11): 7083-7090.

Kalpage FSCP (1974) *Tropical soils: classification, fertility and management*. St. Martin's Press, New York, USA, 294 pp.

Kampfer P (2012) Systematics of prokaryotes: the state of the art. *Antonie Van Leeuwenhoek*, 101(1): 3-11.

Katoh K, Stanley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30: 772-780. doi 10.1093/molbev/mst010

Kidarsa TA, Shaffer BT, Goebel NC, Roberts DP, Buyer JS, Johnson A, Kobayashi DY, Zabriskie TM, Paulsen I, Loper JE (2013) Genes expressed by the biological control bacterium *Pseudomonas protegens* Pf-5 on seed surfaces under the control of the global regulators GacA and RpoS. Environmen Microbiol, 15(3): 716-735

Konstantidinis KT, Tiedje JM (2004) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A*, 102(7): 2567-2572.

Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*, 44(W1): W242-5, dos:10.1093/nar/gkw290

Lo R, Stanton-Cook MJ, Beatson SA, Turner MS, Bansal N (2015) Draft genome sequence of *Pseudomonas fluorescens* SRM1, an isolate from spoiled raw milk. *Genome Announc*, 3(2): e00138-15, doi: 10.1128/genomeA.00138-15.

Loper JE, Hassan KA, Mavrodi DV, Davis EW, Lim CK, Shaffer BT, Elbourne LDH, Stockwell VO, Hartney SL, Breaklwell K, Henkels MD, Tetu SG, Rangel LI, Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS, Pierson EA,

Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE, Paulsen IT (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* 8(7): e1002784.

Lujan AM, Gomez P, Buckling A (2015) Siderophore cooperation of the bacterium *Pseudomonas fluorescens* in soil. *Biol Lett*, 11: 20140934

Manikandan R, Raguchander T (2015) *Pseudomonas fluorescens* (Pf1) mediated chitinolytic activity in tomato plants against *Fusarium oxysporum* f. sp. *lycopersici*. *Afr J Microbiol Res*, 9(19): 1331-1337.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1): 10-12.

Oren A, Garrity GM (2014) Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. *Antonie Van Leeuwenhoek*, 106: 43-56.

Palleroni NJ (1992) Introduction to the Pseudomonadaceae. *The prokaryotes, a handbook on the biology of bacteria, ecophysiology, isolation, identification and applications*, vol. III, 2nd ed (Balows A, Truper HG, Dworkin M, Harder W, Schlepper KH, eds), 3071-3085. Springer, New York.

Patten CL, Glick BR (2002) Role of *Pseudomonas putida* indoleacetic acid in development of the host plant root system. *Appl Env Microbiol*, 68(8): 3795-3801.

Rezzonico F, Binder C, Défago G, Moenne-Locoz Y (2005) The type III secretion system of biocontrol *Pseudomonas fluorescens* KD targets the phytopathogenic Chromista *Pythium ultimum* and promotes cucumber protection. *Mol Plant Microbe Interact*, 18(9): 991-1001

Ross SM (1993) Organic matter in tropical soils: current conditions, concerns and prospects for conservation. *Progress in Physical Geography*. 17(3): 265-305.

Seeman T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14): 2068-2069.

Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW (2011) *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiol Rev*, 35(4): 652-680.

Spiers AJ, Buckling A, Rainey PB (2000) The causes of *Pseudomonas* diversity. *Microbiol*, 146: 2345-2350

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30: 1312-1313.

Taghavi S, Garafola C, Monchy S, Newman L, Hoffman A, Weyens N, Barac T, Vangronsveld J, van der Lelie D (2009) Genome survey and characterization of endophytic bacteria exhibiting a beneficial effect on growth and development of poplar trees. *Appl Env Microbiol* 75(3): 748-757.

Trapet P, Avoscan L, Klinguer A, Pateyron S, Citerne S, Chervin C, Mazurier S, Lemanceau P, Wendehenne D, Besson-Bard A (2016) The *Pseudomonas fluorescens* siderophore pyoverdine weakens *Arabidopsis thaliana* defense in favour of growth in iron-deficient conditions. *Plant Physiol*, pp. 15.01537, doi:http://dx.doi.org/10.1104

Visca P, Imperi F, Lamont IL (2006) Pyoverdine siderophores: from biogenesis to biosignificance. *TRENDS Microbiol*, 15(1): 22-30.

Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, 2009.

5 COMMUNITIES-POPULATIONS-GENES: A MULTIFACETED ANALYSIS TO DEPICT ECOLOGY AND EVOLUTION OF FLUORESCENT *Pseudomonas* IN SOILS

Abstract

Bulk soil and rhizosphere comprise different niches for microbial life. *Pseudomonas* is a bacterial genus hosting agricultural relevant species, and previous studies showed that its community can be distinctly shaped in bulk soil and rhizosphere. However, it is unknown whether the genome content of closely related bacterial populations and genomes are also distinct in these two habitats. In order to address this issue, we isolated 76 fluorescent *Pseudomonas* spp. (40 and 36 from bulk soil and rhizosphere of sugarcane, respectively) and submitted them to whole genome sequencing (WGS). In parallel, the physiological profile of the isolates was assessed using BIOLOG Ecoplates, and the community structure of *Pseudomona*s spp. was assessed by 16S rDNA high-throughput sequencing and qPCR of specific 16S rRNA gene fragment. Results showed that both the *Pseudomonas* spp. communities and genome contents of the isolates are different in rhizosphere or bulk soil. A phylogenomic approach classified the isolates as members of the *P. fluorescens* and *P. putida* groups. In the *P. fluorescens* group, significant enrichments of genes related to phosphatases and use of xylose were found in isolates from rhizosphere and bulk soil, respectively. In the *P. putida* group, we found enrichments of functions related to D-galactonate catabolism (confirmed by phenotype) and inorganic phosphate in isolates from the rhizosphere. We used xylose utilization function (*xut* genes) as a model to verify whether genomic differences also result in phenotypic differences between populations, essential for a distinct ecological behavior. A mutant with a knockout in the *xutA* gene (B26Δ*xut*A) lacked the ability to grow using D-xylose as a carbon source, a feature observed in most bulk soil isolates (including wild type B26), and absent in most rhizosphere isolates. Xylose utilization was therefore shown to be one of the features related to the ecological differentiation of bulk soil and rhizosphere populations in the *P. fluorescens* group. Our results contributed to the understanding of fluorescent *Pseudomonas* spp. ecology on soil, describing the genomic differences in populations prevailing in rhizosphere and bulk soil, and nominated genes related to the differential occurrence of populations in each of these habitats.

Key words: Population genomics, niche partitioning, fluorescent Pseudomonas spp., genome evolution, plant-associated bacteria, soil ecology, xylose isomerase

5.1 Introduction

Soil is the interface between mineral and organic material on land surface, supporting terrestrial life (Paul, 2014). Agriculture is fully dependent on soil, as it provides water, nutrients and physical support for plant development and growth. The interplay among soil physics, chemistry and biological activity is promoted by microbes, which participate in nutrients cycling, act in soil structuring, and associate to plants (Paul, 2014). However, soil is a heterogeneous environment for microbial life, since the physical and chemical characteristics varies in space and time. Microbes strive in soil hotspots, for example, attached to aggregate surfaces, on detritusphere, in biopores, and also in niches generated by the presence of plants, such as rhizosphere (Kuzyakov and Blagodatskaya, 2015). The rhizosphere is a very particular habitat in soils, serving as a hot spot prompted by the roots, that drastically change the physical (oxygen levels, aggregation, water content) and chemical (pH, carbon supply) properties of the surrounding soil (Sokolova, 2015). The abundance of carbon is the main driver of the rhizosphere, what mainly differentiate this habitat from the soil not influenced by plant roots, known as bulk soil, commonly characterized by low amounts of readily available organic contents to sustain microbial activity (Dennis et al., 2010). Therefore, bulk soil and rhizosphere are distinct habitats comprising different niches for microbial life, which occur in a spatial distance of millimeters in the soil environment.

Understanding soil microbial ecology can support a step forward to achieve a more sustainable agriculture, more efficient in the exploration of microbial functions, and less dependent on using scarce or noxious substances that impacts the environment (Liu Y et al., 2015). Previous studies have shown the potential that rhizosphere has to shape microbial community by enriching microbes able to protect plants against stressors (Mendes et al., 2011; Philippot et al., 2013). Several evidences support the conclusion that microbial communities are differentially shaped in rhizosphere and bulk soil on several soil types, crop species and parts of the world (Berg and Smalla, 2009; Uroz et al., 2010; Mendes et al., 2014; Lopes et al., 2016). Moreover, it has been also demonstrated that the efficiency of plants to properly assembly the rhizosphere is dependent on soil biodiversity (Mendes et al., 2011).

However, studies on other levels of microbial life in soil habitats are still demanded. We know very little about the ecology and evolution of bacterial populations that inhabit the different soil habitats. It is also mandatory a better knowledge on the set of genes that can make specific bacterial populations able to colonize each niche. These studies are still dependent on culture based approaches, as culture-independent analyses rarely offer the opportunity to compare several genomes of highly phylogenetically related microbes in a

single sample, as it also does not provide the opportunity to experimentally evaluate the phenotypes of targeted microbes. It has come a moment where the knowledge about microbial communities can be flavored with data from populations and genes dynamics in soil habitats, integrating a more complete picture, which will hopefully improve our capacity to manage agricultural systems in order to improve the sustainable productivity.

The genus *Pseudomonas* is eligible to perform such approach, as it harbors several bacterial groups relatively easy to be cultured, it is widespread in soils, several species are described as plant growth promoting rhizobacteria (PGPR), and some studies suggested that its community is distinct in rhizosphere compared to bulk soil (Costa et al., 2007; Garcia-Salamanca et al., 2012). The fluorescent species are the main targets for studying *Pseudomonas* spp. in agroecosystems, as they are the most related to PGPR activities (Silby et al., 2011). Nineteen fluorescent *Pseudomonas* isolates from rhizosphere and endosphere of *Populus deltoides* were contrasted by comparative genomics revealing differences in genome content, with a remarkable difference on genes related to growth and plant signaling, which were enriched in the rhizosphere and endosphere populations, respectively (Timm et al., 2015). Hence, a similar approach, based on more genomic sequences, comparing bulk soil and rhizosphere, would generate innovative data on the differential selection and/or evolution of *Pseudomonas* spp. in the soil habitats. In combination to that, important crops may serve for such approaches, indicating the impact of this knowledge in plants cultivated in great extensions worldwide. Sugarcane is one of the most cultivated crops in Brazil, the largest producer in the world (Rudorff et al., 2010). Plants harvested in approximately 9 million ha (2016) are destined to sugar and bioethanol production, supplying world demands for food and renewable energy, respectively (Loarie et al., 2011; CONAB, 2017).

Thus, this study assessed the community structure of *Pseudomonas* spp., and the genome sequences of 76 fluorescent *Pseudomonas* spp. isolated from sugarcane rhizosphere and bulk soil. We tested the hypothesis that, besides communities, the populations of each habitat have differences in genome content and are enriched in specific genes/functions. The physiological profile of the isolates was also investigated to infer about the phenotype changes between populations, derived from the genotypic differences. Some differences found in genome content were deeper investigated by the generation of gene-depleted mutants, allowing inferences about the ecological and evolutionary processes related to the enrichment of such genes in the populations of rhizosphere or bulk soil.

5.2 Material and Methods

5.2.1 Samples Collection

Sampling was performed in a sugarcane field cultivated for 10 years in a green-harvest cropping system, located in Piracicaba-SP, Brazil, in 2014 (22°69´S/47°64´W). Soil samples were collected in two below ground compartments: the soil layer of 1-2 mm adhering to plant roots (rhizosphere); and the soil between the cropping rows (0-10 cm) avoiding roots (bulk soil). Six sampling points were selected along the sugarcane plantation, where the bulk soil and rhizosphere samples (approximately 300 g of soil) were collected. Soil was carried to the laboratory and each sample was homogenized. In total, 12 samples were used for further analysis (6 from bulk soil and 6 from the sugarcane rhizosphere).

5.2.2 Analysis of the *Pseudomonas* spp. Communities in Samples of Bulk Soil and Rhizosphere

DNA was extracted from bulk soil and rhizosphere samples by using the Mobio Power Soil DNA isolation kit (Mobio, USA). This approach generated DNA suitable for PCR amplification. We used primers S-D-Bact-0341-b-S-17/S-D-Bact-0785-a-A-21 (Klindworth et al., 2013), which cover the V3-V4 region of the 16S rRNA gene. PCRs were assembled using 1 μL of template DNA (approximately 30 ng), 5 nmol of dNTPs, 75 nmol of MgCl2, Taq Buffer 1.2X, 0.1 μL of each primer and 10 nmol of Taq DNA polymerase. Thermal cycles consisted of an initial denaturation of 95° for 3 min, 30 cycles of 95°C for 45 s (denaturation), 57°C for 1 min and 45 s (annealing) and 72°C for 1 min (extension), followed by a final extension of 72°C for 4 min. Libraries were prepared and sequencing was performed on an Illumina MiSeq platform by a third partner service (Laboratory of Functional Genomics Applied to Agriculture and Agroenergy). Resulting reads were deposited on SRA database (NCBI) as BioProject PRJNA319762 with accession numbers: SAMN04904299 (B1); SAMN04904300 (B2); SAMN04904301 (B3); SAMN04904302 (B4); SAMN04904303 (B5); SAMN04904304 (B6); SAMN04904305 (R1); SAMN04904306 (R2); SAMN04904307 (R3); SAMN04904308 (R4); SAMN04904309 (R5); and SAMN04904310 (R6) (R = Rhizosphere and B = Bulk Soil samples).

The complete analysis of the 16S rRNA gene sequences (for total bacterial community) supported a publication on major distinctions between the bacterial communities in the bulk soil and rhizosphere of sugarcane (Lopes et al., 2016). Here, the dataset was used to a more specific analysis, where the operational taxonomic units (OTUs) classified as

Pseudomonas spp. by the Greengenes database were filtered from the OTU table generated by the UCLUST algorithm (97% cutoff) using the command filter_taxa_from_otu_table.py on QIIME software (Caporaso et al., 2010). Thus, a *Pseudomonas* spp. OTU table was obtained, and representative sequences of each OTU classified as *Pseudomonas* spp. was retrieved from the full set of sequences by using the filter_fasta.py command.

Pseudomonas spp. OTU table derived from the 16S rRNA gene high-throughput sequencing was exported to Primer-6 software (Clarke and Gorley, 2006), where data was normalized and a non-metric multidimensional scaling (NMDS) using the Bray-Curtis similarity index was performed in order to assess the ordination of rhizosphere and bulk soil samples. In addition, whole bacterial community OTU table was exported to STAMP software, where a Welch's t-test was performed using the Benjamini-Hochberg *P*-value correction to also verify the differences in relative abundance of reads classified as *Pseudomonas* spp. (Parks et al., 2014).

Real-time PCR (qPCR) analysis was performed using a primer pair that anneal with high specificity (99%) to a fragment of the 16S rRNA gene of *Pseudomonas* spp. (Bergmark et al., 2012). Thus, the copy numbers of this 16S rRNA gene fragment was quantified to assess whether rhizosphere and bulk soil samples harbor significantly different abundances of *Pseudomonas* spp. Real-time PCR amplification was performed using the specific primers PseF/PseR (385 nM) in a reaction with Sybr Green (1X), 1 μL of template DNA and 385 nM of each primer (PseF/PseR) for a total volume of 25 μL (Bergmark et al., 2012). Amplification was conducted using a StepOne Real-Time System (Applied Biosystems) under the following conditions: 1 cycle of 95°C for 10 min; and 40 cycles of 95°C for 30 s and 60°C for 1 min, followed by a melting curve analysis. For the standard curve, soil-derived amplicons of the same 16S rRNA gene fragment were diluted from 10^{-1} to 10^{-8} and quantified. These dilutions were submitted to amplifications under similar conditions as described above. Results of *Pseudomonas* spp. 16S rRNA gene sequences quantification was exported to Past software (Hammer et al., 2001), where data was normalized and a Tukey pairwise comparison test was performed aiming to assess the existence of significant differences ($P<0.05$) between bulk soil and rhizosphere samples.

5.2.3 Isolation and genome analysis of fluorescent *Pseudomonas* spp. from bulk soil and rhizosphere

Bacterial isolation was performed using Pseudomonas Agar Base selective media (Oxoid, UK) supplemented with the antibiotics cetrimide, fucidin and cephalosporin (10, 10 and 50 mg L^{-1}, respectively). Aliquots of soil suspensions from the same bulk soil and rhizosphere samples previously described were plated and incubated at 28°C, and resulting fluorescent colonies on UV light were selected and stored on glycerol at -80°C. The genomic DNA of the 76 isolates obtained in this process was extracted using the Wizard Kit (Promega, USA) and submitted to whole genome sequencing (WGS). For that, Nextera barcoded libraries were constructed and sequenced using the 150mer paired end kit in a single lane on an Illumina HiSeq 3000. Library preparation and sequencing was performed on the Center for Genome Research and Biocomputing (CGRB) at Oregon State University.

Draft genomes were assembled using SPAdes 3.7.0 and annotations were obtained using the Prokka software (Bankevich et al., 2012; Seeman, 2014). The genome sequences of the 76 isolates were previously used for taxonomical classification and disclosing of the new *Pseudomonas* spp. diversity found in the Brazilian tropical soil (Lopes et al., 2017 - submitted). In summary, genomes varied from 15 to 94 contigs, varying between 5.56 and 7.08 Mbp, but with similar average genome size in populations of bulk soil and rhizosphere. This Whole Genome Shotgun project (BioProject PRJNA383709) has been deposited at DDBJ/ENA/GenBank under the accessions NEHD00000000-NEKA00000000. The versions described in this paper are versions NEHD01000000-NEKA01000000. The accession numbers for all genome sequences are available on Supplementary Table 5.1.

Here we used these genome sequences to infer on ecological differences between bulk soil and rhizosphere populations of fluorescent *Pseudomonas* spp. A phylogenomic approach was used to classify the isolates and infer about their phylogenetic relationship. Putative orthologues present in the genome sequences of the 76 isolates and in other 18 relevant reference strains were inferred by clustering genes using the OrthoMCL algorithm on Get Homologues software (Contreras-Moreira and Vinuesa, 2013). Gene clusters with single copies and present in all 94 genome sequences were individually aligned using Mafft 7.271 (L-insi-i algorithm) and the resulting alignments of gene sequences were concatenated using RAxML 8.2.8, before generating a maximum likelihood phylogenetic tree, which was visualized using the ItoL platform (Stamakis, 2014; Katoh et al., 2013; Letunic and Bork, 2016).

The Get Homologues software was also used to split the gene clusters of the 76 isolates analyzed, and to separate the gene clusters of the 57 isolates classified in the *P. fluorescens* group or the 19 isolates classified in the *P. putida* group. The matrices of gene clusters abundance were used for statistical analyses. To avoid bias from a single interpretation based on gene clusters annotation performed by Get Homologues, InterProScan 5 software was used to identify conserved protein domains in the amino acid sequences based on Pfam and TIGRFAM protein databases (Jones et al., 2014). After the identification of the conserved domains, GO-terms were attributed to each inferred function. The abundance of the GO-terms in each genome sequence was organized into matrices, which were also further used for statistical analyses.

The gene clusters matrices generated on Get Homologues were exported to R software, where a discriminant analysis of principal components (DAPC) was performed using the adegenet package (Jombart et al., 2008; Jombart et al., 2010), aiming to assess whether the bulk soil and rhizosphere genome sequences form two distinct groups based on discriminant factors. The same matrices, as well as those made of GO-terms were also exported to STAMP software, where Welch's t-test was performed using the Bonferroni *P*-value correction, aiming to identify the genes and functions significantly different between rhizosphere and bulk soil populations ($P<0.05$).

The 16S rRNA gene sequences were extracted from the genome assemblies and compared to sequences of the OTUs classified as *Pseudomonas* spp. in the community analysis. Mafft 7.271 (L-insi-i algorithm) was used to align sequences, and the 16S rRNA gene sequences of the assemblies were trimmed to match with the aligned region of the OTU sequences (V3-V4 hypervariable region). A maximum likelihood phylogenetic tree was performed on MEGA 7 (Tamura et al., 2013) and visualized on Itol platform (Letunic and Bork, 2016). In addition, amplicon and genome derived sequences were separated in OTUs using UCLUST algorithm with a 97 or 99% cutoff in the USEARCH 6.0 software (Edgar, 2010), a similar approach used in the community analysis, in order to supply information about the coverage of the genomic analysis upon the total community of *Pseudomonas* spp.

5.2.4 Phenotypic analysis of the isolates

The physiological profile of the isolates was assessed using BIOLOG Ecoplates, which contains 31 wells with different single carbon sources (C-sources). Purified cultures of each isolate were grown in Luria-Bertani (LB) media for 24 hours at 28°C. Cells were washed twice with autoclaved MQ water to remove any remaining carbon of the growth media,

followed by adjusting the optical density (OD = 0.03) of all cultures before inoculating 150 µL in the BIOLOG plates, which were immediately incubated at 28°C. Absorbance readings were performed after 24 hours at 600 nm. Each isolate culture was inoculated in two different plates, aiming to both use replicates and reduce potential bias associated to individual plates.

The absorbance tables of the BIOLOG analyses were exported to Primer-6, where data was normalized and NMDS was performed using the Bray-Curtis index to get the ordination of samples based on the physiological profile of the isolates. SIMPER analysis was also performed to identify the C-sources most contributing to the differences between bulk soil and rhizosphere samples.

5.2.5 Construction of Δ*xut*A mutants and growth curves on D-xylose medium

Twenty-two strains, each representative of a phylogenetic clade pointed by phylogenomics, were used to test their capacity to grow on D-xylose, aiming to verify if the phenotype of the isolates supports the genomic findings regarding xylose-utilization (*xut*) genes, which were significantly enriched in the bulk soil population of the *P. fluorescens* group. Among the analyzed strains, half was isolated from bulk soil and half from the rhizosphere. Despite most part of the strains containing the *xut* operon are from bulk soil, rhizosphere strains containing *xut* genes as well as bulk soil strains lacking *xut* genes were included in the analysis.

Growth curves were measured using a TECAN machine. To test their abilities to grow on D-xylose, a minimal medium (M9) containing D-xylose as the sole C-source was prepared by mixing four individual sterilized solutions: 10.0 mL of 5X M9 salts (autoclaved); 0.1 mL of 1M MgSO4 (autoclaved); 0.005 mL of 1M CaCl2 (autoclaved); 5.0 mL of D-xylose 10% (filter sterilized); and 0.05 mL of a micronutrients solution (filter sterilized) for a final volume of 50 mL. After mixing the solutions, the medium was buffered by addition of 20 g/L 3-(N-morpholino)-propanesulfonic acid (MOPS), adjusted to pH 7, and filter sterilized. The preparation of the M9 salts and micronutrients solution is supplied in Supplementary Table 5.2. The micronutrients solution was prepared based on the preparation of the medium for *E. coli* (Cam et al., 2015).

Overnight cultures of each strain (22) were centrifuged and washed twice with sterilized deionized water in order to remove all carbon residues. After optical density (O.D.) adjustment to 0.03 (600 nm), 10 µL of each culture was inoculated in 90 µL of M9 or LB media in the TECAN plate. O.D. values were recorded every hour at 600 nm for a period of ~65 hours. Triplicates were used for each strain in each treatment (M9+D-xylose;

M9+glucose; and LB medium), and blanks consisted in media without inoculation of any strain.

Two phylogenetically related isolates B26 and R38, isolated from bulk soil and rhizosphere, respectively, were used as models to further assess the capacity to grow on D-xylose, which would derive from the presence (B26) or absence (R38) of the *xut* genes. For that, growth curves of the two isolates were also measured using TECAN using the same protocol described above. Alternatively, growth curves were tested with glucose replacing D-xylose in the M9 medium or with LB medium.

A knockout was performed in the *xutA* gene, present in B26, aiming to test whether the growth in D-xylose of this strain is directly related to the presence of this gene. The Δ*xut*A mutant was generated by in-frame deleting the *xutA* gene from the chromosome of wild-type (WT) strain B26. To this end, a *xutA* deletion construct was made. Two DNA fragments flanking the *xutA* gene were PCR amplified by using the primer pairs xutA-F1/xutA-Ovlp-R1 and xutA-R2/xutA-Ovlp-F1 (Supplementary Table 5.3). These two fragments were fused together by PCR and digested using *Hind*III and *Eco*R1 to generate a 982-bp DNA fragment containing *xutA* with a 1,279-bp internal deletion (wild-type *xutA* has 1,317 bp). This DNA fragment was ligated to suicide vector pEX18Tc to create construct p18Tc-ΔxutA. This depletive construction was transferred into strain *E. coli* S17-1 and introduced into strain B26 WT by bi-parental mating to make the B26ΔxutA mutant. The deletion of *xutA* was confirmed by PCR using oligonucleotide pair xutA-F3/xutA-R3 which generates a 1,402-bp fragment with the genomic DNA of B26WT as template, and a 123-bp fragment with that of the deletion B26ΔxutA mutant. The same protocol described above was followed to assess the growth curve of B26ΔxutA compared to B26WT and R38 in M9 medium with D-xylose, aiming to phenotypically test whether *xutA* gene affect D-xylose utilization and consequently the ability of the bulk soil strain (B26) to grow on D-xylose.

5.3 Results

5.3.1 Shifts in *Pseudomonas* spp. community of rhizosphere and bulk soil

A total of 601 reads classified as *Pseudomonas*, comprised in 134 OTUs, were retrieved from the high-throughput sequencing of the 16S rRNA genes present in the DNA directly extracted from the bulk soil and rhizosphere samples. Multivariate analysis clearly indicated that the community structure of the studied genus is significantly different between bulk soil and rhizosphere (Figure 5.1A), supported by ANOSIM (P=0.002; R=0.74). The

ordination of samples in the NMDS also suggested that rhizosphere communities of *Pseudomonas* spp. are more similar between different sampled plants than the observed in the communities of different bulk soil samples (Figure 5.1A).

In addition, the abundance of *Pseudomonas* was also significantly higher in rhizosphere compared to bulk soil, either in the relative abundance of sequences classified as *Pseudomonas* or in the real-time PCR data (Figure 5.1B and C). The relative abundance of reads classified in the genus *Pseudomonas* was significantly higher in the rhizosphere than in bulk soil samples (P=1.92 e^{-3}, Figure 5.1B), and quantification of specific 16S rRNA gene copies from *Pseudomonas* spp. showed an average of $2.5x10^{4}$ in the bulk soil and $8.7x10^{5}$ copies g soil $^{-1}$ in the rhizosphere samples (Figure 5.1C, P=0.002).

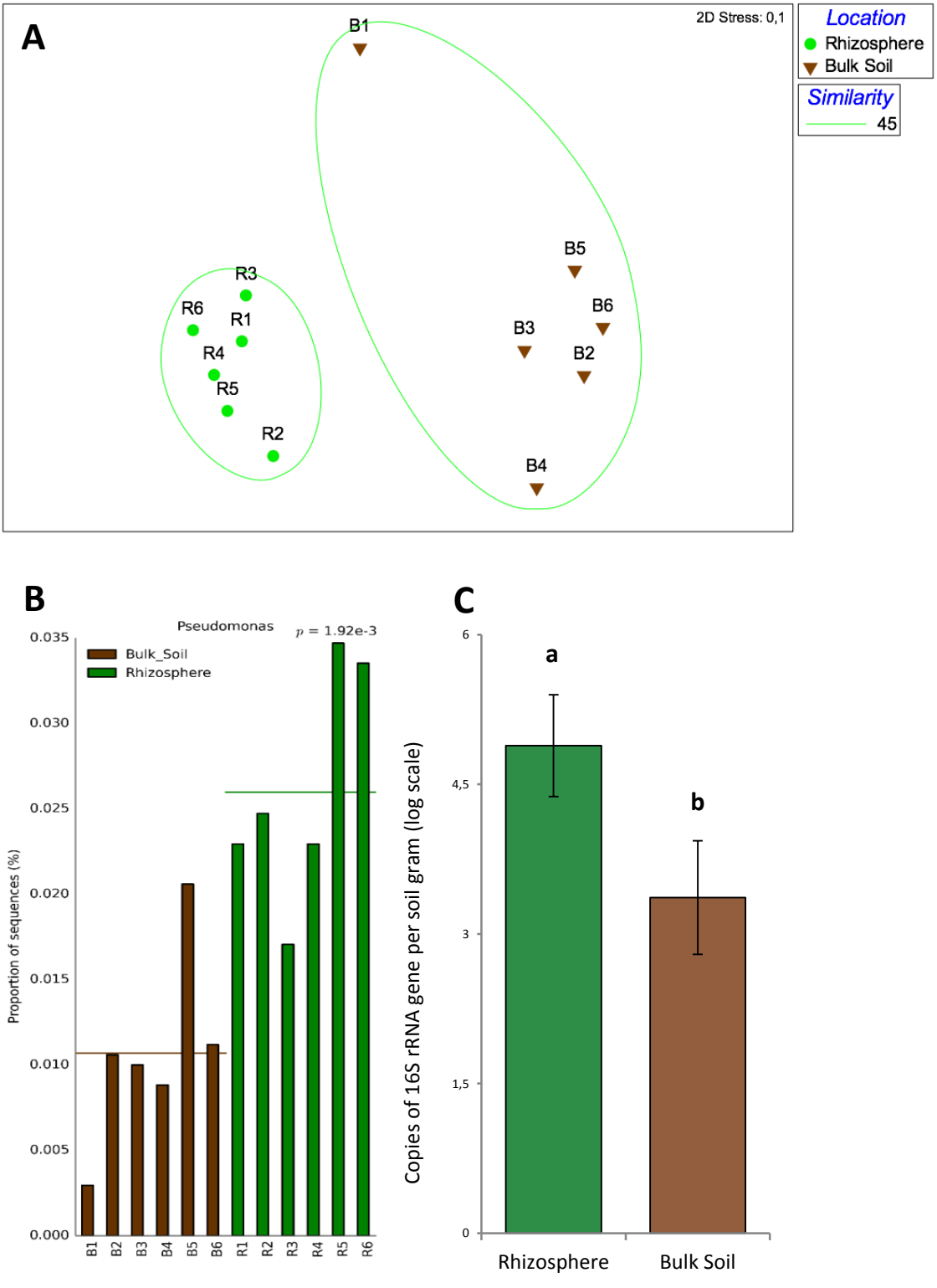


Figure 5.1 A) Changes in *Pseudomonas* community structure shown by NMDS based on the *Pseudomonas* spp. OTU table from 16SrRNA gene high-throughput sequencing; B) Differences in the relative abundance of reads classified as *Pseudomonas* spp. from 16S rRNA gene high-throughput sequencing; C) Differences in 16S rRNA gene copy numbers specific to *Pseudomonas* spp. analyzed by qPCR

5.3.2 Phylogenetic classification of the 76 isolates based on phylogenomics

The 76 genome sequences of the fluorescent *Pseudomonas* spp. isolated from sugarcane rhizosphere (36 isolates) or bulk soil (40 isolates) were combined in a dataset together with other 18 reference genomes of the genus *Pseudomonas*. The comparison of these genome contents found 2,064 shared gene clusters (present in 100% of genome sequences). From this total, 1,540 genes were found in single copies in all the 94 genome sequences, and were therefore used for the phylogenomics analysis. Phylogenomics classified 57 isolates (31 from rhizosphere and 26 from bulk soil) in the group *P. fluorescens*. The remaining 19 isolates (6 from rhizosphere and 13 from bulk soil) were classified in the *P. putida* group (Figure 5.2).

A deeper phylogenetic description was performed by Multi-Locus Sequence Analysis (MLSA) and classified the isolates of the *P. fluorescens* group inside four subgroups (Lopes et al., 2017- submitted). In combination, phylogenomics and MLSA have shown that 53 isolates (~93%) of the *P. fluorescens* group are closely related phylogenetically, covering 2 neighbor *P. fluorescens* subgroups. Isolates B6/R26; and R37/B28 are phylogenetically more distant to all others (Figure 5.2). The phylogenetic classification does not clearly splits the bulk soil and rhizosphere isolates of the *P. fluorescens* group (Figure 5.2). Regarding the isolates classified inside the *P. putida* group, two phylogenetic clades are clearly observed, hosting the isolates from rhizosphere or bulk soil.

In order to estimate the coverage of the *Pseudomonas* community represented in the analyzed genomes, the 16S rRNA gene was extracted from the genome sequences of our 76 isolates and compared to the 16S rRNA gene sequences of each single OTU classified as *Pseudomonas* obtained in the community analysis. The phylogenetic tree using such 16S rRNA gene sequences showed that the isolates were closed to few community OTUs (Supplementary Figure S5.1). The pick OTUs using UCLUST split the isolates in two OTUs by using the 97% cutoff, and in five OTUs by using the 99% cutoff (Supplementary Figure S5.1). From the five OTUs, two contained isolates from rhizosphere or bulk soil classified in the *P. putida* group; two encompassed the pairs B26/R26 or R37/B28; and the last one contained all other 53 isolates of the *P. fluorescens* group (the same closely related in phylogenomics), together with one sequence representing an OTU in the community approach, and the reference *P. fluorescens* Pf0-1. Therefore, the following comparison in the *P. fluorescens* group is mostly based on one single OTU found in the community analysis.

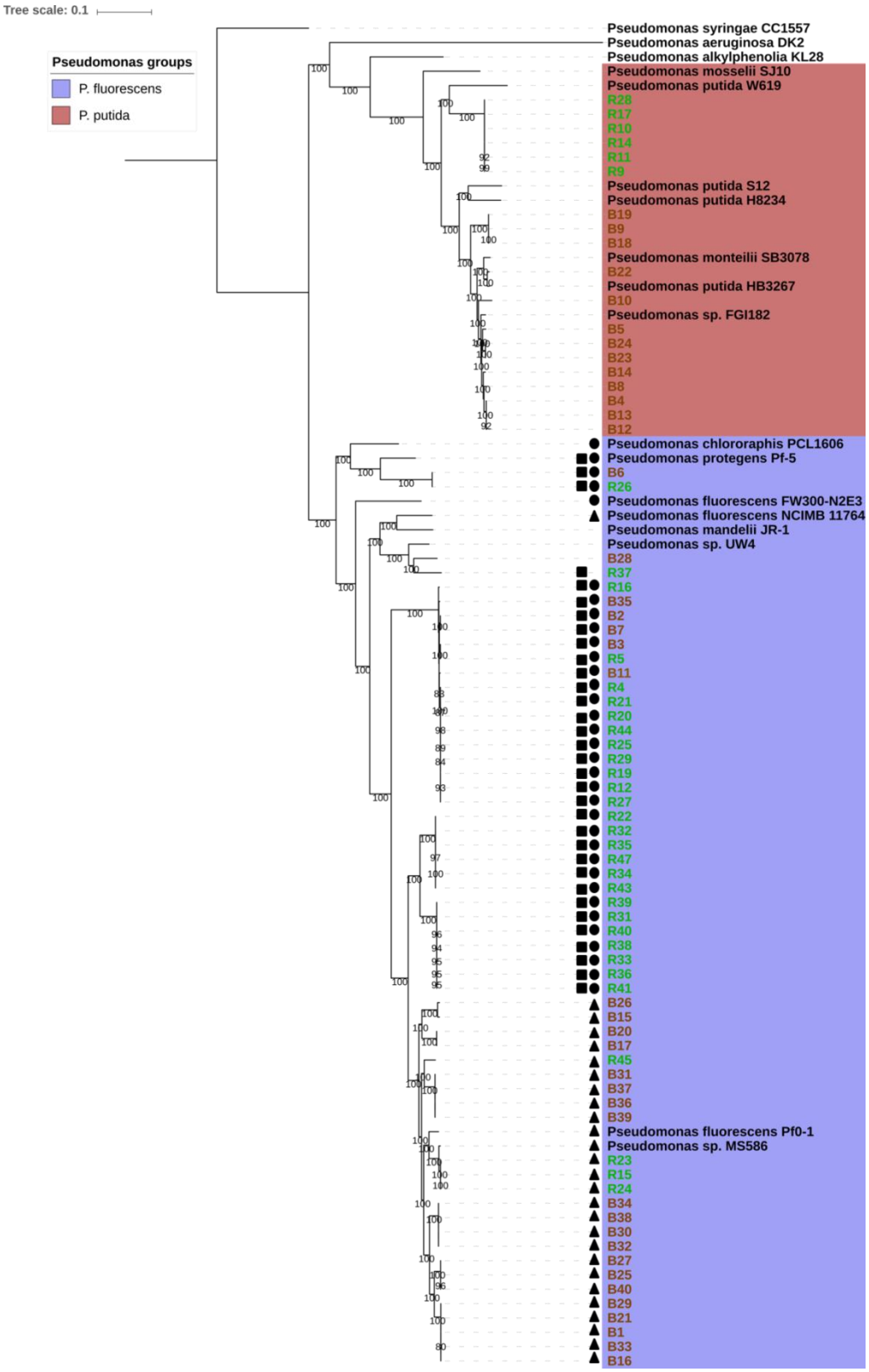


Figure 5.2 Phylogenetic inference based on phylogenomics using all gene sequences shared by all 76 isolates of this study and 18 reference strains. Shared genes with single copies in each genome sequence (1,540) were used for this analysis. The isolates of our study are labeled in green (rhizosphere) or brown (bulk soil), while the reference strains are labeled in black color. Isolates classified in the *P. fluorescens* group with a triangle on its side have the xylose utilization genes; with a circle have the acid phosphatase gene; and with a square have the alkaline phosphatase gene. The numbers below each branch are the bootstrap values (100 bootstrap tests)

5.3.3 Comparative genomic analysis reveals differences between rhizosphere and bulk soil populations

The genome analyses in Get Homologues resulted in almost the same number of shared gene clusters in the populations of rhizosphere (3,363) and bulk soil (3,384) classified in the *P. fluorescens* group (Table 5.1). On the other hand, the populations of the *P. putida* group had a higher number of shared gene clusters in the rhizosphere (5,186) than in the bulk soil (4,026), reflecting the higher phylogenetic relatedness and fewer quantity of rhizosphere isolates in this group (Table 5.1, Figure 5.2). However, in both groups, the number of accessory gene clusters was higher in the bulk soil than in the rhizosphere populations - 7,515 and 7,262 in the *P. fluorescens* group; 3,323 and 64 in the *P. putida* group, respectively - suggesting a higher genetic diversity in the bulk soil than in the rhizosphere populations (Table 5.1).

Table 5.1. Number of gene clusters found in the isolates

	Bulk soil isolates (*P. fluorescens* group) (27)	Rhizosphere isolates (*P. fluorescens* group) (30)	Bulk Soil isolates (*P. putida* group) (13)	Rhizosphere isolates (*P. putida* group) (6)
Shared gene clusters	3,384	3,363	4,026	5,186
Total gene clusters	10,899	10,625	7,349	5,250
Accessory gene clusters	7,515	7,262	3,323	64

Contrasting the set of genome sequences from bulk soil and rhizosphere by DAPC resulted in the separation of populations without any overlap, meaning that there are some factors (gene clusters) able to differentiate the populations from the two habitats (Figure 5.3A and B). This differentiation was more pronounced in populations classified in the *P. putida* than in the *P. fluorescens* group, reflecting the conservative phylogenetic signal of isolates from *P. putida* group found in the bulk soil or rhizosphere. Only 3 PCA eigenvalues were necessary to separate the bulk soil and rhizosphere populations of the *P. putida* group (Figure 5.3B), while in the *P. fluorescens* group it was necessary to reach the saturation of variance (20 eigenvalues) to separate the populations from each habitat (Figure 5.3A). These results indicate that several gene clusters are differentially distributed in the bulk soil or rhizosphere populations of the *P. putida* group, probably linked to their phylogenetic difference. Inversely, DAPC indicated that fewer gene clusters have different frequencies in the bulk soil and rhizosphere populations of the *P. fluorescens* group, but these differences are less related to the phylogeny and more linked to the occupation of each soil habitat. DAPC also suggested

that bulk soil populations have more genetic variation (more peaks in the graphs) than the rhizosphere populations in both groups, supporting the results of Get Homologues (Figure 5.3A and B).

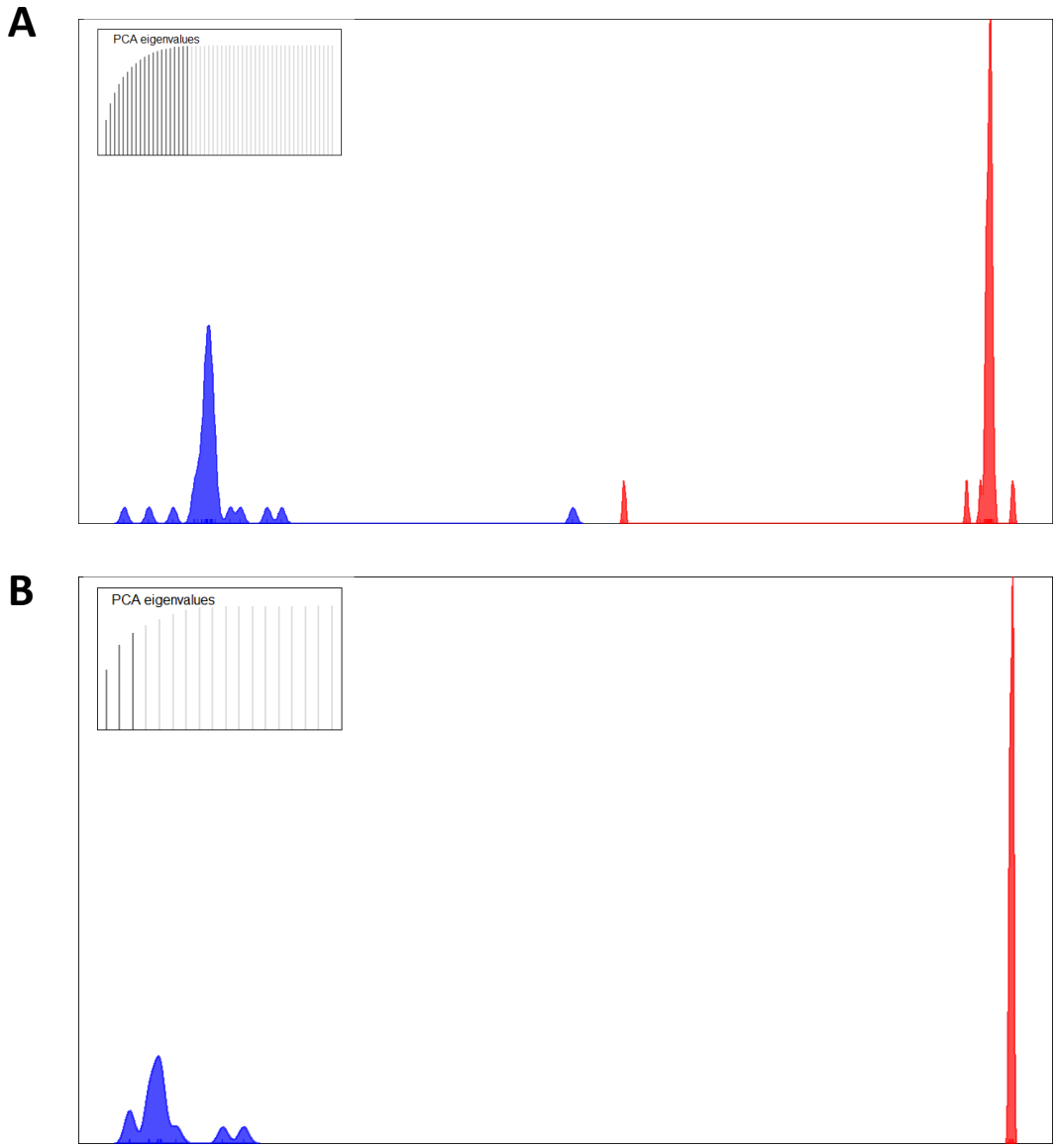


Figure 5.3 Discriminant analysis of principal components (DAPC) between rhizosphere and bulk soil isolates (populations) for each phylogenetic group, *P. fluorescens* (A) and *P. putida* (B). At the top left are the PCA eigenvalues needed to discriminate without overlap the populations of the two habitats. The bulk soil populations are represented in blue, while the rhizosphere populations are represented in red color

In order to identify what are the genomic differences between the rhizosphere and bulk soil populations pointed by DAPC, patterns of gene (gene clusters) and function levels (GO-terms) were compared. In the *P. putida* group, hundreds of differences were observed between the populations of both habitats for genes and functions. In the *P. fluorescens* group,

48 genes were found in distinct frequencies between the populations of the two habitats (P<0.05), from which 32 genes were enriched in the bulk soil populations and 16 genes were enriched in the rhizosphere populations. Similarly, 29 and 22 GO-terms were significantly higher (P<0.05) in the populations of rhizosphere and bulk soil of this group, respectively (Supplementary Figure S5.2). Some genes and functions also revealed different frequencies when analyzing all the 76 *P. fluorescens* and *P. putida* genome sequences together (Supplementary Figure S5.3).

Among the functions and genes differentially enriched in each population, some were remarkably related to the occupation of the rhizosphere or bulk soil. For example, in isolates derived from the rhizosphere, genomes were enriched with genes codifying for alkaline phosphatase in both groups of *Pseudomonas*, and acid phosphatase in the *P. fluorescens* group. In counterpart, the bulk soil populations of the *P. fluorescens group* were highly enriched in genes associated to xylose utilization (Table 5.2). Consistently, the remarkable functions enriched in the rhizosphere populations were GO-terms related to the metabolism of organic (*P. fluorescens* group) and inorganic (*P. putida* group) phosphate; while the GO-terms related to xylose utilization were enriched in the bulk soil populations of the *P. fluorescens* group (Table 5.2). In addition, the GO-term "Hydrolase activity, hydrolysing O-glycosyl compounds" was enriched in bulk soil genome sequences of the *P. fluorescens* group, and "D-galactonate catabolic process" was enriched in the rhizosphere populations of the *P. putida* group (Table 5.2). The five xylose genes enriched in the genomes of the *P. fluorescens* group from bulk soil comprises the xylose isomerase operon (*xutA, xutR, xutG, xutH, xutF*). The acid and alkaline phosphatases genes enriched in the rhizosphere isolates of the same group were classified as hypothetical proteins, detected based on the occurrence of conserved domains for phosphatase, and therefore considered putative new phosphatases yet to be properly described.

Table 5.2 Highlighted gene clusters and GO-terms with significant differences between bulk soil and rhizosphere populations

Enriched in rhizosphere genomes	*P*-value			**Enriched in bulk soil genomes**	*P*-value		
Gene clusters	All together	*P. fluorescens* group	*P. putida* group	**Gene clusters**	All together	*P. fluorescens* group	*P. putida* group
Alkaline phosphatase	2.3E-07	6.5E-04	1.0E-15	Xylose isomerase xylA	ns	1.5E-03	ns
Acid phosphatase	0.036	1.9E-03	ns	Transcription regulator xylR	ns	1.5E-03	ns
GO-terms				D-xylose ABC transporter ATP-binding protein xylG	ns	1.5E-03	ns
Phosphatase activity	0.044	9.3E-06	ns	D-xylose ABC transporter permease xylH	ns	1.5E-03	ns
Dephosphorylation	ns	3.7E-05	ns	D-xylose ABC transporter substrate-binding protein xylF	ns	1.5E-03	ns
Undecaprenyl-diphosphatase activity	ns	3.7E-05	ns	**GO-terms**			
Inorganic phosphate transmembrane transporter activity	ns	ns	2.4E-03	D-xylose-importing ATPase activity	3.5E-03	6.3E-06	ns
Phosphate ion transport	ns	ns	1.8E-03	Xylose isomerase activity	3.5E-03	6.3E-06	ns
Gluconate transmembrane transport	ns	ns	1.0E-15	D-xylose transport	3.5E-03	6.3E-06	ns
Lactate transport	ns	ns	2.5E-15	Hydrolase activity, hydrolyzing O-glycosyl compounds	2.2E-03	2.9E-03	ns
D-galactonate catabolic process	ns	ns	1.2E-11				

5.3.4 General physiological profile reflects the phylogeny of the isolates

The physiological profile of the isolates measured by BIOLOG and analyzed by NMDS revealed a preferential clustering of samples according to the phylogenetic groups instead of their niche in soils (Figure 5.4A). When separating the analysis for each *Pseudomonas* group, it was possible to observe the intra-group variations. A slight trend for separation between bulk soil and rhizosphere isolates was found for those classified in the *P. fluorescens* group, although they had a clear overlap reflecting their phylogeny (Figure 5.4B).

On the other hand, the isolates of the *P. putida* group clearly clustered according to the bulk soil or rhizosphere habitats, also resembling the phylogenetic results (Figure 5.4C). SIMPER analysis pointed D-galactonic acid and D-galacturonic acid the main C sources that explained the physiological differences between the rhizosphere and bulk soil populations of the *P. putida* group, contributing to 8.88% and 8.45% of the dissimilarities, respectively (both more oxidized by the rhizosphere isolates). The levels of oxidation of D-galactonate were dramatically different between rhizosphere and bulk soil isolates, showing the supportive result of the phenotype for the genomic findings (Figure 5.4D).

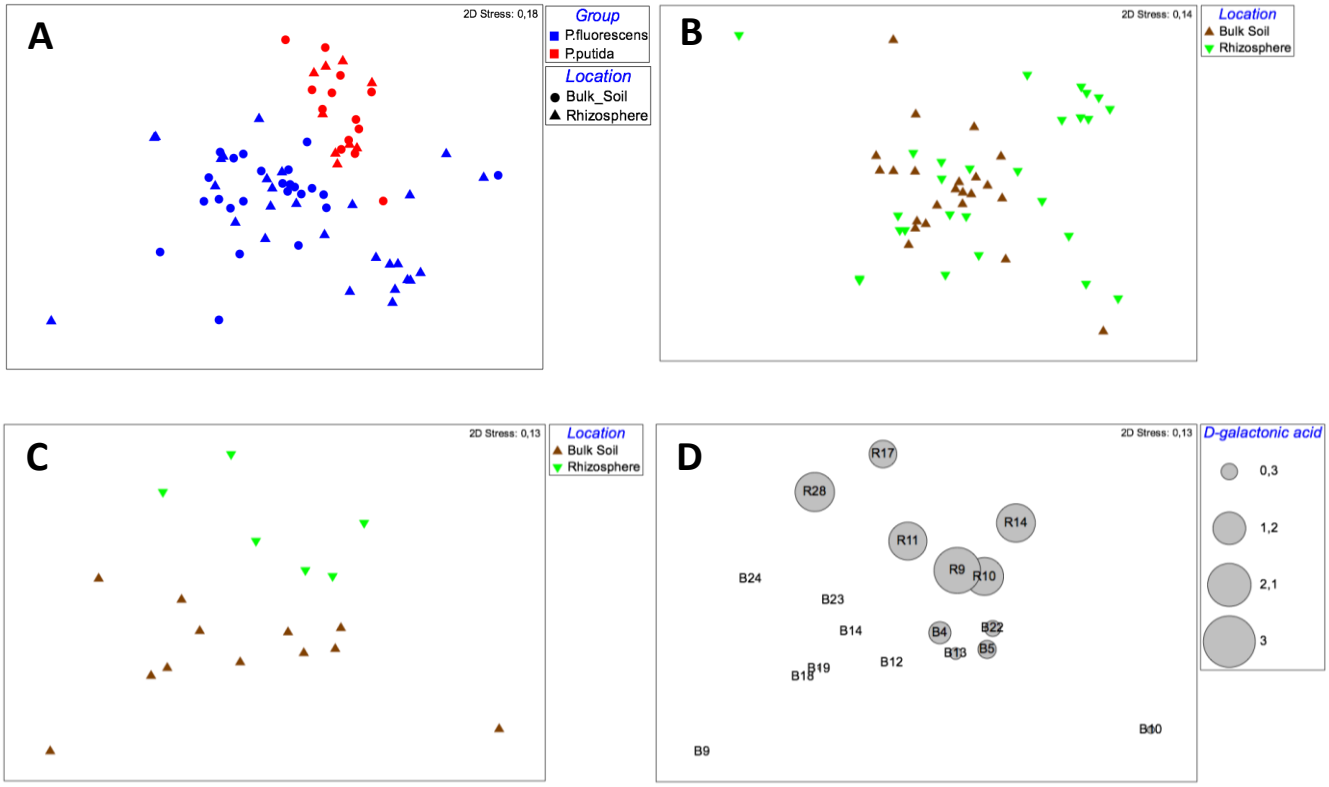


Figure 5.4 Physiological profile of the isolates analyzed by NMDS, based on the BIOLOG analysis considering: A) the profile of all 76 isolates; B) the profile of the isolates classified in the *P. fluorescens* group; and C) the profile of the isolates classified in the *P. putida* group. D) Oxidation levels (absorbance 600 nm) of the carbon source D-galactonic acid by isolates in the *P. putida* group

5.3.5 The connection between the presence of *xutA* and D-xylose utilization

Based on the higher amount (75%) of isolates classified in the *P. fluorescens* group in our study, and supported by the variations found among isolates being not related to the phylogenetic signal, we focused our attention to the ecological differences between rhizosphere and bulk soil populations of the *P. fluorescens* group. One of the most significant differences between the compared populations was the enrichment of xylose utilization genes

in bulk soil isolates. Moreover, these genes are known and well characterized. Thus, the xylose utilization function was chosen for a deeper analysis in order to test the hypothesis that this genotypic difference found in comparative genomics result in a phenotypic difference between the rhizosphere and bulk soil populations, and therefore is ecologically relevant.

The first attempt in this issue was to test the cultivation of 22 isolates (11 from bulk soil and 11 from rhizosphere) - which covers all phylogenetic clades pointed by phylogenomics - by using D-xylose as single C-source in M9 minimal medium. The isolates containing the *xut* operon showed a higher growth than the ones where these genes were absent. The rhizosphere isolates with those genes showed a similar growth level compared to the bulk soil ones, suggesting that independently on the niche occupation, if the *xut* genes are present, its codified functions can be used to assimilate D-xylose when necessary (Figure 5.5A). However, as previously shown, the amount of isolates containing this set of genes is smaller in the rhizosphere, whilst most isolates from bulk soil presented this gene.

In order to go deeper in this issue, two phylogenetically close related isolates of the *P. fluorescens* group, named R38 (depleted of *xut* genes) and B26 (hosting *xut* genes), were chosen as models for better testing the growth in D-xylose, after showing a different growth in the first assay (Figure 5.5A). The TECAN growth curves showed that both isolates show a similar pattern in a carbon rich environment (LB medium), but the isolate from the rhizosphere R38 had a slight higher growth compared to the one from the bulk soil (B26). After approximately 4 hours, they entered in the exponential (log) phase and they reached the stationary phase after approximately 30 hours of growth showing an O.D. of ~1.6 (Figure 5.5B). The similar experiment, but using M9 minimal medium amended with glucose as the single carbon source, revealed similar results, with a more prevalent grow of R38 over B26 (Figure 5.5C), mainly after the log phase. In this case, the exponential phase of the isolates started approximately after 17 hours and reached the stationary phase after ~51 hours with an O.D. of ~0.6 and ~0.4 for R38 and B26, respectively. The similar trends of the previous two experiments completely changed when the C-source was solely D-xylose amended to the minimal medium, confirming the first assay of this topic. In this case, B26 showed a much higher growth than R38 (Figure 5.5A and D). R38 showed virtually no growth on D-xylose (O.D. < 0.1), whereas B26 showed a similar growth in D-xylose compared to glucose (O.D. ~0.4), starting the log phase also after 17 hours and reaching the stationary phase after ~57 hours (Figure 5.5D).

Aiming to test the hypothesis that the xylose utilization genes enriched in the genome sequences of the bulk soil populations are directly related to their higher growth in D-xylose,

the *xutA* gene that codifies for xylose isomerase was knocked out in B26. The mutants with the *xutA* deletion B26ΔxutA 1 and B26ΔxutA 2 lacked the capacity to grow using D-xylose as a C-source, reaching the same low level observed in R38 (Figure 5.5E). These results confirmed that the bulk soil isolate grows significantly more in D-xylose than the rhizosphere isolate, and the presence of xylose utilization genes - mainly the gene coding for the xylose isomerase enzyme (*xutA*) - are responsible for the higher growth in D-xylose.

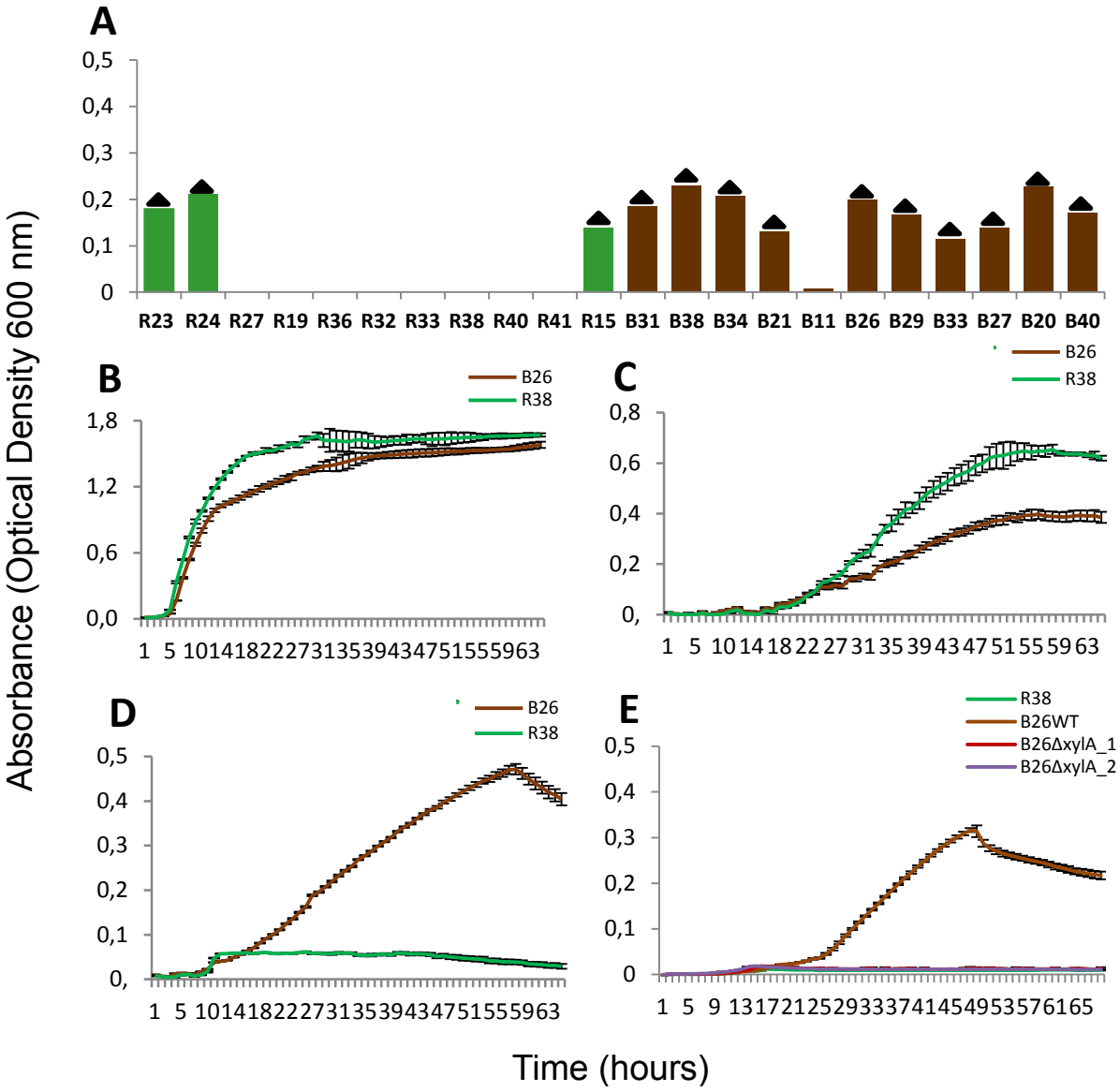


Figure 5.5 Growth curves comparing bulk soil and rhizosphere populations. A) Strains of each phylogenetic clade growing on M9 medium with D-xylose as the single carbon source (C-source). Rhizosphere or bulk soil isolates are represented by green and brown bars, respectively. Triangles indicate the presence of *xut* genes in the strain. The O.D. shown in this graph was measured after 60 hours of growth, when curves were stable (stationary phase). B) Representative isolates containing (B26) or not (R38) the *xut* genes, growing on Luria Bertani (LB) medium; C) M9 minimal medium containing glucose as the single C-source; D) M9 medium containing D-xylose as the single C-source. E) Two *xutA* deletion mutants (B26ΔxutA 1; B26ΔxutA 2) compared to B26 wild type (WT) and R38 on M9 medium with D-xylose as the single C-source

5.4 Discussion

5.4.1 Differential selection of *Pseudomonas* spp. in bulk soil and rhizosphere of sugarcane

Rhizosphere and bulk soil are soil locations differing in its physicochemical characteristics (Li et al., 2016). Such distinctions also create different niches for microbial life, which consequently selects particular microbial communities in each habitat (Berg and Smalla, 2009; Uroz et al., 2010; Mendes et al., 2014). The differential shaping in bacterial community was previously showed in sugarcane rhizosphere compared to bulk soil (Lopes et al., 2016). However, not all bacterial taxa are affected by this selection. In sugarcane soils, only 21% of the sequenced bacterial families were significantly different between bulk soil and rhizosphere samples (Lopes et al., 2016). Here we showed that *Pseudomonas* is a genus under a different selection in rhizosphere and bulk soil, having a greater abundance in the rhizosphere, where it is differentially structured at community level between these habitats. These results suggest that the plant roots favor the growth of some *Pseudomonas* OTUs that are possibly more rhizosphere-competent than others. In addition, the carbon-rich environment of the roots surroundings seem to be a preferential soil spot for those selected *Pseudomonas* spp., as the total abundance of the genus was higher in the rhizosphere than in bulk soil. In contrast, *Pseudomonas* OTUs favored in bulk soil were less abundant, which is in line with the lower carbon availability of this environment.

Consistently with the high soil heterogeneity, the *Pseudomonas* community of bulk soil samples was more distinct to each other than the observed among rhizosphere samples, suggesting that the rhizosphere is a habitat with similar selective pressures between different plants to this group of bacteria. Based on that, and in the number of shared and accessory genes it can be argued that the pan-genomes of fluorescent *Pseudomonas* spp. is bigger in the bulk soil, and could serve as a reservoir of biodiversity for the hotspots in soils, such as the rhizosphere. The characteristics of each habitat might select for specific genes acting like an environmental filter for genetic diversity.

Studies using other approaches also suggested the rhizosphere and bulk soil distinct selection on *Pseudomonas* spp. in other plants. Both tomato and flax roots were shown to harbor different genotypes and phenotypes of culturable fluorescent *Pseudomonas* spp. strains when compared to bulk soil (Lemanceau, 1994). *Pseudomonas* genus community of strawberry and oilseed rape also shown to be different between rhizosphere and bulk soil

samples (Costa et al., 2007). At this point, we have just added sugarcane in the list of plants that shapes *Pseudomonas* spp. community in their roots vicinity.

5.4.2 Selection based on genome content of closely related isolates

Although the selection in the *Pseudomonas* genus community is clear and stated, little is known about the selection of rhizosphere or bulk soil upon specific populations or in an intra-population level, based on the genome content of individuals. In this case, the better fitness of different genotypes within populations is supposed to be selected. Approaches for studying bacterial communities have no resolution to assess these intra-species differences and therefore much can be found using population genomics. In our study, most isolates of the *P. fluorescens* group (53) were phylogenetically related, belonging to two neighboring subgroups (Lopes et al., 2017 - submitted). It was also shown that these 53 isolates were classified in a single 16S rRNA gene OTU (>99% similarity), the basic taxonomic unit commonly used in microbial community studies. Even so, such comparison among phylogenetically related bacteria was able to show genomic differences regarding bulk soil and rhizosphere isolates, revealing important ecological distinctions between the populations of the two habitats.

Other studies showed distinctions in the populations of fluorescent *Pseudomonas* spp. inhabiting different soil or plant habitats. A phenotypic screening of 698 fluorescent *Pseudomonas* spp. suggested that strains with few beneficial properties for plant growth are preferentially selected in the maize rhizosphere compared to bulk soil (Vacheron et al., 2016). On the other hand, genomic and phenotypical differences were found when comparing 19 rhizosphere and endosphere fluorescent *Pseudomonas* spp. isolates of *Populus deltoides*, where more functions related to growth - i.e. phosphate solubilization, protease activity, denitrification and root growth promotion - were observed in the rhizosphere isolates; while more functions related to plant signaling and metabolic versatility were found in the endosphere isolates (Timm et al., 2015). Consistently, our results revealed more genes and functions related to plant growth in genome sequences from isolates obtained from the rhizosphere. The functions related to phosphorus cycle were one of the most significant, enriched in the rhizosphere populations of both groups of *Pseudomonas*. In the *P. fluorescens* group, we found an enrichment of genes and functions related to acid and alkaline phosphatases in the rhizosphere isolates, while in the *P. putida* group, an enrichment of transporters of phosphate ions and organic acids related to phosphate solubilization were found. Interestingly, one of the main limitations for plant productivity in several tropical soils

is the phosphate deficiency – promoted by the chemical properties of such soils (Cherubin et al., 2016). Therefore, our results suggest that one of the main criteria for fluorescent *Pseudomonas* spp. selection in the sugarcane rhizosphere could be the ability to make phosphate available.

In the rhizosphere isolates of the *P. putida* group, we found the enrichment of D-galactonate catabolism GO-term, and the higher oxidation of D-galactonic/D-galacturonic acid in the BIOLOG assay. D-galacturonate is the major component of pectin, and is found in root exudates (Zhang et al., 2011; Tawaraya et al., 2015). The ability to use these compounds could be useful for colonizing rhizosphere. A similar result was observed when analyzing total microbial community of the same soil samples, where both community level physiological profile (CLPP) and metagenome prediction indicated that this function is significantly higher in the rhizosphere microbiome (Lopes et al., 2016). Our results suggest that isolates in the *P. putida* group follow the pattern observed in the rhizosphere microbiome.

In the bulk soil population of the *P. fluorescens* group, we observed a significantly higher frequency of a gene set related to D-xylose assimilation, which comprises the xylose utilization (*xut*) operon. D-xylose is the major component of hemicellulose (Lachke, 2002), the most labile polymer of lignocellulose, which is also composed by the recalcitrant fractions cellulose and lignin (Mood et al., 2013). The sugarcane soil used in this approach was managed in a green-harvest system, where the field is permanently covered with higher amounts of sugarcane biomass (predominantly composed of lignocellulose). As bulk soil is commonly characterized as oligotrophic, with limited carbon sources (Dennis et al., 2010), the capacity to assimilate xylose could be a great advantage in the bulk soil.

Although some genes and functions were enriched in the populations of rhizosphere or bulk soil, the general physiological profile showed the clustering of isolates according to the phylogeny (Figure 5.4). It indicates that most C-sources analyzed in this approach are not under different selective pressures between rhizosphere and bulk soil populations of the *P. fluorescens* group, and therefore the general metabolism is not driven by the occupation of ecological niches, which is highly consistent with the low number of genes/functions significantly different between the studied habitats for this group. We also found a prevalent separation of metabolic patterns driven by phylogeny in the *P. putida* group. Hence, few features appear to be under distinct selective pressures in each soil habitat targeted in this study. Therefore, the niche partitioning between rhizosphere and bulk soil populations in our study is related to specific genomic changes, which is in agreement with previous findings in populations of *Vibrio* spp. from ocean habitats, where gene-specific rather than genome-wide

selection led to the differential occupation of niches by the subpopulations (Shapiro et al., 2012).

In combination, our results argue for a niche occupation related to features based on phylogeny (*P. putida* group) or intra-population variation of a few functions and genes (*P. fluorescens* group). In the latter case, the dynamics of acquisition or loss of these features in the isolates remains to be properly addressed. Horizontal gene transfer (HGT), gene loss and genomic rearrangements generates and multiply genetic diversity, and contribute to the occurrence of variants within a population that can have advantage for occupying specific niches. In this study, we have screened the genome sequences for plasmids (which were absent) and other markers of HGT, but results did not support for clear differences between rhizosphere and bulk soil regarding these issues.

It is also noteworthy that the comparison of the 16S rRNA gene sequences between the community OTUs and the sequenced isolates showed that: I) our isolation and culturing capacity is still low, since only 5 OTUs hosted sequences from the 76 isolates, out of the 134 present in the *Pseudomonas* spp. community, and II) ecological studies of microbial communities based on OTUs derived from partial 16S rRNA sequences cannot separate populations with distinct ecological behavior, and therefore much care should be taken when making functional inferences based on classification of OTUs. For example, the majority of isolates in the *P. fluorescens* group (53) are considered the same OTU in community analysis, but their subpopulations show clear ecological distinctions related to their niches.

5.4.3 Xylose utilization supports ecological differentiation between populations

Comparative genomics indicate the use of D-xylose as a feature prevailing in isolates from bulk soil in comparison to those from the rhizosphere in the populations of the *P. fluorescens* group. It corroborates previous results, where analysis of CLPP patterns elected D-xylose as the C-source most significantly oxidized by the bulk soil community (Lopes et al., 2016). As described above, such capacity would constitute an advantage to capture carbon in oligotrophic environments. In counterpart, as the rhizosphere is characterized by a high and diverse input of labile carbon sources from the root exudation (Dennis et al., 2010), the selective pressures for D-xylose utilization in the bacteria inhabiting this compartment might be reduced, making this capacity not required for bacterial survival.

The shift in xylose utilization genes frequencies between the bulk soil and rhizosphere populations can be inferred as a micro-evolutionary process related to the ecological niches of each population (Falush, 2009). Both selection for D-xylose catabolism may have promoted

the accumulation of those genes in the bulk soil population and genetic drift might have decreased the frequency of those genes in the rhizosphere population. Another possibility is that the lack in the *xut* operon caused a gain of fitness which was selected in the rhizosphere environment, where fast growth rate was shown to be important for fluorescent *Pseudomonas* spp. (Simons et al., 1996). Although having a different frequency between bulk soil and rhizosphere populations, the *xut* operon was not totally excluded from the rhizosphere population, and the genes did not reached the fixation in the bulk soil population. One hypothesis to explain it is the lack of true geographical barrier between bulk soil and rhizosphere, which allows bacteria that lives in one habitat to migrate or be accidentally transported to another. The alternative or complementary hypothesis is that the shifting selective pressures of the two habitats are not strong enough to allow the existence of only the most adapted genotype in each environment, but allowing the co-existence of both genotypes in different frequencies in each habitat.

In the bacterial xylose isomerase pathway, D-xylose is firstly isomerized to D-xylulose, which is phosphorylated by a xylulokinase and then enters in the pentose phosphate pathway, where energy and precursor molecules for biosynthesis are generated (Liu et al., 2015). In *E. coli*, the genes related to this pathway contain the transcriptional activator *xylR*; the ABC transporter genes *xylF*, *xylG* and *xylH*; the isomerase-coding gene *xylA* and the xylulokinase-coding gene *xylB* (Lawlis et al., 1984). In *Pseudomonas*, those genes were characterized as *xut* (for D-xylose utilization) genes, and the *xut* operon lacks *xutB*, but other pentose kinases were suggested to perform the same function (Liu et al., 2015). The knockout in the *xutA* gene successfully showed a decrease in D-xylose growth phenotype in our bulk soil isolate (B26), consistently to what was observed in other *Pseudomonas* strain (SBW25) isolated from the phyllosphere of sugar beet (Liu et al., 2015). The analysis of several other strains covering the phylogenetic clades found in our study (pointed by phylogenomics) made it clear that this difference is not only restricted to B26 and R38, but observed in the whole populations of the *P. fluorescens* group. The phenotype confirmation of the genomic difference regarding D-xylose utilization gives another evidence for the suggestion that this function is one of the most important ecological distinctions between rhizosphere and bulk soil populations of the *P. fluorescens* group in our study.

Liu et al. (2015) suggested that *xut* genes might be important for the colonization of plants, both in the shoots and in the rhizosphere of sugar beet. However, the *xutA* mutant was only compared to its wild type, but not to other strains of the rhizosphere population that potentially could not have those genes. In addition, strain SBW25 was isolated from the

phyllosphere and not rhizosphere, where the physicochemical characteristics are different and the evolution of the bacteria could have been dependent on D-xylose utilization, as the leaf surface is poor in nutrients compared to rhizosphere (Turner et al., 2013). Nevertheless, the plant exudation (sugar beet), soil and climate of that study can be very different compared to the conditions of the sampled area in the present study. Taking these results together, we call attention to the need for study populations instead of single isolates in order to make ecological inferences, since genetic variation was observed inside populations and a single observation is not enough to depict what is prevalent in a population of a given habitat. In sum, our results suggest that bulk soil environment selects genotypes able to catabolize D-xylose, while in the rhizosphere this capacity might be not essential.

5.5 Conclusions

This study showed that sugarcane rhizosphere and bulk soil differentially shape the community of *Pseudomonas* spp., by changing its composition and increasing its abundance in the rhizosphere. Similarly, the genome content of closely related isolates is distinct between the habitats. Populations of rhizosphere are enriched in genes/functions related to phosphorus cycle in both phylogenetic groups analyzed (*P. putida* and *P. fluorescens*); D-galactonate catabolism and D-galactonic acid oxidation is significantly higher in the rhizosphere population of the *P. putida* group. Xylose utilization genes are enriched in the bulk soil populations of the *P. fluorescens* group. Consistently, the growth in D-xylose is higher for isolates from bulk soil than those from rhizosphere. The knockout of the *xutA* gene of the bulk soil isolate B26 confirmed that the D-xylose isomerase pathway and the *xut* operon are responsible for the higher ability to grow on D-xylose. Although some genes/functions were changed in the rhizosphere and bulk soil populations, the general physiological profile resembles the phylogeny of the isolates, showing that only specific traits were differentially selected by the rhizosphere and bulk soil environments. In combination, these results inaugurate population and gene-based studies of microbial ecology in soils, and highlight that understanding communities is the first step, but very shallow, to provide a better exploration of microbial functions in soils, and its benefits for agriculture.

5.6 Supplementary Material

Supplementary Table 5.1 Accession numbers of the 76 genome sequences available in the Genbank/DDBJ/ENA databases

SUBID	BioProject	BioSample	Accession	Organism
SUB2594650	PRJNA383709	SAMN06771400	NEHD00000000	Pseudomonas sp. R9(2017)
SUB2594650	PRJNA383709	SAMN06771399	NEHE00000000	Pseudomonas sp. R5(2017)
SUB2594650	PRJNA383709	SAMN06771398	NEHF00000000	Pseudomonas sp. R4(2017)
SUB2594650	PRJNA383709	SAMN06771397	NEHG00000000	Pseudomonas sp. R47(2017)
SUB2594650	PRJNA383709	SAMN06771396	NEHH00000000	Pseudomonas sp. R45(2017)
SUB2594650	PRJNA383709	SAMN06771395	NEHI00000000	Pseudomonas sp. R44(2017)
SUB2594650	PRJNA383709	SAMN06771394	NEHJ00000000	Pseudomonas sp. R43(2017)
SUB2594650	PRJNA383709	SAMN06771393	NEHK00000000	Pseudomonas sp. R41(2017)
SUB2594650	PRJNA383709	SAMN06771392	NEHL00000000	Pseudomonas sp. R40(2017)
SUB2594650	PRJNA383709	SAMN06771391	NEHM00000000	Pseudomonas sp. R39(2017)
SUB2594650	PRJNA383709	SAMN06771390	NEHN00000000	Pseudomonas sp. R38(2017)
SUB2594650	PRJNA383709	SAMN06771389	NEHO00000000	Pseudomonas sp. R37(2017)
SUB2594650	PRJNA383709	SAMN06771388	NEHP00000000	Pseudomonas sp. R36(2017)
SUB2594650	PRJNA383709	SAMN06771387	NEHQ00000000	Pseudomonas sp. R35(2017)
SUB2594650	PRJNA383709	SAMN06771386	NEHR00000000	Pseudomonas sp. R34(2017)
SUB2594650	PRJNA383709	SAMN06771385	NEHS00000000	Pseudomonas sp. R33(2017)
SUB2594650	PRJNA383709	SAMN06771384	NEHT00000000	Pseudomonas sp. R32(2017)
SUB2594650	PRJNA383709	SAMN06771383	NEHU00000000	Pseudomonas sp. R31(2017)
SUB2594650	PRJNA383709	SAMN06771382	NEHV00000000	Pseudomonas sp. R29(2017)
SUB2594650	PRJNA383709	SAMN06771381	NEHW00000000	Pseudomonas sp. R28(2017)
SUB2594650	PRJNA383709	SAMN06771380	NEHX00000000	Pseudomonas sp. R27(2017)
SUB2594650	PRJNA383709	SAMN06771379	NEHY00000000	Pseudomonas sp. R26(2017)
SUB2594650	PRJNA383709	SAMN06771378	NEHZ00000000	Pseudomonas sp. R25(2017)
SUB2594650	PRJNA383709	SAMN06771377	NEIA00000000	Pseudomonas sp. R24(2017)
SUB2594650	PRJNA383709	SAMN06771376	NEIB00000000	Pseudomonas sp. R23(2017)
SUB2594650	PRJNA383709	SAMN06771375	NEIC00000000	Pseudomonas sp. R22(2017)
SUB2594650	PRJNA383709	SAMN06771374	NEID00000000	Pseudomonas sp. R21(2017)
SUB2594650	PRJNA383709	SAMN06771373	NEIE00000000	Pseudomonas sp. R20(2017)
SUB2594650	PRJNA383709	SAMN06771372	NEIF00000000	Pseudomonas sp. R19(2017)
SUB2594650	PRJNA383709	SAMN06771371	NEIG00000000	Pseudomonas sp. R17(2017)
SUB2594650	PRJNA383709	SAMN06771370	NEIH00000000	Pseudomonas sp. R16(2017)
SUB2594650	PRJNA383709	SAMN06771369	NEII00000000	Pseudomonas sp. R15(2017)
SUB2594650	PRJNA383709	SAMN06771368	NEIJ00000000	Pseudomonas sp. R14(2017)
SUB2594650	PRJNA383709	SAMN06771367	NEIK00000000	Pseudomonas sp. R12(2017)
SUB2594650	PRJNA383709	SAMN06771366	NEIL00000000	Pseudomonas sp. R11(2017)
SUB2594650	PRJNA383709	SAMN06771365	NEIM00000000	Pseudomonas sp. R10(2017)
SUB2594650	PRJNA383709	SAMN06771364	NEIN00000000	Pseudomonas sp. B9(2017)
SUB2594650	PRJNA383709	SAMN06771363	NEIO00000000	Pseudomonas sp. B8(2017)
SUB2594650	PRJNA383709	SAMN06771362	NEIP00000000	Pseudomonas sp. B7(2017)
SUB2594650	PRJNA383709	SAMN06771361	NEIQ00000000	Pseudomonas sp. B6(2017)

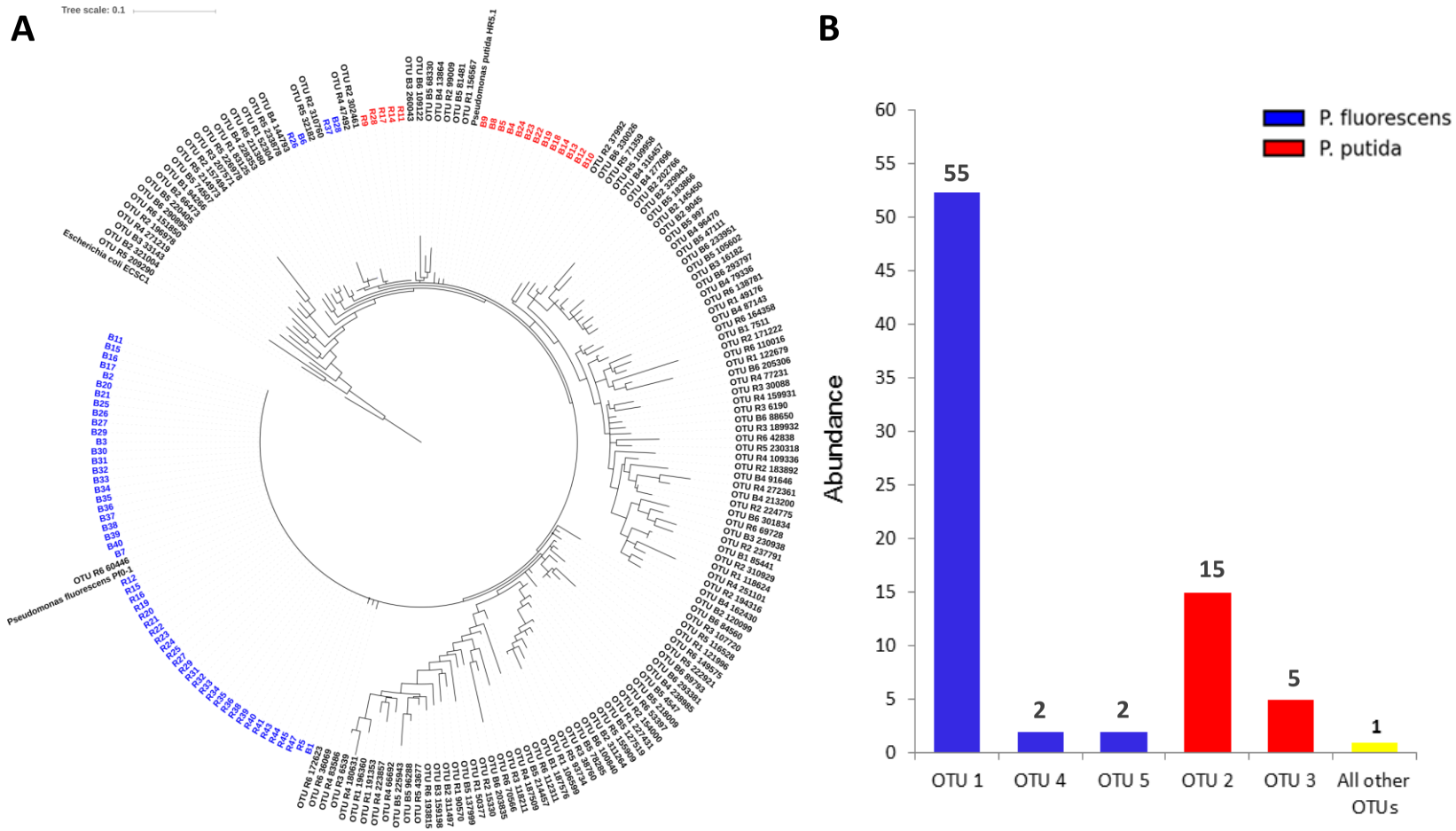
SUB2594650	PRJNA383709	SAMN06771360	NEIR00000000	Pseudomonas sp. B5(2017)
SUB2594650	PRJNA383709	SAMN06771359	NEIS00000000	Pseudomonas sp. B4(2017)
SUB2594650	PRJNA383709	SAMN06771358	NEIT00000000	Pseudomonas sp. B40(2017)
SUB2594650	PRJNA383709	SAMN06771357	NEIU00000000	Pseudomonas sp. B3(2017)
SUB2594650	PRJNA383709	SAMN06771356	NEIV00000000	Pseudomonas sp. B39(2017)
SUB2594650	PRJNA383709	SAMN06771355	NEIW00000000	Pseudomonas sp. B38(2017)
SUB2594650	PRJNA383709	SAMN06771354	NEIX00000000	Pseudomonas sp. B37(2017)
SUB2594650	PRJNA383709	SAMN06771353	NEIY00000000	Pseudomonas sp. B36(2017)
SUB2594650	PRJNA383709	SAMN06771352	NEIZ00000000	Pseudomonas sp. B35(2017)
SUB2594650	PRJNA383709	SAMN06771351	NEJA00000000	Pseudomonas sp. B34(2017)
SUB2594650	PRJNA383709	SAMN06771350	NEJB00000000	Pseudomonas sp. B33(2017)
SUB2594650	PRJNA383709	SAMN06771349	NEJC00000000	Pseudomonas sp. B32(2017)
SUB2594650	PRJNA383709	SAMN06771348	NEJD00000000	Pseudomonas sp. B31(2017)
SUB2594650	PRJNA383709	SAMN06771347	NEJE00000000	Pseudomonas sp. B30(2017)
SUB2594650	PRJNA383709	SAMN06771346	NEJF00000000	Pseudomonas sp. B2(2017)
SUB2594650	PRJNA383709	SAMN06771345	NEJG00000000	Pseudomonas sp. B29(2017)
SUB2594650	PRJNA383709	SAMN06771344	NEJH00000000	Pseudomonas sp. B28(2017)
SUB2594650	PRJNA383709	SAMN06771343	NEJI00000000	Pseudomonas sp. B27(2017)
SUB2594650	PRJNA383709	SAMN06771342	NEJJ00000000	Pseudomonas sp. B26(2017)
SUB2594650	PRJNA383709	SAMN06771341	NEJK00000000	Pseudomonas sp. B25(2017)
SUB2594650	PRJNA383709	SAMN06771340	NEJL00000000	Pseudomonas sp. B24(2017)
SUB2594650	PRJNA383709	SAMN06771339	NEJM00000000	Pseudomonas sp. B23(2017)
SUB2594650	PRJNA383709	SAMN06771338	NEJN00000000	Pseudomonas sp. B22(2017)
SUB2594650	PRJNA383709	SAMN06771337	NEJO00000000	Pseudomonas sp. B21(2017)
SUB2594650	PRJNA383709	SAMN06771336	NEJP00000000	Pseudomonas sp. B20(2017)
SUB2594650	PRJNA383709	SAMN06771335	NEJQ00000000	Pseudomonas sp. B1(2017)
SUB2594650	PRJNA383709	SAMN06771334	NEJR00000000	Pseudomonas sp. B19(2017)
SUB2594650	PRJNA383709	SAMN06771333	NEJS00000000	Pseudomonas sp. B18(2017)
SUB2594650	PRJNA383709	SAMN06771332	NEJT00000000	Pseudomonas sp. B17(2017)
SUB2594650	PRJNA383709	SAMN06771325	NEJU00000000	Pseudomonas sp. B10(2017)
SUB2594650	PRJNA383709	SAMN06771326	NEJV00000000	Pseudomonas sp. B11(2017)
SUB2594650	PRJNA383709	SAMN06771327	NEJW00000000	Pseudomonas sp. B12(2017)
SUB2594650	PRJNA383709	SAMN06771328	NEJX00000000	Pseudomonas sp. B13(2017)
SUB2594650	PRJNA383709	SAMN06771329	NEJY00000000	Pseudomonas sp. B14(2017)
SUB2594650	PRJNA383709	SAMN06771330	NEJZ00000000	Pseudomonas sp. B15(2017)
SUB2594650	PRJNA383709	SAMN06771331	NEKA00000000	Pseudomonas sp. B16(2017)

Supplementary Table 5.2 Solutions used for making the M9 minimal medium

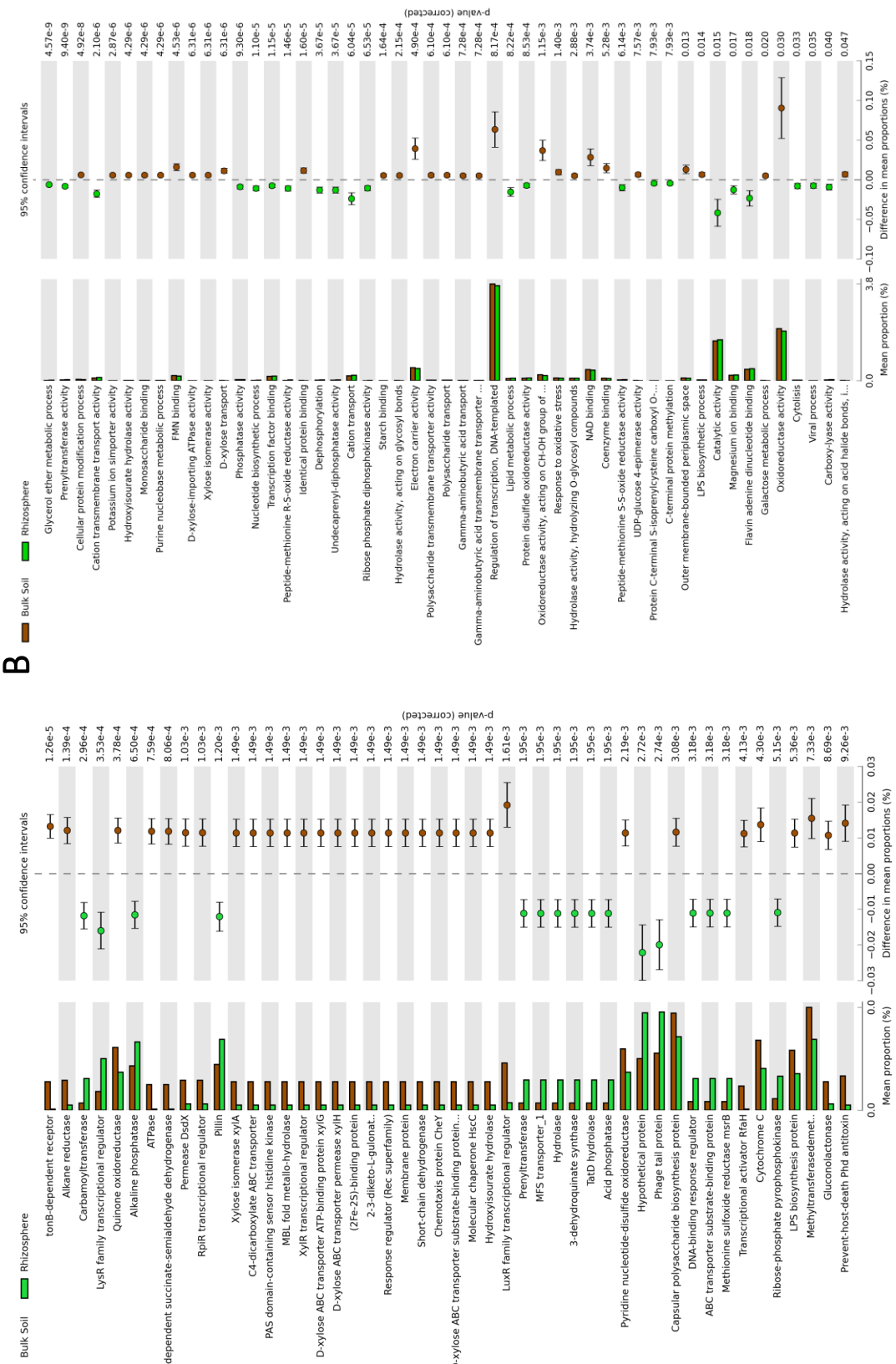
Micronutrients solution	(g/L)	M9 salts solution	(g/L)
$FeCl_3$	0.1	$Na_2HPO_4.7H2O$	64.0
Thiamine HCl	0.06	KH_2PO_4	15.0
$NaEDTA.2H_2O$	0.4	NaCl	2.5
$CoCl_2.6H2O$	1.8	NH_4Cl	5.0
$ZnClSO_4.7H2O$	1.8		
$NaMoO_4.2H2O$	0.4		
H_3BO_3	0.1		
$MnSO_4.H_2O$	1.2		
$CuCl2.2H_2O$	1.2		

Supplementary Table 5.3 Sequences of primers used for making the *xutA* deletion mutants

Oligonucleotides	DNA sequences*
xutA-F1	TATAAGCTTTGGGATAGGCCCGCTCATCC
xutA-Ovlp-R1	CGTGGGCGATCAGCGATAGATGAGTCGACATCGGGGAAGTACGG
xutA-Ovlp-F1	CCGTACTTCCCCGATGTCGACTCATCTATCGCTGATCGCCCACG
xutA-R2	ATAGAATTCGGCTACCGCATTGGTCAGC
xutA-F3	TATAAGCTTGCTAGTATCGGGACACGCCCC
xutA-R3	ATGGTACCTCGTGGGCGATCAGCGATAGATG



Supplementary Figure S5.1 Comparison between the 16S rRNA gene sequences extracted from the genome assemblies of the 76 isolates and the 16S rRNA gene sequences of the operational taxonomic units (OTUs) classified as *Pseudomonas* spp. in the high-throughput 16S rRNA gene sequencing of the soil samples. A) Maximum likelihood phylogenetic tree containing the 16S rRNA gene sequences of the 76 isolates and the 134 *Pseudomonas* spp. OTUs found in the community analysis; B) split OTUs of the 286 sequences (76 isolates + 134 community OTUs) using the UCLUST algorithm at 99% similarity cutoff



Supplementary Figure S5.2 List of gene clusters (A) and GO-terms (B) showing significant differences between rhizosphere and bulk soil populations of the *P. fluorescens group* (57 isolates). Statistical analysis were performed using the Welch's t-test with the Bonferroni *P*-value correction

Supplementary Figure S5.3 List of gene clusters (A) and GO-terms (B) showing significant differences between rhizosphere and bulk soil populations considering all 76 isolates. Statistical analysis were performed using the Welch's t-test with the Bonferroni *P*-value correction

References

Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov A, Lesin V, Nikolenko S, Pham S, Prjibelski A, Pyshkin A, Sirotkin A, Vyahhi N, Tesler G, Ale- kseyev M, Pevzner P (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477 DOI 10.1089/cmb.2012.0021

Berg G, Smalla K (2009) Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol*, 68: 1-13

Bergmark L, Poulsen PHB, Al-Soud WA, Norman A, Hansen LH, Sorensen SJ (2012) Assessment of the specificity of *Burkholderia* and *Pseudomonas* qPCR assays for detection of these genera in soil using 454 pyrosequencing. *FEMS Microbiol Lett*, 333: 77-84

Cam Y, Alkim C, Trichez D, Trebosc V, Vax A, Bartolo F, Besse P, Francois JM, Walther T (2016) Engineering of a synthetic metabolic pathway for the assimilation of (D)-D-xylose into value-added chemicals. *ACS Synth Biol*, 5: 607-618

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zanevald J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7 335–336. 10.1038/nmeth.f.303

Cherubin MR, Franco ALC, Cerri CEP, Karlen DL, Pavinatto PS, Rodrigues M, Davies CA, Cerri CC (2016) Phosphorus pools responses to land-use change for sugarcane expansion in weathered Brazilian soils. Geoderma, 265: 27-38

Clarke KR, Gorley RN (2006) *PRIMER v6: User Manual/Tutorial*. Plymouth: PRIMER-E.

CONAB (2017) *Acompanhamento da safra brasileira de cana-de-açúcar* , v.2 – Safra 2016/17, n.4

Contreras-Moreira B, Vinuesa P (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol*, 79(24): 7696-7701

Costa R, Gomes NCM, Krogerrecklenfort E, Opelt K, Berg G, Smalla K (2007) *Pseudomonas* community structure and antagonistic potential in the rhizosphere: insights gained by combining phylogenetic and functional gene-based analyses. *Environ Microbiol*, 9(9): 2260-2273.

Dennis PG, Miller AJ, Hirsch PR (2010) Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiol Ecol*, 72: 313-327

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19): 2460-2461

Falush D (2009) Toward the use of genomics to study microevolutionary change in bacteria. *PLoS Genet*, 5(10): e1000627. doi:10.1371/journal. pgen.1000627

Garcia-Salamanca A, Molina-Henares MA, Dillewijn P, Solano J, Pizarro-Tobias P, Roca A, Duque E, Ramos JL (2012) Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil. *Microb Biotechnol*, 6(1): 36-44

Hammer Ø, Harper DAT, Ryan PD (2001). PAST: paleontological statistics software package for education and data analysis. *Paleontol. Electron.* 4 1–9.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24(11): 1403-1405

Jombart T, Devilled S, Bollox F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*, 11: 94

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9): 1236-1240.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780 DOI 10.1093/molbev/mst010.

Klindworth A., Pruesse E., Schweer T., Peplies J., Quast C., Horn M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1 10.1093/nar/gks808

Kuzyakov Y, Blagodatskaya E (2015) Microbial hotspots and hot moments in soil: concept & review. *Soil Biol Biochem*, 83: 184-199

Lachke A (2002) Biofuel from D-xylose: the second most abundant sugar. *Resonance*, 7(5): 50-58

Lawlis VB, Dennis MS, Chen EY, Smith DH, Henner DJ (1984) Cloning and sequencing of the D-xylose isomerase and xylulose kinase genes of *Escherichia coli*. *Appl Environ Microbiol* 47: 15–21.

Lemanceau P, Corberand T, Gardan L, Latour X, Daguerre G, Boeufgras JM, Alabouvette C (1995) Effect of two plant species, flax (*Linum usitatissinum* L.) and tomato (*Lycopersicon esculentum* Mill.), on the diversity of soil borne populations of fluorescent Pseudomonads. *Appl Environ Microbiol*, 61(3): 1004-1012

Letunic I, Bork P (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44(W1): W242- 5, doi:10.1093/nar/gkw290

Li Z, Zu C, Wang C, Yang J, Yu H, Wu H (2016) Different responses of rhizosphere and non-rhizosphere soil microbial communities to consecutive *Piper nigrum* L. monoculture. *Sci Rep*, 6: 35825

Liu Y, Pan X, Li J (2015) A 1961-2010 record of fertiliser use, pesticide application and cereal yields: a review. *Agron Sustain Dev*, 35: 83-93

Loarie SR, Lobell DB, Asner GP, Mu Q, Field CB (2011) Direct impacts on local climate of sugar-cane expansion in Brazil. *Nature Climate Change*, 1: 105-109

Loper JE, Hassan KA, Mavrodi DV, Davis EW, Lim CK, Shaffer BT, Elbourne LDH, Stockwell VO, Hartney SL, Breaklwell K, Henkels MD, Tetu SG, Rangel LI, Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE, Paulsen IT (2012) Comparative genomics of plant-associated Pseudomonas spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genet* 8(7): e1002784.

Lopes LD, Pereira e Cassia MS, Andreote FD (2016) Bacterial abilities and adaptation toward the rhizosphere colonization. *Front Microbiol*, 7:134, doi: 10.3389/fmicb.2016.01341

Mendes LW, Kuramae EE, Navarrete AA, van Even JA, Tsai SM (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J*, 8: 1577–1587. doi: 10.1038/ismej.2014.17

Mendes R, Kruijt M, Bruijn I, Dekkers E, Voort M, Schneider JHM, Piceno YM, DeSantis T, Andersen GL, Bakker PAHM, Raaijmakers JM (2011) Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science*, 332(6033): 1097-100

Mood SH, Golfeshan AH, Tabatabaei M, Jouzani GS, Najafi GH, Gholami M, Ardjmand M (2013) Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. *Renewable and Sustainable Energy Reviews*, 27: 77-93

Paul EA (2014) *Soil microbiology, ecology and biochemistry*. Academic press.

Parks DH, Tyson GW, Hugenholtz P, Beiko RG (2014) STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30 3123–3124. 10.1093/bioinformatics/btu494

Rudorff BFT, Aguiar DA, da Silva WF, Sugawara LM, Adami M, Moreira MA (2010) Studies on the rapid expansion of sugarcane for ethanol production in São Paulo state (Brazil) using landsat data. *Remote Sens*, 2(4): 1057-1076

Seeman T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14): 2068-2069

Shapiro BJ, Friedman J, Cordero OX, Preheim SP, Timberlake SC, Szabó G, Polz MF, Alm EJ (2012) Population genomics of early events in the ecological differentiation of bacteria. *Science*, 336: 48-51.

Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW (2011) Pseudomonas genomes: diverse and adaptable. *FEMS Microbiol Rev*, 35(4): 652-680

Simons M., van der Bij A. J., Brand I., de Weger L. A., Wijffelman C. A., Lugtenberg B. J. J. (1996). Gnotobiotic system for studying rhizosphere colonization by plant-growth promoting *Pseudomonas* bacteria. *Mol. Plant Microbe Interact.* 9 600–607. 10.1094/MPMI-9-0600

Sokolova TA (2015) Specificity of soil properties in the rhizosphere: analysis of literature data. *Eurasian Soil Sci*, 9: 1097-1111

Stamatakis A (2014) RAxMLversion 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312-1313

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*, 30(12): 2725-2729

Tawaraya K., Horie R., Saito A., Shinano T., Wagatsuma T., Saito K., et al. (2015). Metabolite profiling of shoot extracts, root extracts, and root exudates of rice plant under phosphorus deficiency. *J. Plant Nutr.* 36 1138–1159. 10.1080/01904167.2013.780613

Timm CM, Campbell AG, Utturkar SM, Jun SR, Parales RE, Tan WA, Robeson MS, Lu TYS, Jawdy S, Brown SD, Ussery DW, Schadt CW, Tuskan GA, Doktycz MJ, Weston DJ, Pelletier DA (2015) Metabolic functions of *Pseudomonas fluorescens* strains from *Populus deltoides* depend on rhizosphere or endosphere isolation compartment. *Front Microbiol*, 6: 1118, doi: 10.3389/fmicb.2015.01118.

6 FINAL CONSIDERATIONS

6.1 Thesis in Brief

The information captured and scientifically digested in this thesis constitute a step forward in the studies of microbial ecology in sugarcane fields, and possibly a milestone in the assessment of bacterial life in rhizosphere and bulk soil.

The article derived from **Section 3** described the sugarcane rhizosphere effect in the microbial community, by revealing differential bacterial communities in the roots vicinity and in the bulk soil, and also named the main taxa selected in the rhizosphere. In addition, this approach also identified key functions for rhizosphere colonization, with a remark for the utilization of pectin compounds (D-galacturonic acid). In the **Section 4**, genome sequences were used to disclose a new diversity of fluorescent *Pseudomonas* spp. in the tropical soil prospected, claiming for new species and proposing new subgroups of *P. fluorescens*. This finding might be related to biogeography and endemicity of *Pseudomonas* spp. in soils. The **Section 5** encompass an innovative and deep survey on bacterial populations between bulk soil and rhizosphere, based on comparison of genomic sequences, metabolic profiling, and culminating with a gene-centric analysis, where mutants were used to prove the inferences primarily based on gene enrichments in the populations of each soil habitat. This study showed that phylogenetically related strains inhabiting rhizosphere or bulk soil have differences in genome content. Genes related to phosphorus cycling were significantly enriched in rhizosphere; and genes related to the utilization of D-xylose were enriched in the bulk soil population of the *P. fluorescens* group. Here we found two matches between **Sections 3 and 5**: the catabolism of D-galactonate was significantly higher in rhizosphere, either based on sequencing or metabolic profiling analyses; and D-xylose was more efficiently used by the total community or by the *P. fluorescens* population found in bulk than the ones found in rhizosphere.

6.2 The Exploration of HGT in Bulk Soil and Rhizosphere

In **Section 3**, we found a higher prediction of genes related to HGT events in the rhizosphere than in bulk soil community. This information was firstly considered as an evidence that this event should be more frequent to bacteria inhabiting rhizosphere. However, it was not observed for the 76 genomes of fluorescent *Pseudomonas* spp. analyzed in the previous sections. Despite data was not included in the article of **Section 5**, we have attempted to detect horizontally transferred genes in the genome sequences of the 76

Pseudomonas spp. isolates. However, even using several approaches, no general evidences for distinctions were found in rhizosphere and bulk soil populations, for both phylogenetic groups of fluorescent *Pseudomonas* spp., only single differences not broadly applicable (Figure 6.1).

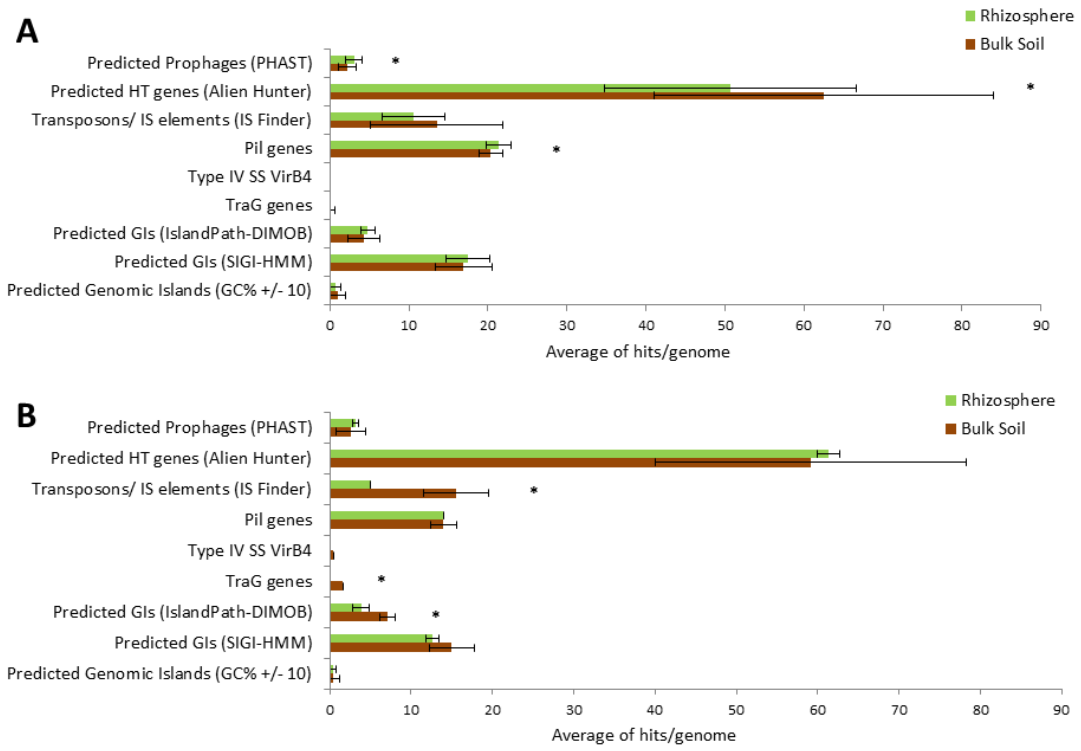


Figure 6.1 Exploration for differences in horizontally transferred genes between the bulk soil and rhizosphere populations of the (A) *P. fluorescens* (57 isolates) and (B) *P. putida* groups (19 isolates). In parenthesis are listed the softwares or methods used to detect each category. The categories without parenthesis are specific genes related to HGT, manually searched by BLAST. The average of hits in the genome sequences of each population (bulk soil or rhizosphere) was statistically compared using the Tukey test (*$P < 0.05$) for each category

In addition, four independent methods (i.e. mapping to reference genomes; BLASTs for essential plasmid genes; *de novo* assembly of plasmids; search for outlier fragments with distinct GC content and genome coverage compared to the chromosome) were used and none detected any plasmids in the genome sequences analyzed, corroborating the literature, where few plasmids were identified in *P. fluorescens* species (https://www.ncbi.nlm.nih.gov/genome/genomes/150?). Therefore, we conclude that in general there was no higher frequency of HGT processes related to rhizosphere population of fluorescent *Pseudomonas* spp., as found in the **Section 3** for the whole microbiome analysis. Since *Pseudomonas* genus is one among thousands of bacterial genera in soils, there is a need to further investigate this issue in other bacterial groups in order to have more general

statements on this issue. Moreover, the metagenome prediction (**Section 3**) pointed to other bacterial families as related to the enrichment of HGT predicted genes in the rhizosphere microbiome (mainly Koribacteraceae) (Table 6.1).

Table 6.1 Classification of OTUs that most contributed for the inferred functions predicted by PICRUSt highlighted in the Section 3. Functions significantly different (*P*<0.05 after Bonferroni correction) between bulk soil and rhizosphere

Protein/Enzyme	P-value	Major 5 contributor taxa (phylum - families)
Pilus assembly protein CpaE	9.58 e^{-3}	Proteobacteria- Bradyrhizobiaceae, Syntrophorhabdaceae; Nitrospirae- 0319-6A21; Acidobacteria- Solibacteraceae, Acidobacteriaceae
Type IV pilus assembly protein PilV	0,018	Proteobacteria- Entotheonellaceae, Syntrophorhabdaceae, Comamonadaceae, Sinobacteraceae; Nitrospirae- 0319-6A21
Type IV secretion system protein VirB4	2.81 e^{-4}	Acidobacteria- Koribacteraceae, Solibacteraceae, Acidobacteriaceae; Proteobacteria- Bradyrhizobiaceae, Hyphomicrobiaceae
Type IV secretion system protein VirB5	4.69 e^{-3}	Proteobacteria- Bradyrhizobiaceae, Hyphomicrobiaceae; Acidobacteria- Koribacteraceae, Solibacteraceae, Acidobacteriaceae
Type IV secretion system protein VirB6	7.53 e^{-3}	Acidobacteria- Koribacteraceae, Solibacteraceae, Acidobacteriaceae; Proteobacteria- Bradyrhizobiaceae, Hyphomicrobiaceae
Type IV secretion system protein VirB9	7.78 e^{-4}	Proteobacteria- Bradyrhizobiaceae, Hyphomicrobiaceae; Acidobacteria- Koribacteraceae, Solibacteraceae, Acidobacteriaceae

6.3 The Origin of Genes Enriched in Bulk Soil and Rhizosphere Populations

The main genes significantly enriched in the bulk soil or rhizosphere populations of the *P. fluorescens* group, *i.e.* xylose utilization genes and phosphatases, respectively, (**Section 5**) were further analyzed to provide more evidences for the discussion about HGT. Three independent approaches were used in order to infer about the processes related to the presence or absence of those genes in the sequenced genomes, *i.e.* GC% of the genes compared to the whole genome; position of the gene in the genome sequences of different isolates; and comparison between the phylogenetic structure of the enriched genes to the phylogeny depicted by phylogenomics. The GC percentages of the phosphatases and xylose-utilization genes were highly consistent to the GC content of the whole genomes, and the position of the genes among the genome sequences of different isolates were virtually the same. These two evidences argue for the absence of HGT processes in these genes among the isolates analyzed,

and support the process of recent gene losses. In order to properly confirm this hypothesis, the occurrence of enriched genes was contemplated along the phylogenetic distribution of the isolates.

The phosphatases-coding genes were present in most clades of the *P. fluorescens* group, but the final branches (where most part of the bulk soil isolates are) showed the absence of those genes. Based on the pattern of this tree, there is a higher probability of the phosphatases genes to have been lost in a common ancestor of the isolates in these clades, as suggested by the previous analyses (Figure 5.2 of **Section 5**). However, an opposite pattern was observed for the xylose utilization genes, which were mostly absent in the isolates sharing the common ancestor of the *P. fluorescens* group, and suddenly appeared in the final branches of the tree. Therefore, there is a probability that those genes were horizontally transferred (Figure 5.2 of **Section 5**), but from phylogenetic related taxa, as markers for HGT were not found. Aiming to further investigate this issue, the *xutA* gene was retrieved from all genome sequences (24) and compared to other *Pseudomonas* spp. reference strains that also contain those genes (8). Interestingly, the phylogenetic relationship between isolates in both the *xutA* and phylogenomic tree was very similar (Figure 6.2A and Figure 5.2 of **Section 5**).

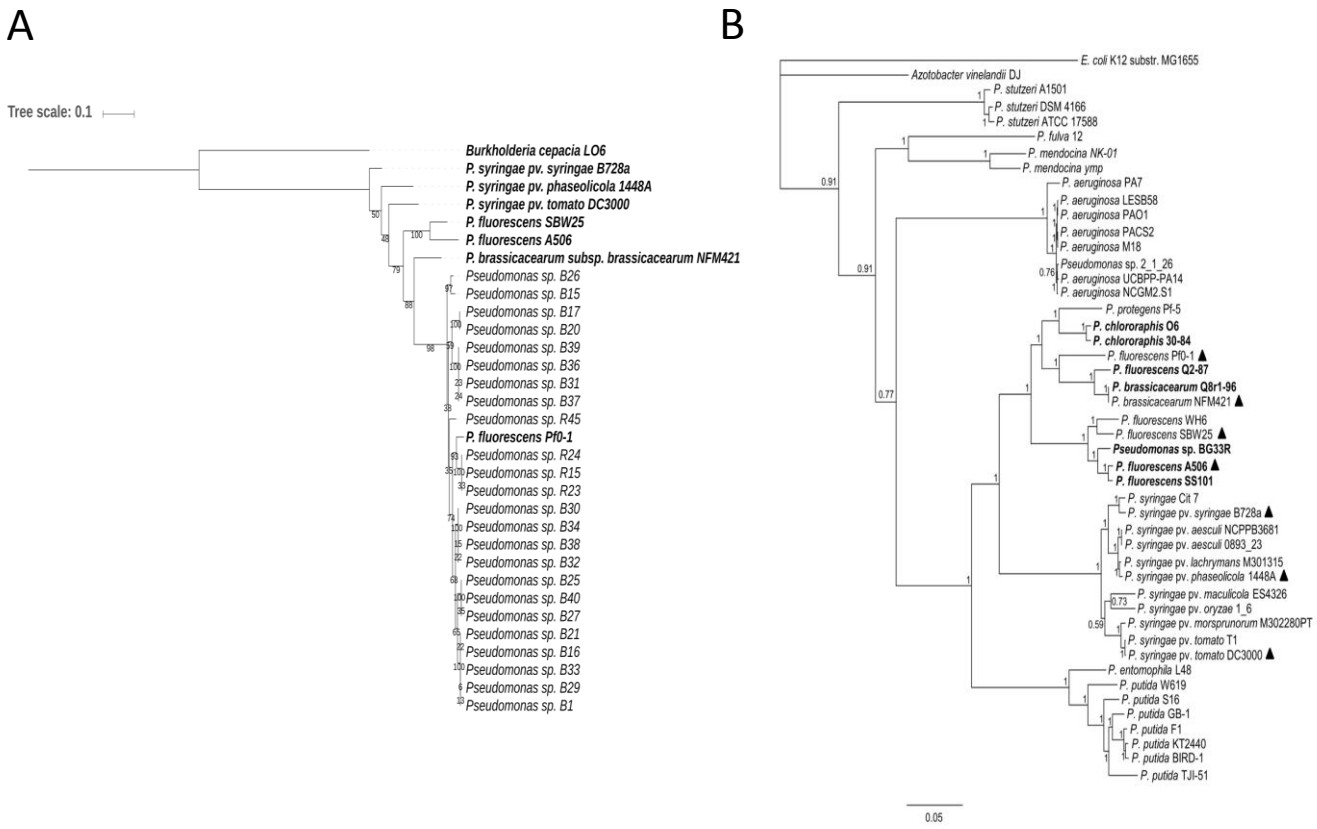


Figure 6.2 A) Phylogenetic tree of the *xylA* gene extracted from the 24 genome sequences that have the xylose utilization genes. B) Comparison of the previous tree to the phylogenetic inference performed by Loper et al. (2012) by phylogenomics, where strains with triangles have the xylose utilization genes

A phylogenomic analysis including the same reference strains mentioned was performed by Loper et al. (2012 – reference in **Section 5**) and a relatively high compatibility with the *xutA* tree was also observed (Figure 6.2A and B). In this tree, the reference genomes of the *P. fluorescens* group showed the same structure observed in our *xutA* tree. However, the references of the *P. syringae* group did not. For example, *P. syringae* pv. tomato DC3000 clustered to the isolates in the *P. fluorescens* group and not to the isolates in the *P. syringae* group (Figure 6.2B). These results suggest that HGT possibly occurred for the xylose utilization genes in the long-term evolution of the *Pseudomonas* genus, but in the shorter-term diversification of the *P. fluorescens* group and mainly in the diversification of the studied isolates, the lack of xylose utilization genes was indeed probably a result of gene loss. Therefore, we conclude that gene losses were responsible for the lack of phosphatases or xylose utilization genes in some isolates of our study, and the following selection caused a different enrichment in the populations of rhizosphere and bulk soil.

6.4 Final Remarks

This thesis was innovative to improve the resolution of approaches to study bacterial life in soils. We started from microbial community analyses; passing through population genomic analysis; to gene-level exploration. In combination, the outcomes claim for the importance of such approaches to improve our comprehension about soil microbial ecology, with a remarkable importance for the connections between bacteria and plants. Therefore, the higher resolution proposed in this study can increase the possibility to unravel ecological processes. The evolutionary thought was directly applied to the interpretation of our results and the correlation of ecology and evolution was suggested to be highly important for understanding microbial processes. Besides the theoretical and basic science contribution, the data generated in this thesis is expected to contribute for an eventual practical use in agriculture. More specifically, the investigation of rhizosphere and bulk soil of sugarcane fields disclosed new knowledge that advanced our understanding of their differences. Key bacterial features for colonizing rhizosphere or survive in bulk soil were suggested and experimentally supported. The accumulation of information on soil microbial ecology will hopefully be of great importance for practical purposes in agricultural management aiming to favor the sustainable food production.