UNIVERSIDADE DE SÃO PAULO
CENTRO DE ENERGIA NUCLEAR NA AGRICULTURA

Metagenome of Amazon forest conversion: impacts on soil-borne microbial diversity and functions

Lucas William Mendes

Piracicaba
2014

Lucas William Mendes

Metagenome of Amazon forest conversion: impacts on soil-borne microbial diversity and functions

Tese apresentada ao Centro de Energia Nuclear na Agricultura da Universidade de São Paulo para obtenção do título de Doutor em Ciências

Área de Concentração: Biologia na Agricultura e no Ambiente

Orientadora: Profa. Dra. Siu Mui Tsai

Co-orientadora: Dra. Eiko Eurya Kuramae

Piracicaba

2014

Dados Internacionais de Catalogação na Publicação (CIP)

Seção Técnica de Biblioteca - CENA/USP

Mendes, Lucas William

Metagenome of Amazon forest conversion: impacts on soil-borne microbial diversity and functions / Lucas William Mendes; orientadora Siu Mui Tsai; co-orientadora Eiko Eurya Kuramae. - - Piracicaba, 2014.

97 f. : il.

Tese (Doutorado – Programa de Pós-Graduação em Ciências. Área de Concentração: Biologia na Agricultura e no Ambiente) – Centro de Energia Nuclear na Agricultura da Universidade de São Paulo.

1. Biologia molecular 2. Bactérias 3. Bioinformática 4. Ecologia microbiana 5. Ecologia molecular 6. Florestas tropicais – Amazônia 7. Desmatamento – Amazônia 8. Microbiologia do solo 9. Manejo do solo 10. Uso do solo I. Título

CDU 579.26

To my family

Acknowledgments

God | my lovely **wife** di | my **family** edison, cidinha, rodrigo and paulinha | the **funding agencies** fapesp (proc. 2010/20353-7) and capes/wageningen (bex 1240/11-0) | my **supervisor** dra. siu mui tsai | my **sandwich supervisors** dra. eiko eurya kuramae and dr. johannes antonie van veen | to **cena-usp** and **nioo-knaw** | my **teachers** from cena and esalq | my **lab friends** acacio, ademir, aline, aline m, ana, amanda, andressa, bia, bruna, caio, camila, carol, cleiziane, clóvis, dani, daniel, dennis, enéas, fabi, felippe, fernanda, fernanda c, fran, gustavo, helena, janne, joão, lina, lucas p, maju, marcela, mari, marília, marina, milena, patrícia, paula, rafael, raquel, rosi, taketani, and many others that i met on may way | to **technical staff** of the cell and molecular biology laboratory, josé elias gomes, wagner picinini, fábio duarte and francisco (*in memoriam*) | to my friend ludmila | to agronomic engineer castro | people from **cpg cena**, neuda, sônia, daiane, fábio, alzira and gilson | to all people from **nioo** agata, alexandra, almudena, anne, emilia, gerda, julia, kowalchuk, marion, mattias, max, nardy, olaf, paloma, paolina, paul, paulien, ruth, sarah, sascha, saskia, sotos, thomas, yan yan, yani, yao | to my **wageningen** friends anderson-zani, hudson-andrea-marcela, joseph-geisa, anderson-cris, nick | to all people that complete my history

Lucas William Mendes

"We have to remember that what we observe is not nature itself,

but nature exposed to our method of questioning"

***(Werner Heisenberg** | 1901 – 1976)*

ABSTRACT

MENDES, L. W. **Metagenome of Amazon forest conversion**: impacts on soil-borne microbial diversity and functions. 2014. 97 f. Tese (Doutorado) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2014.

The Amazon rainforest is considered the world's largest reservoir of plant and animal biodiversity, but in recent years has been subjected to high rates of deforestation for the conversion of native areas into agricultural fields and pasture. The understanding of the effects of land-use change on soil microbial communities is essential, taking into account the importance that these organisms play in the ecosystem. In this context, this thesis evaluated the effect of these changes on microorganism communities in soils under different land-use systems. In the first study, the microbial communities were analyzed using the next-generation sequencing Illumina Hiseq2000, considering samples from native forest, deforested area, agriculture and pasture. From the analysis of approximately 487 million sequences was possible to show that microbial communities respond differently in each land-use system, with changes in both taxonomic and functional diversity. Also, we suggested that ecosystem function in forest soils is maintained by the abundance of microorganisms, while in disturbed areas such functioning is maintained by high diversity and functional redundancy. In the second study, we assessed the extent to which a particular plant species, *i.e.* soybean, is able to select the microbial community that inhabits the rhizosphere. From the metagenomic sequencing by the 454 GS FLX Titanium platform we investigated the taxonomic and functional diversities of soil and rhizosphere communities associated to soybean, and also tested the validity of neutral and niche theories to explain rhizosphere community assembly process. The results suggest that soybean selects a specific microbial community inhabiting the rhizosphere based on functional traits, which may be related to benefits to the plant, such as growth promotion and nutrition. This process of selection follows largely the niche –based theory indicating the selection power of the plant and other environmental variables in shaping the microbial community both at the taxonomic and functional level. This thesis highlights the importance of microbial ecology studies in the context of the Amazon to a better understanding of the effects of deforestation on microorganisms, and provides information that can be suitable for future development of sustainable approaches for the ecosystem use.

Keywords: Microbial ecology. Molecular ecology. Tropical Amazon soils. Land-use changes. High-throughput sequencing.

RESUMO

MENDES, L. W. **Metagenoma da conversão da floresta Amazônica**: impactos na diversidade taxonômica e funcional dos micro-organismos do solo. 2014. 97 f. Tese (Doutorado) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2014.

A floresta Amazônica é considerada o maior reservatório de biodiversidade vegetal e animal do mundo, porém, nos últimos anos tem sido submetida à altas taxas de desmatamento para a conversão de áreas de mata nativa em campos de agricultura e pastagem. A compreensão sobre os efeitos dessa mudança de uso da terra sobre as comunidades microbianas do solo é fundamental, levando-se em consideração a importância que esses organismos desempenham no ecossistema. Neste contexto, este trabalho de tese avaliou o efeito dessas mudanças sobre as comunidades de micro-organismos em solos sob diferentes sistemas de uso. No primeiro estudo, as comunidades microbianas foram analisadas por meio do sequenciamento de nova geração Illumina Hiseq2000, sendo consideradas amostras de áreas de floresta nativa, área desmatada, agricultura e pastagem. A partir da análise de aproximadamente 487 milhões de sequências foi possível mostrar que as comunidades microbianas respondem diferentemente em cada sistema de uso do solo, com alterações tanto na diversidade taxonômica quanto funcional. Também, sugere-se que o funcionamento do ecossistema em solos de floresta é mantido pela abundância dos micro-organismos presentes, enquanto nas áreas alteradas esse funcionamento é mantido pela alta diversidade e redundância funcional. No segundo estudo foi avaliado até que ponto uma espécie de planta, *i.e.* soja, é capaz de selecionar a comunidade habitante de sua rizosfera. A partir do sequenciamento metagenômico pela plataforma 454 GS FLX Titanium da Roche foi investigado a diversidade taxonômica e funcional das comunidades de solo e rizosfera associadas à soja, e testou-se a validade das teorias neutras e de nicho para explicar o processo de formação das comunidades microbianas. Os resultados sugerem que a soja seleciona uma comunidade específica que habita sua rizosfera com base em atributos funcionais, os quais podem estar relacionados com benefícios à planta, como promoção do crescimento e nutrição. Esse processo de seleção segue a teoria de nicho, indicando o poder de seleção da planta e de outras variáveis ambientais em moldar as comunidades microbianas tanto de forma taxonômica quanto funcional. Esta tese destaca a importância de estudos em ecologia microbiana no contexto da Amazônia para uma melhor compreensão dos efeitos do desmatamento sobre os microrganismos e disponibiliza informações que podem ser futuramente utilizados para o desenvolvimento de metodologias mais sustentáveis para o uso do ecossistema.

Palavras-chave: Ecologia microbiana. Ecologia molecular. Microbiologia do solo. Solos tropicais da Amazônia. Mudança de uso da terra. Sequenciamento em larga escala.

SUMMARY

1 INTRODUCTION

> *"I may say with truth that whenever I consider in my thoughts the beautiful order, how one thing issues out of and is derived from another, then it is as though I had read a divine text, written into the world itself, not with letters but rather with essential objects, saying: Man, stretch thy reason hither, so that thou mayest comprehend these things"*
>
> (**Johannes Kepler** | 1571 – 1630)

The Amazon forest harbors 40% of all remaining tropical rainforests, playing a fundamental role in biodiversity conservation, biogeochemical cycles, and climate regulation. Despite its global importance, the Amazon's "arc of deforestation" has been the world's most active deforestation frontier in recent decades. Deforestation in the Brazilian Amazon can be defined as clear cutting and conversion of the native forest cover to other land-uses, mainly agriculture and cattle pasture. According to the National Institute for Space Research (INPE, 2011) the land-use change has resulted in annual forest area loss of 18,918 Km^2 from 1998 to 2007, being the frontier states of Mato Grosso, Rondônia, and Pará accounted for 85% of all its deforestation. This region includes several types of land-uses, since the extraction performed on native vegetation, until the cultivation with different levels of technology for agriculture and livestock. In agriculture, stand out the crop of soybeans, cotton, corn and rice, grown in no-till or conventional form; in livestock, there are pasture areas well managed and others more degraded. In pasture areas the use of *Brachiaria brizantha* is predominant; while in agriculture, is common the use of soybean crop (*Glicine max* L. Merril) preceded by the succession of corn (*Zea mayz* L.).

The growing demand for agricultural and livestock production has led to the opening of new areas of native forest. However, when conservation practices are not adopted, the intensive land-use affects negatively both the environment and agricultural productivity (CERRI et al., 2004; FOLEY et al., 2005). In soils, the reduction of the amount of soil organic matter is accompanied by the emission of gases to the atmosphere, especially CO_2, CH_4 and N_2O, increasing the global warming (KNORR et al., 2005). For the soil sustainability, besides the reduction of amount there is also a change in the quality of the remaining organic matter.

These changes are recorded in the breakdown and disintegration of soil with losses by erosion, reducing the availability of plant nutrients and lower water storage capacity. These factors have negative effects on agricultural productivity, with reduction in food production, and sustainability of the soil-plant-atmosphere (LAL, 2003; SIX et al., 2004; KNORR et al., 2005).

A better understanding of the ecosystem functioning is of paramount importance in the pursuit for sustainable practices, and the microorganisms that inhabit the soil should be taken into consideration. Soil-borne microorganisms represent the largest biodiversity pool on Earth, with $>10^{30}$ microbial cells, and estimates of 10^{4}-10^{6} different species, per gram of soil (WHITMAN et al., 1998; TORSVIK et al., 2002; ROESCH et al., 2007). With their enormous numbers and large biomass, and demonstrated involvement in numerous key functions, soil-borne microbial communities hold a central place in terrestrial ecosystems. The soil microbial communities carry out numerous essential ecosystem functions (BARDGETT et al., 2008), including nutrient cycling, facilitating plant nutrition, disease suppression, water purification, and biological attenuation of pollutants. Nowhere are soil microbial communities likely to be more complex than under tropical rain forests, which house the majority of plant diversity on Earth (DIRZO; RAVEN, 2003).

One assumption often made is that biodiversity loss is happening more rapidly in the tropics due to agricultural activities. The process of land conversion and agricultural intensification is one of the most significant causes of biodiversity loss, with negative effects both on the environment and to the sustainability of agricultural production. Deforestation and agricultural intensification are common land-use changes in tropical regions such as Southeast Asia, Costa Rica, and the Amazon, which are major hot spots of biodiversity (ERWIN, 1983; MYERS et al., 2000). Due to anthropogenic activities and dramatic changes in land-use, these regions have the highest deforestation rate (SODHI et al., 2004; HANSEN et al., 2008), which has impacted their rich and unique biodiversity (BROOK et al., 2003; SODHI et al., 2004; HOFFMANN et al., 2010). Transformations in the environment due to land-use changes have a direct effect on the community of microorganisms inhabiting the soil. The advances in agriculture in these regions have transformed natural areas into agricultural fields, where studies on microbial ecology are needed to increase our knowledge about the effects of land-use changes over the microbial community structure, composition, and function. Despite the increased appreciation of belowground microbial diversity in the tropics, little is known about bacterial taxa responses and much less about the functional responses to alterations in

soil chemical properties and fertility in consequence of the deforestation and agricultural management of forest soils.

Despite the importance of microbial ecology studies in soils, the Amazon region received attention only in recent years. The first description of the microbial diversity in Amazon soils was made by Borneman and Triplett (1997), who used a Sanger sequencing approach to analyze sequences from forests and pastures. In this first glimpse, they showed an immense microbial diversity and differences in composition between forest and pasture. Sequences affiliated to *Clostridia* were related to forest samples while *Bacillus* were related to pasture. In a more recent study, Jesus et al. (2009) showed that changes in land-use alter the structure of bacterial communities in the Amazon region. Phylogenetic analysis revealed that the groups *Bacteroidetes*, *Actinobacteria*, *Firmicutes,* and *Proteobacteria* responded to differences related to the land-use. Also, Navarrete et al. (2010) showed that land-use change alters the structure of microbial communities in Amazon soils. Other studies in the Brazilian Amazon have also contributed to our understanding of the effects of land-use changes on the soil microbial communities (CENCIANI et al., 2009; O'NEILL et al., 2009; GROSSMAN et al., 2010; PAZINATO et al., 2010; TAKETANI; TSAI, 2010; NAVARRETE et al., 2011; 2013; GERMANO et al., 2012; TAKETANI et al., 2013).

In the last two decades, molecular tools have been allied to the classical microbiological methods to provide new insights into microbial ecology of soils. The rapid increase on molecular ecology issues has emerged as a result of the advancement of molecular biology. The molecular ecology applied to environmental microbiology should include not only diversity studies, but should address studies of functional traits related to microbial processes. Powerful molecular tools are becoming available to allow the examination of microbial communities through analysis of microbial DNA and RNA (PCR, gene probes, DNA sequencing, and metagenomics). Recombinant DNA techniques have provided a means whereby many of the obstacles associated with cultivation and description can be overcome and subsequently has allowed many new insights into the complexity of natural microbial communities. For example, molecular approaches based on 16S ribosomal RNA (rRNA) sequence analysis allow direct investigation of the community structure, diversity (richness and abundance), and phylogeny of microorganisms in almost any environment, while quantification of the individual types of microorganisms or entire microbial communities may be addressed by nucleic acid hybridization techniques. Molecular sequence analysis of communities' DNA allows a new perception of the microbial diversity and function in a broad

range of environments. The sequencing approach is a powerful tool for the study of microbial communities inhabiting soil and could be useful to predict changes in soil properties and quality.

The assessment of the microbial diversity will be advanced by the development of new technologies that answer some key questions about the "who, what, when, where, why, and how" of microbial communities (KNIGHT et al., 2012). Although new technologies are being used to better understand microbial communities in soil, there is a lack of data addressing this issue in tropical soils. The next-generation sequencing (NGS) technologies have allowed microbial ecologists to advance from analyzing a few hundred sequences to millions per study. The advance in sequencing technologies from Sanger to 454 pyrosequencing and Illumina has opened new possibilities in microbial community analysis by making it possible to collect millions of sequences, spanning hundreds of samples. The increase in the number of sequences per run from parallel pyrosequencing technologies such as the Roche 454 GS FLX (5×10^5) to Illumina GAIIx (1×10^8) is of the order of 1,000-fold and greater than the increase in the number of sequences per run from Sanger (1×10^3 through 1×10^4) to 454 (CAPORASO et al., 2011). In addition, the use of barcode strategies allows the analysis of thousands of samples in a single run. With the advance of such technologies the read length has increased, although they are far shorter than the desirable length or the read length obtained from traditional Sanger sequencing (~1000 bp) (LUO et al., 2012). The 454 pyrosequencing was the first next-generation sequencing technology available as a commercial product (MARGULIES et al., 2005) and can be considered the cornerstone of the sequencing revolution. The development of the pyrosequencing method allowed an advance of metagenome studies by increasing the number of reads and decreasing costs per sequence, enabling a deep phylogenetic community analysis.

In recent years, the use of metagenomics in the studies of soil microbial communities has enabled researchers to have an overview not only of the diversity, but also the functional traits, which are also an important approach to defining microbiological parameters for monitoring soil quality in land-use systems. The analysis of functional diversity can also provide information on how adaptive microorganisms may influence the fertility of soils. The rapid advance of sequence technologies allied to bioinformatics tools are increasing the possibility of massive studies on microbial ecology for a deep comprehension of the composition and function that soil microorganisms play in a wide range of ecosystems. The

new information available will be useful for a better understanding of soil quality and improve the sustainable use of the ecosystem.

Considering the importance of microbial ecology studies in tropical soils and the need of gathering new information about the structure of microbial community in Amazon region, this thesis was defined based on the use of new technologies, as 454 pyrosequencing and Illumina, to provide new insights not only about the diversity, but also about microbial functional traits. Thus, the Study 1 sought understands the response of microbial taxonomic and functional groups to the land-use change in Amazon soils. The Study 2 focused on the microbial community inhabiting the bulk soil and the rhizosphere of soybean, in order to determine the extent to which a particular plant species is able to select a rhizospheric microbial community from the bulk soil reservoir.

1.1 Hypothesis

The Study 1 presented in this thesis sought to test the hypothesis that microbial community structure in Amazon soils respond to land-use change by altering the abundance of specific taxonomic and functional groups in each soil type, as well as their correlations soil abiotic factors. The Study 2 tested the hypothesis that niche-based mechanisms explain the microbial community assembly in the rhizosphere through a selective power of the plant in the rhizospheric environment.

1.2 Objectives

1.2.1 General objective

The general objective of this thesis was to assess the effects of land-use change in Amazon region on the microbial community structure and taxonomic and functional composition. In addition, assess the power of soybean in selecting a specific microbial community inhabiting the rhizosphere.

To achieve the objectives were used shotgun metagenomics with next-generation sequencing technologies, such as Roche 454 GS FLX Titanium and Illumina HiSeq 2000 platforms. Statistical and bioinformatics tools were used to integrate the data for ecological interpretations.

1.2.2 Specific objectives

To achieve the general objective of this thesis the following specific objectives were considered:

- To determine how the community composition and the functional capabilities varies across different land-use systems and across time, by applying a DNA shotgun metagenomics approach to analyze samples collected in native forest, deforested site, agriculture (soybean) and pasture.
- To understand the process of community selection and assembly in the soybean rhizosphere by applying a DNA shotgun metagenomic approach to analyze the microbial community inhabiting the bulk soil and the rhizosphere in agricultural fields with 1-year and 5-year of soybean cultivation.

1.3 Structure of the thesis

This thesis comprises an introductory initial text followed by two studies presented in scientific manuscript format written in English language. The supplementary materials indicated in each chapter are available in the Appendix section.

1.4 INTRODUÇÃO

A floresta Amazônica abriga 40% da área de florestas tropicas remanescentes, desempenhando um papel fundamental na conservação da biodiversidade, nos ciclos biogeoquímicos e na regulação do clima. Apesar de sua importância global, o "arco do desmatamento" da Amazônia tem sido a fronteira de desmatamento mais ativa do mundo nas últimas décadas. Em geral, o desmatamento na Amazônia brasileira pode ser definido como o corte e a conversão da cobertura florestal nativa para outros usos da terra, principalmente a agricultura e pastagem. De acordo com o Instituto Nacional de Pesquisas Espaciais (INPE, 2011), as mudanças do uso da terra resultou em perda da área florestal anual de 18.918 Km^2 de 1998 a 2007, sendo a fronteira dos estados de Mato Grosso, Rondônia e Pará responsável por 85% de todo o desmatamento. Esta região inclui vários tipos de usos da terra, desde a extração realizada na vegetação nativa até o cultivo com diferentes níveis de tecnologia para agricultura e pecuária. Na agricultura destacam-se o cultivo de soja, algodão, milho e arroz, sendo cultivados em forma de plantio direto ou convencional. Na pecuária existem áreas de pastagens bem manejadas e outras mais degradadas. Nas áreas de pastagem é predominante o uso de *Brachiaria brizantha*, enquanto na agricultura é comum o uso da cultura da soja (*Glicine max* L. Merril), precedido pela sucessão de milho (Zea *mays* L.).

A crescente demanda por produção agrícola e pecuária leva à abertura de novas áreas de floresta nativa. No entanto, quando práticas de conservação não são adotadas, a utilização intensiva do solo afeta negativamente o ambiente e a produtividade agrícola (CERRI et al., 2004; FOLEY et al., 2005). A redução da quantidade de matéria orgânica do solo é acompanhada pela emissão de gases para a atmosfera, especialmente CO_2, CH_4 e N_2O, aumentando o potencial de aquecimento global (KNORR et al., 2005). Além da redução da quantidade de matéria orgânica também existe uma alteração na qualidade restante. Essas mudanças são notadas na quebra e desintegração do solo com perdas por erosão, reduzindo a disponibilidade de nutrientes para as plantas e diminuindo a capacidade de armazenamento de água. Esses fatores têm efeitos negativos sobre a produtividade agrícola, com redução na produção de alimentos e na sustentabilidade do sistema solo-planta-atmosfera (LAL, 2003; SIX et al., 2004; KNORR et al., 2005).

Uma melhor compreensão do funcionamento do ecossistema é importante para a busca de novas práticas sustentáveis, e, neste contexto, os micro-organismos que habitam o solo devem ser levados em consideração. Os micro-organismos do solo representam a maior reserva de biodiversidade da Terra, com aproximadamente $>10^{30}$ células microbianas, sendo estimado 10^4 a 10^6 espécies diferentes por grama de solo (WHITMAN et al., 1998; TORSVIK et al., 2002; ROESCH et al., 2007). Com essa enorme quantidade de biomassa e seu envolvimento demonstrado em inúmeras funções importantes no ambiente, as comunidades microbianas do solo ocupam um lugar central nos ecossistemas terrestres. Eles realizam inúmeras funções essenciais do ecossistema (BARDGETT et al., 2008), incluindo a ciclagem de nutrientes, facilitando a nutrição de plantas, a supressão de doenças, purificação da água e atenuação biológica de poluentes. Em nenhum lugar as comunidades microbianas do solo tendem a ser mais complexas do que em solos de florestas tropicais, as quais abrigam a maior parte da diversidade vegetal na Terra (DIRZO; RAVEN, 2003).

Uma suposição feita frequentemente é que a perda da biodiversidade está acontecendo mais rapidamente em regiões tropicais devido às atividades agrícolas. O processo de conversão de terras e intensificação agrícola é uma das mais importantes causas de perda de biodiversidade, com efeitos negativos tanto sobre o meio ambiente quanto para a sustentabilidade da produção agrícola. O desmatamento e a intensificação agrícola são as mudanças do uso da terra mais comuns em regiões tropicais, como o Sudeste Asiático, Costa Rica e Amazônia, que são os principais *hotspots* de biodiversidade (ERWIN, 1983; MYERS et al., 2000). Devido às atividades antrópicas e as mudanças dramáticas no uso da terra, estas regiões têm a maior taxa de desmatamento (SODHI et al., 2004; HANSEN et al., 2008), o que tem impactado a sua biodiversidade rica e única (BROOK et al., 2003; SODHI et al., 2004; HOFFMANN et al., 2010). Transformações no ambiente devido às mudanças do uso da terra têm um efeito direto sobre a comunidade de micro-organismos que habitam o solo. Os avanços da agricultura nessas regiões têm transformado áreas naturais em áreas agrícolas, onde são necessários estudos em ecologia microbiana para aumentar nosso conhecimento sobre os efeitos dessas mudanças do uso do solo sobre a estrutura, composição e função das comunidades microbianas. Apesar do aumento da valorização da diversidade microbiana dos solos nas regiões tropicais, pouco ainda se sabe sobre a resposta de alguns táxons, e muito menos sobre aspectos funcionais, para as alterações nas propriedades químicas e fertilidade do solo em consequência do desmatamento e manejo agrícola dos solos de floresta.

Apesar da importância dos estudos em ecologia microbiana em solos, a Amazônia recebeu atenção apenas nos últimos anos. A primeira descrição da diversidade microbiana em solos foi feita por Borneman e Triplett (1997), que utilizaram uma abordagem de sequenciamento Sanger para analisar sequências de florestas e pastagens. Neste primeiro vislumbre eles mostraram uma imensa diversidade microbiana, com diferenças na composição entre a floresta e a pastagem. Sequências afiliadas à *Clostridium* foram relacionadas com amostras de floresta enquanto *Bacillus* foram relacionados à pastagem. Jesus et al. (2009) demonstraram que as mudanças do uso do solo alteram a estrutura de comunidades microbianas em solos Amazônicos. A análise filogenética revelou que os grupos *Bacteroidetes*, *Actinobacteria*, *Firmicutes* e *Proteobacteria* responderam às diferenças relacionadas ao uso da terra. Além disso, Navarrete et al. (2010) também demonstraram que mudanças no uso da terra alteram a estrutura das comunidade microbianas nos solos da Amazônia. Outros estudos realizados na Amazônia brasileira também contribuíram para uma melhor compreensão dos efeitos das mudanças do uso da terra sobre as comunidades microbianas do solo (CENCIANI et al., 2009; O'NEILL et al., 2009; GROSSMAN et al., 2010; PAZINATO et al., 2010; TAKETANI; TSAI, 2010; NAVARRETE et al., 2011; 2013; GERMANO et al., 2012; TAKETANI et al., 2013).

Nas últimas duas décadas, ferramentas moleculares têm se aliado aos métodos de microbiologia clássicos para fornecer novos insights sobre a ecologia microbiana dos solos. O rápido aumento nas questões de ecologia molecular emergiu como resultado do avanço da biologia molecular. A ecologia molecular aplicada à microbiologia ambiental deve incluir não só estudos sobre a diversidade, mas deve também abordar aspectos funcionais relacionados aos processos microbianos. Ferramentas moleculares poderosas estão se tornando cada vez mais disponíveis para permitir o estudo de comunidades microbianas através da análise de DNA e RNA (PCR, sondas genéticas, sequenciamento de DNA e metagenômica). Técnicas de DNA recombinante têm proporcionado um meio pelo qual muitos dos obstáculos associados ao cultivo podem ser superados e, consequentemente, têm permitido novas descobertas sobre a complexidade das comunidades microbianas naturais. Por exemplo, as abordagens moleculares baseadas na sequência do gene ribossomal 16S rRNA permitem a investigação direta da estrutura, diversidade (riqueza e abundância) e filogenia de micro-organismos em quase qualquer ambiente. A análise molecular de sequências de DNA permite uma nova percepção da diversidade taxonômica e funcional microbiana em uma ampla variedade de ambientes. O sequenciamento é uma ferramenta poderosa para o estudo de comunidades

microbianas que habitam o solo e pode ser útil para prever mudanças nas propriedades e qualidade do solo.

Os estudos da diversidade microbiana avançará com o desenvolvimento de novas tecnologias que respondam questões importantes sobre "quem, o quê, quando, onde, porquê e como" de comunidades microbianas (KNIGHT et al., 2012). Embora as novas tecnologias estejam sendo aplicadas para uma melhor compreensão das comunidades microbianas do solo, há uma notável falta de dados sobre os solos tropicais. As tecnologias de próxima geração de sequenciamento (NGS, do inglês *Next-Generation Sequencing*) têm permitido ecólogos microbianos avançar a partir da análise de algumas sequências para milhões por estudo. O avanço nas tecnologias de sequenciamento de Sanger para 454 e Illumina abriu novas possibilidades na análise da comunidade microbiana, tornando possível a coleta de milhões de sequências, incluindo centenas de amostras. O aumento no número de sequências por corrida de piroseqüenciamento Roche 454 GS FLX ($5x10^5$) para Illumina GAIIx ($1x10^8$) é da ordem de 1000 vezes a mais do que o aumento do número de sequências por corrida de Sanger ($1x10^3$ a $1x10^4$) para o 454 (CAPORASO et al., 2011). Além disso, a utilização das estratégias de *barcode* permite a análise de milhares de amostras num único ensaio. O avanço destas tecnologias também tem aumentado o comprimento da sequência, apesar de ainda serem muito mais curto do que o tamanho desejado ou até mesmo o obtido pela técnica de Sanger tradicional (~1000 pb) (LUO et al., 2012). O pirosequenciamento 454 foi a primeira tecnologia de sequenciamento de nova geração disponível como um produto comercial (MARGULIES et al., 2005) e pode ser considerado a pedra angular da revolução do sequenciamento. O desenvolvimento do método de pirosequenciamento permitiu um grande avanço nos estudos metagenômicos, por aumentar o número de sequências e diminuir os custos por corrida, permitindo maior profundidade nas análises filogenéticas das comunidades.

Nos últimos anos, a utilização da metagenômica nos estudos de comunidades microbianas do solo permitiu aos pesquisadores ter não só uma visão geral da biodiversidade, mas também sobre as características funcionais, as quais constituem uma abordagem importante para a definição de parâmetros biológicos de monitoramento da qualidade do solo em sistemas de uso da terra. A análise da diversidade funcional também pode fornecer informações sobre como micro-organismos adaptativos podem influenciar a fertilidade dos solos. O rápido avanço dessas tecnologias, aliadas a ferramentas de bioinformática, estão aumentando as possibilidades de estudos maciços em ecologia microbiana para uma

compreensão profunda da composição e função que os micro-organismos desempenham nos solos de uma grande variedade de ecossistemas. As novas informações disponíveis serão úteis para uma melhor compreensão da qualidade do solo e para um uso mais sustentável do ecossistema.

Considerando a importância dos estudos de ecologia microbiana em solos tropicais, e a crescente necessidade de reunir novas informações sobre a estrutura das comunidades microbianas na região Amazônica, esta tese foi definida com base na utilização das novas tecnologias de sequenciamento, como as plataformas Roche GS FLX Titanium 454 e Illumina Hiseq 2000, para fornecer novos insights não apenas sobre a diversidade, mas também sobre os aspectos funcionais microbianos. Assim, o Estudo 1 procurou entender a resposta taxonômica e funcional dos micro-organismos face às mudanças do uso da terra em solos da Amazônia. O Estudo 2 foi voltado para a comunidade microbiana que habita o solo e a rizosfera de soja, de modo a determinar até que ponto uma espécie vegetal específica é capaz de selecionar uma comunidade microbiana rizosférica e quais as implicações deste fato.

1.5 Hipóteses

O Estudo 1 apresentado nesta tese buscou testar a hipótese de que a estrutura de comunidades microbianas em solos Amazônicos responde às mudanças do uso do solo, alterando a abundância de grupos taxonômicos e funcionais específicos em cada tipo de solo, como também altera as suas relações com fatores abióticos do solo. O Estudo 2 testou a hipótese de que mecanismos baseados em teoria de nicho explicam a montagem da comunidade microbiana por meio de um poder de seleção da planta no ambiente de rizosfera.

1.6 Objetivos

1.6.1 Objetivos geral

O objetivo geral desta tese foi acessar os efeitos da mudança do uso do solo na região Amazônica sobre a estrutura taxonômica e funcional das comunidades microbianas. Também,

acessar o poder da planta de soja em selecionar uma comunidade especifica habitante de sua rizosfera.

Para atingir o objetivo foi realizado sequenciamento metagenômico por meio de plataformas de nova geração, como Roche 454 GS FLX Titanium e Illumina HiSeq 2000. Ferramentas de estatística e bioinformática foram utilizadas para integrar os dados para uma interpretação ecológica.

1.6.2 Objetivos específicos

Para alcançar os objetivo geral desta tese, os seguintes objetivos específicos foram considerados:

- Determinar como a composição e as capacidades funcionais das comunidades variam entre diferentes sistemas de uso da terra e durante o tempo, por meio da realização de sequenciamento metagenômico de DNA para analisar amostras coletadas em áreas de floresta nativa, área desmatada, agricultura (soja) e pastagem.
- Entender os processos de seleção e montagem de comunidades na rizosfera de soja, por meio do sequenciamento metagenômico de DNA para analisar a comunidade habitante do solo e da rizosfera em campos de plantio com 1 e 5 anos de cultivo.

1.7 Estrutura da Tese

Esta tese compreende um texto inicial introdutório seguido por dois estudos escritos em Inglês, no formato de manuscritos científicos. Os materiais suplementares indicados em cada capitulo estão disponíveis na seção Apêndice.

REFERENCES

BARDGETT, R.D.; FREEMAN, C.; OSTLE, N.J. Microbial contributions to climate change through carbon cycle feedbacks. **The ISME Journal**, London, v. 2, p. 805-814, 2008.

BORNEMAN, J.; TRIPLETT, E.W. Molecular microbial diversity in soils from Eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. **Applied Environmental Microbiology**, Baltimore, v. 63, p. 2647-2653, 1997.

BROOK, B.W.; SODHI, N.S.; NG, P.K.L. Catastrophic extinctions follow deforestation in Singapore. **Nature**, London, v. 424, p. 420-423, 2003.

CAPORASO, J.G.; LAUBER, C.L.; WALTERS, W.A.; BERG-LYONS, D.; LOZUPONE, C.A.; TUMBAUGH, P.J.; FIERER, N.; KNIGHT, R. **Proceedings of the National Academy of Science of the USA,** Washington, DC, v. 108, p. 4516-4522, 2011.

CENCIANI, K.; LAMBAIS, M.R.; CERRI, C.C.; DE AZEVEDO, L.C.B.; FEIGL, B.J. Bacteria diversity and microbial biomass in forest, pasture and fallow soils in the southwestern Amazon Basin. **Revista Brasileira de Ciência do Solo**, Campinas, v. 33, p. 907-916, 2009.

CERRI, C.C.; BERNOUX, M; CERRI, C.E.P; FELLER, C. Carbon cycling and sequestration opportunities in South America: the case of Brazil. **Soil Use and Management**, New York, v. 20, p. 248-254, 2004.

DIRZO, R.; RAVEN, P.H. Global state of biodiversity and loss. **Annual Review of Environment and Resources**, Palo Alto, v. 28, p. 137-167, 2003.

ERWIN, T.L. Beetles and other arthropods of the tropical forest canopies at Manaus, Brazil, sampled with insecticidal fogging techniques. In: SUTTON S.L.; WHITMORE, T.C.; CHADWICK, A.C. (Ed.). **Tropical rain forests**: ecology and management. Oxford: Blackwell Scientific Publications, 1983. p. 59–75.

FOLEY, J.A.; DEFRIES, R.; ASNER, G.P.; BARFORD, C.; BONAN, G.; CARPENTER, S.R.; CHAPIN, F.S.; COE, M.T.; DAILY, G.; GIBBS, H.K.; HELKOWSKI, J.H.; HOLLOWAY, T.; HOWARD, E.A.; KUCHARIK, C.J.; MONFREDA, C.; PATZ, J.A.; PRENTICE, I.C.; RAMANKUTTY, N.; SNYDER, P.K. Global Consequences of land use. **Science**, Washington, DC, v. 309, p. 570-574, 2005.

GERMANO, M.G.; CANNAVAN, F.S.; MENDES, L.W.; LIMA, A.B.; TEIXEIRA, W.G.; PELLIZARI, V.H.; TSAI, S.M. Functional diversity of bacterial genes associated with aromatic hydrocarbon degradation in anthropogenic dark earth of Amazonia. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 47, p. 654-664, 2012.

GROSSMAN, J.M.; SHEAFFER, C.; WYSE, D.; BUCCIARELLI, B.; VANCE, C.; GRAHA, P.H. An assessment of nodulation and nitrogen fixation in inoculated *Inga oestediana*, a nitrogen-fixing tree shading organically grown coffee in Chiapas, Mexico. **Soil Biology and Biochemistry**, Oxford, v. 38, n. 4, p. 769-784, 2006.

JESUS, E.C.; MARSH, T.L.; TIEDJE, J.M.; MOREIRA, F.M.S. Changes in land use alter the structure of bacterial communities in Western Amazon soils. **The ISME Journal**, London, v. 3, p. 1004-1011, 2009.

HANSEN, M.C.; STEHMAN, S.V.; POTAPOV, P.V.; LOVELAND, T.R.; TOWNSHEND, J.R.G.; DEFRIES, R.S.; PITTMAN, K.W.; ARUNARWATI, B.; STOLLE, F.; STEININGER, M.K.; CARROLL, M.; DIMICELI, C. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 105, p. 9439-9444, 2008.

HOFFMANN, M. et al. The impact of conservation on the status of the world's vertebrates. **Science**, Washington, DC, v. 330, p. 1503-1509, 2010.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Projeto Prodes:** Monitoramento da Floresta Amazônica Brasileira por Satélite Prodes. São José dos Campos, 2011. Disponível em: http://www.obt.inpe.br/prodes/index.html.

KNIGHT, R.; JANSSON, J.; FIELD, D.; FIERER, N.; DESAI, N.; FUHRMAN, J.A.; HUGENHOLTZ, P.; VAN DER LELIE, D.; MEYER, F.; STEVENS, R.; BAILEY, M.J.; GORDON, J.I.; KOWALCHUCK, G.A.; GILBERT, J.A. Unlocking the potential of metagenomics through replicated experimental design. **Nature Biotechnology**, London, v. 30, p. 513-520, 2012.

KNORR, W.; PRENTICE, I.C.; HOUSE, J.I.; HOLLAND, E.A. Long-term sensitivity of soil carbon turnover to warming. **Nature**, London, v. 433, p. 298-301, 2005.

LAL, R. Global potential of soil carbon sequestration to mitigate the greenhouse effect. **Critical Reviews in Plant Sciences,** Boca Raton, v. 22, p. 151-184, 2003.

LUO, C.; TSEMENTZI, D.; KYRPIDES, N.; READ, T.; KONSTANTINIDIS, K.T. Direct comparison of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. **PLoS One**, New York, v. 7, e30087, 2012.

MARGULIES, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, London, v. 437, p. 376-380, 2005.

MEYER, F.; PAARMAN, D.; D'SOUZA, M.; OLSON, R.; GLASS, E.M.; KUBAL, M.; PACZIAN, T.; RODRIGUEZ, A.; STEVENS, R.; WILKE, A.; WILKENING, J.; EDWARDS, R.A. The Metagenomics RAST server – A public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, London, v. 9, p. 386-393, 2008.

MYERS, N.; MITTERMEIER, R.A.; MITTERMEIER, C.G.; DA FONSECA, G.A.B.; KENT, J. Biodiversity hotspots for conservation priorities. **Nature**, London, v. 403, p. 853-858, 2000.

NAVARRETE, A.A.; CANNAVAN, F.S.; TAKETANI, R.G.; TSAI, S.M. A molecular survey of the diversity of microbial communities in different Amazonian agricultural model systems. **Diversity**, Bethesda, v. 2, p. 787-809, 2010.

NAVARRETE, A.A.; TAKETANI, R.G.; MENDES, L.W.; CANNAVAN, F.S.; MOREIRA, F.M.S.; TSAI, S.M. Land-use systems affect archaeal community structure and functional diversity in Western Amazon soils. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 35, p. 1527-1540, 2011.

NAVARRETE, A.A.; KURAMAE, E.E; DE HOLLANDER, M.; PIJL, A.S.; VAN VEEN, J.A.; TSAI, S.M. Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. **FEMS Microbiology Ecology**, Amsterdam, v. 83, p. 607-621, 2013.

O'NEILL, B.; GROSSMAN, J.; TSAI, M.T.; GOMES, J.E.; LEHMANN, J.; PETERSON, J.; NEVES, E.; THIES, J.E. Bacterial community composition in Brazilian Anthrosols and adjacent soils characterized using culture and molecular identification. **Microbial Ecology**, Heidelberg, v. 58, p. 23-35, 2009.

PAZINATO, J.M.; MENDES, L.W.; PAULO, E.N.; VAZOLLER, R.F.; TSAI, S.M. Molecular characterization of the archaeal community in an Amazonian Wetland Soil and culture-dependent isolation of methanogenic Archaea. **Diversity**, Bethesda, v. 2, p. 1026-1047, 2010.

ROESCH, L.F.W.; CAMARGO, F.A.O.; BENTO, F.M.; TRIPLETT, E.W. Biodiversity of diazotrophic bacteria within the soil, root and stem of field-grown maize. **Plant and Soil**, Dordrecht, v. 302, p. 91-104, 2007.

SIX, J.; OGLE, S.M.; BREIDT, F.J.; CONANT, R.T.; MOSIER, A.R.; PAUSTIAN, K. The potential to mitigate global warming with no-tillage management is only realized when practised in the long term. **Global Change Biology**, Malden, v. 10, p. 155-160, 2004.

SODHI, N.S.; KOH, L.P.; BROOK, B.W.; NG, P.K.L. Southeast Asian biodiversity: an impending disaster. **Trends in Ecology and Evolution**, Cambridge, v. 19, p. 654-660, 2004.

TAKETANI, R.G.; TSAI, S.M. The influence of different land uses on the structure of Archaeal communities in Amazonian Anthrosols based on 16S rRNA and *amo*A genes. **Microbial Ecology**, Heidelberg, v. 59, p. 734-743, 2010.

TAKETANI, R.G.; LIMA, A.B.; JESUS, E.C.; TEIXEIRA, W.G.; TIEDJE, J.M.; TSAI, S.M. Bacterial community composition of anthropogenic biochar and Amazonian anthrosols assessed by 16S rRNA gene 454 pyrosequencing. **Antonie van Leeuwenhoek**, Dordrecht, v. 104, n. 2, p. 233-242, 2013.

TORSVIK, V.; OVREAS, L.; THINGSTAD, T.F. Prokaryotic diversity - Magnitude, dynamics, and controlling factors. **Science**, Washington, DC, v. 296, p. 1064-1066, 2002.

WHITMAN, W.B.; COLEMAN, D.C.; WIEBE, W.J. Prokaryotes: the unseen majority. **Proceedings of the National Academy of Science of the USA**, Washington, DC, v. 95, p. 6578- 6583, 1998.

2

CONSEQUENCES OF AMAZON RAINFOREST CONVERSION: METAGENOMIC ANALYSIS OF SOIL-BORNE MICROBIAL COMMUNITY AND THEIR FUNCTIONAL ATTRIBUTES

Abstract

This study focused on the impact of the land-use change on the structure of soil-borne microbial communities in Amazon region, where areas of native forest are converted to agriculture field and cattle pasture. Here, we used shotgun metagenomics approach to investigated the taxonomic and potential functional structures, composition and diversity of microbial communities in soils from different land-uses types in Southeastern Amazon, *i.e.* native forest, deforested area, agriculture and pasture. As a response to the land-use change the soil microbial communities presented different structure among sites, differences that were larger in magnitude than temporal variability. Major differences were observed in community composition as shown by the differential distribution of *Proteobacteria*, *Acidobacteria*, *Verrucomicrobia*, *Actinobacteria*, *Chloroflexi*, *Bacteroidetes*, *Firmicutes* and *Planctomycetes*. Genes related to nutrient cycle, respiration and resistance to stress were found abundantly in forest. The altered soil presented high abundance of genes related to metabolism of DNA, RNA and Protein. Agriculture soils and pasture were among the most diverse taxonomically and functionally. In forest, the ecosystem function is maintained by the microbial abundance, while in agriculture and pasture is maintained by high diversity and functional redundancy. Taken together, our findings clearly show that land-use change in Amazon forest affects the soil-borne microbial communities in both composition and functional attributes.

Keywords: Soil microbial ecology. Tropical rainforest. Land-use change. Shotgun metagenomics.

2.1 Introduction

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

(**Leonardo da Vinci** | 1452 – 1519)

The Amazon rainforest is the world's largest reservoir of plant and animal diversity, and it has been subjected to especially high rates of land-use conversion (RODRIGUES et al., 2013). Areas of native forest are cleared through slash-and-burn prior to being converted to agricultural field and cattle pasture. The process of land conversion in the tropics is one of the most significant causes of biodiversity loss, with negative effects both on the environment and on the sustainability of agricultural production. The anthropogenic activities can affect both the biodiversity and functionality of soil microorganisms, resulting in a reduction of microbial functions and species loss (FAO, 2012). The knowledge regarding the impact of forest conversion to the soil-borne microbial communities is of paramount importance because microorganisms represent the largest biodiversity pool on Earth and hold a central place in terrestrial ecology, playing numerous biogeochemical functions to the ecosystem (BARDGETT et al., 2008).

The size and structure of bacterial communities in soils is shaped by biotic and abiotic factors, such as soil physical and chemical properties (FIERER et al., 2012; KURAMAE et al., 2012; NAVARRETE et al., 2013). Several studies have been shown that the conversion of Amazon rainforest affects the microbial communities and functional diversity by altering the structure and composition in soils (BORNEMAN; TRIPLETT, 1997; JESUS et al., 2009; TAKETANI; TSAI, 2010; NAVARRETE et al., 2010; 2011; GERMANO et al., 2012; RODRIGUES et al., 2013). However, despite the increasing appreciation of the belowground microbial community in Amazon soils, there is a lack of information about the responses of the microorganisms to the deforestation and subsequent land-use conversion, for both taxonomic and functional microbial structure and composition.

The recent advance in the next-generation DNA sequencing methods, such as pyrosequencing (MARGULIES et al., 2005) and shotgun metagenome (VENTER et al., 2004), boosts scientific interests to understanding the complexity of microbial community in a wide range of environments. Focuses are not only addressed on the composition, but also on

functional traits and the relationship between community and external drivers including environmental factors (LANGENHEDER et al., 2010). These approaches, together with bioinformatics tools are increasing the possibility of massive studies on microbial ecology for a deep comprehension of the responses to environmental gradients, such as land-uses changes.

The current study focus understanding the consequences of the Amazon rainforest conversion over the taxonomic and functional composition and structure of the soil-borne microbial community. We used shotgun metagenomics sequencing from four sites representing the most common land-use types in the Amazon region, *i.e.* native forest, deforested site, agriculture and pasture. The samples were collected at the beginning and end of rainy season along two years, in order to determine how the community composition and the functional capabilities vary across the land-use and time. Here, we hypothesize that shifts in the structure of microbial communities might be related to the conversion of the forest to land-use systems. We addressed three basic questions: (i) how the diversity is altered across different land-uses? (ii) what are the dominant groups in each type of soil? (iii) what are the cores of functions prevalent in each soil?

2.2 Material and Methods

2.2.1 Sampling sites

Bulk soil samples were collected at four different sites located in the Southeastern Brazilian Amazon, in the state of Mato Grosso, Brazil, in the Ipiranga do Norte municipality, as follow: native forest, deforested site, agriculture and pasture (Table 2.1). Oxisol is the predominant soil order in the sampling sites (MATO GROSSO, 2001), and the climate in the region is classified as Am (Koppen's classification), with annual average temperature of 28ºC and average precipitation of 2000 mm.

The sampling areas were selected according to the vegetation cover and soil use. All the sites were previously covered with native tropical rainforest, being cleared through slash-and-burn and subsequently converted into agricultural field or pasture. At each sampling site, the soil samples were collected from five points in four different sampling periods: October 2009, April 2010, November 2010 and March 2011, comprising the beginning and end of the

rainy season. However, the sampling in the deforested site was performed only in two periods (October 2009 and April 2010), because deforested sites do not remain unused for a long time. In the subsequent sampling (November 2010), this area had been used for agricultural purposes. The samples were collected in one central sampling point and other four points directed toward the north, south, east, and west of the central point (at least 50 m apart from the central point). Soil samples were taken from the 0- to 20-cm topsoil layer. First, the litter layer was removed, and then, the soil samples were collected. A total of 80 bulk soil samples were collected in field (4 sites x 5 sampling points per site x 4 sampling periods). Samples were transported to the laboratory under ice and stored at -20ºC until processing.

Table 2.1 - Sampling areas description

Site	Latitude	Longitude	Land-use time	Dominant plant species
Forest	-11º40'54.97"	-55º50'8.79"		
Deforested	-11º41'0.17"	-55º50'3.55"	1 year	no plants
Agriculture	-11º41'5.08"	-55º50'13.54"	5 years	*Glicine max*
Pasture	-11º43'1.94"	-55º47'41.48"	>10 years	*Brachiaria brizanta*

2.2.2 Soil physicochemical parameters

Soil chemical and physical properties were determined for each sample based on 400 g of soil, performed at the Laboratory of Soil Analysis at "Luiz de Queiroz" College of Agriculture (ESALQ/USP, Piracicaba, Brazil). Soil pH was measured in a 1:2.5 soil/water suspension. Exchangeable Al, Ca, and Mg were extracted with KCL 1 M. Calcium and Mg were determined by atomic absorption spectrometry and Al by acid-base titration. Phosphorus and K were extracted by ion-exchange resin. Potential acidity (H+Al) was estimated by an equation based on the pH determined in SMP buffer solution (pH SMP). Available micronutrients (Fe, Mn, Zn, and Cu) were extracted by Mehlich 1 and determined by atomic absorption spectrometry. Boron was extracted with hot water and determined by spectrophotometry with azomethine-H at 420 nm. Some of the results allowed the calculation of other parameters such as exchangeable bases (SB), the sum of Ca, Mg, and K; cation exchange capacity (CEC), the sum of Ca, Mg, K, Al, and H; base saturation (V), the percentage relation between SB and CEC; and Al saturation (m%), the percentage relation between exchangeable Al and CEC. Soil texture was determined using Bouyoucos densimeter after shaking the soil vigorously with NaOH 1 M as dispersant. Total nitrogen was determined by Kieldahl method; NH^{4+} and NO^{3-} by Raney/Kieldahl.

2.2.3 DNA extraction and sequencing

DNA extraction from 250 mg of soil samples was carried out using PowerSoil DNA Isolation Kit (Mobio Laboratories, Carlsbad, CA, USA), according to manufacturer's protocol. DNA quality and concentration were measured by 1% TSB (BRODY; KERN, 2004) agarose gel electrophoresis and NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, USA). From the five samples collected per site, two were randomly mixed for two replicates per site in each sampling period. In total, 28 DNA samples were sequenced on an Illumina HiSeq2000 platform (Illumina, Inc.) at Macrogen Inc. Company, South Korea.

2.2.4 Shotgun metagenomic analysis

Unassembled DNA sequences were annotated with the Metagenomics Rapid Annotation (MG-RAST) pipeline version 3.3.3.3 (MEYER et al., 2008) for downstream analyses. Taxonomic and functional profiles were generated using the normalized abundance of sequences matches to the M5NR and SEED database, respectively. A table of the frequency of hits to each individual taxon (taxonomy) or subsystem (function) for each metagenome was generated and normalized by dividing by the total number of hits to remove bias indifference in read lengths and sequencing efforts. To identify hits, BlastX was used with a minimum alignment length of 15 bp and an *E*-value cut-off of $E<1x10^{-2}$ (FIERER et al., 2012). The data matrices generated were used for statistical analyses.

2.2.5 Statistical analyses

Redundancy Analysis (RDA) was used to determine the correlation between community structure and soil physicochemical properties. All matrices were initially analyzed using Detrended Correspondence analysis (DCA) to evaluate the gradient size of the specie distribution; this analysis indicated linearly distributed data (length of gradient <3), revealing that the best-fit mathematical model for the data was RDA. Forward selection (*FS*) and the Monte Carlo permutation test were applied with 1,000 random permutations to verify the significance of soil chemical properties upon a microbial community. RDA plots were

generated from the 28 samples by using Canoco 4.5 (Biometrics, Wageningen, The Netherlands). We used analysis of similarities (ANOSIM) to test whether sample categories harbored significantly different metagenomes or microbial communities. Alpha and Beta diversities were calculated from a matrix of richness using Shannon's index and Whittaker, respectively. ANOSIM was calculated with the software PRIMER (CLARK; GORLEY, 2006) and Alpha and Beta diversities with the software PAST (HAMMER et al., 2001).

To determine statistical differences among soil samples, the Statistical Analysis of Metagenomic Profiles (STAMP) software package was used (PARKS; BEIKO, 2010). For this, a table of the frequency of hits of taxa and functional subsystem for each metagenome was generated from MG-RAST and used as input. *P*-values were calculated using the two sided Fischer's Exact test (FISCHER, 1958), while confidence intervals were calculated using the Newcombe-Wilson method (NEWCOMBE, 1998) and correction was made using Benjamini-Hochberg FDR (BENJAMINI; HOCHBERG, 1995).

Network analyses were performed to better understand the correlation between the taxonomical and functional profiles and soil physicochemical parameters. For network inference, we calculated all possible Spearman's rank correlation coefficient. To filtering the data for reduced network complexity, we considered high correlations with cut-off of coefficient $r > 0.7$ and statistically significant P-value < 0.05. The nodes in the reconstructed network represent taxa (class level), functional categories (subsystem level 1) and physicochemical parameters, and edges represent high and significant correlation between nodes. The topology of the network graph was made based of a set of measures calculated, as average node connectivity, average path length, diameter, and cumulative degree distribution. Statistical analyses were carried out in the R environment (http://www.r-project.org/) and networks visualization with the interactive platform Gephi (BASTIAN et al., 2009).

2.3 Results

2.3.1 Soil physicochemical characteristics

The soil pH ranged among sites from 4.22 to 5.23, with forest and deforested site presenting lower values and agriculture the highest. The higher pH in the agricultural field is due to soil correction by liming for soybean planting. The soils contained 42-47% sand, 4-5%

silt, and 46-53% clay, except for pasture, which presented higher percentage of sand. Forest and deforested sites were different from agriculture and pasture in some micro and macronutrients (Table 2.2). Also due to soil preparation for cultivation, the agriculture area presented higher values of P, K, Ca, and Mg. Pasture soils presented low values of organic matter. The precipitation in the four collecting periods varied between 40-50 mm and the temperature 28-33 ºC (Figure 2.1). According to RDA followed by Monte Carlo analysis, base saturation ($F = 16.62$, $p = 0.002$), NO^{3-} ($F = 8.16$, $p = 0.022$), organic matter ($F = 5.50$, $p = 0.004$), H+Al (potential acidity) ($F = 4.56$, $p = 0.004$), Cu ($F = 3.60$, $p = 0.01$), pH ($F = 4.27$, $p = 0.004$), total nitrogen ($F = 2.57$, $p = 0.018$) showed a significant correlation with general community structure (Figure 2.2).

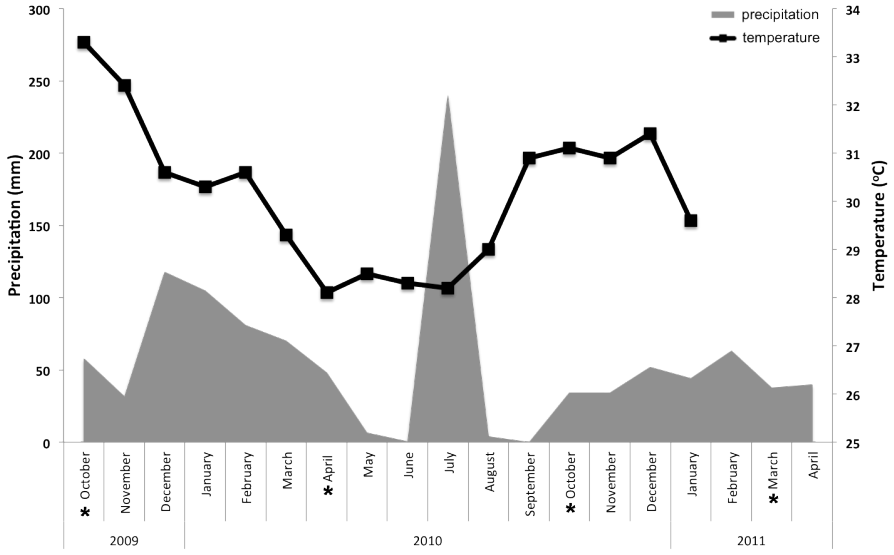


Figure 2.1 - Average temperature and precipitation in the sampling sites at Southeastern Amazon.
* indicate the time of sampling

Table 2.2 - Average soil chemical factors of soil sampling sites

	Forest	*Deforested*	*Agriculture*	*Pasture*
pH	4.53±0.14 a	4.22±0.37 a	5.23±0.20 c	4.83±0.19 b
OM	36.3±5.18 a	35.5±8.63 a	35.4±6.03 a	17.1±3.21 b
P	2.80±2.08 a	1.60±0.85 a	17.8±10.79 b	2.00±0.78 a
K	0.55±0.20 a	0.59±0.33 a	2.20±1.34 b	0.58±0.28 a
Ca	1.40±0.67 a	1.30±0.14 a	16.2±5.83 b	2.35±1.81 a
Mg	1.10±0.20 a	1.20±0.28 a	10.3±3.80 b	1.70±1.01 a
Al	15.0±3.24 a	14.7±4.38 a	2.55±0.97 b	7.85±0.10 c
H+Al	69.7±9.05 a	58.3±12.8 a	51.3±15.7 a	32.0±6.02 b
SB	2.94±0.57 a	2.85±0.13 a	28.9±10.8 b	4.56±2.93 a
CEC	72.6±8.65 a	61.1±13.0 a	80.2±25.9 a	36.6±8.51 b
V	4.15±1.18 a	4.6±0.85 a	36.2±3.81 c	12.0±4.31 b
m	83.6±4.22 a	83.6±3.68 a	9.05±4.40 c	67.1±9.00 b
B	0.19±0.02 a	0.19±0.02 a	0.17±0.03 a	0.16±0.02 a
Cu	0.17±0.05 a	0.20±0.00 a	0.30±0.17 ab	0.35±0.06 ab
Fe	119±24.5 a	111±44.4 a	60.6±8.00 b	114±23.8 a
Mn	1.46±0.48 a	1.41±0.89 a	0.85±0.35 a	1.00±0.24 a
Zn	0.18±0.08 a	0.14±0.03 a	1.52±0.62 b	0.24±0.15 a
Sand	47.9±1.50 a	44.4±3.19 a	42.5±3.07 a	79.1±1.17 b
Silt	5.30±2.15 a	4.69±3.52 a	4.43±2.45 a	2.40±0.68 a
Clay	46.7±1.07 a	50.8±6.71 a	53.0±5.39 a	18.4±1.22 b
N	1760±254 a	1719±158 a	1642±213 a	1193±293 a
$NH4^+$	17.7±2.39 a	17.0±3.00 a	16.8±2.90 a	12.0±3.5 ab
NO^-3	12.4±4.60 a	17.0±2.90 a	12.7±4.95 a	10.4±4.05 a
C	21.7±3.00 a	24.2±3.30 a	20.2±4.13 a	9.4±2.35 b

Values are mean ±SE; n = 4. Within columns, means followed by the same letter are not significantly different ($P < 0.05$) based on Tukey HSD test. Ca, Mg, K, Al, potential acidity (H+Al), sum of bases (SB) are expressed in $nmol.kg^{-1}$; OM and C is expressed in $g.Kg^{-1}$; P is expressed in $mg.kg^{-1}$; B, Fe, Mn, Zn, K, Cu, and cation exchange capacity in pH 7 (CEC) are expressed in $mg.dm^{-3}$; Sand, silt and clay are expressed in %; N (total), NH^{4+} and NO^{3-} are expressed in $mg.kg^{-1}$. P and K – Mehlich 1 extractor. Ca, Mg, and Al – Kcl 1N. H+Al – SMP extractor. N – Kjeldahl. NH4+ and NO-3 – Raney/Kjeldahl. OM – organic matter. m – Al saturation index. V – base saturation index; H+Al = potential acidity

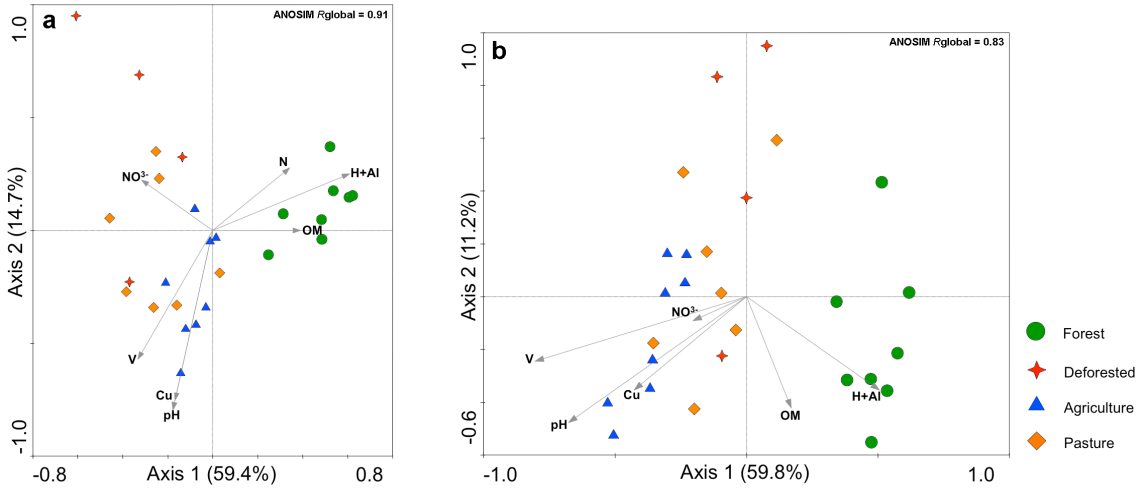


Figure 2.2 - Redundancy analysis of microbial community patterns and soil characteristics from samples of four different land use systems in the Amazon regions. **(a)** Taxonomic analysis using relative abundance based on M5NR database at class level. **(b)** Functional analysis using relative abundance based on SEED database at subsystem level 1. Analysis of similarity (ANOSIM) is indicated in the upper right of each graph. OM = organic matter; V = base saturation; H+Al = potential acidity

2.3.2 Sequencing effort

A total of 487 million of 150-bp sequences was obtained by shotgun metagenomic approach, with average of 18 millions sequences per sample. A total of 21-39% of the sequences could be annotated in MG-RAST server (Supplementary Table 2.1). Of the initial 28 samples, one failed to sequence, yielding a total of 27 samples that were used for downstream analyses. The sequences are deposited in MG-RAST server (accession number indicated in Supplementary Table 2.1).

2.3.3 Diversity comparisons across sites and time

The alpha diversity patterns were highly variable across the four sites and the agriculture site consistently harbored the most diversity of bacterial communities across all sampling times (Figure 2.3A). The temporal variability in alpha diversity within a given land-use was lower than the differences between land-use types. Furthermore, the temporal alpha diversity on the pasture was more variable than other sites. On the other hand, the forest presented less variability across the sampling time. There were significant correlations between taxonomic and functional diversities (Figure 2.3B), where soils with higher taxonomic diversity presented higher functional diversity. In contrast with alpha diversity results, the temporal beta diversity was significantly higher in forest, indicating dissimilar bacterial composition across time.

2.3.4 Structure of microbial communities across different land-uses

Soils were collected in native rainforest, deforested site, agriculture and pasture. The sites were selected to span the different land-uses found at the Southeastern Amazon region after the deforestation. The majority of shotgun metagenomic reads were derived from *Bacteria* (92% hits to M5NR database) with sequences also matching *Eukaryote* (6.5%) and *Archaea* (1.5%). The three most abundant phyla were *Proteobacteria*, *Actinobacteria* and *Acidobacteria* (Figure 2.4).

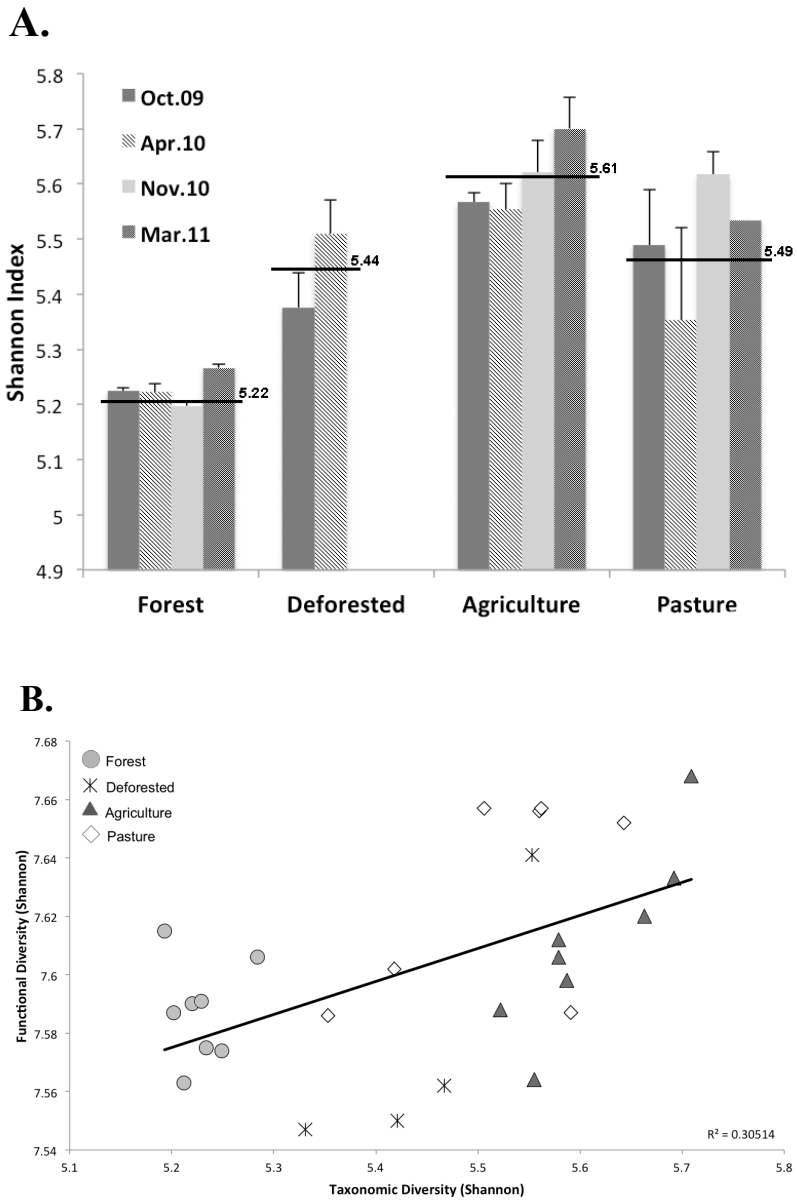


Figure 2.3 - (A) Levels of alpha diversity based on Shannon index for the different land use systems across four sampling times. The black lines represent the diversity average. **(B)** Linear regression showing differences in alpha diversity levels across different land use systems. X-axis shows taxonomic diversity and Y-axis shows functional diversity, both based on Shannon index

In order to visualize the differences in community structure among different land-uses, abundance matrices of taxonomical and functional profiles were used for Redundancy Analysis (RDA) (Figure 2.2 and Supplementary Figure 2.1). The forest soil-borne communities were clearly taxonomically and functionally distinct from the altered soils, being clustered apart from deforested, agriculture and pasture soils (ANOSIM $R = 0.98$, $p < 0.01$).

The microbial taxa were more variable among the samples than the functional categories (Supplementary Figure 2.2). ANOSIM pointed that the community structure varies more across land-use than across time (Table 2.3).

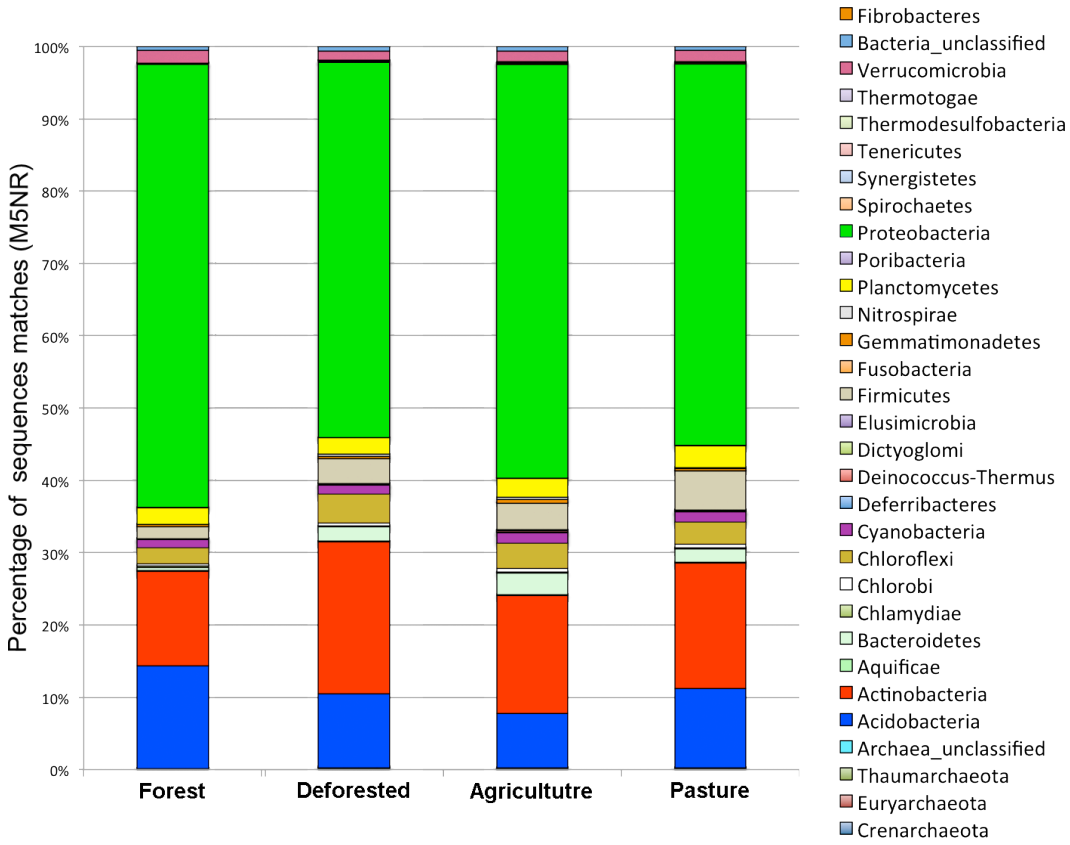


Figure 2.4 - Taxonomic affiliation of metagenomic reads. Results for complete datasets evaluated by BLASTX analysis against the M5NR database using MG-RAST v3.3 software

Table 2.3 - ANOSIM comparing the taxonomical and functional profiles across and within different land-use systems

	Taxonomical[a]		Functional[b]	
	R values	*p* values	*R* values	*p* values
across env.				
Forest x Deforested	**1**	**<0.01**	**0.98**	**<0.01**
Forest x Agriculture	**1**	**<0.01**	**1**	**<0.01**
Forest x Pasture	**0.98**	**<0.01**	**0.98**	**<0.01**
Deforested x Agriculture	**0.81**	**<0.01**	**0.72**	**<0.01**
Deforested x Pasture	**0.46**	**<0.02**	**0.40**	**<0.03**
Agriculture x Pasture	**0.80**	**<0.01**	**0.55**	**<0.01**
whithin env.				
Forest x Forest	0.58	>0.3	0.43	>0.3
Deforested x Deforested	1	>0.3	1	>0.3
Agriculture x Agriculture	0.85	>0.3	0.2	>0.3
Pasture x Pasture	0.88	>0.3	0.07	>0.4

[a] Samples compared with an abundance matrix at class level
[b] Samples compared with an abundance matrix of functional categories at subsystem level 1
In bold are shown the values statistically significant.

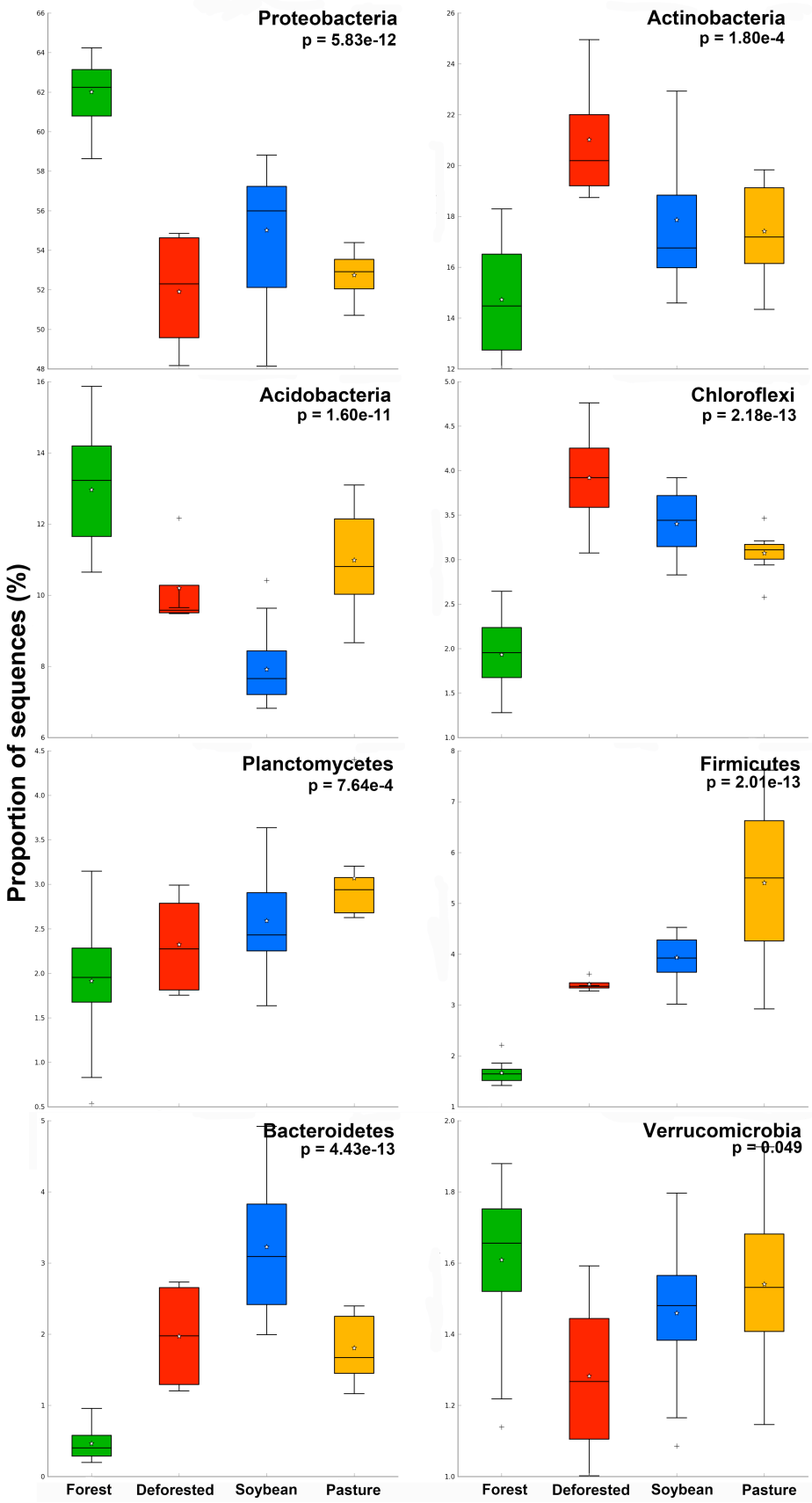


Figure 2.5 - Box plot showing the distribution in the proportion of the eight most abundant phyla assigned to samples from forest, deforested site, agriculture and pasture comprising all sampling periods. Boxes indicate the IQR (75th to 25th of the data). The median value is shown as a line within the box and the mean value as a star. Whiskers extend to the most extreme value within 1.5*IQR. Outliers are shown as crosses

The abundance of most bacterial and archaeal groups, as analyzed by STAMP, was not statistically different between the four sites. However, each land-use system presented dominance of some specific phylum when compared to all sites. Forest samples presented high abundance of the phyla *Proteobacteria*, *Acidobacteria* and *Verrucomicrobia*, accounted for 61%, 14% and 2% of the total sequences, respectively. The deforested site presented high abundances of *Actinobacteria* and *Chloroflexi* accounted for 10% and 3% of total sequences in this area, respectively. Agriculture presented high abundances of *Bacteroidetes*, which represent 1% of the sequences. And when compared to all sites, pasture presented an overrepresentation of the phyla *Firmicutes* and *Planctomycetes*, accounted for 2% and 10% of sequences, respectively (Figure 2.5).

The functional profiles of the soils were analyzed according to the SEED database and compared with STAMP, and the most prevalent core of functions in forest soils were related to respiration, nutrient cycles (sulfur, potassium, phosphorus and nitrogen) and resistance to stress. On the other hand, we found that genes associated to DNA/RNA and protein metabolism, and cell division were abundant in the altered sites (Figure 2.6).

In order to predict the functional patterns of the microbial communities in each site, nine functional categories were selected based on the relation with nutrient cycle (metabolism of nitrogen, phosphorus, potassium, sulfur, and respiration) and the relation to stress (stress response, defense and metabolism of DNA and RNA). The comparison was made based on the number of genera and the abundance of genes of each function. In a general overview, agriculture and pasture samples presented more genera related to metabolism of nitrogen, sulfur, DNA and RNA, respiration and stress response. Forest and deforested samples presented more genera related to metabolism of phosphorus and potassium (Table 2.4). Despite the low number of genera playing some specific functions in forest, these functional genes were 10-fold more abundant than the altered soils (Table 2.5).

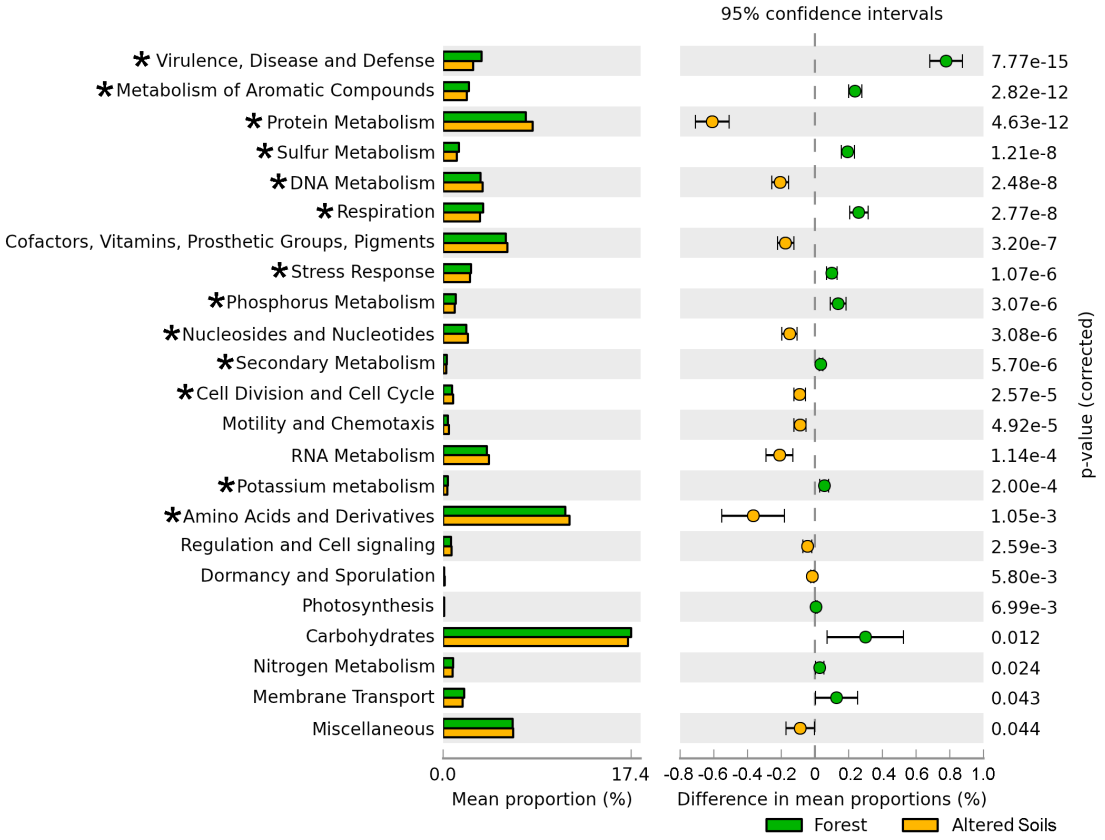


Figure 2.6 - STAMP analysis of functional profiles at subsystem level 1 (SEED database) of forest and altered soils (deforested, agriculture and pasture). Groups overrepresented in forest (green) correspond to positive differences between proportions and groups overrepresented in altered soils (yellow) correspond to negative differences between proportions. The comparison was accomplished including all replicates and sampling. Corrected *p*-values were calculated using Benjamini-Hochberg FDR approach ($p < 0.05$). * indicate the subsystem that was significative in at least 3 sampling periods. For detailed analysis in each sampling period see supplementary Figure 3

Table 2.4 - Number of genera linked to functional categories in the four land-use systems

	Forest	Deforested	Agriculture	Pasture
Nitrogen Met.	11	9	12	11
Phosphorus Met.	5	6	3	4
Potassium Met.	4	4	3	3
Sulfur Met.	6	6	7	7
Respiration	11	11	15	14
Defense	5	6	6	8
DNA Met.	9	11	14	14
RNA Met.	9	8	11	12

Total of genera presented in the four sampling period.
Met = Metabolism

Table 2.5 - Abundance of genes related to functional categories in the four land-use systems

	Forest	Deforested	Agriculture	Pasture
Nitrogen Met.	1.71E+05	4.65E+04	6.89E+04	8.07 E+04
Phosphorus Met.	1.63E+05	4.51E+04	5.90E+04	7.16 E+04
Potassium Met.	1.46E+05	2.95E+04	5.91E+04	7.28 E+04
Sulfur Met.	1.71E+05	4.63E+04	6.87E+04	8.01 E+04
Respiration	1.71E+05	5.88E+04	6.89E+04	7.05 E+04
Stress	1.71E+05	4.63E+04	6.87E+04	7.06 E+04
Defense	1.55E+05	4.46E+04	6.65E+04	7.90 E+04
DNA Met.	1.70E+05	4.64E+04	6.88E+04	8.03 E+04
RNA Met.	1.71E+05	4.65E+04	6.80E+04	8.00 E+04

Average abundance for 4 sampling periods.
Met = Metabolism

2.3.5 Network analysis

In order to gain a more integrated understanding of the microbial communities and to compare the complexity of correlations operating in the studied soils, a network analysis were conducted to verify all the connections among taxonomic profile, functional traits, and chemical properties. All positive and negative correlations were measured: between taxonomical profile (tax-tax), between functional categories (func-func), between chemical properties (chem-chem), between taxonomic and functional profiles (bact-func), between taxonomic and chemical parameters (tax-chem), and between functional categories and chemical parameters (func-chem) (Figure 2.7 and Table 2.6). The results showed that soils under forest and agriculture presented a more complex network. Pasture soils presented less number of correlations. Deforested soils presented a very low number of correlations and were excluded from the comparison. In general, forest soils presented slightly higher complex correlations than agriculture soils. The network of forest had 113 nodes and 639 edges, and the modularity was 3.7 with 10 communities. For agriculture soils, the network presented 131 nodes and 600 edges, and the modularity was 2.99 with 16 communities. Pasture soil network had 116 nodes and 494 edges, and the modularity was 4.65 with 10 communities. Forest soils presented higher number of “func-func”, “chem-chem”, “tax-chem” and “func-chem” correlations. On the other hand, the number of “tax-tax” and “tax-func” correlations was higher in agriculture. In forest and pasture the number of positive correlations was higher than negatives, while in agriculture the number of negatives correlations was higher than positives.

Table 2.6 - Number of correlations as inferred by Spearman

		tax	func	chem	tax-func	tax-chem	func-chem	Total
Forest	***positive***	102	25	79	33	71	42	352
	negative	47	24	57	40	73	46	287
	total	149	49	136	73	144	88	639
Agriculture	***positive***	127	13	17	67	33	22	279
	negative	45	11	85	83	73	24	321
	total	172	24	102	150	106	46	600
Pasture	***positive***	95	11	54	32	64	16	272
	negative	45	8	41	59	56	13	222
	total	140	19	95	91	120	29	494

Total numbers of strong ($r > 0.07$) and significant ($p < 0.05$) pairwise correlations among taxonomic profile (class level), functional profile (subsystem level 1), and chemical properties. All combinations of correlations were measured. tax: taxonomic; func: functional; chem: chemical.

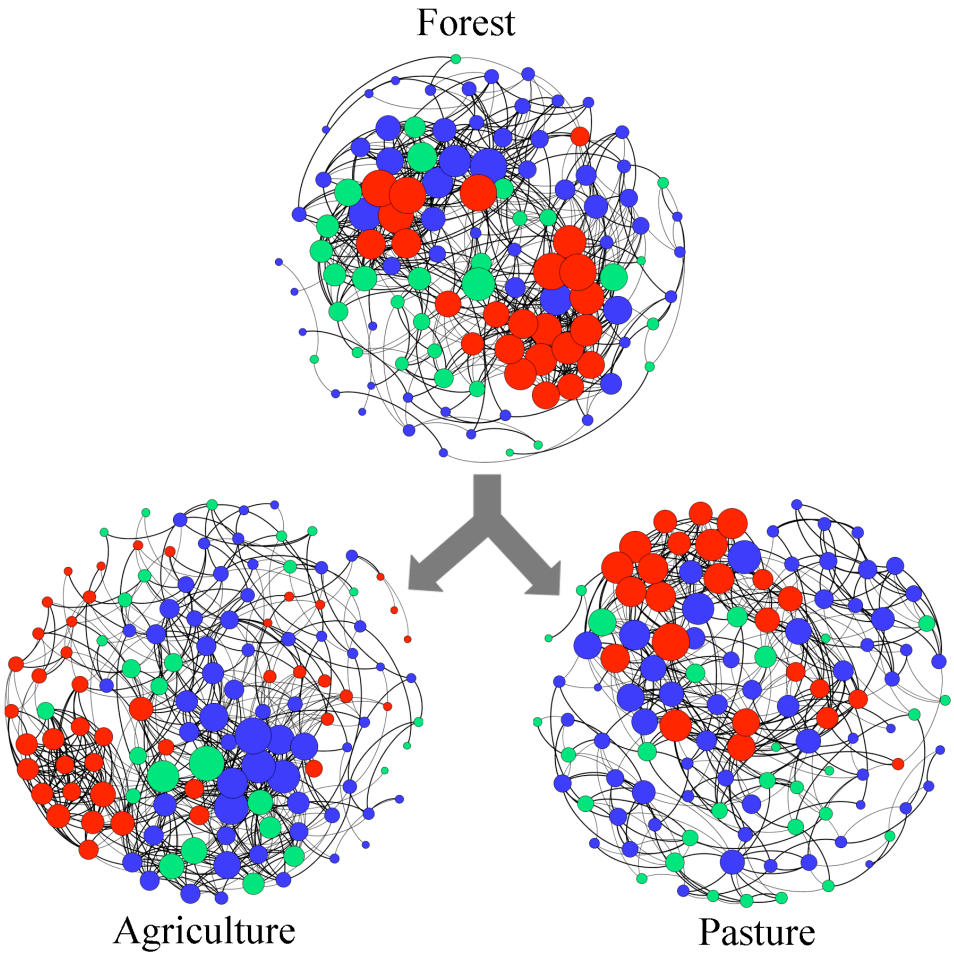


Figure 2.7 - Network representation of forest, agriculture and pasture soils, based on correlation analysis from taxonomic and functional profiles, and soil physicochemical parameters. A connection stands for strong (Spearman's $r>0.7$) and significant ($p<0.05$) correlation. Blue nodes indicate taxonomic affiliation at class level, green nodes indicate functional categories based on subsystems at level 1 (SEED Database) and red nodes indicate soil physicochemical parameters. The size of each node is proportional to the number of connections

2.4 Discussion

Several studies in Amazon forest region have shown that this environment harbors high diverse microbial communities, also demonstrating that land-use changes have altered the structure and composition of microbial communities and functional diversity of the soils (BORNEMAN; TRIPLETT, 1997; JESUS et al., 2009; NAVARRETE et al., 2010; 2011; 2013; TAKETANI; TSAI, 2010; GERMANO et al., 2012). However, in this study we used shotgun metagenome to provide new information integrating taxonomic e functional data to a better understanding of the effects of the land-use change over the soil-borne microbial communities. We hypothesized that the shifts in the communities might be related to conversion of the forest to different land-use systems, which have effects on the soil properties. Based on this hypothesis, distinct land-use may show distinct microbial community structure, both taxonomically and functionally. Here, we depicted the taxonomic and potential function diversities of the microorganisms inhabiting soils from native forest, deforested area, agriculture (soybean) and pasture.

The conversion of native forest to different land-uses through slash-and-burn has effects on the soil properties. Forest burning causes significant removal of organic matter, deterioration of soil structure, loss of nutrient, leaching and erosion, and marked alteration in both quantity and composition of the soil-borne microbial communities (CERTINI, 2005). In our sites, the forest conversion to agriculture follows the annually rotational production order: millet – soybean – maize, under no-tillage. It has been shown that no-tillage management contributes to soil conservation driven mainly by the crop residues on the surface, which increases soil organic matter content and improves soil physical properties (FRANCHINI et al., 2007; CALEGARI et al., 2008). Considering that most of microbial communities activity occurs within a few centimeters of the soil surface (BABUJIA et al., 2010), the management system, along with forest conversion affects microbial organisms. Our RDA result showed that samples grouped according to land-use type, with forest being clustered apart from the other sites. Souza et al. (2013) also used shotgun metagenomics approach to show that different agricultural management affects the structure and composition of microbial communities. Also, we detected that the differences were more across land-use system than across the chronosequence. Rather, the composition of microbial communities was strongly correlated with specific soil properties. The community of forest samples were correlated to organic matter content and H+Al, while the altered soils, *i.e.* deforested, agriculture and

pasture, were correlated to pH, Cu and NO^{3-} content, and base saturation. Although the pH has been considered to be the most important factor that explains the variation of microbial communities in soils (FIERER et al., 2006; LAUBER et al., 2009), other factors have also influence on the microbial structure in soil, for example organic matter content, moisture, and nutrient availability (FIERER et al., 2007; 2012; KURAMAE et al., 2012; NAVARRETE et al., 2012).

The land-use change affected the microbial diversity of the soils, with forest sites presenting lower levels of alpha diversity. These data are consistent with other studies using the 16S rRNA gene, which showed an increase in the taxonomy diversity after the conversion of Amazon forest (CENCIANI et al., 2009; JESUS et al., 2009; RODRIGUES et al., 2013). Although these studies showed higher alpha diversity in the altered soils, Rodrigues et al. (2013) demonstrated that the conversion of Amazon forest to agriculture decreased the beta diversity, which means that the microbial communities were more similar in composition across space. The authors suggest that the decrease of beta diversity is an indication of biotic homogenization, where the community similarity increases over time and space (OLDEN et al., 2004). Although our sampling methodology does not allow the calculation of the spatial beta diversity, the temporal beta diversity was higher in forest soil, which indicates that the community in the forest soil was more variable over time. In addition, the temporal variability in alpha diversity between land-use types was greater than the differences within a given land-use. This is contrary to what was found by Lauber et al. (2013), which studied the temporal variability between two agricultural management fields and one grassland site. However, they suggest that temporal variability needs to be carefully assessed when comparing microbial diversity across different soils. Also, our data showed a significant correlation between taxonomic and functional diversities, suggesting that the overall functional diversity is predictable from the taxonomic diversity of the microbial communities. In a metagenomic study across different biomes, Fierer et al. (2012) suggest that the types of taxa found in a community are also important in predict functional diversity.

As a response to land-use change in Amazon region here investigated, we found differences in abundance of some Phyla in each land-use types. Forest soils presented higher relative abundances of *Proteobacteria*, *Acidobacteria* and *Verrucomicrobia*. Within the *Bacteria* Domain, the phylum *Proteobacteria* represented the dominant phylum in all four land-use systems, but in a comparison to the altered sites, the abundance was higher in forest, with a decrease after the deforestation, which might be related to the drastic change made by

the slash-and-burn practice. The phylum *Acidobateria* was also representative in forest, presenting a decrease after the deforestation. Similar to our findings, Navarrete et al. (2013) studying different land-use system in the same Amazon region found that *Acidobacteria* respond to the deforestation, decreasing in relative abundance in soils of cropland and pasture. *Verrucomicrobia* are important members of the rhizosphere (CHOW et al., 2002), which explain higher abundance in forest soils and a decrease after deforestation. In the deforested site, there was an increase of *Actinobacteria* and *Chloroflexi.* Studies have shown higher abundance of *Actinobacteria* in desert soils (FIERER et al., 2012) and deforested area (NAVARRETE, 2012). As a consequence of the forest conversion through slash-and-burn, there is a soil temperature increasing and the deposit of high amount of carbon from the ashes. This alteration explain higher abundance of *Actinobacteria* in this area, since they are important decomposers and are also capable to produce spores, which allow them to resist to perturbations events (VENTURA et al., 2007). Some members of *Chloroflexi* are aerobic thermophiles, which grow well in high temperatures, and have important role in the decomposition of organic matter (YAMADA et al., 2005). In agriculture soils, used for soybean cropping, the phylum *Bacteroidetes* was overrepresented in comparison to the other land-use types. Interestingly, this phylum includes plant-growth promoting and cellulose decomposing (VERKHOVTSEVA et al., 2007; SOLTANI et al., 2010), which might be related to soybean cultivation. In pasture soils, we observed an increase of *Firmicutes*, which is similar to what was found by Rodrigues and colleagues (2013), who studied the conversion forest-to-pasture in Amazon and found an increase of *Firmicutes* in pasture soils. They suggested that this abundance is due to the notable resistance to desiccation and extremes of environmental variation (BATTISTUZZI; HEDGES, 2009), being resistant to variation of soil temperature throughout the day, such as pasture sampled in our study.

The study of functional prediction of microbial communities is very important to a better understand of the response to alterations in soils. In a recent study, Burke et al. (2011) suggest that the understanding of a community structure may not be based only on taxonomy but rather the more functional level of genes. In our study, the shotgun approach revealed that sequences related to respiration, nutrient cycles and resistance to stress were more abundant in forest. On the other hand, sequences affiliated to cell division and to metabolism of DNA, RNA, and protein, were differently accounted for the altered soils. These results show the impact of the deforestation on the microbial functional traits in soils, which indicate the adaptations of the organisms to survive under stress conditions. The genome stability is

challenged by endogenous and environmental agents, which induce DNA lesions and genome rearrangements. Also, the agriculture soils and pasture presented more functional diversity and more richness of genera playing specific functions. For these soils, the functional redundancy is very important to keep the ecosystem functioning. But while the functional diversity of agriculture and pasture is higher than forest, the abundance of genes related to important functions is 10-fold higher in forest. Therefore, our results suggest that the functional equilibrium in forest is maintained based on the abundance rather than the diversity.

As a response to the land-use change we found some differences in the network associations for each studied soil. Forest presented slightly more complex network, which can be based on the correlation between functional and chemical parameters. On the other hand, agriculture and pasture had higher correlations related to the taxonomical diversity. Once again, our results point to the importance of the taxonomic and functional diversity in maintain soil functioning after a stress event. Tardy et al. (2013), in an experiment with heat stress, showed that the soil functional stability is very dependent on the microbial diversity. In our study, a higher taxonomical and functional diversity was found in agriculture and pasture soils. Considering the stress event of transformation of the forest through slash-and-burn, a higher diversity in the altered soils would contribute to maintain the soil function. Some studies have shown that the resistance and resilience of microbial process may be increased with increasing microbial diversity (GRIFFITHS et al., 2000; DEGENS et al., 2001; GIRVAN et al., 2005). Resilience is defined as the rate at which microbial composition returns to its original composition after being disturbed (ALLISON; MARTINY, 2008). Our results showed that soils of agriculture and pasture presented higher diversity, along with higher richness of different genera playing some specific functions, which led to functional redundancy of soil microbial communities. Together, functional redundancy and biodiversity may buffer functional shifts induced by environmental variations, acting as an “ecological insurance” for the ecosystem (YACHI; LOREAU, 1999).

2.5 Conclusion

This study highlights that the conversion of the Amazon forest towards agricultural practices can have strong effects on soil microbiota, effects that were larger in land-use type than temporal variability. We have shown that different land-use systems affect soil attributes, with consequences on microbial taxonomic and potential functional diversities and structures. Agriculture and pasture soils were among the most diverse, suggesting that higher diversity and functional redundancy is important to maintain the ecosystem functioning. On the other hand, the equilibrium in forest is maintained based on the abundance of organisms. Thus, our work contributed to increase the knowledge about the response of microbial communities to land-use change in Amazon soils and demonstrate that the metagenomic approach is a powerful tool to assess taxonomic and potential functional diversities of soil microorganisms.

REFERENCES

ALLISON, S.T.; MARTINY, J.B.H. Resistance, resilience, and redundancy in microbial communities. **Proceedings of the National Academy of Sciences of the USA,** Washington, DC, v. 105, p. 11512-11519, 2008.

BABUJIA, L.C.; HUNGRIA, M.; FRANCHINI, J.C.; BROOKES, P.C. Microbial biomass and activity at various soil depths in a Brazilian Oxisol after two decades of no-tillage and convetional tillage. **Soil Biology and Biochemistry**, Oxford, v. 42, p. 2174-2181, 2010.

BARDGETT, R.D.; FREEMAN, C.; OSTLE, N.J. Microbial contributions to climate change through carbon cycle feedbacks. **The ISME Journal**, London, v. 2, p. 805-814, 2008.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. In: INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA, 3, 2009, San Jose, California. **Proceedings…** Menlo Park: AAAI Press, 2009. Available at: http://aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009.

BATTISTUZZI, F.U.; HEDGES, S.B. A major clade of prokaryotes with ancient adaptations to life on land. **Molecular Biology and Evolution**, Oxford, v. 26, p. 335-343, 2009.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society, Series B**, London, v. 57, p. 289-300, 1995.

BORNEMAN, J.; TRIPLETT, E.W. Molecular microbial diversity in soils from eastern Amazonia: Evidence for unusual microorganisms and microbial population shifts associated with deforestation. **Applied Environmental Microbiology**, Baltimore, v. 63, p. 2647-2653, 1997.

BRODY, J.R.; KERN, S.E. Sodium boric acid: Atriz-less, cooler conductive medium for DNA electrophoresis. **BioTechniques**, New York, v. 36, p. 214-216, 2004

BURKE, C.; STEINBERG, P.; RUSCH, D.; KJELLEBERG, S.; THOMAS, T. Bacterial community assembly based on functional genes rather than species. **Proceedings of the National Academy of Sciences of the USA,** Washington, DC, v. 108, p. 14288-14293, 2011.

CALEGARI, A.; HARGROVE, W.L.; RHEINHEIMER, D.D.S.; RALIS, H.R.; TESSIER, D.; TOURDONNET, S.; GUIMARÃES, M.F. Impact of long-term no-tillage and cropping system management on soil organic carbon in an Oxisol: A model for sustainability. **Agronomy Journal**, Madison, v. 100, p. 1013-1019, 2008.

CENCIANI, K.; LAMBAIS, M.R.; CERRI, C.C.; BASILIO DE AZEVEDO, L.C.; FEIGL, B.J. Bacteria diversity and microbial biomass in forest, pasture and fallow in the southwestern Amazon Basin. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 33, p. 907-916, 2009.

CERTINI, G. Effects of fire on properties of forest soils: a review. **Oecologia**, Berlin, v. 143, p. 1-10, 2005.

CHOW, M.L.; RADOMSKI, C.C.; McDERMOTT, J.M.; DAVIES, J.; AXELROOD, P.E. Molecular characterization of bacterial diversity in Lodgepole pine (*Pinus contorta*) rhizosphere soils from British Columbia forest soils differing in disturbance and geographic source. **FEMS Microbiology Ecology**, Amsterdam, v. 42, p. 347-357, 2002.

CLARK, K.; GORLEY, R. **PRIMER**. Primer-E. Plymouth, UK, version 6, 2006.

DEGENS, B.P.; SCHIPPER, L.A.; SPARLING, G.P.; DUNCAN, L.C. Is the microbial community in a soil with reduced catabolic diversity less resistant to stress or disturbance? **Soil Biology and Biochemistry**, Oxford, v. 33, p. 1143-1153, 2001.

FAO. Agriculture and Consumer Protection Department. **Conservation agriculture**. **Rome,** 2012. Available at: http://www.fao.org/nr/cgrfa/cthemes/cgrfa-micro-organisms/en//.

FIERER, N.; JACKSON, R.B. The diversity and biogeography of soil bacterial communities. **Proceedings of the National Academy of Sciences of the USA,** Washington, DC, v. 103, p. 626-631, 2006.

FIERER, N.; BRADFORD, M.A.; JACKSON, R.B. Toward an ecological classification of soil bacteria. **Ecology**, Brooklyn, v. 88, p. 1354-1364, 2007.

FIERER, N.; LEFF, J.W.; ADAMS, B.J.; NIELSEN, U.N.; BATES, S.T.; LAUBER, C.L.; OWENS, S.; GILBERT, J.A.; WALL, D.H.; CAPORASO, J.G. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. **Proceedings of the National Academy of Sciences of the USA,** Washington, DC, v. 109, p. 21390-21395, 2012.

FISHER, W.D. On grouping for maximum homogeneity. **Journal of the American Statistical Association**, Boston, v. 53, p. 789-798, 1958.

FRANCHINI, J.C.; CRISPINO, C.C.; SOUZA, R.A.; TORRES, E.; HUNGRIA, M. Microbiological parameters as indicators of soil quality under various soil management and crop rotation system in southern Brazil. **Soil and Tillage Research**, Amsterdam, v. 92, p. 18-29, 2007.

GERMANO, M.G.; CANNAVAN, F.S.; MENDES, L.W.; LIMA, A.B.; TEIXEIRA, W.G.; PELLIZARI, V.H.; TSAI, S.M. Functional diversity of bacterial genes associated with aromatic hydrocarbon degradation in anthropogenic dark earth of Amazonia. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 47, p. 654-664, 2012.

GIRVAN, M.S.; CAMPBELL, C.D.; KILHAM, K.; PROSSER, J.I.; GLOVER, L.A. Bacterial diversity promotes community stability and functional resilience after perturbation. **Environmental Microbiology**, Oxford, v. 7, p. 301-313, 2005.

GRIFFITHS, B.S.; RITZ, K.; BARDGETT, R.D.; COOK, R.; CHRISTENSES, S.; EKELUND, F.; SØRENSEN, S.J.; BÅÅTH, E.; BLOEM, J.; de RUITER, P.C.; DOLFING, J.; NICOLARDOT, B. Ecosystem response of pasture soil communities to fumigation-induced microbial diversity reductions: an examination of the biodiversity – ecosystem function relationship. **Oikos**, Copenhagen, v. 2, p. 279-294, 2000.

HAMMER, Ø.; HARPER, D.A.T.; RYAN, P.D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. **Palaeontologia Electronica**, College Station, v. 4, n. 1, art. 4, 9 p. 2001.

JESUS, E.D.; MARSH, T.L.; TIEDJE, J.M.; MOREIRA, F.M.D. Changes in land use alter the structure of bacterial communities in Western Amazon soils. **The ISME Journal**, London, v. 3, p. 1004-1011, 2009.

KURAMAE, E.E.; YERGEAU, E.; WONG, L.C.; PIJL, A.S.; van VEEN, J.A.; KOWALCHUK, G.A. Soil characteristics more strongly influence soil bacterial communities than land-use type. **FEMS Microbiology Ecology**, Amsterdam, v. 79, p. 12-24, 2012.

LANGENHEDER, S.; BULLING, M.T.; SOLAN, M.; PROSSER, J.I. Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity. **PLoS One**, New York, v. 5, e10834, 2010.

LAUBER, C.L.; RAMIREZ, K.S.; AANDERUD, Z.; LENNON, J.; FIERER, N. Temporal variability in soil microbial communities across land-use types. **The ISME Journal**, London, v. 7, p. 1641-1650, 2013.

LAUBER, C.L.; KNIGHT, R.; HAMADY, M.; FIERER, N. Soil pH as a predictor of soil bacterial community structure at the continental scale: A pyrosequencing-based assessment. **Applied Environmental Microbiology**, Baltimore, v. 75, p. 5111-5120, 2009.

MATO GROSSO (Estado). Secretaria de Estado de Planejamento d Coordenação Geral – SEPLAN. **Mapa de solos do Estado de Mato Grosso.** Mato Grosso, 2001. Available at: www.seplan.mt.gov.br. Accessed at: March 20 2012.

MARGULIES, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, London, v. 437, p. 376-380, 2005.

MEYER, F.; PAARMAN, D.; D'SOUZA, M.; OLSON, R.; GLASS, E.M.; KUBAL, M.; PACZIAN, T.; RODRIGUEZ, A.; STEVENS, R.; WILKE, A.; WILKENING, J.; EDWARDS, R.A. The Metagenomics RAST server – A public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, London, v. 9, p. 386-393, 2008.

NAVARRETE, A.A. **Bacterial ecology in Amazonian soils under deforestation and agricultural management**. 2012. 136 p. Tese (Doutorado em Ciências) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2012.

NAVARRETE, A.A.; CANNAVAN, F.S.; TAKETANI, R.G.; TSAI, S.M. A molecular survey of the diversity of microbial communities in different Amazonian agricultural model systems. **Diversity**, Bethesda, v. 2, p. 787-809, 2010.

NAVARRETE, A.A.; TAKETANI, R.G.; MENDES, L.W.; CANNAVAN, F.S.; MOREIRA, F.M.S.; TSAI, S.M. Land-use systems affect archaeal community structure and functional diversity in western Amazon soils. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 35, p. 1527-1540, 2011.

NAVARRETE, A.A.; KURAMAE, E.E.; de HOLLANDER, M.; PIJL, A.S.; van VEEN, J.A.; TSAI, S.M. Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. **FEMS Microbiology Ecology**, Amsterdam, v. 83, p. 607-621, 2013.

NEWCOMBE, R.G. Improved confidence intervals for the difference between binomial proportions based on paired data. **Statistics in Medicine**, Chichester, v. 17, p. 2635-2650, 1998.

OLDEN, J.D.; POFF, N.L. Toward a mechanism understanding and prediction of biotic homogenization. **The American Naturalist**, Chicago, v. 162, p. 442-460, 2003.

PARKS, D.H.; BEIKO, R.G. Identifying biologically relevant differences between metagenomic communities. **Bioinformatics**, Oxford, 26, p. 715-721, 2010.

RODRIGUES, J.L.M.; PELLIZARI, V.H.; MUELLER, R.; BAEK, K.; JESUS, E.D.; PAULA, F.S.; MIRZA, B.; HAMAOUI, G.S.; TSAI, S.M.; FEIGL, B.; TIEDJE, J.M.; BOHANNAN, B.J.M.; NÜSSLEIN, K. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 110, p. 988-993, 2013.

SOLTANI, A.A.; KHAVAZI, K.; ASADI-RAHMANI, H.; OMIDVARI, M.; DAHAJI, P.; MIRHOSEYNI, A.H. Plant growth promoting characteristics in some *Flavobacterium* spp. isolated from soils of Iran. **Journal of Agricultural Science**, Toronto, v. 2, p. 106-115, 2010.

SOUZA, R.C.; CANTÃO, M.E.; VASCONCELOS, A.N.T.; NOGUEIRA, M.A.; HUNGRIA, M. Soil metagenomics reveals differences under conventional and no-tillage with crop rotation and succession. **Applied Soil Ecology**, Amsterdam, v. 72, p. 49-61, 2013.

TAKETANI, R.G.; TSAI, S.M. The influence of different land uses on the structure of archaeal communities in Amazon anthrosols based on 16S rRNA and *amo*A genes. **Microbial Ecology**, New York, v. 59, p. 734-743, 2010.

TARDY, V.; MATHIEU, O.; LÉVÊQUE, J.; TERRAT, S.; CHABBI, A.; LEMANCEAU, P.; RANJARD, L.; MARON, P.A. Stability of soil microbial structure and activity depends on microbial diversity. **Environmental Microbiology Reports**, Hoboken, 2013. DOI: 10.1111/1758-2229.12126.

VENTER, J.C.; REMINGTON, K.; HEIDELBERG, J.F.; HALPERN, A.L.; RUSCH, D.; EISEN, J.A.; WU, D.; PAULSEN, I.; NELSON, K.E.; NELSON, W.; FOUTS, D.E.; LEVY, S.; KNAP, A.H.; LOMAS, M.W.; NEALSON, K.; WHITE, O.; PETERSON, J.; HOFFMAN, J.; PARSONS, R.; BADEN-TILLSON, H.; PFANNKOCH, C.; ROGERS, Y.H.; SMITH, H.O. Environmental genome shotgun sequencing of the Sargasso sea. **Science**, Washington, DC, v. 304, p. 66-74, 2004.

VENTURA, M.; CANCHAYA, C.; TAUCH, A.; CHANDRA, G.; FITZGERALD, G.F.; CHATER, K.F.; van SINDEREN, D. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. **Microbiology and Molecular Biology Reviews**, Washington, DC, v. 71, p. 495-548, 2007.

VERKHOVTSEVA, N.; KUBAREV, E.; MINEEV, V. Agrochemical agents in maintaining the structure of the soil microbial community. **Russian Agricultural Science**, Heidelberg, v. 33, p. 100-102, 2007.

YACHI, N.H.; LOREAU, M. Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. **Proceedings of the National Academy of Sciences of the USA,** Washington, DC, v. 96, p. 1463-1468, 1999.

YAMADA, T.; SEKIGUCHI, Y.; IMACHI, H.; KAMAGATA, Y.; OHASHI, A.; HARADA, H. Diversity localization, and physiological properties of filamentous microbes belonging to Chloroflexi sub phylum in mesophilic and thermophilic methanogenic sludge granules. **Applied Environmental Microbiology**, Baltimore, v. 71, p. 7493-7503, 2005.

3

TAXONOMICAL AND FUNCTIONAL MICROBIAL COMMUNITY SELECTION IN SOYBEAN RHIZOSPHERE[1]

Abstract

This study addressed the selection of the rhizospheric microbial community from the bulk soil reservoir under agricultural management of soybean in Amazon forest soils. We used a shotgun metagenomics approach to investigate the taxonomic and functional diversities of microbial communities in bulk soil and rhizosphere associated to soybean and tested the validity of neutral and niche theories to explain rhizosphere community assembly processes. Our results showed a clear selection at both taxonomic and functional levels operating in the assembly of the soybean rhizosphere community. The taxonomic analysis revealed that the rhizosphere community is a subset of the bulk soil community. Species abundance in rhizosphere fits the log-normal distribution model, which is an indicator of the occurrence of niche-based processes. In addition, the data indicate that the rhizosphere community is selected based on functional cores related to the metabolisms of nitrogen, iron, phosphorus and potassium, which are related to benefits to the plant, such as growth promotion and nutrition. The network analysis including bacterial groups and functions was less complex in rhizosphere, suggesting the specialization of some specific metabolic pathways. We conclude that the assembly of microbial community in the rhizosphere is based on niche-based processes as a result of the selection power of the plant and other environmental factors.

Keywords: Niche theory. Neutral theory. Microbial ecology. Land use. Metagenomics. Brazilian Amazon.

[1] Aceito para publicação em The ISME Journal.

3.1 Introduction

"The most beautiful experience we can have is the mysterious. It is the fundamental emotion that stands at the cradle of true art and true science".

(**Albert Einstein** | 1879 – 1955)

It is well known that microorganisms are essential for the health of our planet by playing key roles in the major biogeochemical cycles (FALKOWSKI et al., 2008; DELMOND et al., 2012). In particular in soils, their abundance and diversity are huge surpassing 10^9 bacterial cells per gram, ranging from thousands of different taxa (TORSVIK et al., 1990; ROESCH et al., 2007). The size and structure of bacterial communities in soils is shaped by biotic and abiotic factors, such as soil physical and chemical factors (FIERER et al., 2012; KURAMAE et al., 2012; NAVARRETE et al., 2013), climate and plants species (BERG; SMALLA, 2009). The development of new technologies, such as culture independent techniques, boosts scientific interests to understanding microbial community (TORSVIK; ØVREÅS, 2002; DINSDALE et al., 2008; SPOR et al., 2011). Focuses are not only addressed on community compositions (FROSTEGARD, 2011; NEMERGUT, 2011), but also on community assembly (LINDSTROM; LANGENHENDER, 2012) and the relationship between community and external drivers including environmental factors (LANGENHEDER et al., 2010). Several hypotheses have been raised regarding microbial community assemblies, including "niche hypothesis" and "neutral theory" (DUMBRELL, 2010). The neutral theory predicts that the structure and composition of species will be related to geographic distance between samples as a result of dispersal limitation and that abundance will follow a zero-sum multinomial (ZSM) distribution (McGILL et al., 2006). The ZSM is the unique species abundance distribution model predicted by neutral theory (HUBBELL, 2001). On the other hand, the niche-based theory predicts that changes in the species community will be related to changes in environmental variables (JONGMAN et al., 1995) and that species abundances will follow pre-emption, broken stick, log-normal and Zipf-Mandelbrot models (MOTOMURA, 1932; MACARTHUR, 1957; McGILL et al., 2007). Both theories are well connected with environmental factors, but neither suggest how the microbial community assembly in the main hot spot of life, *i.e.* the rhizosphere, is driven.

Within the soil system the immediate surroundings of the plant root, *i.e.* rhizosphere, is a hot spot of microbial activity. Several studies regarding rhizospheric communities have been focused on the impact of plant species including *Arabidopsis* (BULGARELLI et al., 2012; LUNDBERG et al., 2012), rice (KNIEF et al., 2011), oak (UROZ et al., 2010), norway spuce (CALVARUSO et al., 2009), wild oats (DeANGELIS et al., 2008), potato (RASCH et al., 2006), tobacco (ROBIN et al., 2006), and soybean (XU et al., 2009), among others. However, information with deep taxonomic and functional analysis in the rhizosphere of soybean plants is scarce. Soybean is one of the most important cultivated crops mainly used in food and fodder production. The southeast Amazon region in Brazil has become one of the largest producing regions of the world. In order to achieve this, large areas of native forest in the Amazon region have been deforested and subsequent converted into cattle pasture and arable land. These land use changes have altered the structure and composition of microbial communities and functional diversity of the soils in the Amazon region (BORNEMAN; TRIPLETT, 1997; JESUS et al., 2009; NAVARRETE et al., 2010; TAKETANI; TSAI, 2010; NAVARRETE et al., 2011; GERMANO et al., 2012; NAVARRETE et al., 2013).

In order to assess the consequences of land use changes on microbial communities and their functioning and thus on ecosystem behavior, it is necessary to understand the process of community selection and assembly in the soybean rhizosphere. However, detailed information based on deep taxonomic and functional analyses of microbial community assembly in the rhizosphere of soybean plants is scarce. Therefore, we applied a DNA shotgun metagenomic approach to analyze the microbial community inhabiting the bulk soil and the rhizosphere in agricultural crop fields of soybean, so to determine the extent to which a particular plant species, *i.e.* soybean, is able to select a rhizospheric microbial community from the bulk soil reservoir. We hypothesize that niche-based mechanisms will explain the community assembly in the rhizosphere through a selective power of the plant in the rhizospheric environment. For this purpose we compared the taxonomic and functional profiles among 24 independent samples from soils and rhizosphere of first and fifth years of soybean cultivation.

Although there are reports addressing the composition and structure of rhizospheric community from different plant species, this is the first study that used deep metagenome sequencing and integrating taxonomic and functional data to examine whether neutral or niche-based mechanisms would best explain the microbial community assembly in the rhizosphere.

3.2 Material and Methods

3.2.1 Sampling sites

Bulk soil samples were collected in agricultural fields in two different sites located in the Southeastern Brazilian Amazon, in the state of Mato Grosso, Brazil, in the municipalities of Ipiranga do Norte (11°40'54.97" S and 55°50'8.79" W) and of Porto dos Gaúchos (11º44'29.62" S and 56°15'44.52" W). Oxisol is the predominant soil order of the sampling sites (MATO GROSSO, 2001), and the climate in the region is classified as Am (Koppen's classification), with annual average temperature of 28ºC and average precipitation of 2000 mm.

The sampling sites were selected according to vegetation cover, soil use, and management practices. In the Porto dos Gaúchos municipality, areas covered with native tropical rainforest were cleared in 2008 and subsequently converted into agricultural land. Since 2004, forest conversion to agricultural use occurred in areas located in Ipiranga do Norte. In both field locations, after forest conversion to arable land, the annual crop rotation was: millet, soybean, and maize, under no-tillage. After deforestation, fertilizers, pesticides, and a liming treatment were applied to the cropland fields at both locations. The cropland fields received different amounts of lime to increase soil pH to 5 and 6.

At each sampling site, the soil samples were collected from five points before sowing in two different sampling periods (November 2009 and November 2010). Sampling was performed by selecting one central point and four other sampling points (at least 50 m apart from the central point) directed towards the north, south, east, and west of the central points. Soil samples were taken from the 0- to 20-cm topsoil layer. First, the litter layer was removed, and, then, the soil samples were collected. A total of 20 bulk soil samples were collected in field (2 field locations x 2 sampling periods x 5 samplings per site). Samples were transported to the laboratory within 72h after sampling for the preparation of the mesocosms.

3.2.2 Mesocosm experiments

The soil samples collected at the field were used in mesocosm experiments, where soybean plants were grown in greenhouse at CENA-University of São Paulo (USP), Piracicaba, Brazil. The experiments were carried out in the greenhouse in order to normalize the influence of environmental parameters (such as moisture regime and temperature) on the growth conditions for the plants. To simulate the field conditions the same cultivar of soybean *Glycine max* (L.) Merril (Cultivar M-SOY 8866) was used in the experiments and, before sowing, the seeds were inoculated with *Bradyrhizobium japonicum*, in a concentration of 10^{10} viable cells per kg of seed, which is a common practice for soybean cultivation in Brazil.

The mesocosms consisted of ceramic pots (30 cm high x 20 cm diameter) with a stone layer of 5 cm on the bottom. The pots were filled with approximately 8 kg of soil and three seeds of soybean were sowed in each pot. Each soil sample consisted of a composite sample by mixing six subsamples collected from the 0- to 20-cm topsoil layer. Fifteen pots for each site were used in the experiment, including control pots without seed sown. The plants germinated at 28/19^{o}C (day/night) with 12h photoperiod. Temperature and moisture were regularly adjusted to create optimal growth conditions for the plants. Soil samples were collected after 80 days of plant growth, which period corresponds to prior plant maturity. Plants were collected and the roots with attached soil were removed from the pots and transported on ice to the laboratory. The roots were shaken to remove the loose soil and the remaining attached soil, considered to be the rhizosphere soil, was collected using sterile brushes. Soil samples from the control pots, were considered as bulk soil. The experiments were carried out two times during tropical summer period of November-January 2009/2010 and November-January 2010/2011.

3.2.3 DNA extraction and sequencing

DNA extraction from 250 mg of soil samples was carried out using PowerSoil DNA Isolation Kit (Mobio Laboratories, Carlsbad, CA, USA), according to the manufacturer's protocol. DNA quality and concentration were measured by 1% TSB (BRODY; KERN, 2004) agarose gel electrophoresis and NanoDrop 1000 spectrophotometer (Thermo Scientific,

Waltham, USA). In total, 24 DNA samples were sequenced (Macrogen Inc. Company, South Korea) on a Roche 454 automated sequencer GS-FLX system Titanium series reagent (454 Life Sciences, Brandford, CT, USA).

3.2.4 Annotation of metagenomic sequences and data analysis

Unassembled DNA sequences were annotated with the Metagenomics Rapid Annotation (MG-RAST) pipeline version 3.3 (MEYER et al., 2008). Taxonomic and functional profiles were generated using the normalized abundance of sequences matches to the SEED database (AZIZ et al., 2008). A table of the frequency of hits to each individual taxa (taxonomy) or subsystem (function) for each metagenome was generated and normalized by dividing by the total number of hits to remove bias indifference in read lengths and sequencing efforts. To identify hits, BlastX was used with a minimum alignment length of 50 bp and an E-value cut-off of $E<1x10^{-5}$ (DINSDALE et al., 2008; SMITH et al., 2012). Nonmetric multidimensional scaling (NMDS) and Principal Component Analysis (PCA) plots were used to visualize the structure among samples, using the taxonomic and functional abundance matrix generated as described above. NMDS plots were generated from Bray-Curtis similarity index matrices of the 24 samples by using the PAST software (HAMMER, 2001) and PCA by using Canoco 4.5 (Biometrics, Wageningen, The Netherlands).

All statistical analyses were performed for the two years of the experiments, and when convenient, the data were presented together. To determine statistical differences between the rhizosphere and the bulk soil samples, the Statistical Analysis of Metagenomic Profiles (STAMP) software package was used (PARKS; BEIKO, 2010). For this, a table of the frequency of hits of taxa and functional subsystem (SEED database) for each metagenome was generated from MG-RAST and used as input. P-values were calculated using the two sided Fischer's Exact test (FISHER, 1958), while confidence intervals were calculated using the Newcombe-Wilson method (NEWCOMBE, 1998) and correction was made using Benjamini-Hochberg FDR (BENJAMINI; HOCHBERG, 1995).

Network analyses were performed to better understand the taxonomic and functional relations within the microbial community. For analyzing the networks we calculated all possible Spearman's rank correlation coefficient. To filtering the data for reduced network complexity, we considered high correlations with cut-off of coefficient $r > 0.7$ and statistically

significant *P*-value <0.01 and <0.001 for taxonomy and function, respectively taking into account all the replicates. The nodes in the reconstructed network represent taxa and functional groups, and edges represent high and significant correlation between nodes. The topology of the network graph was made based of a set of measures calculated, as average node connectivity, average path length, diameter, and cumulative degree distribution. Statistical analyses were carried out in the R environment (http://www.r-project.org/) and networks visualization with the interactive platform Gephi (BASTIAN, 2009).

To test whether neutral or niche-based mechanism best explain the composition and structure of the microbial community we examined the rank abundance distribution. The neutral theory predicts that rank abundance distribution will be consistent with zero-sum multinomial model (ZSM) (HUBBEL, 2001). On the other hand, niche-based theory assumes that the rank abundance distribution would fit the pre-emption, broken stick, log-normal and Zipf-Mandlebrot models (MOTOMURA, 1932; MacARTHUR, 1957; McGILL et al., 2007). The species rank abundance for each metagenomic sample were fit to broken stick, pre-emption, log-normal, Zipf and Zipf-Mandelbrot rank abundance models using the command "radfit" found in the R package vegan (OKSANEN, 2010; R Development Core Team, 2007), and the ZSM model using TeTame (JABOT et al., 2008). AIC values for generated models were calculated based on the equation AIC = -2log-likehood+2*npar, where npar represents the number of parameters in the fitted model (FEINSTEIN; BLACKWOOD, 2012). The AIC values were compared to determine which model provided the best fit to empirical data (DUMBRELL et al., 2010).

3.3 Results

3.3.1 Soil characteristics

The soil samples used in the mesocosm experiments were collected in agricultural fields under soybean crop in arable systems of different age, *i.e.* 1-year and another with 5-year of harvesting. Both sites have the same classification and historic origin (slash-and-burn deforestation). The pH of the 1-year soybean site was 4.1±0.3, and the texture of 50:3:47 (sand:silt:clay) while the pH of the 5-year soybean site was 5.0±0.2 and the texture of 40:3:57. Among all sites and replicates the organic matter content varied between 34-38 $g.kg^{-1}$, total nitrogen varies between 1.43-1.82 $g.kg^{-1}$, and organic carbon between

16-24 $g.kg^{-1}$ (Supplementary Table 3.1; for more detailed soil characteristics discussion see NAVARRETE et al., 2013).

3.3.2 Taxonomic and functional profiling of metagenomes

Overall, sequencing yielded more than 3,2 million reads. After quality trimming, a total of 2,472,359 sequences, with average length of 523 bp, were obtained for 24 samples (Supplementary Table 3.2). Using cut-off of $E<1x10^{-5}$ and 50 bp minimal align length on MG-RAST server an average of 75% of sequences were predicted as protein. Metagenomic libraries were dominated by Bacteria (96% of hits to SEED) with sequences also matching Eukaryote (3%) and Archaea and Virus (1%). *Proteobacteria* represented the highest percentage of matches to the SEED database for all samples with average of 47% of all sequences, followed by *Actinobacteria* (23%), *Firmicutes* (6%) and *Acidobacteria* (5%) (Supplementary Figure 3.1A). When focusing on the most abundant phylum *Proteobacteria*, the numbers of sequences affiliated with distinct classes presented subtly differences among rhizosphere and bulk soil (Supplementary Figure 3.1B). Rarefaction curve analysis shows that sequencing has detected most of the diversity present in all samples for the two years of experiments (Supplementary Figure 3.2) In general 28% of total sequences could not be assigned to known sequences in the database.

3.3.3 Bulk soil vs. soybean rhizosphere

In order to visualize the differences in community structure and function between bulk soil and soybean rhizosphere samples, the taxonomic abundance profiles were used to compute a Bray-Curtis similarity matrix, coordinated into two dimensions by using NMDS (Figure 3.1). Samples were grouped according to site and age of harvest; rhizosphere samples were more similar to each other than to bulk soil samples and the bulk soil samples themselves. This analysis revealed clear differences in the microbial community structure between rhizosphere and bulk soil. When samples were compared using the STAMP software, there was an overrepresentation of the phyla *Actinobacteria*, *Acidobacteria*, *Chloroflexi*, *Cyanobacteria*, *Chlamydiae*, *Tenericutes, Deferribacteres, Chlorobi,*

Verrucomicrobia and *Aquificae* ($p<0.01$), in the rhizosphere samples (Supplementary Figure 3.3). At class taxonomical level, the rhizosphere presented an overrepresentation of *Bacilli*, *Mollicutes*, *Clostridia*, *Gammaproteobacteria*, *Epsilonproteobacteria*, *Chlamydiae* and *Thermomicrobia* ($p<6.9e^{-4}$) (Supplementary Figure 3.4).

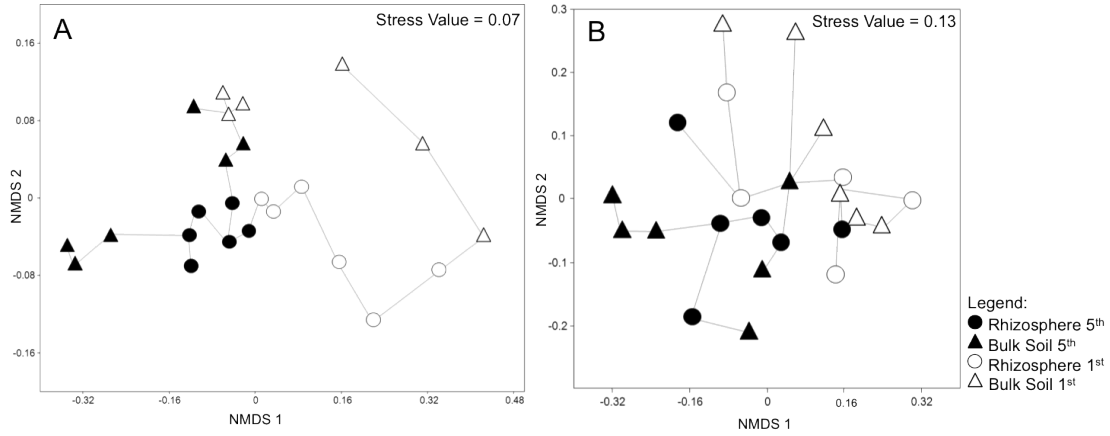


Figure 3.1 - NMDS of Bray-Curtis similarity matrix among 24 samples from mesocosms experiments with soybean. Taxonomic (A) and functional (B) analyses using relative abundance based on SEED bacterial matches at phylum level and subsystem level 2. The lines between dots represent the minimal spanning tree, which connects all points with minimal total length, based on similarity index. Stress values are shown in the upper right of the graphs

The functional profiles of bulk soil and soybean rhizosphere samples were analyzed according to the SEED database, and the most prevalent core of functions for all samples was "carbohydrates", while a high abundance of sequences matching "membrane transport", "iron acquisition and metabolism" occurred in the rhizosphere (Supplementary Table 3.3). The comparison of the functional profiles by STAMP revealed an overrepresentation of membrane transport ($p<1e^{-15}$) and functional cores related to metabolism of nitrogen, phosphorus, potassium and iron (Supplementary Figure 3.5 and 3.6).

The network analysis was markedly different for rhizosphere and bulk soil samples (Figure 3.2 and Table 3.1). In general, the rhizosphere presented slightly less complex correlations than bulk soil. In bulk soil, the network had 181 nodes and 194 edges (positive correlations), and the modularity was 0.828 with 55 communities while for rhizosphere, the network presented 182 nodes and 143 edges, and the modularity was 0.914 with 59 communities. All positive and negative correlations were measured: between bacterial groups (bact-bact), between functional groups (func-func) and between bacterial and functional groups (bact-func) (Table 3.1). The number of positive correlations was higher than the

negative correlations in bulk soil. In general, the number of correlations was higher in the bulk soil than rhizosphere, except for function-function negative correlations, which showed the highest number in rhizosphere. For bulk soil samples, the five bacterial groups that presented more correlations were *Deltaproteobacteria*, *Bacteroidia*, *Chloroflexi*, *Planctomycetacia* and *Sphingobacteria*, whilst *Chloroflexi*, *Deltaproteobacteria*, *Solibacteres*, *Sphingobacteria* and *Gammaproteobacteria* were the groups with more correlations in the rhizosphere (as indicated in Figure 3.2 and Supplementary Figure 3.7).

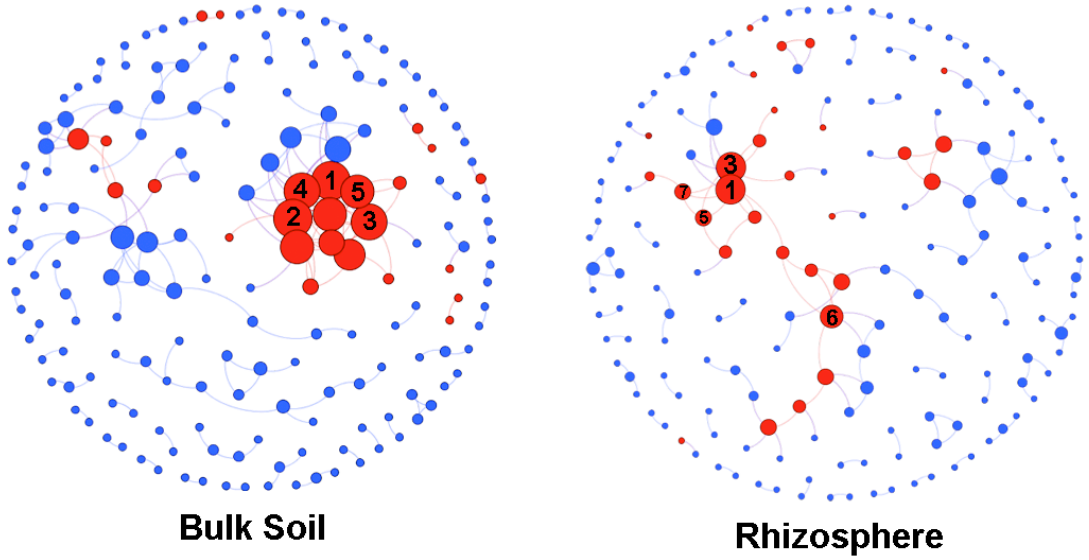


Figure 3.2 - Network of bulk soil and rhizosphere based on correlation analysis from taxonomic and functional profiles. A connection stands for strong (Spearman's r>0.7) and significant (p<0.01) correlation. Red nodes indicate taxonomic affiliation at class level and blue nodes indicate functional categories based on subsystems at level 2 (SEED Database). The size of each node is proportional to the number of connections. The numbers inside red nodes indicate the phyla with more correlations, as follow: 1. *Deltaproteobacteria*; 2. *Bacteroidetes*; 3. *Chloroflexi*; 4. *Planctomycetacia*; 5. *Sphingobacteria*; 6. *Solibacteres*; 7. *Gammaproteobacteria*

Table 3.1 - Number of correlations as inferred by Spearman

		Phylogenetic	Functional	Phyl X Func	Total
Rizhosphere	pairwise correlations	739	97,957	17,278	115,974
	significant correlations	109 (45)	5,355 (140)	1,240 (51)	**6,704 (236)**
	significant positive correlations	79 (28)	2792 (84)	675 (31)	**3,546 (143)**
	significant negative correlations	30 (17)	2563 (56)	565 (20)	**3,158 (93)**
Bulk Soil	pairwise correlations	732	98,503	17,256	116,491
	significant correlations	121 (62)	5871 (184)	1333 (64)	**7,325 (310)**
	significant positive correlations	86 (52)	3044 (109)	623 (33)	**3,753 (194)**
	significant negative correlations	35 (10)	2827 (75)	710 (31)	**3,572 (116)**

The total numbers of pairwise correlations as well as significant correlations (p<0.05) among phylogenetic profile (class level), among functional profile (subsystem level 2), and between phylogenetic and functional profiles are shown. The numbers in parentheses indicate higher correlation coefficient for phylogenetic (r>0.7 and p<0.01), functional (r>0.7 and p<0.001), and Phyl x Func (r>0.7 and p<0.001) data.

The data of all samples were fitted to theoretical species abundance distribution to test whether neutral or niche-based mechanisms best explain the composition and structure of microbial communities from bulk soil and rhizosphere. The comparison of different rank abundance distribution models based on AIC values, indicated the ZSM was closest fit with samples from bulk soil, which is consistent with neutral theory dynamics. On the other hand, rhizosphere samples were closest fit with log-normal model, indicating niche-based theory. When we analyze the data based on the time of harvesting, 1-year samples fitted neutral models, while 5-year samples fitted niche-based model (Supplementary Table 3.4).

3.3.4 Soybean rhizosphere 1-year vs. 5-year

The taxonomical and functional profiles of the rhizosphere were compared for the 1-year and 5-year of soybean cultivation. Principal Component Analysis (PCA) applied to visualize the differences among samples showed a clear separation according to time, revealing a distinct community composition, for both phylum (ANOSIM R=0.525 $p<0.01$) and class (R=0.52 $p<0.01$) taxonomical levels (Figure 3.3A and 3.3B). On the other hand, the functional profile did not present a clear separation between 1 and 5-year, for both Subsystem level 1 (R=0.12 $p>0.05$) and Subsystem level 2 (R=0.24 $p>0.05$) (Figure 3.3C).

Although the diversity was not statistically different ($p>0.05$) between 1-year (Shannon-Wiener H'=2.49) and 5-year (H'=2.27), the community structure became more homogeneous in the 5-year, as seen in the NMDS plot with points more disperse for samples of 1-year (Figure 3.1A).

The STAMP analysis showed an enrichment of the phyla *Proteobacteria* ($p<0.001$), *Verrucomicrobia* ($p<0.001$), and *Planctomycetes* ($p<0.01$) in the 5-year samples and a decrease of *Actinobacteria* ($p<0.001$) (Figure 3.4). Regarding the functional profile, the subsystem "protein metabolism" ($p<0.01$) was more abundant in the sample of the 1-year crop site, whilst the subsystem "regulation and cell signaling" ($p<0.05$) was more abundant in 5-year sites (Figure 3.5).

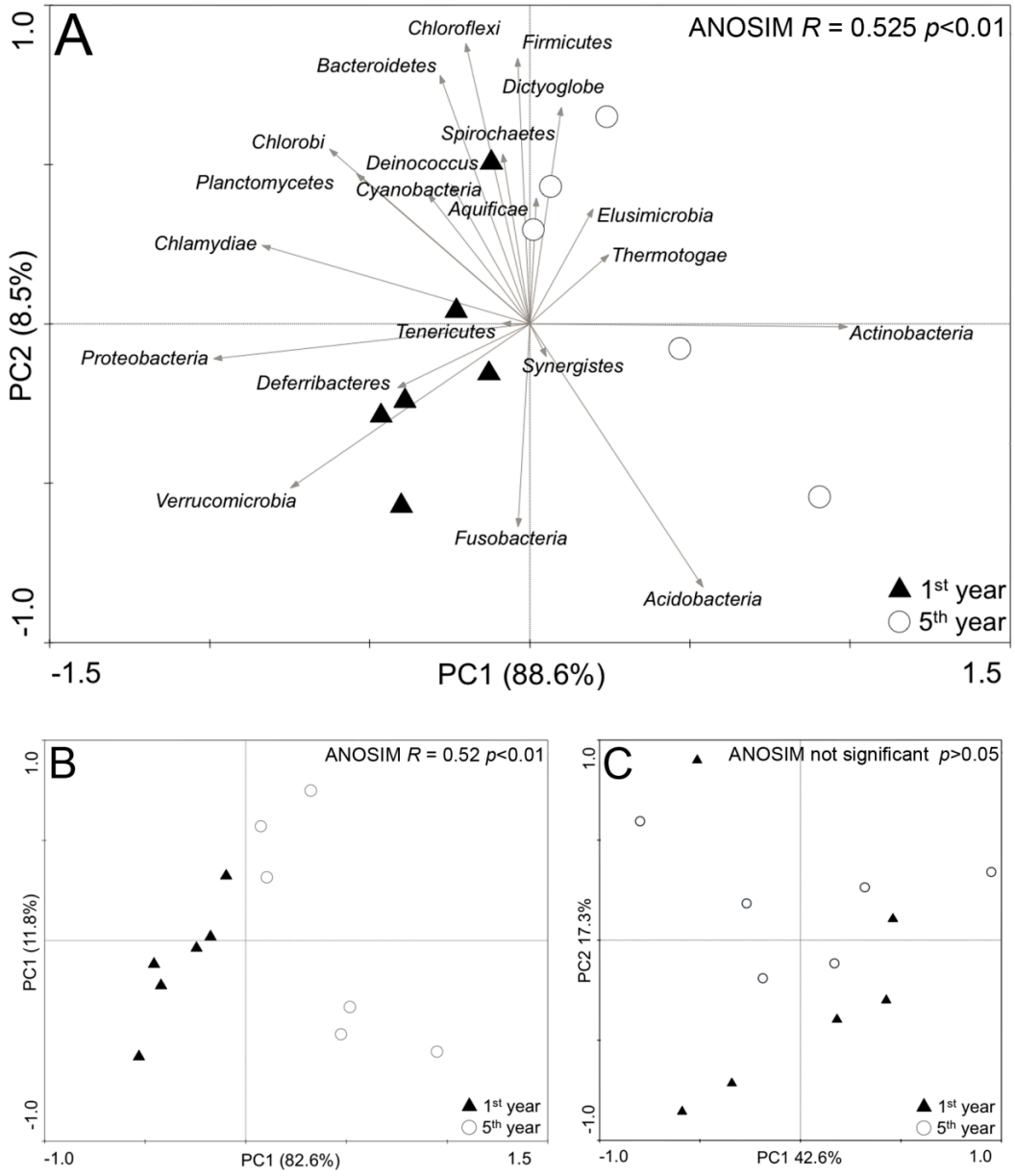


Figure 3.3 - PCA performed on taxonomic and functional profiles (SEED Database) between 1-year and 5-year soybean rhizosphere samples. (A) Taxonomic biplot at phylum level. (B) Taxonomic plot at class level. (C) Functional plot at level 1. Similarity values (ANOSIM) are shown in the upper right of each plot

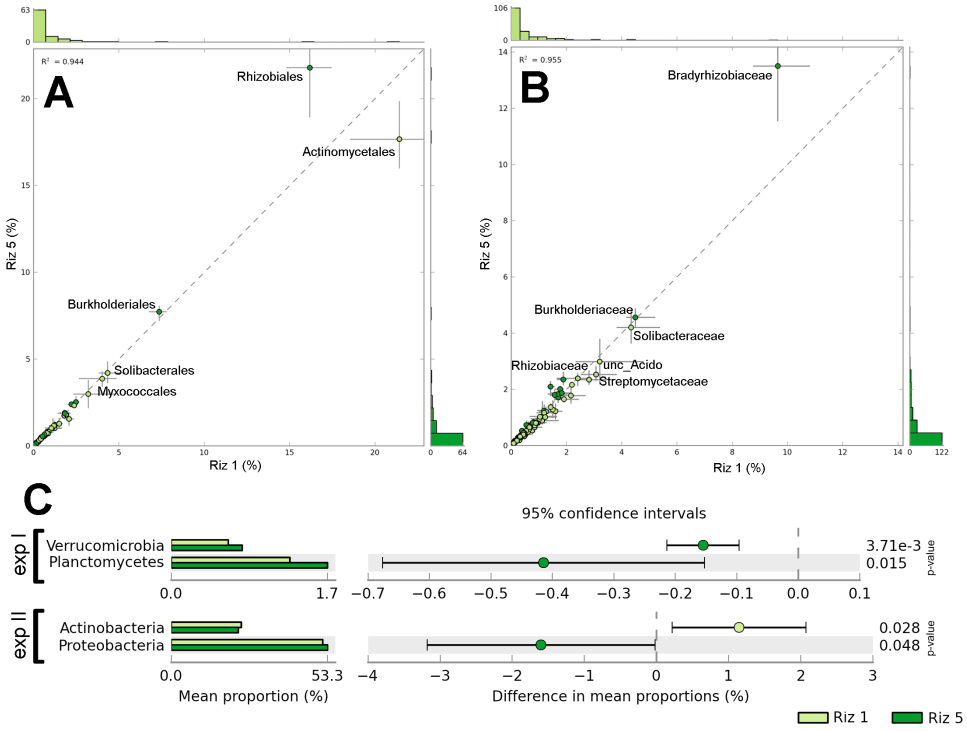


Figure 3.4 - Comparison of taxonomic profile between 1-year (light green) and 5-year (dark green) soybean rhizosphere samples. (A) Scatter-plot showing differences at order level ($p<0.05$). (B) Scatter-plot showing differences at family level ($p<0.05$). (C) Differences in phylum abundance between 1-year and 5-year rhizosphere samples for the two years of experiments (I and II). Corrected *P*-values were calculated using Benjamini-Hochberg FDR approach ($p<0.05$)

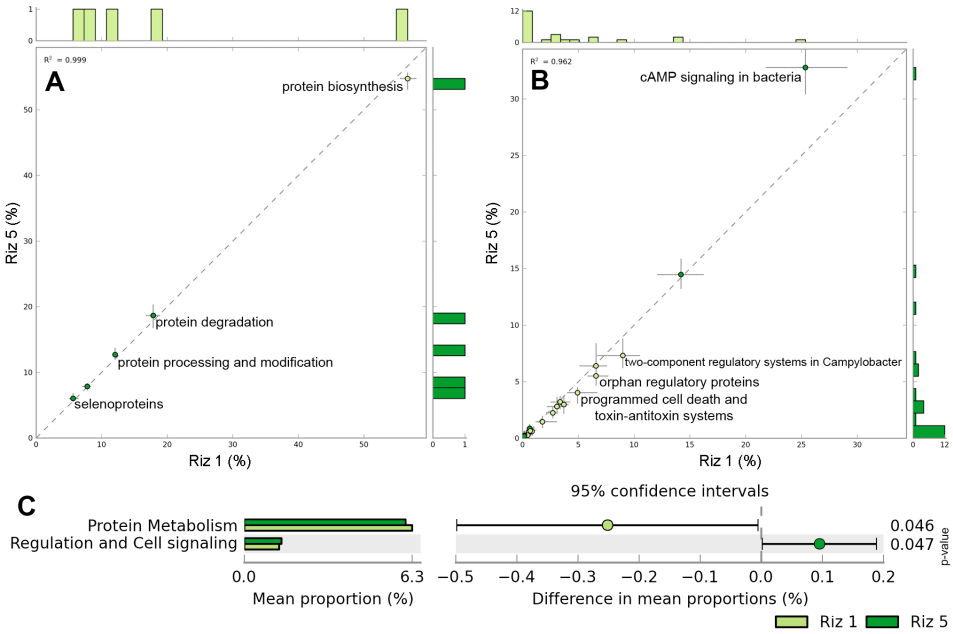


Figure 3.5 - Comparison of functional profile between 1-year (light green) and 5-year (dark green) soybean rhizosphere samples. (A) Scatter-plot showing differences for protein metabolism (SEED level 2). (B) Scatter-plot showing differences for regulation and cell signaling (SEED level 2). (C) Functional groups (SEED level 1) statistically different between 1-year and 5-year rhizosphere. Corrected *P*-values were calculated using Benjamini-Hochberg FDR approach ($p<0.05$)

3.4 Discussion

The ordination of the taxonomic profiles revealed a clear separation between rhizosphere and bulk soil samples, which indicates a selective change in the bacterial community structure in the rhizosphere as has been shown by others (BERG; SMALLA, 2009; MARSCHNER et al., 2001; UROZ et al., 2010). This influence is related to plant species and soil characteristics (BERG; SMALLA, 2009; MARSCHNER et al., 2001; KOWALCHUK et al., 2002). Remarkably, the NMDS for functional profiles did not show a clear separation between rhizosphere and bulk soil samples. This suggests that the taxonomy does not reflect the functional profile of a community. In a recent study on bacterial communities associated with the green macroalga, Burke et al. (2011) suggested that the understanding of a community structure should not be based only on taxonomy but rather on the functional level. However, it is known that the community present in the rhizosphere is originated from the bulk soil, which serves as species reservoir. In a previous study, Ridder-Duine et al. (2005) showed that the bacterial community structure within the rhizosphere is determined by the bulk soil microbial composition. Also, plant species may affect the indigenous microbial community in the bulk soil, selecting for a specific population. Thus, the rhizospheric community is a subset filtered via niche utilization (LUNDBERG et al., 2012).

In order to gain a more integrated understanding of microbial community composition and functional traits, a network analysis was accomplished. Network analyzes have been used by biologists, mathematicians, social scientists, and computer scientists to explore interactions between entities. However, only recently network analysis has been applied to microbial ecology, exploring co-occurrence patterns between microbial taxa in complex communities (BARBERAN et al., 2012; FAUST; RAES, 2012). Our analysis was constructed using all positives and negatives correlations between taxonomic and functional groups. In general, the number of correlations was lower in rhizosphere when compared to bulk soil. Considering the presence of a specific microbial community inhabiting the rhizosphere, a less complex network would confirm the occurrence of a selection process in the rhizosphere at taxonomic level.

From the top five bacterial groups that presented most correlations with other bacterial groups and functional categories, Deltaproteobacteria, Chloroflexi and Sphingobacteria are representatives in both soybean rhizosphere and bulk soil. On the other hand, exclusively in

the soybean rhizosphere the groups *Solibacteres* and *Gammaproteobacteria* presented high number of correlations, which interestingly are not the most abundant groups. Thus, these data suggest an important role of rare groups in community assembly, by keeping important connections on a larger scale with other groups and displaying important functional traits. Within the class *Gammaproteobacteria*, the two most abundant orders were *Enterobacteriales* and *Pseudomonadales*, which were showed to preferentially inhabit the rhizosphere of some plants and are known for their beneficial effects on plant growth and/or protection against pathogens (HAICHAR et al., 2008).

In our study, some specific taxonomic and functional groups were more representative in the rhizosphere than in the bulk soil. The affiliation of the sequences in the SEED database and further analysis in STAMP indicated five functional cores overrepresented in the rhizosphere, *i.e.* 'membrane transport', 'nitrogen metabolism', 'phosphorus metabolism', 'potassium metabolism' and 'iron acquisition and metabolism'. The ecological relevance of the traits are:

Membrane transport. We found a high abundance of sequences affiliated to "secretion system type IV". The "secretion system type IV" is associated to symbiotic interactions between bacterial community and other organisms (BURKE et al., 2011). Interestingly, even with the inoculation of *Bradyrhizobium* (*Proteobacteria*) we still found "secretion system type IV" binned to other bacterial phyla as *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Proteobacteria*.

Nitrogen Metabolism. We found a high abundance of genes involved with nitrogen fixation, denitrification and nitrite and nitrate ammonification in the rhizosphere. Nitrogen is generally the most limiting element to plant growth, and most of the nitrogen taken up by plants is from the soil in the forms of nitrate. The "nitrosative stress" function was also found representative in the rhizosphere. This function is related to a de-regulated synthesis or overproduction of NO and NO-derived, products that can have toxic physiological consequences to the plant (CORPAS et al., 2011). These sequences assigned to nitrogen metabolism could be related to the effect of the fertilizer treatment in the field and plant nutrition. Although we have inoculated *Bradyrhizobium* in the experiment, other groups related to nitrogen fixation were abundant, i.e. *Rhizobium*, *Sinorhizobium*, *Azorhizobium* and *Mesorhizobium*.

Phosphorus Metabolism. The annotation of sequences related to phosphorus metabolism in rhizosphere revealed a predominance of genes related to P uptake and alkylphosphonate utilization indicating enhancement of plant P availability by P solubilization and mineralization and decreasing the pH (MARSCHNER et al., 2011).

Potassium Metabolism. Potassium plays an important role in the growth and development of plants, consisting one of the essential macronutrient needed. Microorganisms play a key role in the K cycle, with specific groups, as *Pseudomonas*, *Bacillus*, *Rhizobium*, and *Flavobacterium* capable of mobilizing potassium in accessible form in soils (HU et al., 2006).

Iron acquisition and metabolism. Genes related to "iron acquisition in vibrio", "heme and hemin uptake and utilization system in gram negatives" were observed, which might be related to an increase of iron acquisition by the plant. The acquisition of iron is poorly available in most of soil and, in the rhizosphere, Fe is mobilized by plant- or microbe-derived compounds, and there is an intense competition for uptake (MARSCHNER et al., 2011). The rhizosphere microorganisms may enhance or decrease plant Fe uptake in a several ways (for review see MARSCHNER et al., 2011).

Thus, the enrichment of these five functional traits in the rhizosphere indicates a selection of the microbial community based on functional categories that is related to plant nutrition. As a response to a long-term land use, the rhizospheric community composition of 5-year of soybean harvest showed some degree of a homogenization when compared to the 1-year (Figure 3.1A). This process is a common result of ecosystem conversion and is driven by the increase of community similarity over time (OLDEN; POFF, 2003). Also, Rodrigues et al. (2013) showed that the conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. One of the reasons of homogenization is the increase in the ranges of existing species. Our results showed an enrichment of *Verrucomicrobia* ($p<0.001$), *Proteobacteria* and *Planctomycetes* ($p<0.01$) and a decrease in abundance of *Actinobacteria* ($p<0.01$) in the 5-year samples (Figure 3.4C).

The phylum *Actinobacteria* was more abundant in the 1-year soybean rhizosphere. This group is well known for antimicrobial secondary metabolites production, playing an important role in the decomposition of organic materials, as organic matter turnover and carbon cycle (VENTURA et al., 2007). Studies have shown a higher abundance of Actinobacteria in desert soils (FIERER et al., 2012) and deforested areas (NAVARRETE,

2012), while *Verrucomicrobia* decreased. In our experiments, areas of native forest were deforested through slash-and-burn and then prepared to agriculture. As a consequence of deforestation, a high amount of carbon from the ashes is deposited in the soil. Also, there is an increase of soil temperature. This event could explain the higher abundance of *Actinobacteria* in the 1-year samples, since they are important decomposers; also, when the soil dries out, they are capable to produce spores, which allow them to resist to perturbations events (VENTURA et al., 2007). Interestingly, in the 5-year rhizosphere samples the phylum *Verrucomicrobia* increased, indicating a recovery after the deforestation.

The phylum *Proteobacteria* was significantly abundant in 5-year soybean rhizosphere. Specific groups of *Proteobacteria* contain plant-growth-promoting members, as *Gammaproteobacteria*, which was abundant in rhizosphere samples, and respond chemotactically to exudates and are efficient in the utilization of plants exudate products (GARCÍA-SALAMANCA et al., 2012). The class *Alphaproteobacteria* presented a significant enrichment in the 5-year soybean rhizosphere. Within this class, groups related to the nitrogen cycle were abundant in the rhizosphere. Genera related to nitrogen fixing found were *Rhizobium*, *Sinorhizobium*, *Bradyrhizobium*, *Azorhizobium*, *and Mesorhizobium*. Therefore, the long-term land use had a directed effect towards enrichment of specific bacterial groups.

In relation to functional traits, in the 5-year rhizosphere samples the "protein degradation, processing and modification" were abundant, which indicates intense bacterial activity in the rhizosphere. Also, within "regulation and cell signaling", the "cAMP signaling in bacteria" was abundant in 5-year samples. The cAMP signaling system regulates important metabolic pathways through signaling cascade used in cell communication (AGARWAL; BISHAI, 2009), which is part of a complex system that drives basic cellular activities and coordinates cell actions. The high abundance of sequences related to cAMP signaling in 5-year rhizosphere may indicate a mechanism in which some species perceive and correctly respond to the microenvironment of the rhizosphere.

Ultimately, the study of complex communities gives a better understanding of microbial community assembly, which is important to clarify our comprehension about the ecology of microorganisms. Our results showed a power of selection over the rhizospheric community assembly. It is well known that community assembly is highly dependent on a multitude of trophic influences, which is dependent of the environmental biological diversity (CARUSO et al., 2011). Two theories have been raised to explain the microbial community

assembly: niche theory, which considers the importance of deterministic processes, and neutral theory, which focuses on stochastic process (LEIBOLD; McPEEK, 2006). Both theories, albeit viewed as contradictory, are not mutually exclusive. In some systems, both deterministic (niche theory) and stochastic (neutral theory) processes are responsible for structuring ecological communities (CHAVE, 2004). Our data pointed a microbial community selection in the rhizosphere via niche filtering. Niche-based process appeared to play a major role in the microbial community assembly in the rhizosphere, while the bulk soil composition and structure seemed to be regulated by neutral process. Also the selection at functional level in the rhizosphere seems to be based on process of the niche based theory. This trend is more evident for the 5-year samples, which indicate that the power of selection increased after 5 years of soybean harvesting.

Although most of rhizosphere samples fitted niche-based model, some samples fitted the ZSM, indicating a contribution from neutral process, mainly for the first year of soybean harvesting. Some studies also have been shown that both deterministic and neutral processes are evident in structuring microbial communities (CARUSO et al., 2011; DUMBRELL et al., 2010; FERRENBERG et al., 2013). The long-term land use drives to an increased selective power of bacterial groups linked to functions that are beneficial to the plant.

3.5 Conclusion

Our results suggest that soybean selects a specific microbial community inhabiting the rhizosphere based on functional traits, which may be related to benefits to the plant, as growth promotion and nutrition. This selection follows largely the niche –based theory indicating the selection power of the plant and other environmental variables in shaping the microbial community both at the taxonomic and functional level. Long-term cultivation strengthens the selective power of the crop to communities and functions that are beneficial to the plant. Further analysis are needed to better understand the mechanisms by which the plant selects the rhizospheric community, whereby the study of roots exudates and its influence in shaping microbial communities is of prime importance.

REFERENCES

AGARWAL, N.; BISHAI, W.R. cAMP signaling in Mycobacterium tuberculosis. **Indian Journal of Experimental Biology**, New Delhi, v. 47, p. 393-400, 2009.

AZIZ, R.K.; BARTELS, D.; BEST, A.A.; DEJONGH, M.; DISZ, T.; EDWARDS, R.A.; FORMSMA, K.; GERDES, S.; GLASS, E.M.; KUBAL, M.; MEYER, F.; OLSEN, G.J.; OLSON, R.; OSTERMAN, A.L.; OVERBEEK, R.A.; MCNEIL, L.K.; PAARMANN, D.; PACZIAN, T.; PARRELLO, B.; PUSCH, G.D.; REICH, C.; STEVENS, R.; VASSIEVA, O.; VONSTEIN, V.; WILKE, A.; ZAGNITKO, O. The RAST Server: Rapid annotation using Subsystems Technology. **BMC Genomics**, London, v. 9, p. 75, 2008.

BARBERÁN, A.; BATES, S.T.; CASAMAYOR, E.O.; FIERER, N. Using network analysis to explore co-occurrence patterns in soil microbial communities. **The ISME Journal**, London, v. 6, p. 343-351, 2012.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. In: INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA, 3, 2009, San Jose, California. **Proceedings…** Menlo Park: AAAI Press, 2009. Available at: http://aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society, Series B**, London, v. 57, p. 289-300, 1995.

BERG, G.; SMALLA, K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. **FEMS Microbiology Ecology**, Amsterdam, v. 68, p. 1-13, 2009.

BORNEMAN, J.; TRIPLETT, E.W. Molecular microbial diversity in soils from eastern Amazonia: Evidence for unusual microorganisms and microbial population shifts associated with deforestation. **Applied Environmental Microbiology**, Baltimore, v. 63, p. 2647-2653, 1997.

BRODY, J.R.; KERN, S.E. Sodium boric acid: Atriz-less, cooler conductive medium for DNA electrophoresis. **BioTechniques**, New York, v. 36, p. 214-216, 2004

BULGARELLI, D.; ROTT, M.; SCHLAEPPI, K.; VAN THEMAAT, E.V.L.; AHMADINEJAD, N.; ASSENZA, F.; RAUF, P.; HUETTEL, B.; REINHARDT, R.; SCHMELZER, E.; PEPLIES, J.; GLOECKNER, F.O.; AMANN, R.; EICKHORST, T.; SCHULZE-LEFERT, P. Revealing structure and assembly cues for Arabidopsis roo-inhabiting bacterial microbiota. **Nature**, London, v. 488, p. 91-95, 2012.

BURKE, C.; STEINBERG, P.; RUSCH, D.; KJELLEBERG, S.; THOMAS, T. Bacterial community assembly based on functional genes rather than species. **Proceedings of the National Academy of Sciences of the USA,** Washington, DC, v. 108, p. 14288-14293, 2011.

CALVARUSO, C.; MARESCHAL, L.; TURPAULT, M.P.; LECLERC, E. Rapid clay weathering in the rhizosphere of Norway spruce and oak in an acid forest ecosystem. **Soil Science Society of America Journal**, Madison, v. 73, p. 331-338, 2009.

CARUSO, T.; CHAN, Y.; LACAP, D.C.; LAU, M.C.Y.; MCKAY, C.P.; POINTING, S.B. Stochastic and deterministic processes interact in the assembly of desert microbial communities on a global scale. **The ISME Journal**, London, v. 5, p. 1406-1413, 2011.

CHAVE, J. Neutral theory and community ecology. **Ecology Letters**, Oxford, v. 7, p. 241-253, 2004.

CORPAS, F.J.; LETERRIER, M.; VALDERRAMA, R.; AIRAKI, M.; CHAKI, M.; PALMA, J.M., BARROSO, J.B. Nitric oxide imbalance provokes a nitrosative response in plants under abiotic stress. **Plant Science**, Shannon, v. 181, p. 604-611, 2011.

DeANGELIS, K.M.; BRODIE, E.L.; DESANTIS, T.Z.; ANDERSEN, G.L.; LINDOW, S.E.; FIRESTONE, M.K. Selective progressive response of soil microbial community to wild oat roots. **The ISME Journal,** London, v. 3, p. 168-178, 2008.

DELMOND, T.O.; PRESTAT, E.; KEEGAN, K.P.; FAUBLADIER, M.; ROBE, P.; CLARK, I.M.; PELLETIER, E.; HIRSCH, P.R.; MEYER, F.; GILBERT, J.A.; LE PASLIER, D.; SIMONET, P.; VOGEL, T.M. Structure, fluctuation and magnitude of a natural grassland soil metagenome. **The ISME Journal**, London, v. 6, p. 1677-1687, 2012.

DINSDALE, E.A.; EDWARDS, R.A.; HALL, D.; ANGLY, F.; BREITBART, M.; BRULC, J.M.; FURLAN, M.; DESNUES, C.; HAYNES, M.; LI, L.; MCDANIEL, L.; MORAN, M.A.; NELSON, K.E.; NILSSON, C.; OLSON, R.; PAUL, J.; BRITO, B.R.; RUAN, Y.; SWAN, B.K.; STEVENS, R.; VALENTINE, D.L.; THURBER, R.V.; WEGLEY, L.; WHITE, B.A.; ROHWER, F. Functional metagenomic profiling of nine biomes. **Nature**, London, v. 452, p. 629-632, 2008.

DUMBRELL, A.J.; NELSON, M.; HELGASON, T.; DYTHAM, C.; FITTER, A.H. Relative roles of niche and neutral process in structuring a soil microbial community. **The ISME Journal**, London, v. 4, p. 337-345, 2010.

FALKOWSKI, P.G.; FENCHEL, T.; DELONG, E.F. The microbial engines that drive Earth's biogeochemical cycles. **Science**, Washington, DC, v. 320, p. 1034-1039, 2008.

FAUST, K.; RAES, J. Microbial interactions: from networks to models. **Nature Reviews Microbiology**, London, v. 10, p. 538-550, 2012.

FEINSTEIN, L.M.; BLACKWOOD, C.B. Taxa-area relationship and neutral dynamics influence the diversity of fungal communities on senesced tree leaves. **Environmental Microbiology**, Oxford, v. 14, p. 1488-1499, 2012.

FERRENBERG, S.; O'NEILL, S.P.; KNELMAN, J.E.; TODD, B.; DUGGAN, S.; BRADLEY, D.; ROBINSON, T.; SCHMIDT, S.K.; TOWNSEND, A.R.; WILLIAMS, M.W.; CLEVELAND, C.C.; MELBOURNE, B.A.; JIANG, L.; NEMERGUT, D.R. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. **The ISME Journal**, London, 2013. doi:10.1038/ismej.2013.11.

FIERER, N.; LEFF, J.W.; ADAMS, B.J.; NIELSEN, U.N.; BATES, S.T.; LAUBER, C.L.; OWENS, S.; GILBERT, J.A.; WALL, D.H.; CAPORASO, J.G. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. **Proceedings of the National Academy of Sciences of the USA,** Washington, DC, v. 109, p. 21390-21395, 2012.

FISHER, W.D. On grouping for maximum homogeneity. **Journal of the American Statistical Association**, Boston, v. 53, p. 789-798, 1958.

FROSTEGARD, A.; TUNLID, A.; BÅÅTH, E. Use and misuse of PFLA measurements in soils. **Soil Biology and Biochemistry**, Oxford, v. 43, p. 1621-1625, 2011.

GARCÍA-SALAMANCA, A.; MOLINA-HENARES, M.A.; van DILLEWIJN, P.; SOLANO, J.; PIZARRO-TOBÍAS, P.; ROCA, A.; DUQUE, E.; RAMOS, J.L. Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil. **Microbial Biotechnology**, Oxford, v. 6, p. 36-44, 2012.

GERMANO, M.G.; CANNAVAN, F.S.; MENDES, L.W.; LIMA, A.B.; TEIXEIRA, W.G.; PELLIZARI, V.H.; TSAI, S.M. Functional diversity of bacterial genes associated with aromatic hydrocarbon degradation in anthropogenic dark earth of Amazonia. **Pesquisa Agropecuária Brasileira**, Brasília, DF, v. 47, p. 654-664, 2012.

HAICHAR, F.Z.; MAROL, C.; BERGE, O.; RANGEL-CASTRO, J.I.; PROSSER, J.I.; BALESDENT, J.; HEULIN, T.; ACHOUAK, W. Plant host habitat and root exudates shape soil bacterial community structure. **The ISME Journal**, London, v. 2, p. 1221-1230, 2008.

HAMMER, Ø.; HARPER, D.A.T.; RYAN, P.D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. **Palaeontologia Electronica**, College Station, v. 4, p. 9, 2001.

HU, X.; CHEN, J.; GUO, J. Two phosphate- and potassium-solubilizing bacteria isolated from Tianmu Mountain, Zhejiang, China. **World Journal of Microbiology and Biotechnology**, Oxford, v. 22, p. 983-990, 2006.

HUBBELL, S.P. **The unified neutral theory of biodiversity and biogeography**. Princeton, NJ: Princeton University Press, 2001. (Monographs in Population Biology, 32).

JABOT, F.; ETIENNE, R.F.; CHAVE, J. Reconciling neutral community models and environmental filtering: theory and an empirical test. **Oikos**, Copenhagen, v. 117, p. 1308-1320, 2008.

JESUS, E.D.; MARSH, T.L.; TIEDJE, J.M.; MOREIRA, F.M.D. Changes in land use alter the structure of bacterial communities in Western Amazon soils. **The ISME Journal**, London, v. 3, p. 1004-1011, 2009.

JONGMAN, R.H.G.; Ter BRAAK C.F.J.; Van TONGEREN, O.F.R. **Data Analysis in Community and Landscape Ecology**. Cambridge, UK: Cambridge University Press, 1995.

KNIEF, C.; DELMOTTE, N.; CHAFFRON, S.; STARK, M.; INNEREBNER, G.; WASSMANN, R.; von MERING, C.; VORHOLT, J.A. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. **The ISME Journal**, London, v. 6, p. 1378-1390, 2011.

KOWALCHUK, G.A.; BUMA, D.S.; de BOER, W.; KLINKHAMER, P.G.L.; van VEEN, J.A. Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. **Antonie Van Leeuwenhoek**, Dordrecht, v. 81, p. 509-520, 2002.

KURAMAE, E.E.; YERGEAU, E.; WONG, L.C.; PIJL, A.S.; van VEEN, J.A.; KOWALCHUK, G.A. Soil characteristics more strongly influence soil bacterial communities than land-use type. **FEMS Microbiology Ecology**, Amsterdam, v. 79, p. 12-24, 2012.

LANGENHEDER, S.; BULLING, M.T.; SOLAN, M.; PROSSER, J.I. Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity. **PLoS One**, New York, v. 5, e10834, 2010.

LEIBOLD, M.A.; MCPEEK, M.A. Coexistence of the niche and neutral perspectives in community ecology. **Ecology**, Brooklyn, v. 87, p. 1399–1410, 2006.

LINDSTRÖM, E.S.; LANGENHENDER, S. Local and regional factors influencing bacterial community assembly. **Environmental Microbiology Reports**, Hoboken, v. 4, p. 1-9, 2012.

LUNDBERG, D.S.; LEBEIS, S.L.; PAREDES, S.H.; YOURSTONE, S.; GEHRING, J.; MALFATTI, S.; TREMBLAY, J.; ENGELBREKTSON, A.; KUNIN, V.; del RIO, T.G.; EDGAR, R.C.; EICKHORST, T.; LEY, R.E.; HUGENHOLTZM, P.; TRINGE, S.G.; DANGL, F.L. Defining the core Arabidopsis thaliana root microbiome. **Nature**, London, v. 488, p. 86-90, 2012.

MACARTHUR, R. On the relative abundance of bird species. **Proceedings of the National Academy of Sciences of the USA,** Washington, DC, v. 43, p. 293-295, 1975.

MARSCHNER, P.; YANG, C.H.; LIEBEREI, R.; CROWLEY, D.E. Soil and plant specific effects on bacterial community composition in the rhizosphere. **Soil Biology and Biochemistry**, Oxford, v. 33, p. 1437-1445, 2001.

MARSCHNER, P.; CROWLEY, D.; RENGEL, Z. Rhizosphere interactions between microorganisms and plants govern iron and phosphorus acquisition along the root axis – model and research methods. **Soil Biology and Biochemistry**, Oxford, v. 43, p. 883-894, 2011.

MATO GROSSO (Estado). Secretaria de Estado de Planejamento d Coordenação Geral – SEPLAN. **Mapa de solos do Estado de Mato Grosso.** Mato Grosso, 2001. Available at: www.seplan.mt.gov.br. Accessed at: March 20 2012.

MCGILL, B.J.; RAMPAL, S.; ETIENNE, J.S.; GRAY, D.A.; ANDERSON, M.J.; BENECHA, H.K.; DORNELAS, M.; ENQUIST, B.J.; GREEN, J.L.; HE, F.; HURLBERT, A.H.; MAGURRAN, A.E.; MARQUET, P.A.; MAURER, B.A.; OSTLING, A.; SOYKAN, C.U.; UGLAND, K.I.; WHITE, E.P. Species abundance distribution: moving beyond single prediction theories to integration within an ecological framework. **Ecology Letters**, Oxford, v. 10, p. 995-1015, 2007.

MCGILL, B.J.; MAURER, B.A.; WEISER, M.D. Empirical evaluation of neutral theory. **Ecology**, Ithaca, v. 87, p. 1411-1423, 2006.

MEYER, F.; PAARMAN, D.; D'SOUZA, M.; OLSON, R.; GLASS, E.M.; KUBAL, M.; PACZIAN, T.; RODRIGUEZ, A.; STEVENS, R.; WILKE, A.; WILKENING, J.; EDWARDS, R.A. The Metagenomics RAST server – A public resource for the automatic phylogenetic and functional analysis of metagenomes. **BMC Bioinformatics**, London, v. 9, p. 386, 2008.

MOTURA, I. On the statistical treatment of communities. **Zoological Magazine**, Tokyo, v. 44, p. 379-383, 1932.

NAVARRETE, A.A.; KURAMAE, E.E.; de HOLLANDER, M.; PIJL, A.S.; van VEEN, J.A.; TSAI, S.M. Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. **FEMS Microbiology Ecology**, Amsterdam, v. 83, p. 607-621, 2013.

NAVARRETE, A.A. **Bacterial ecology in Amazonian soils under deforestation and agricultural management**. 2012. 136 p. Tese (Doutorado em Ciências) – Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, 2012.

NAVARRETE, A.A.; TAKETANI, R.G.; MENDES, L.W.; CANNAVAN, F.S.; MOREIRA, F.M.S.; TSAI, S.M. Land-use systems affect archaeal community structure and functional diversity in western Amazon soils. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 35, p. 1527-1540, 2011.

NAVARRETE, A.A.; CANNAVAN, F.S.; TAKETANI, R.G.; TSAI, S.M. A molecular survey of the diversity of microbial communities in different Amazonian agricultural model systems. **Diversity**, Bethesda, v. 2, p. 787-809, 2010.

NEMERGUT, D.R.; COSTELLO, E.K.; HAMADY, M.; LOZUPONE, C.; JIANG, L.; SCHMIDT, S.K.; FIERER, N.; TOWNSEND, A.R.; CLEVELAND, C.C.; STANISH, L.; KNIGHT, R. Global patterns in the biogeography of bacterial taxa. **Environmental Microbiology**, Oxford, v. 13, p. 135-144, 2011.

NEWCOMBE, R.G. Improved confidence intervals for the difference between binomial proportions based on paired data. **Statistics in Medicine**, Chichester, v. 17, p. 2635-2650, 1998.

OKSANEN, P. **Vegan 1.17-0 in R version 2.10.1.** London, 2009. Available at: http://cc.oulu.fi~jarioksa/softhelp.vegan.html, 2010.

OLDEN, J.D.; POFF, N.L. Toward a mechanism understanding and prediction of biotic homogenization. **The American Naturalist**, Chicago, v. 162, p. 442-460, 2003.

PARKS, D.H.; BEIKO, R.G. Identifying biologically relevant differences between metagenomic communities. **Bioinformatics**, Oxford, v. 26, p. 715-721, 2010.

R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2007. Available at: http://www.R-project.org.

RASCH, F.; HÖDL, V.; POLL, C.; KANDELER, E.; GERZABEK, M.H.; van ELSAS, J.D.; SESSITSCH, A. Rhizosphere bacteria affected by transgenic potatoes with antibacterial activities compared with the effects of soil, wild-type potatoes, vegetation stage and pathogen exposure. **FEMS Microbiology Ecology**, Amsterdam, v. 56, p. 219-235, 2006.

RIDDER-DUINE, A.S.; KOWALCHUK, G.A.; GUNNEWIEK, P.J.A.K.; SMANT, W.; van VEEN, J.A.; de BOER, W. Rhizosphere bacterial community composition in natural stands of *Carex arenaria* (sand sedge) is determined by bulk soil community composition. **Soil Biology and Biochemistry**, Oxford, v. 37, p. 349-357, 2005.

ROBIN, A.; MOUGEL, C.; SIBLOT, S.; VANSUYT, G.; MAZURIER, S.; LEMANCEAU, P. Effect of ferritin overexpression in tobacco on the structure of bacterial and pseudomonad communities associated with the roots. **FEMS Microbiology Ecology**, Amsterdam, v. 58, p. 492-502, 2006.

RODRIGUES, J.L.M.; PELLIZARI, V.H.; MUELLER, R.; BAEK, K.; JESUS, E.D.; PAULA, F.S.; MIRZA, B.; HAMAOUI, G.S.; TSAI, S.M.; FEIGL, B.; TIEDJE, J.M.; BOHANNAN, B.J.M.; NÜSSLEIN, K. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. **Proceedings of the National Academy of Sciences of the USA**, Washington, DC, v. 110, p. 988-993, 2013.

ROESCH, L.F.; FULTHORPE, R.R.; RIVA, A.; CASELLA, G.; HADWIN, A.K.M.; KENT, A.D.; DAROUB, S.H.; CAMARGO, F.A.; FARMERIE, W.G.; TRIPLETT, E.W. Pyrosequencing enumerates and contrasts soil microbial diversity. **The ISME Journal**, London, v. 1, p. 283-290, 2007.

SMITH, R.J.; JEFFRIES, T.C.; ROUDNEW, B.; FITCH, A.J.; SEYMOUR, J.R.; DELPIN, M.W.; NEWTON, K.; BROWN, M.H.; MITCHELL, J.G. Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems. **Environmental Microbiology**, Oxford, v. 14, p. 240-253, 2012.

TAKETANI, R.G.; TSAI, S.M. The Influence of Different Land Uses on the Structure of Archaeal Communities in Amazonian Anthrosols Based on 16S rRNA and amoA Genes. **Microbial Ecology**, New York, 59, p. 734-743, 2010.

TORSVIK, V.; ØVREÅS, L. Microbial diversity and function in soil: from genes to ecossystems. **Current Opinion in Microbiology**, London, v. 5, p. 240-5, 2002.

TORSVIK, V.; GOKSOYR, J.; DAAE, F.L. High diversity in DNA of soil bacteria. **Applied and Environmental Microbiology**, Washington, DC, v. 56, p. 782-787, 1990.

UROZ, S.; BUEE, M.; MURAT, C.; FREY-KLETT, P.; MARTIN, F. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. **Environmental Microbiology Reports**, Hoboken, v. 2, p. 281-288, 2010.

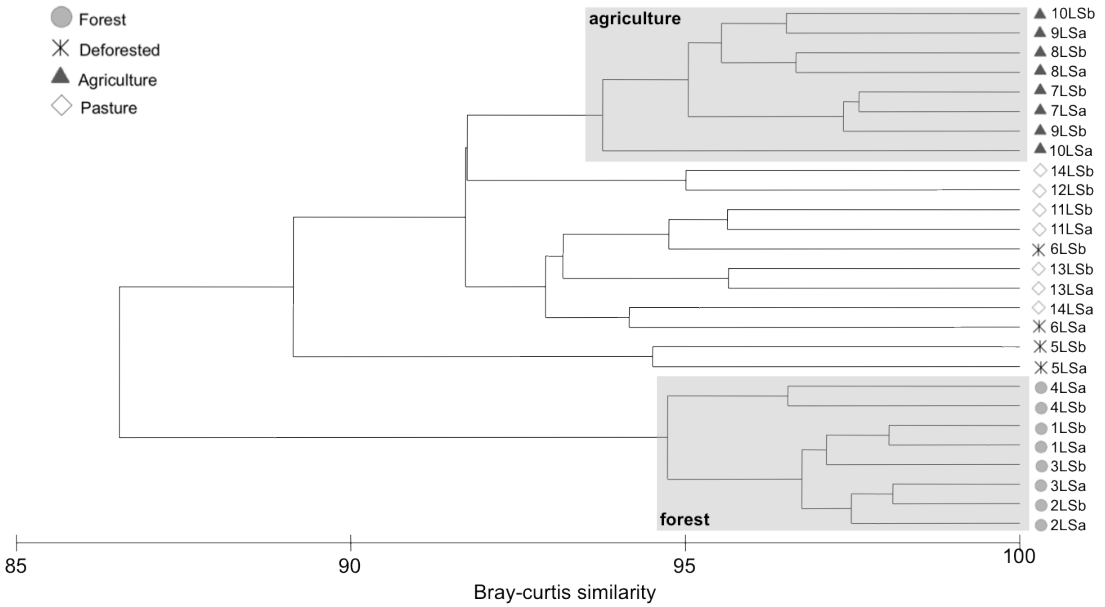
VENTURA, M.; CANCHAYA, C.; TAUCH, A.; CHANDRA, G.; FITZGERALD, G.F.; CHATER, K.F.; van SINDEREN, D. Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. **Microbiology and Molecular Biology Reviews**, Washington, DC, v. 71, p. 495-548, 2007.

XU, Y.; WANG, G.; JIN, J.; LIU, J.; ZHANG, Q.; LIU, X. Bacterial communities in soybean rhizosphere in response to soil type, soybean genotype, and their growth stage. **Soil Biology and Biochemistry**, Oxford, v. 41, p. 919-925, 2009.

APPENDICES

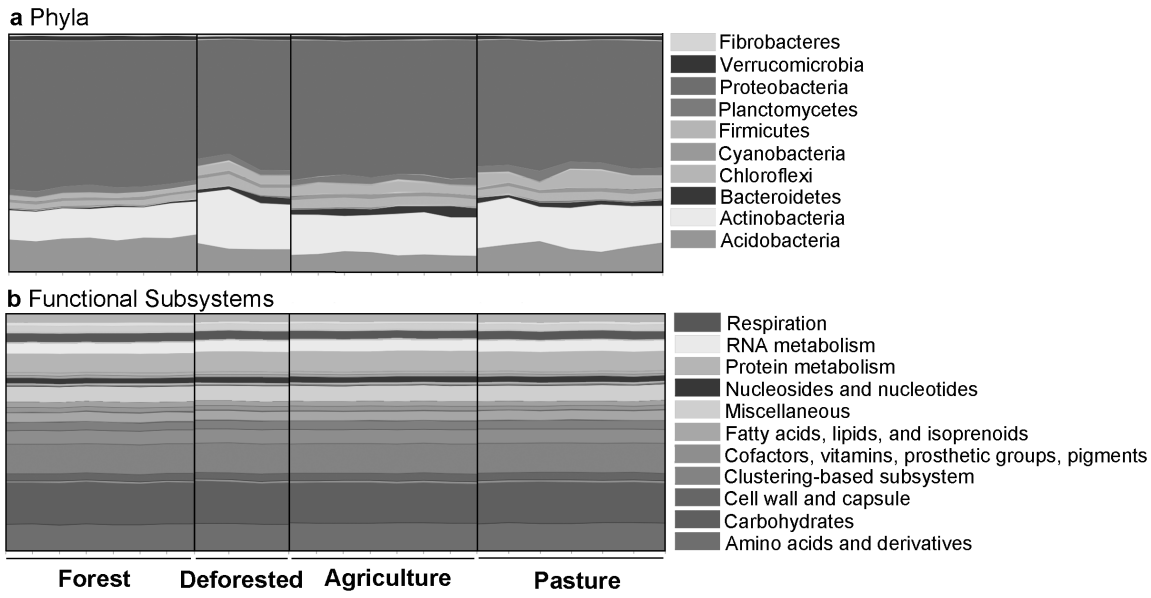
APPENDIX A: Supplementary Material of Chapter 2

Supplementary Figure 2.1.



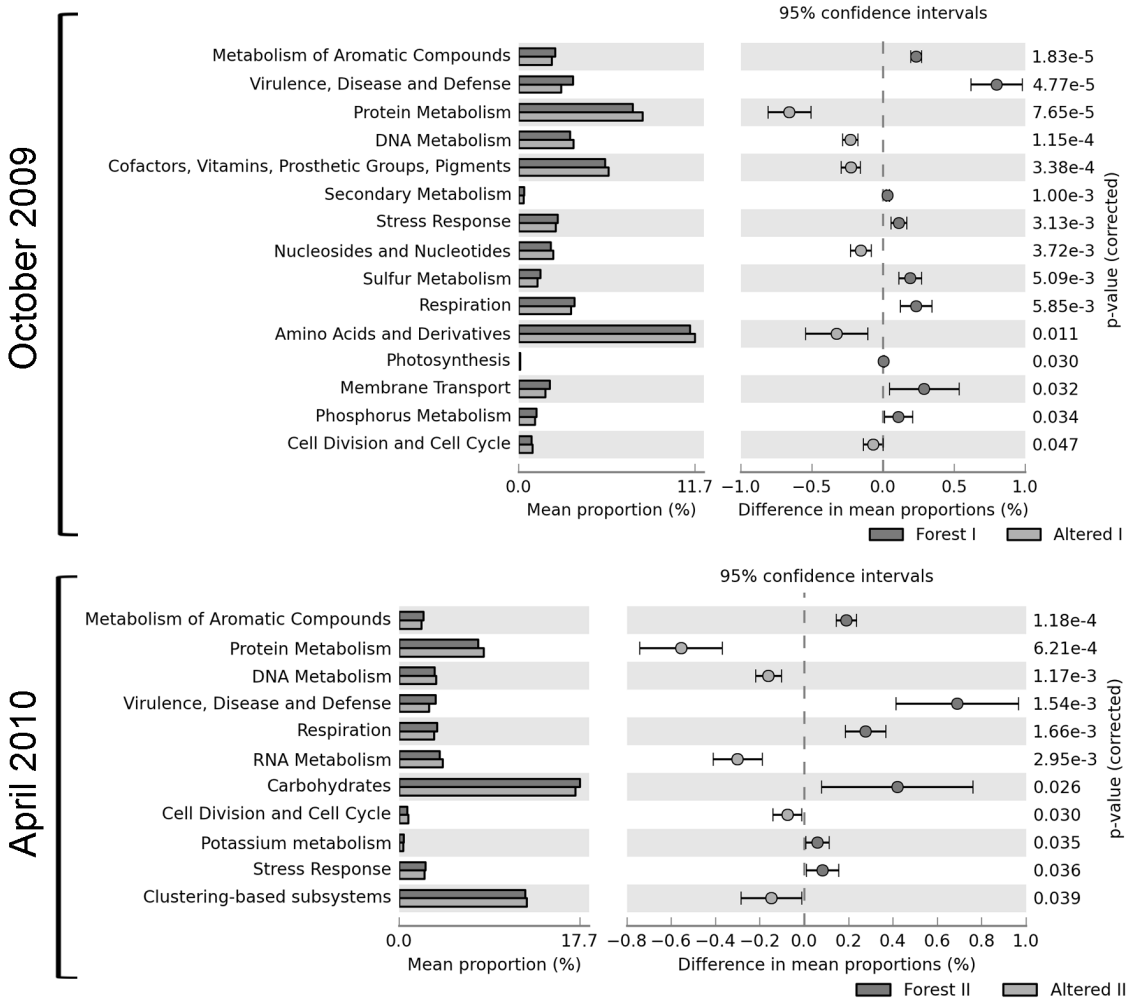
SF 2.1. Cluster dendrograms of taxonomical profile created based on Bray-Curtis similarity for microbial communities of soils from four different land-use systems. Shadowed area indicates the two main clusters for agriculture and forest samples.

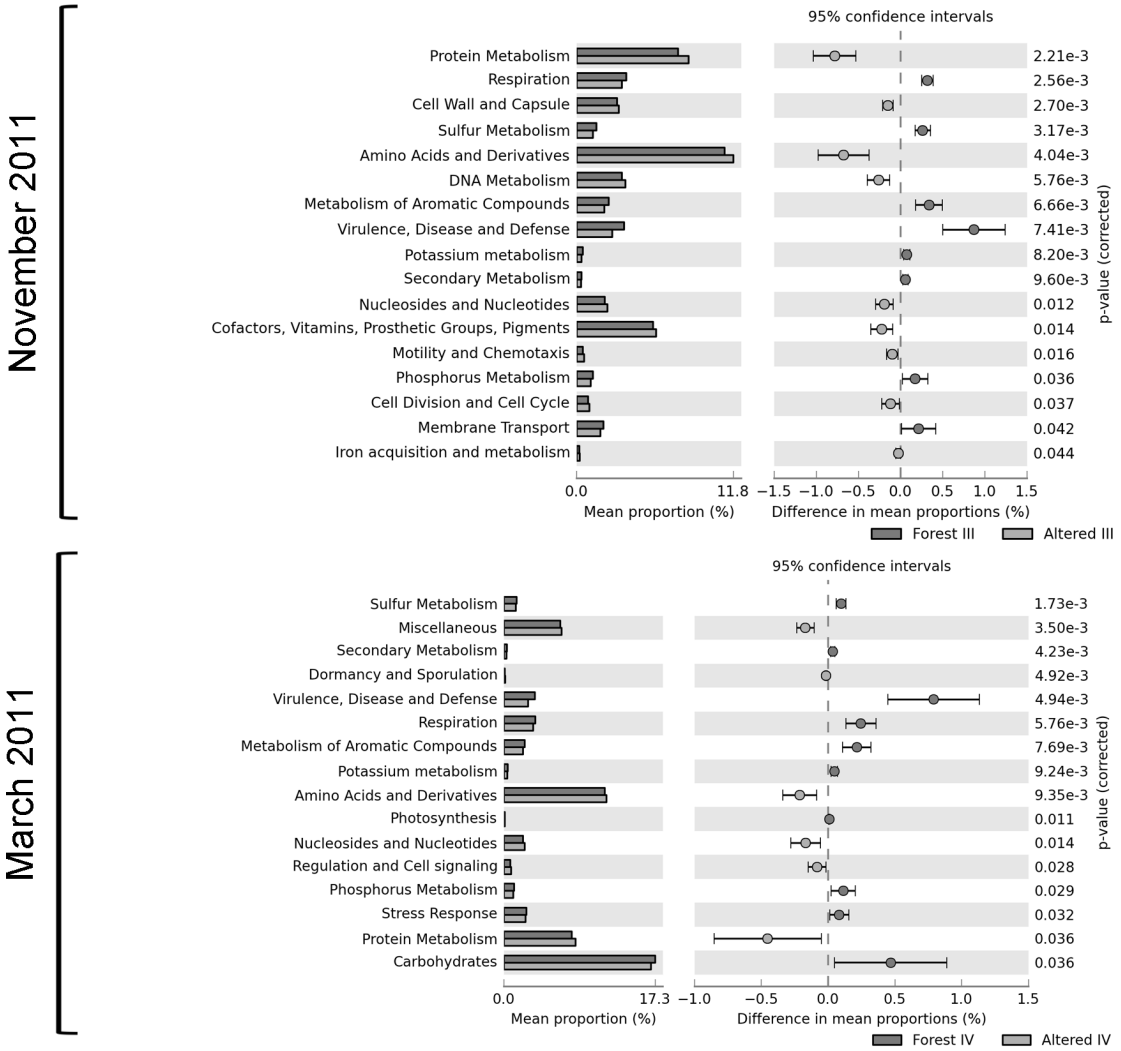
Supplementary Figure 2.2.



SF 2.2. Variation of microbial taxa (a) and functional subsystem (b) among different sites across time. Vertical bars represent microbiome samples from four land-uses type of shotgun data; bars indicate relative abundances colored by microbial phyla (a) and functional subsystems (b). Top 10 abundant phyla/subsystems are shown.

Supplementary Figure 2.3.





SF 2.3. STAMP analysis of functional profiles at subsystem level 1 (SEED database) of forest and altered soils (deforested, agriculture and pasture) for each sampling period. Groups overrepresented in forest (green) correspond to positive differences between proportions and groups overrepresented in altered soils (yellow) correspond to negative differences between proportions. Corrected *p*-values were calculated using Benjamini-Hochberg FDR approach ($p < 0.05$).

Supplementary Table 2.1. Number of sequencing reads, base pairs, reads assigned to SEED Subsystems and percentages of predict proteins after quality control on MG-RAST pipeline.

MG-RAST ID	Sample	Environment Feature	Sampling Time	Total bp	N. of sequence reads	N. of predict protein features	% of annotated proteins
4497389.3	1LSa	Native Forest	October 2009	1,55,671,061	16,373,853	13,678,973	36.9
4497390.3	1LSb	Native Forest	October 2009	1,917,707,251	20,423,704	17,100,917	29.6
4497397.3	2LSa	Native Forest	April 2010	1,717,616,735	18,337,617	15,224,358	36.7
4497398.3	2LSb	Native Forest	April 2010	1,368,742,061	14,554,642	12,238,056	37.7
4497402.3	3LSa	Native Forest	November 2010	2,091,680,663	22,245,990	18,535,746	33.8
4497400.3	3LSb	Native Forest	November 2010	1,652,011,068	17,618,750	6,686,745	37.9
4497401.3	4LSa	Native Forest	March 2011	1,232,841,286	13,127,073	10,983,520	36.3
4497402.3	4LSb	Native Forest	March 2011	1,567,594,935	16,687,613	13,973,697	35.5
4497403.3	5LSa	Deforested Site	October 2009	980,702,897	10,477,278	8,749,054	37.3
4497404.3	5LSb	Deforested Site	October 2009	1,661,893,955	17,817,773	14,770,784	37.9
4497405.3	6LSa	Deforested Site	April 2010	1,349,881,717	14,427,611	11,993,794	34.8
4497406.3	6LSb	Deforested Site	April 2010	1,835,174,499	19,593,526	16,324,269	36.4
4497407.3	7LSa	Agricultural Soil	October 2009	1,691,682,287	18,119,871	15,159,223	38.5
4497408.3	7LSb	Agricultural Soil	October 2009	2,985,513,423	31,966,967	26,511,665	35.7
4497409.3	8LSa	Agricultural Soil	April 2010	1,449,085,730	15,455,238	12,872,776	37.2
4497410.3	8LSb	Agricultural Soil	April 2010	2,340,921,723	25,017,093	9,116,427	36.4
4497411.3	9LSa	Agricultural Soil	November 2010	1,297,860,799	13,908,400	11,613,267	37.7
4497412.3	9LSb	Agricultural Soil	November 2010	2,224,899,417	23,867,305	19,782,081	37.7
4497370.3	10LSa	Agricultural Soil	March 2011	1,452,524,300	15,521,548	12,995,684	39.0
4497371.3	10LSb	Agricultural Soil	March 2011	1,585,023,687	16,901,918	14,272,549	38.6
4497372.3	11LSa	Pasture Soil	October 2009	1,641,076,813	17,445,400	14,529,960	35.1
4497373.3	11LSb	Pasture Soil	October 2009	1,958,519,064	20,851,107	17,406,645	35.4
-	12LSa	Pasture Soil	April 2010	fail	fail	fail	fail
4497374.3	12LSb	Pasture Soil	April 2010	1,440,853,866	15,240,438	12,811,658	35.9
4497375.3	13LSa	Pasture Soil	November 2010	1,938,015,010	20,635,352	16,932,446	32.6
4497376.3	13LSb	Pasture Soil	November 2010	1,420,896,267	15,155,481	12,697,834	-
4497377.3	14LSa	Pasture Soil	March 2011	1,994,863,535	21,256,135	17,846,111	37.2
4497378.3	14LSb	Pasture Soil	March 2011	1,305,482,377	13,926,576	11,678,649	21.0

APPENDIX B: Supplementary Material of Chapter 3

Supplementary Table 3.1. Soil chemical factors of sampling sites.

Soil Factors	1-year crop	5-year crop
pH	4.1 ± 0.3	5.0 ± 0.2
OM	25.8 ± 3.8	35.2 ± 2.4
Al	9.8 ± 3.6	2.0 ± 0.7
H+Al	42.6 ± 13.7	50.4 ± 8.6
m	52.4 ± 14.5	8.0 ± 4.41
P	11.2 ± 18.8	33.6 ± 20.9
K	0.8 ± 0.3	1.7 ± 0.7
Ca	5.2 ± 1.8	13.6 ± 3.0
Mg	2.8 ± 3.6	9.0 ± 2.1
SB	8.8 ± 3.1	24.3 ± 5.4
CEC	51.4 ± 13.9	74.7 ± 12.0
V	18.2 ± 7.7	32.4 ± 4.6
B	0.2 ± 0.03	0.1 ± 0.04
Cu	0.3 ± 0	0.3 ± 0.06
Fe	93.2 ± 49.5	47.6 ± 6.2
Mn	1.6 ± 0.5	0.5 ± 0.2
Zn	0.9 ± 1.3	2.6 ± 2.1
N	1520.0±117.7	1434.6±99.7
$NH4^{+}$	16.6 ± 1.69	15.0 ± 1.63
$NO^{-}3$	13.3 ± 3.39	7.33 ± 0.94
C	15.0 ± 1.63	16.0 ± 0.81
Sand	50.3±1.88	39.6±1.24
Silt	1.66±0.94	1.33±0.47
Clay	48.0±0.0	59.0±1.41

The values are averages based on quintuplicate sampling points in each site. Standard deviations are shown in the table.

Ca, Mg, K, Al, potential acidity (H+Al), sum of bases (SB) are expressed in $nmol.kg^{-1}$; OM and C is expressed in $g.Kg^{-1}$; P is expressed in $mg.kg^{-1}$; B, Fe, Mn, Zn, K, Cu, and cation exchange capacity in pH 7 (CEC) are expressed in $mg.dm^{-3}$; Sand, silt and clay are expressed in %; N (total), NH4+ and NO-3 are expressed in $mg.kg^{-1}$. P and K – Mehlich 1 extractor. Ca, Mg, and Al – Kcl 1N. H+Al – SMP extractor. N – Kjeldahl. NH4+ and NO-3 – Raney/Kjeldahl. OM – organic matter. m – Al saturation index. V – base saturation index.

Supplementary Table 3.2. Number of sequencing reads, base pairs, reads assigned to SEED Subsystems and percentages of predict proteins after quality control on MG-RAST pipeline.

Sample ID	MG-RAST ID	Sample	N. of sequence reads	Mean sequence length	Total bp*	N. of predict Subsystems functions*	% of predicted proteins*
1LM	4477751.3	Riz_5a_1	141,552	523±46	61,451,072	98,237	74.7
2LM	4477749.3	Riz_5a_2	151,054	522±42	63,711,128	102,880	75.4
3LM	4477755.3	Riz_5a_3	140,546	520±45	58,624,166	94,250	75.1
4LM	4477757.3	Bulk_5a_1	144,908	520±46	62,571,569	94,952	71.0
5LM	4477789.3	Bulk_5a_2	137,109	523±42	58,459,177	87,173	69.9
6LM	4477790.3	Bulk_5a_3	144,587	521±41	57,485,200	87,555	71.4
7LM	4478030.3	Riz_1a_1	145,214	518±68	58,031,403	94,180	76.4
8LM	4478037.3	Riz_1a_2	124,076	527±56	51,095,519	80,186	73.8
9LM	4478934.3	Riz_1a_3	113,953	520±36	9,678,615	9,788	75.9
10LM	4478038.3	Bulk_1a_1	140,141	526±46	58,816,390	96,369	76.8
11LM	4478222.3	Bulk_1a_2	142,020	526±49	50,166,332	77,447	72.7
12LM	4478283.3	Bulk_1a_3	140,261	527±56	60,814,070	98,142	75.8
13LM	4478290.3	Riz_5b_1	147,028	525±45	65,025,641	108,156	77.4
14LM	4478292.3	Riz_5b_2	129,829	526±45	55,750,716	88,769	74.5
15LM	4478291.3	Riz_5b_3	112,056	527±50	36,967,527	56,999	72.9
16LM	4478294.3	Bulk_5b_1	138,556	525±47	55,060,604	83,310	71.2
17LM	4478936.3	Bulk_5b_2	137,939	523±41	58,658,766	92,450	73.6
18LM	4478937.3	Bulk_5b_3	146,385	524±43	55,611,272	85,403	71.9
19LM	4478938.3	Riz_1b_1	126,119	523±47	35,784,579	59,398	77.9
20LM	4478939.3	Riz_1b_2	133,146	525±47	51,791,233	89,065	80.1
21LM	4479311.3	Riz_1b_3	151,804	524±40	68,089,631	116,414	79.5
22LM	4478940.3	Bulk_1b_1	138,073	523±42	55,607,597	92,786	77.8
23LM	4478941.3	Bulk_1b_2	143,858	525±45	55,511,047	89,673	75.6
24LM	4478943.3	Bulk_1b_3	123,137	526±50	51,204,159	83,520	76.5

* Post Quality Control

Supplementary Table 3.3. Relative proportion of matches to a given subsystem hierarchy level 1.

Subsystem hierarchy level 1	Rhizosphere		Subsystem hierarchy level 1	Bulk Soil
Carbohydrates	0.133		Clustering-based subsystems	0.143
Clustering-based subsystems	0.112		Carbohydrates	0.141
Cofactors, Vitamins, Prosthetic Groups, Pigments	0.101		Amino Acids and Derivatives	0.110
Miscellaneous	0.091		Miscellaneous	0.078
Membrane Transport	0.069		Cofactors, Vitamins, Prosthetic Groups, Pigments	0.066
Amino Acids and Derivatives	0.058		Protein Metabolism	0.065
Iron acquisition and metabolism	0.045		RNA Metabolism	0.049
Cell Wall and Capsule	0.044		Fatty Acids, Lipids, and Isoprenoids	0.046
RNA Metabolism	0.042		Cell Wall and Capsule	0.042
Protein Metabolism	0.030		DNA Metabolism	0.033
Respiration	0.030		Respiration	0.025
Metabolism of Aromatic Compounds	0.028		Virulence, Disease and Defense	0.024
Stress Response	0.027		Membrane Transport	0.024
Fatty Acids, Lipids, and Isoprenoids	0.027		Nucleosides and Nucleotides	0.023
Virulence, Disease and Defense	0.026		Stress Response	0.023
Nucleosides and Nucleotides	0.023		Metabolism of Aromatic Compounds	0.021
Regulation and Cell signaling	0.020		Regulation and Cell signaling	0.013
Phosphorus Metabolism	0.018		Sulfur Metabolism	0.012
DNA Metabolism	0.017		Cell Division and Cell Cycle	0.011
Phages, Prophages, Transposable elements, Plasmids	0.015		Motility and Chemotaxis	0.010
Motility and Chemotaxis	0.012		Phosphorus Metabolism	0.010
Nitrogen Metabolism	0.010		Phages, Prophages, Transposable elements, Plasmids	0.010
Sulfur Metabolism	0.006		Nitrogen Metabolism	0.007
Cell Division and Cell Cycle	0.005		Iron acquisition and metabolism	0.004
Potassium metabolism	0.005		Potassium metabolism	0.003
Dormancy and Sporulation	0.004		Secondary Metabolism	0.003
Secondary Metabolism	0.003		Dormancy and Sporulation	0.001
Photosynthesis	0.000		Photosynthesis	0.001

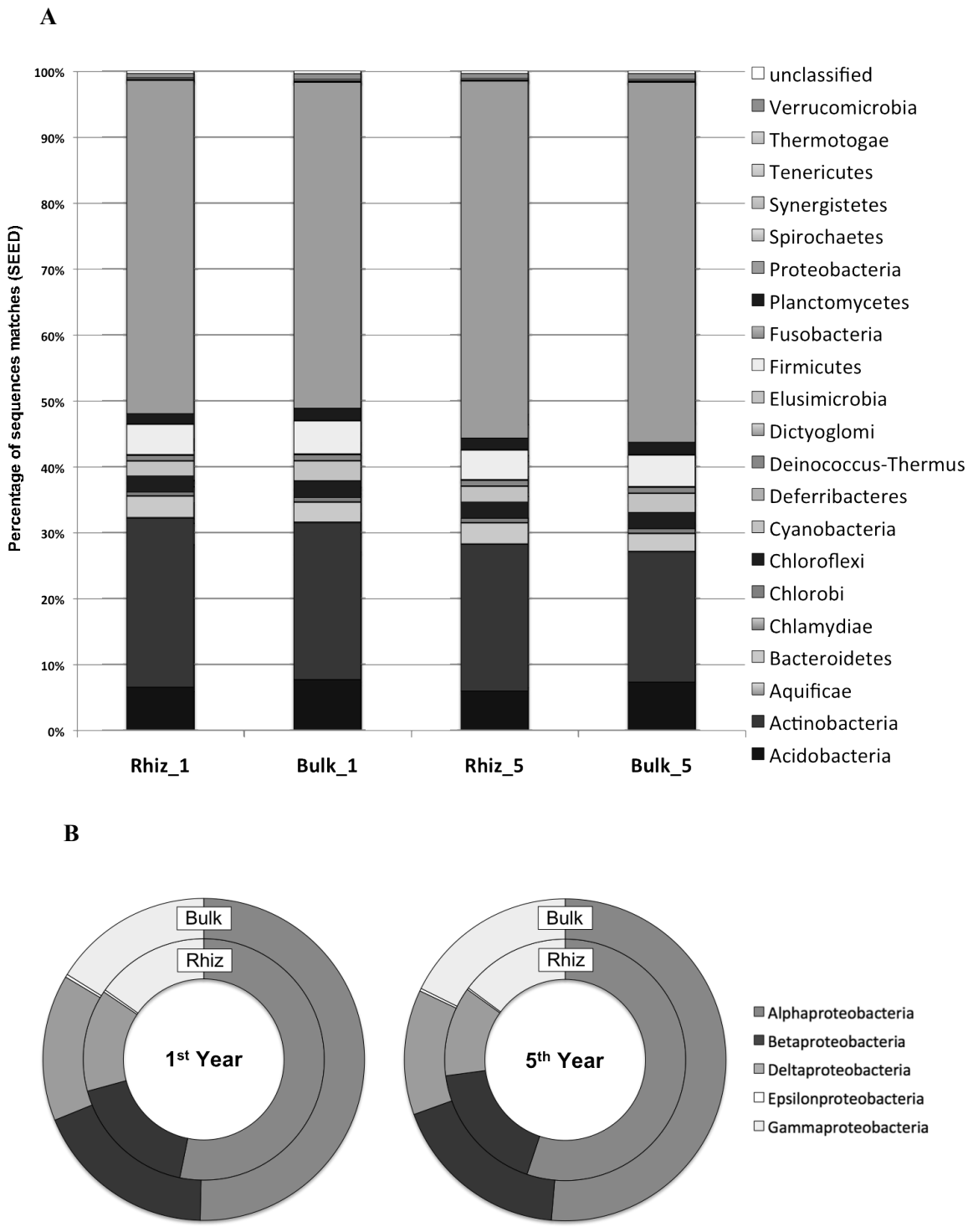
Hits were generated by blasting sequences of 24 metagenomes (12 rhizosphere and 12 bulk soil) to the MG-RAST subsystem database with a minimum alignment length of 50 bp and an E value cut-off of $1x10^{-5}$. Relative representation in the metagenomes was calculated by dividing the number of hit to each category by the total number of hits to all categories.

Supplementary Table 3.4. AIC values for 6 rank abundance distribution models. Lowest AIC values for each sample represents the best fit model.

				AIC[1]					
Sample	**Environ**	**Year**	**Exp.**	**Broken-stick**	**Pre-Emption**	**Log-Normal**	**Zipf**	**Zipf-Mandelbrot**	**ZSM**
10LM	bulk soil	1-year	A	16566.6	14433.6	3544.1	10722.1	*nd*	**3011.52***
11LM	bulk soil	1-year	A	15453.4	13136.4	3608.7	9253.8	*nd*	**3165.38***
12LM	bulk soil	1-year	A	18072.4	17130.7	3353.5	9213.3	4932.9	**3187.30***
22LM	bulk soil	1-year	B	12397.6	11266.7	3030.8	10558.3	*nd*	**2901.26***
23LM	bulk soil	1-year	B	12033.53	11631.04	2848.88	9408.63	4589.85	**2830.94***
24LM	bulk soil	1-year	B	9897.69	9333.39	**2878.37***	9506.42	4312.78	2905.78
4LM	bulk soil	5-years	A	23629.23	20288.77	**2845.67***	8706.34	4902.39	2931.52
5LM	bulk soil	5-years	A	20612.1	18608.4	3168.4	8213.9	4571.0	**2904.84***
6LM	bulk soil	5-years	A	20844.6	19224.9	3034.6	7691.1	4558.6	**2860.14***
16LM	bulk soil	5-years	B	12388.80	12005.43	**2482.06***	8162.28	4123.71	2960.74
17LM	bulk soil	5-years	B	15095.98	14049.54	**2599.84***	9048.28	4728.68	2914.10
18LM	bulk soil	5-years	B	13229.33	12020.41	**2763.61***	9433.87	4816.68	2952.98
7LM	rhizosphere	1-year	A	17797.50	14802.7	3453.7	10613.0	*nd*	**2984.68***
8LM	rhizosphere	1-year	A	17306.3	13818.2	3920.6	10024.1	*nd*	**2994.92***
9LM	rhizosphere	1-year	A	*nd*	*nd*	*nd*	*nd*	*nd*	*nd*
19LM	rhizosphere	1-year	B	7931.75	7375.81	**2433.63***	7143.04	2418.04	2882.70
20LM	rhizosphere	1-year	B	12679.91	11604.07	**2742.40***	9591.34	4776.32	2961.08
21LM	rhizosphere	1-year	B	16487.3	14630.0	3438.7	12509.2	*nd*	**2971.08***
1LM	rhizosphere	5-years	A	17254.95	15507.25	**2955.34***	10061.52	*nd*	2966.68
2LM	rhizosphere	5-years	A	20012.0	17093.2	3487.8	3495.4	*nd*	**3042.70***
3LM	rhizosphere	5-years	A	17708.09	15857.41	**2721.70***	9287.75	4968.54	3014.40
13LM	rhizosphere	5-years	B	16146.35	14662.35	**2865.83***	11147.48	*nd*	2938.24
14LM	rhizosphere	5-years	B	15409.74	13701.41	**2753.47***	9417.29	5076.29	2848.16
15LM	rhizosphere	5-years	B	9383.1	8949.1	**2271.80***	6210.9	3528.3	2762.82

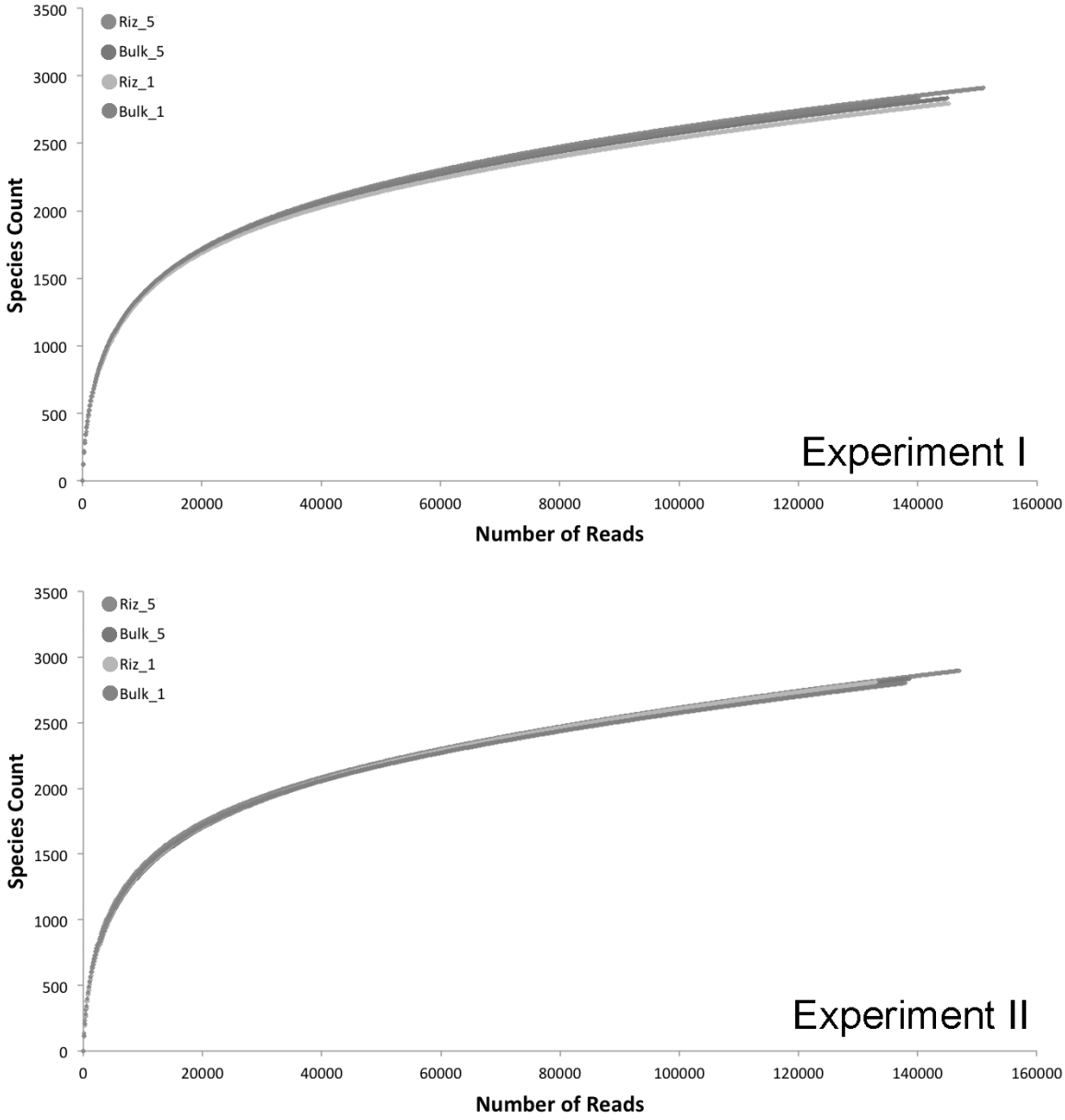
[1]AIC for radfit-generated models calculated the equation AIC = -2log-likehood + 2*npar. AIC was calculated in the same way for the zero-sum model, from the minimum of log-likehood reported by TeTame, then multiplied by -1 to obtain the maximum log-likehood value. * The best fit model with lowest AIC value. *nd* not determined.

Supplementary Figure 3.1.



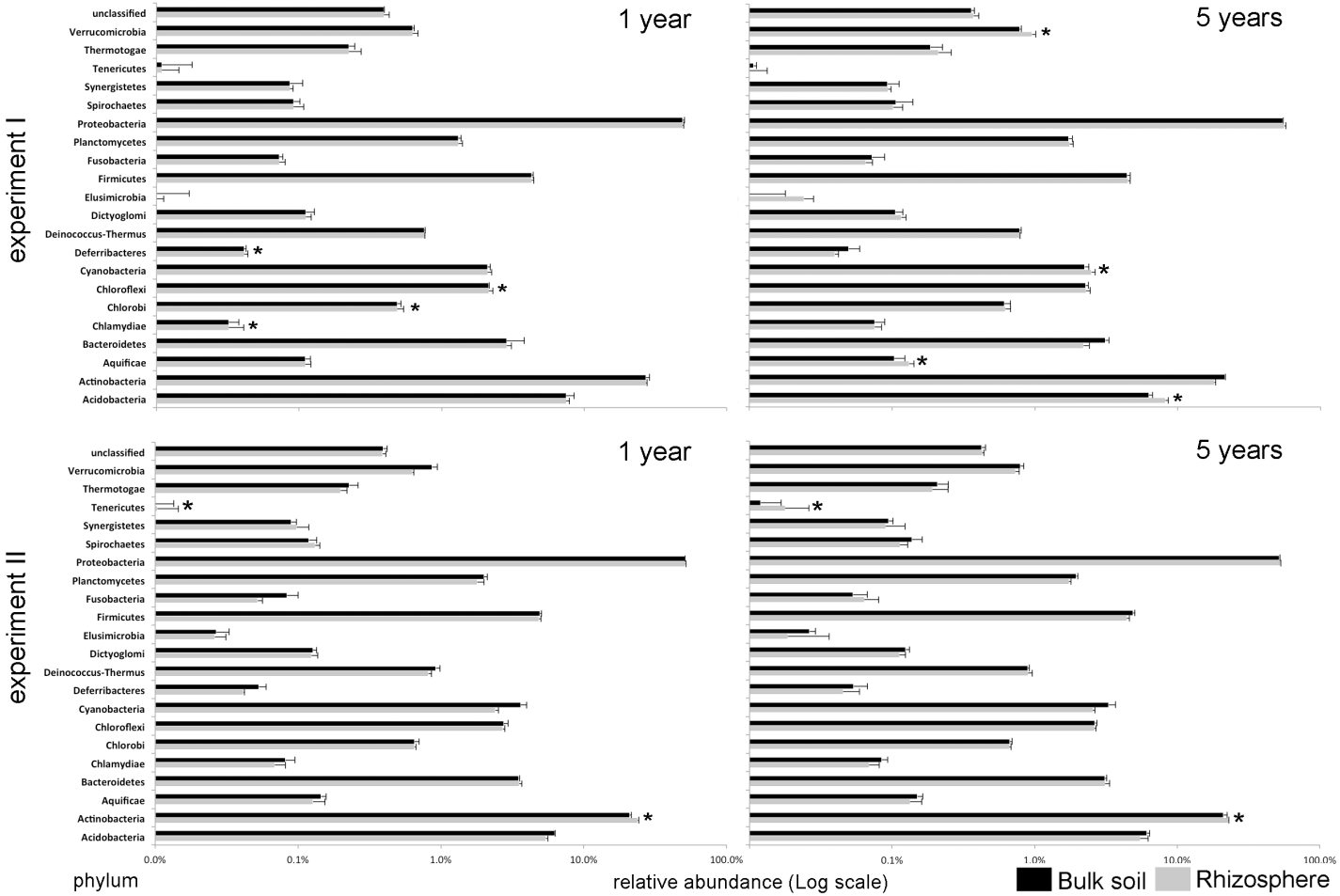
SF 3.1. Taxonomic affiliation of metagenomic reads. (A) Results for complete datasets evaluated by BLASTX analysis against the SEED database using MG-RAST v 3.3 software. (B) Differential proportion of sequences assigned within the phylum *Proteobacteria* for rhizosphere and bulk soil of first and fifth years of crop.

Supplementary Figure 3.2.



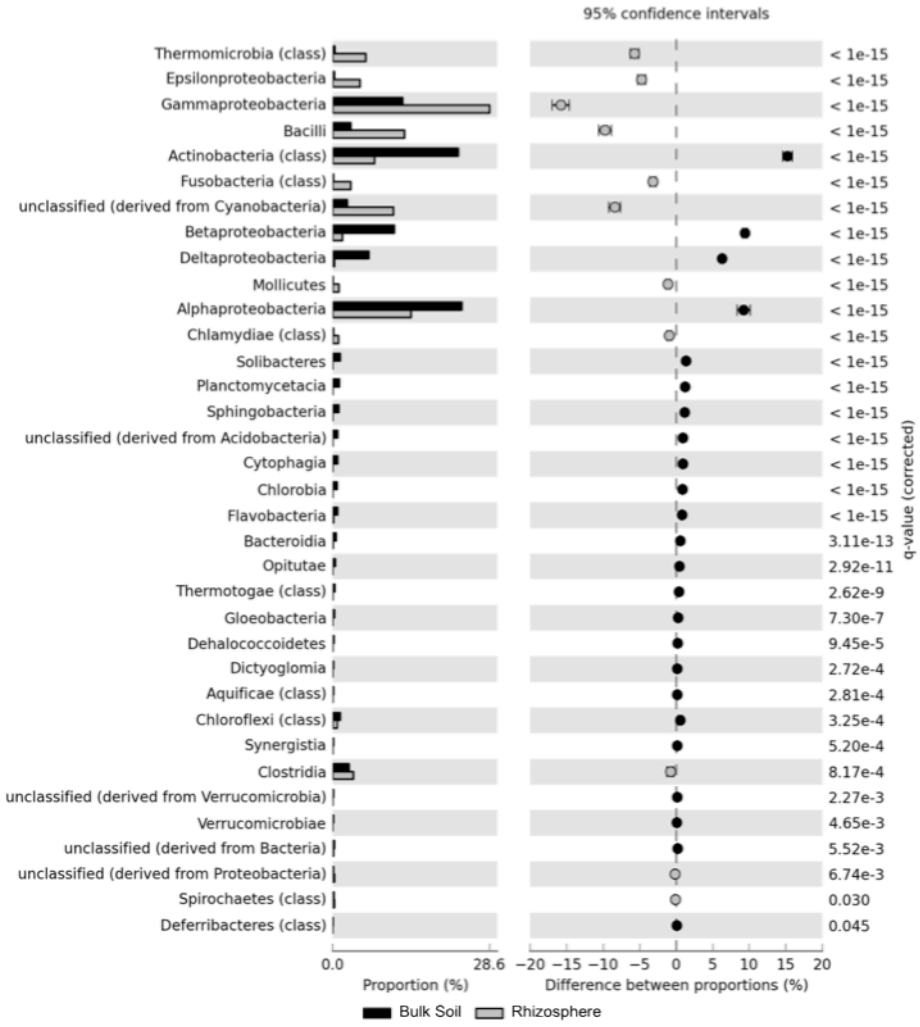
SF 3.2. Rarefaction curves of annotated species richness generated for samples from bulk soil and rhizosphere of first and fifth of soybean cultivation. The curves represent the average number of different species annotations for subsamples of the complete dataset.

Supplementary Figure 3.3.



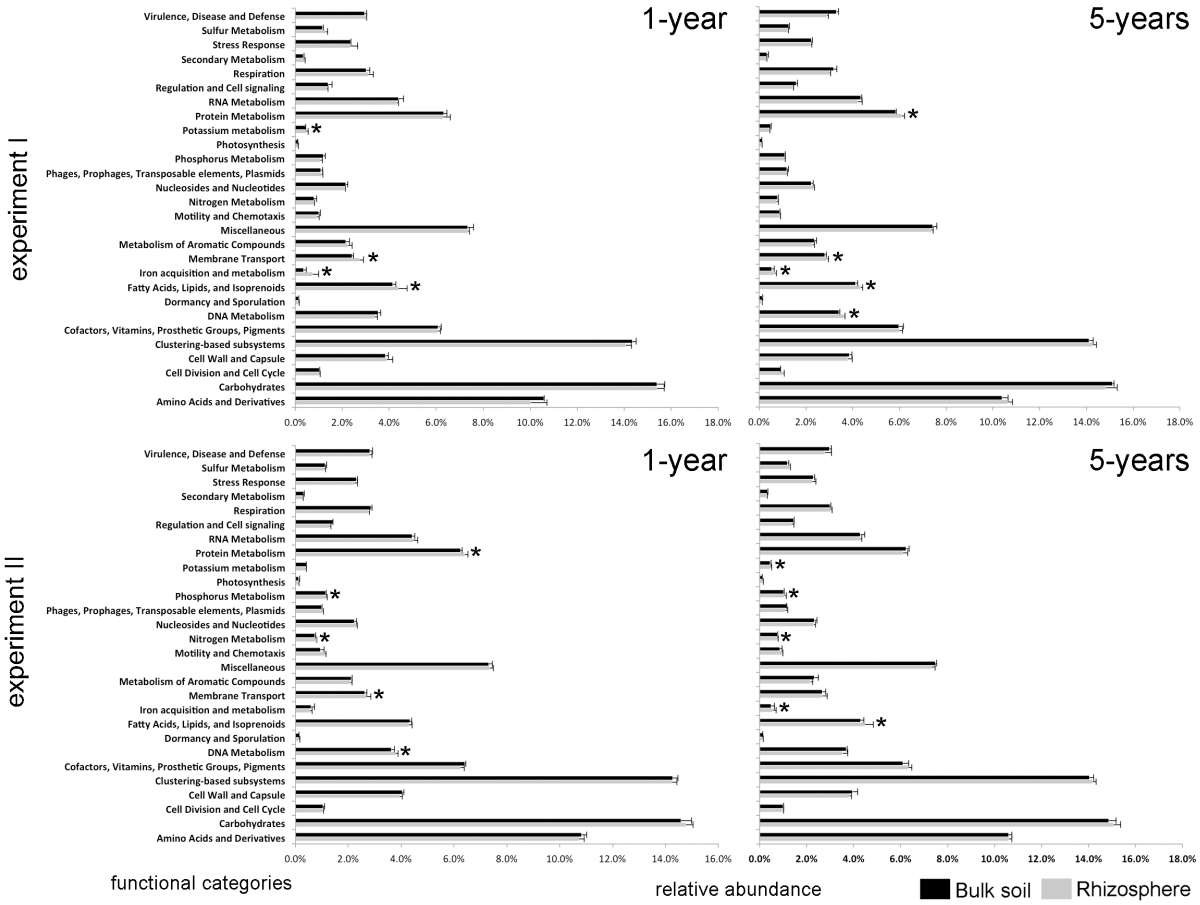
SF 3.3. Relative abundance of bacteria at phylum level based on shotgun metagenomics data. Percentage of total sequence reads in samples from bulk soil and rhizosphere of 1-year and 5-years of soybean harvesting is presented here for the 2 years of experiments (I and II). The error bars show calculated standard variation of triplicate samples and * indicate more abundant phyla in rhizosphere (*P*-value <0.05). Corrected *P*-values were calculated using Benjamini-Hochberg FDR approach (p<0.05).

Supplementary Figure 3.4.



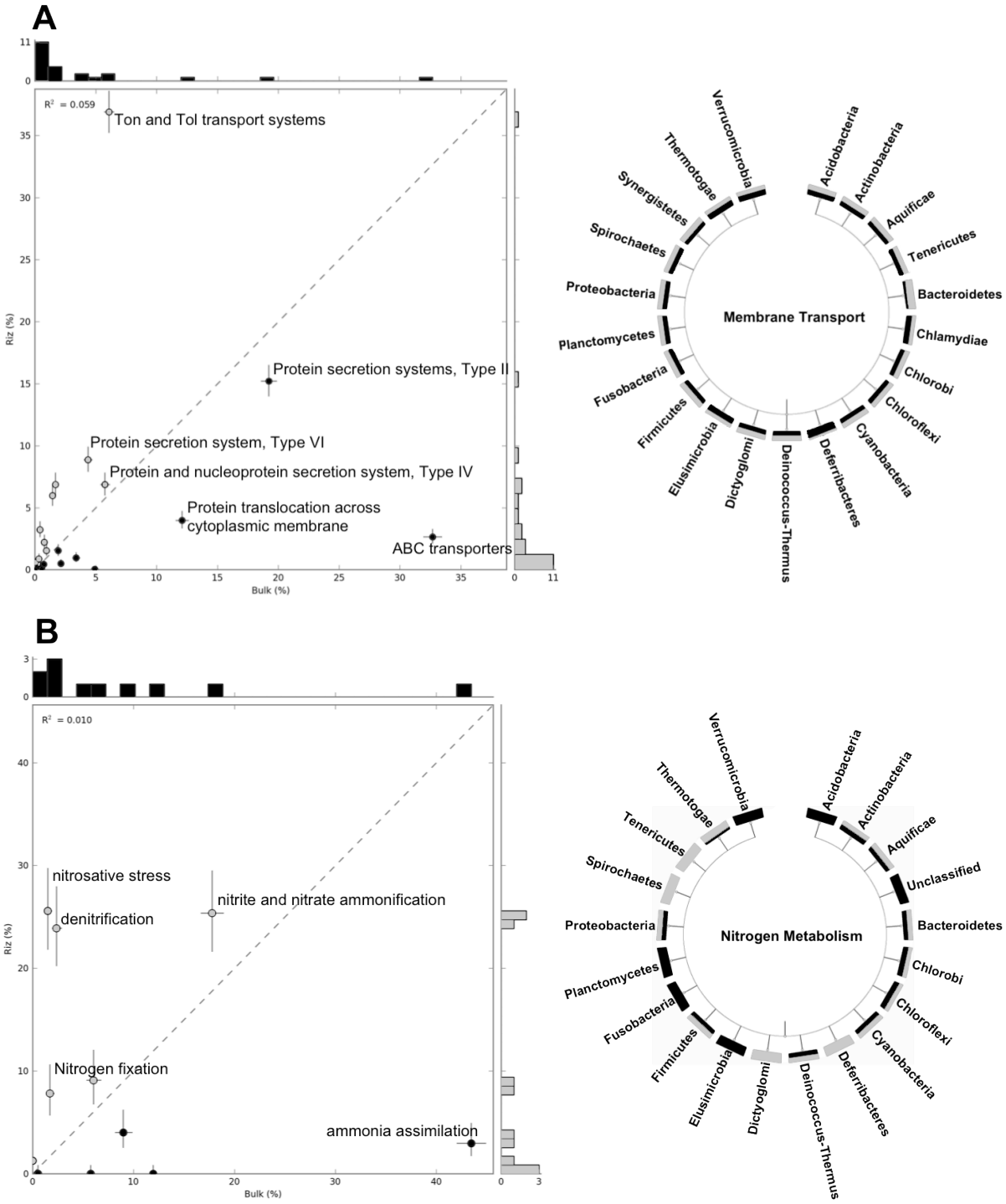
SF 3.4. STAMP analysis of taxonomic profiles at class level between rhizosphere and bulk soil samples. Groups overrepresented in the bulk soil (black) correspond to positive differences between proportions and groups overrepresented rhizosphere (gray) correspond to negative differences between proportions. Corrected *P*-values (q-values) were calculated using Benjamini-Hochberg FDR approach ($p<0.05$).

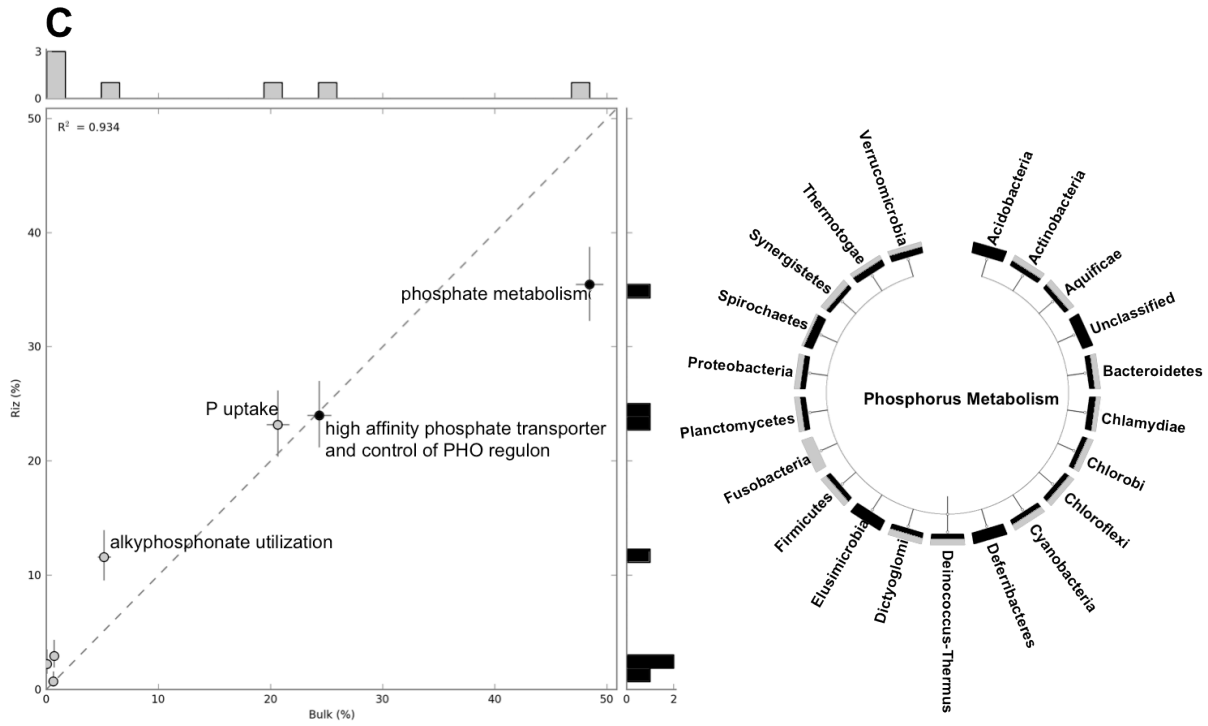
Supplementary Figure 3.5.

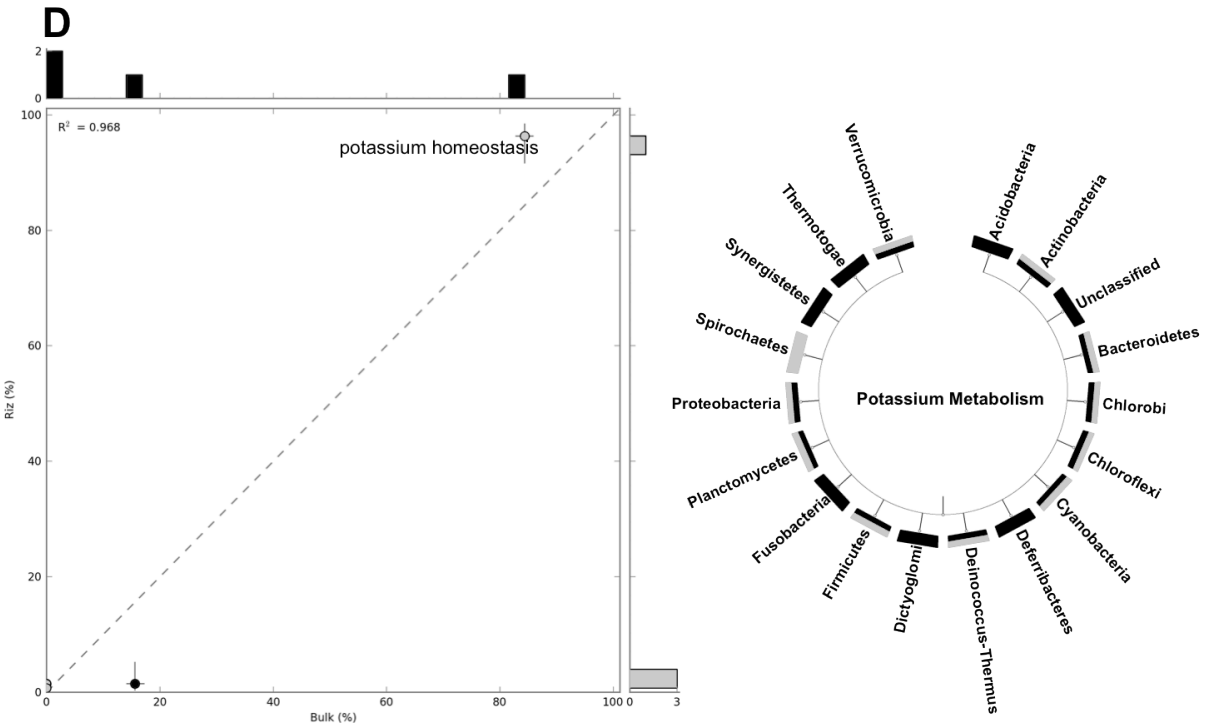


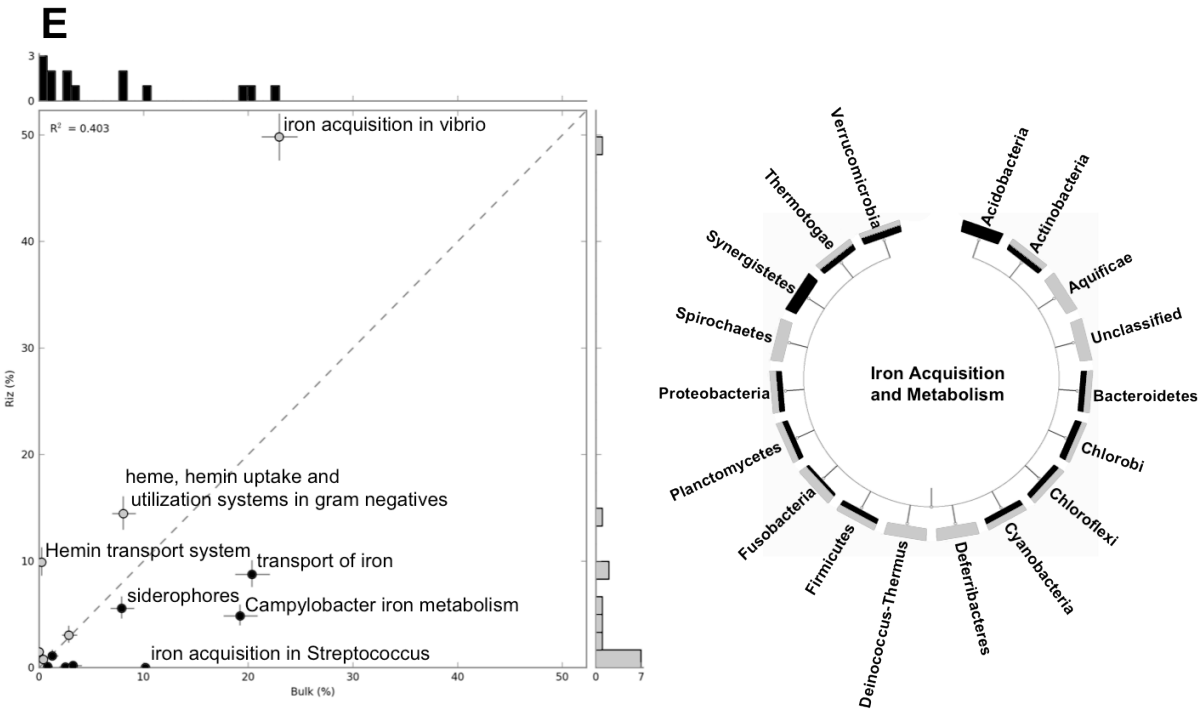
SF 3.5. Relative abundance of functional categories (SEED subsystem level 1) based on shotgun metagenomics data. Percentage of total sequence reads in samples from bulk soil and rhizosphere of 1-year and 5-years of soybean harvesting is presented here for the 2 years of experiments (I and II). The error bars show calculated standard variation of triplicate samples and * indicate categories more abundant in rhizosphere (*P*-value <0.05). Corrected *P*-values were calculated using Benjamini-Hochberg FDR approach (p<0.05).

Supplementary Figure 3.6.



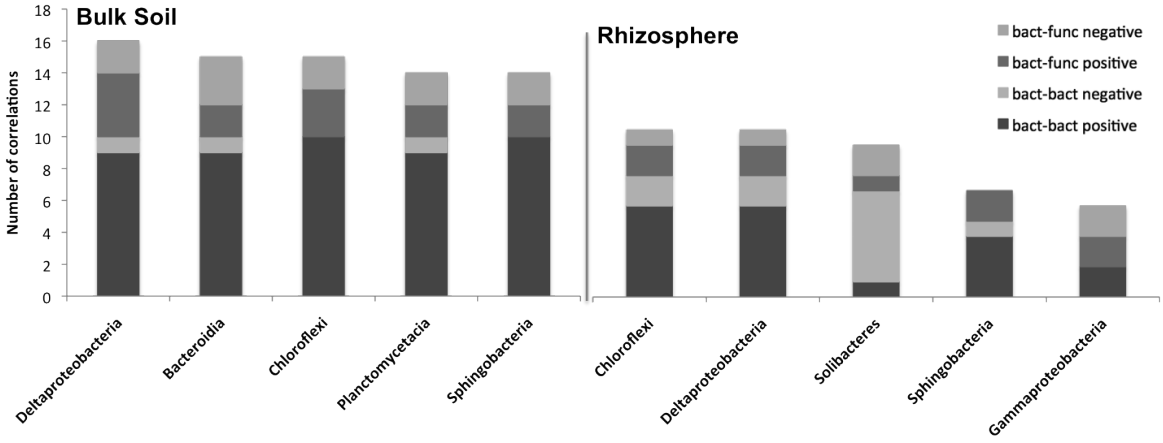

C
R² = 0.934
phosphate metabolism
P uptake
high affinity phosphate transporter
and control of PHO regulon
alkyphosphonate utilization
Riz (%)
Bulk (%)
Phosphorus Metabolism
Verrucomicrobia
Thermotogae
Synergistetes
Spirochaetes
Proteobacteria
Planctomycetes
Fusobacteria
Firmicutes
Elusimicrobia
Dictyoglomi
Deinococcus-Thermus
Deferribacteres
Cyanobacteria
Chloroflexi
Chlorobi
Chlamydiae
Bacteroidetes
Unclassified
Aquificae
Actinobacteria
Acidobacteria


D
R² = 0.968
potassium homeostasis
Riz (%)
Bulk (%)
Potassium Metabolism
Verrucomicrobia
Thermotogae
Synergistetes
Spirochaetes
Proteobacteria
Planctomycetes
Fusobacteria
Firmicutes
Dictyoglomi
Deinococcus-Thermus
Deferribacteres
Cyanobacteria
Chloroflexi
Chlorobi
Bacteroidetes
Unclassified
Actinobacteria
Acidobacteria



SF 3.6. Functional and taxonomic profiles of bulk soil samples (red) and rhizosphere (green) of (A) membrane transport, (B) nitrogen metabolism, (C) phosphorus metabolism, (D) potassium metabolism and (E) iron acquisition and metabolism. The data was calculated for metagenomes from bulk soil and rhizosphere and compared to SEED database using a maximum *E*-value of $1e^{-5}$ and a minimum alignment length of 50 bp.

Supplementary Figure 3.7.



SF 3.7. Top five bacterial groups with higher positive and negative correlations in bulk soil and rhizosphere samples. Numbers of correlations (positives and negatives) between bacteria and bacteria (bact-bact), bacteria and function (bact-func), and function and function (func-func) in bulk soil and in rhizosphere samples.