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

Spatial and temporal dynamics of the microbial communities in soils cultivated with sugarcane

Thiago Gumiere

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

Piracicaba
2017

Thiago Gumiere
Agronomist

Spatial and temporal dynamics of the microbial communities in soils cultivated with sugarcane

versão revisada de acordo com a resolução CoPGr 6018 de 2011

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

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

Piracicaba
2017

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

Gumiere, Thiago

Spatial and temporal dynamics of the microbial communities in soils cultivated with sugarcane / Thiago Gumiere. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. Piracicaba, 2017.

75 p.

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

1. Cana-de-açúcar 2. Comunidades microbiana 3. Comunidade core 4. Modelo de probabilidade I. Título

I dedicate my thesis to

my parents, Eva and Edimilson, who teach me the right path;

my brother Silvio and my sister Marina, who sustain me in their shoulders;

my love and future wife Annelise, who takes care of my heart and my mind;

and to my Godson Henrique, who teaches me the beauty and simplicity of the life.

Everything I have and everything I am, I owe to you.

ACKNOWLEDGEMENTS

To God, the Great Architect of the Universe;

My family, for the support and encouragement that have always helped me to continue walking;

FAPESP for the scholarships granted (2013/18529-8 and 2014/22845-5), which allowed me to develop my work, and improved my abroad experience;

Prof. Dr. Fernando Dini Andreote, for the support and friendship, which guided my professional development over the last six years;

Friends from the Soil Microbiology Laboratory, Danielle, Diogo, Kelly, Armando, Bruna, Dorotéia, Caroline, Arthur, Pole, Mauricio, Maiele, Alessandra, Sonia, Juliana, Pedro, Nicolle and Yasmin. In particular, Denise and Fernandinho for helping me in this long journey of work;

My friend Ademir, for the great scientific discussions, which helped me to dream higher and at the same time kept my feet on the ground;

Prof. Dr. Brendan Bohannan, for helping me to further develop my personal and scientific skills. A friend who taught me that the wisest man is the most modest man;

Kyle, Johanna, Jarrett, Anya, Felipe, and Clarisse, my especial friends, that during one of the most difficult moments of my life, made me feel at home;

Prof. Dr. Silvio José Gumiere, my scientist example, who helped me to stand firm in the face of difficulties, and my friend Dr. Adam Burns, who helped me to develop the probabilistic model;

Friends from EMBRAPA Environment, especially Dr. Rodrigo Taketani for the incredible help with microbial analyses;

The National Laboratory of Science and Technology of Bioethanol (CTBE) team, especially Dr. João Luiz Nunes Carvalho for providing the samples and the metadata;

The coordination of the PPG-Soil and Nutrition plants and the Escola Superior de Agricultura "Luiz de Queiroz" for the opportunity to carry out this work;

To all those who directly or indirectly helped me in the development and completion of this work.

“my roots are in the air
my home is anywhere
If it depends on me, I'm going through”
Até o fim – Engenheiros do Hawaii

“It’s a long way to the top if you wanna Rock 'N' Roll”
AC/DC

SUMMARY

RESUMO

Dinâmica espaço-temporal da comunidade microbiana de solos cultivados com cana-de-açúcar

As condições ambientais que podem modular a dinâmica da comunidade microbiana em solos de culturas são pouco conhecidas. O presente trabalho foi dividido em três partes essenciais, onde *i)* discutiu-se modelos e teorias ecológicas para a exploração microbiana em solo agrícolas, argumentando-se que os modelos ecológicos que particionam as comunidades microbianas, poderiam aumentar a resolução entre interações microbianas e o ambiente, *ii)* desenvolveu-se um modelo probabilístico baseado na freqüência de ocorrência de micro-organismos através de sistema identificando a comunidade microbiana "core". O modelo baseou-se na distribuição de Poisson, sendo este testado em quatro conjuntos de dados disponíveis no Projeto "Earth Microbiome", e *iii)* identificou-se as comunidades bacterianas e fúngicas core em solos cultivados com cana-de-açúcar, verificando-se quais componentes abióticos poderiam modular a composição dos grupos. Com isso, elevou-se a resolução das interações ambiental e microbiana, indicando que o core microbiano e as comunidades microbianas variáveis são moduladas por componentes abióticos distintos. Observou-se também que as comunidades core e variável possuem funcionalidade potencial distinta, como fixação de nitrogênio mais predita para o core bacteriano e processo de nitrificação para a comunidade variável de bactérias. Os resultados do presente trabalho elevam o conhecimento da dinâmica e funcionalidade microbiana, ajudando a revelar e explorar o microbioma do sistema de cultivo.

Palavras-chave: Cana de açúcar; Comunidade microbiana; Distribuição de Poisson; Modelos ecológicos

ABSTRACT

Spatial and temporal dynamics of the microbial communities in soils cultivated with sugarcane

The environmental conditions driving the microbial community dynamics in crop soils remain unclear. Here, we focused on the spatial and temporal dynamics of microbial communities in soils cultivated with sugarcane under different soil managements, during two years. Our work was divided into three essential parts, where *i*) we discuss ecological models and theories for the microbial exploration in crop soils, arguing that those ecological models, which partitioned the microbial communities, may increase the resolution of the environmental and the microbial interactions; *ii*) we developed a probabilistic model based on the occurrence frequency of microorganisms across systems identifying the core microbial community. The model is based on the Poisson distribution, and it was tested in four datasets available in the Earth Microbiome Project; *iii*) we identified the core bacterial and fungal communities across soils cultivated with sugarcane, verifying which abiotic components could drive the composition of groups. We increased the resolution of the environmental and the microbial interactions, showing that the core and the variable microbial communities are driven by distinct abiotic components. We also observed that the core and variable microbial communities harbor distinct potential functionality, as nitrogen fixation being more predicted to the core bacterial commmunity, and nitrification process for the variable bacterial community. Our finds increase the knowledge of microbial dynamics and functionality, helping to reveal and explore the crop system microbiome.

Keywords: Sugarcane; Microbial community; Poisson distribution; Ecological models

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The soil is considered one of the most important sources of microbial diversity, showing vastly functionality in this system (Bhat, 2013). The crop systems have been described as artificial systems, where the most of the soil and environmental conditions are adjusted for the crop development (Odum and Barrett, 2005). It is known that the management of soil triggers differentiation in the structure and composition of the microbial communities (Ciccolini *et al*., 2015). However, the distribution of microbial communities linked with environmental conditions is poorly understood.

For sugarcane, one of the most important crops in Brazil, the microbial diversity associated has been described, as endophytes (Stuart *et al*., 2010) and associated with roots (Luvizotto *et al*., 2010; Yeoh *et al*., 2015). Souza *et al*. (2016) identified a core bacterial and fungal communities across soil and sugarcane organs. Durrer *et al*. (2016) indicated that soil characteristics (chemistry and physical), and crop managements were the main drivers of the bacterial community in soils cultivated with sugarcane. Besides the importance of these strategies of drivers identification, the understanding of microbial distribution across space and time, and also their connections with abiotic components, such climate or soil characteristics are necessary.

Here, we focused on the spatial and temporal dynamics of microbial communities in soils cultivated with sugarcane under different soil managements, sampled along a period of two years. Aiming to clarify the correlation between the microbial community and abiotic components, we applied an ecological model based on the occurrence frequency of microorganisms across the geographical space and along the time. The “core community” is a term used for animal and plants, defining a group of organisms that are ubiquitous in a given habitat (Hanski, 1982). In microbial ecology, the term core microbial community could refer to microbial genes or taxa occurring in a set of samples in a given ecosystem (Shade and Handelsman, 2012). However, the identification of the microorganisms that belong to the core microbial community is based on arbitrary cutoffs, which could limit the knowledge of system correlations.

Our work was divided into three essential parts. The first part aimed to discuss ecological models and theories for the microbial exploration in crop soil, arguing that the ecological models, which partitioned the microbial communities, may increase the resolution of the environmental and the microbial interactions. In the second part, we developed a probabilistic model to identify the core microbial community. The model is based on the

Poisson distribution and it was tested in four datasets available in Earth Microbiome Project (EMP; http://www.earthmicrobiome.org). In the third part, we used this method to identify the core bacterial and fungal communities across soils cultivated with sugarcane, and also verified what abiotic components, such soil characteristics (chemistry and physical, crop management, and climate) could modulate the core and variable microbial communities.

References

Bhat A. (2013). Preserving Microbial diversity of Soil ecosystem: A key to sustainable productivity. Int J Curr Microbiol App Sci 2: 85–101.

Ciccolini V, Bonari E, Pellegrino E. (2015). Land-use intensity and soil properties shape the composition of fungal communities in Mediterranean peaty soils drained for agricultural purposes. Biol Fertil Soils 51: 719–731.

Hanski I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. Oikos 38:210–221.

Luvizotto DM, Marcon J, Andreote FD, Dini-Andreote F, Neves AAC, Araújo WL, *et al*. (2010). Genetic diversity and plant-growth related features of Burkholderia spp. from sugarcane roots. World J Microbiol Biotechnol 26: 1829–1836.

Odum E, Barrett G. (2005). Fundamentals of ecology. 5th ed. Thomson Brooks/Cole, 624p.

Shade A, Handelsman J. (2012). Beyond the Venn diagram: The hunt for a core microbiome. Environ Microbiol 14: 4–12.

Souza RSC de, Okura VK, Armanhi JSL, Jorrín B, Lozano N, Silva MJ da, *et al*. (2016). Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. Sci Rep 1–15.

Stuart RM, Romão AS, Pizzirani-Kleiner AA, Azevedo JL, Araújo WL. (2010). Culturable endophytic filamentous fungi from leaves of transgenic imidazolinone-tolerant sugarcane and its non-transgenic isolines. Arch Microbiol 192: 307–313.

Yeoh YK, Paungfoo-Lonhienne C, Dennis PG, Robinson N, Ragan M A., Schmidt S, *et al*. (2015). The core root microbiome of sugarcanes cultivated under varying nitrogen fertiliser application. Environ Microbiol 18: 1338–135.

2. THE ESSENTIALITY OF ECOLOGICAL MODELS AND THEORIES FOR THE MICROBIAL EXPLORATION IN CROP SOILS

Abstract

The importance of microbial community structure and functionality in crop systems is becoming better recognized. The list of functions attributed to microbes in soils is growing, and several drivers of microbial community composition and functioning have been suggested. In this scenario, the improvement of the knowledge about microbial system must also lie on ecological theories, as those derived and extensively used for communities of animals and plants. Here, we reviewed the current ecological therioes applied for microbial communities, discussing the importance of ecological models based on abundance (rare and dominant), frequency occurrence (core and variable) and generalist-specialist, and argued that the link between the most significant environmental drivers and the assemblage and functioning of microbial communities in crop systems may be better achieved using ecological theories. We also indicated the limitations of ecological models application in microbial systems. We believe that these approaches, based on ecological theories and modeling, would better reveal the integration of microbial communities with abiotic components of the crop systems.

Keywords: Crop systems; Ecological models; Ecological theories; Soil microbiology; Environmental drivers

2.1. Introduction

Different from natural systems, the crop systems could be considered as artificial systems, where the most of the soil and environmental conditions are adjusted to promote the crop development (Odum and Barrett, 2005). The soil is considered one of the most colonized systems by microorganisms, hosting approximately 10^9 microbial cells per gram of soil (Torsvik and Øvreås, 2002), distributed along 10 to 30 thousand distinct microbial genomes (Roesch *et al*., 2007). The microbial functionality in crop soils is vastly described (*e.g.* decomposing organic matter, promoting plant growth, and harnessing biogeochemical cycles). It is also known that the management of soil triggers differentiation in the structure and composition of the microbial communities (Ciccolini *et al.*, 2015).

Previous research has established that soil affect the microbial communities in crop soils. For example, soil chemistry and physics, including pH, potassium, phosphorus, and granulometry; or crop managements composed by the addition of vinasse, charcoal or the mechanical harvesting were the main factors which modulate the bacterial community in soils cultivated with sugarcane (Durrer *et al.*, 2016 *in press*). Thilagar *et al.* (2016) suggested a microbial consortia composed by *Funneliformis mosseae* and *Bacillus sonorensis* which could

reduce by 50% the application of chemical fertilizer in chilly crop, and also improved the soil health. These strategies, however, does not provide solid results, exportable to any cropped area, limiting the development of general approaches to manage microbial communities in crop soils. It is possible to conceive the idea that it might be achieved by the use of strategies better focused in general characteristics and responses of microbial systems, as those provided in ecology-based approaches.

Most of the ecological theories were developed to study communities of animals and plants, but their application on microbial community studies have been increased (Prosser *et al.*, 2007). The diversity-stability hypothesis, the intermediate disturbance hypothesis (IDH), and the taxa-area relationships are important examples of ecological models applied for microbial communities (Ogilvie, 2012). Besides the importance of the results reported by the use of ecological theories, the capacity for prediction in microbial ecology is unexplored, stating the major challenge for the improvement of the exploration of microbial communities in crop systems. It is believed that efforts on the modeling applied to microbial communities may improve the knowledge of the system functioning (Wade *et al*., 2016).

Here, we argued that the link between the most significant environmental drivers with the microbial community across the crop systems may be better achieved using ecological theories integrating the microbial partition, such as the differential analysis of the core and variable members of the microbial community. Next to it, we ask for new approaches that allow the proper determination of these fractions of microbial communities, as others, to improve the ecological concepts for microbes.

2.2. Establishing background for crop soils

Before deeping in the description of the soil microbial community importance, it is essential to properly address the characteristics of cropping systems. Also defined as agroecosystems, crop systems are dominantly composed of animals and plants under artificial selection, aiming to increase the production of specific food crops and other products (Odum and Barrett, 2005). The artificial selection could be expressed as humankind control such animal labor, fertilizers, pest controls, and irrigation (Odum and Barrett, 2005).

The differentiation of crop systems begins with de source and quality of litter, which selected specific decomposer communities, resulting in more homogeneous litter chemistry during the decomposition, possibly directly affecting the turnover and stabilization of soil carbon (Wickings *et al.*, 2012). The number of different crop plant in the field also affect the

carbon and nitrogen ratio of the system, which has been used to estimate the kinetic of organic matter mineralization (Manzoni and Porporato, 2009). For example, the increment of one or more crops in rotation increased the total soil carbon in 8.5% and nitrogen in 12.8%, what also increases the microbial biomass in soils (McDaniel *et al.*, 2014).

Another important characteristic of the crop systems is the management of soil and the fertilization process, which adapt the system for the crop cultivation, targeting the highest possible production. For example, the conventional cultivation of sugarcane includes types of tillage, such plowing, harrowing, and subsoiling, which aimed to reduce the physical resistance of soil structure on the plant root (Grant and Lafond, 1993). The sugarcane is harvested every 12 to 18 months of cultivation, along a cycle of five years. In the harvested process, the sugarcane straw is separated from the steam and left under the soil surface (Bell *et al.*, 2007). The soil chemistry is adjusted annually to attend the sugarcane demand.

The main nutrients applied in soils for sugarcane cultivations are nitrogen, phosphorus, potassium, calcium, magnesium, and sulfur, besides the use of micronutrients like the bore, cupper, manganese, molybdenum and zinc (Malavolta *et al.*, 1997). The hydrogen potential (pH) is one of the main environmental factors mandatory for crop cultivation. According to Dick *et al.* (2000), the pH affects the nutrients availability to microbial community activity. The pH has been also described as a predictor of soil bacterial community distribution (Lauber *et al.*, 2009). Thus, the relevance of the microorganisms in the crop system cultivation has to continue increase, leading agriculture to new levels of productivity and sustainability.

2.3. The relevance of microbial community for soils

Microorganisms are the greatest source of biodiversity on the planet, present in all three domains of the tree of life, encompassing the Bacteria, Archaea and Fungi (Eukaryote) (Hug *et al.*, 2016). The great microbial diversity sustains the occurrence of microbes in every terrestrial environment (Prosser *et al.*, 2007). Among them, soil appears as one of the most abundant in microbial cells and listed among the most diverse systems on Earth (Torsvik and Øvreås, 2002; Roesch *et al*., 2007).

The soil is the major source of the microbiota associated with plants. For example, Zarraonaindia *et al.*, (2015) evaluated the bacterial communities associated with grapevine organs - such leaves, flowers, grapes and roots – and compared it to those in soils, along two growing seasons. The authors found a shared bacterial community between the soil and the

grapevine organs. The authods also indicated pH and C:N ratio as the major environmental drivers of the soil bacterial community, directly reflecting on the bacterial community associated with grapevine organs.

Regarding the microbial functions in the soil, different communities are linked to the catabolism of organic compounds, and subsequent formation of new compounds, driving the development of interactions, as those between plants and microbes (Griffiths *et al*., 2006). The microbial functionality in crop soils is vastly described in the literature, where it is remarked its role in the decomposition of organic matter (Schultz and Urban, 2008), promotion of plant growth (Gumiere *et al.*, 2014), suppression of plant diseases (Mendes *et al.*, 2011), degradation of contaminants (Glick, 2010), stabilization of soil structure (Wakelin *et al.*, 2008), and especially in the biogeochemical cycles, such carbon (Hussain *et al.*, 2011), nitrogen (Nelson *et al.*, 2016), phosphorus (Wakelin *et al.*, 2007) and sulfur (İnceoğlu *et al.*, 2013).

In crop systems, the management of soil drives the structure and composition of the microbial communities, triggering a series of events that interfere in the crop plant development. For example, the denitrification process is considered one of the major losses of nitrogen in form of nitrous oxide, in a process associated with bacteria and fungal activities (Herold *et al.*, 2012). Some conditions in soils are known to promote higher rates of denitrification, as for example, the positive correlation between the potential denitrification rate of bacterial community and the soil pH (Herold *et al.*, 2012). In Table 1, several studies with similar observations are listed, identifying the most important drivers of the microbial community and their relevance for the cropping system. Thus, climate conditions such temperature and precipitation; soil chemistry, physics, crop management, and also the historical use of the lands may affect and shape the soil microbial community. Overall, there is no dominant driver connected to variations of different microbial communities across the crop systems.Then, we can argue that better understanding of the link between the most significant drivers with the microbial community across the crop systems may be achieved using mathematical modeling based on ecological theories (*see* Wade *et al.*, 2016; Prosser *et al.*, 2007).

Table 2.1 - Different crop systems and the drivers of the microbial communities composition

Crop system	Microbial Community	Main results	Reference
Wheat-maize	Abundance of ammonia oxidizers and denitrifiers	The application of nitrogen, phosphorus and potassium fertilizer increased the N_2O emission	(Dong *et al.*, 2015)
Wheat	Rhizobacterial community	The moisture availability was the major driver of soil bacterial community dynamics	(Yang *et al.*, 2013)
Intercropped durum wheat and faba bean	Total bacterial and fungal communities	The fertilization of phosphorus was the main driver of microbial communities in this field trial	(Tang *et al.*, 2016)
Mediterranean peaty	Arbuscular mycorrhizal fungi (AMF), and total fungal community	The land-use intensification resulted in changes of the AMF and total fungi communities. The base saturation and exchangeable calcium were significant drivers of the total fungi composition. The base saturation was also correlated with AMF variance.	(Ciccolini *et al.*, 2015)
Sugarcane	Total fungal community	The soils characteristics and climate variables explained respectively 2.88% and 2.93%. The most significant driver identified was the geographical distance explaining 50.75% of fungi variance.	(Gumiere *et al.*, 2016)
Rice	Methanogenic microbial community	The temperature was the main driver identified which defines the structure and function of the methanogenic microbial community.	(Conrad *et al.*, 2009)
Arabica coffee	Arbuscular mycorrhizal fungi (AMF)	The soil pH, nitrogen, and phosphorus availability were the significant drivers of the AMF community composition.	(De Beenhouwer *et al.*, 2015)
Tomato	Total bacterial and archaeal communities	The antecedent cover crop showed a significant effect on the microbial community of tomato rhizosphere.	(Maul *et al.*, 2014)

2.4. Ecological theories and models for microbial systems in crops

Ecology aims to classify organisms, understand their environment and identify forces of nature which can act upon these groups (Odum and Barrett, 2005). The majority of ecological theories were developed based in studies of animals and plants, and later adapted for microorganisms (Prosser *et al.*, 2007). However, several distinctions are clear among microbes and other groups of living organism. The first is associated with the required adaptation of the species concept, primarily defined as a group of interbreeding natural populations, that is reproductively isolated from other groups (Mayr, 1970). As microbes are mostly asexual, a species can be considered as made of individuals with identical ecological properties, *i.e.* occupy the same niche or developing the same function based on the same biological system (Cohan, 2002). We know little on microbial ecology about simple things in

the species concept of plants and animals, as for example, intraspecific variation, yet poorly explored as bacterial individuality. As the most of the ecological theories depend on species concept, we can suggest that their application on the microbial communities requires a different perspective from that applied for macro-organisms.

The diversity-stability hypothesis, intermediate disturbance hypothesis (IDH), and taxa-area relationships, are important examples of microbial ecological theories (Ogilvie, 2012). The diversity-stability hypothesis states that the species richness (averaging effect) or functional diversity (diversity effect) sustain the ecosystem stability (*reviewed by* McCann, 2000). It has been observed that decreasing the microbial diversity of pasture soils resulted in a decrease of plant decomposition resilience, when soils are exposed to heat or contaminated by copper (Griffiths *et al.*, 2000). A study comparing 21-year of bioorganic and conventional farming systems indicated that the increased biodiversity in organic plots may reduce the dependence of external inputs, as fertilizers (34 to 53%) and pesticide (97%) (Mäder *et al.*, 2002). Thus, the crop systems based on conventional managements would simplify the ecological communities, making the system more suitable for invasions of exogenous organisms (McCann, 2000). Also, in this scenario, plant nutrition and protection would be more dependent of human inputs.

Another hypothesis also applied for microbial communities is the IDH. Firstly presented by Connell (1978), the IDH states that given a gradient of disturbance, the "intermediate" position would have a maximum diversity. Analyzing the dataset provided by Mendes *et al.*, (2011), van der Voort *et al.*, (2016) observed that the suppressiveness of soils was reduced by a heat impact, and the alpha diversity was higher in those samples. The authors suggested that the IDH explained the higher alpha diversity in heated samples, possibly promoted along the reassembly of the disturbed community. At this moment, heat tolerant or fast growers and temperature sensitive or slow growers may differentially multiply, leading to a distinct assemblage of final stable community. The IDH is possibly a system undergoing in crop soils, which are constantly subjected to managements and distinct inputs, generating a continuous intermediate disturbance of the system.

Focusing on the microbial community distribution, the taxa-area relationships is used to plot the distribution of microorganisms across different size of area (Horner-Devine *et al.*, 2004). It has been verified, for bacteria (Horner-Devine *et al.*, 2004) and fungi (Gumiere *et al.*, 2016), that microbial communities located close each other are more similar than those located farther. This distribution, named as distance-decay, was also observed more

pronounced in soils under the Amazon forest than those converted into areas for cattle pasture (Rodrigues *et al.*, 2013). These authors suggested that, besides the increased of alpha diversity, the establishment of pasture systems resulted in homogenization of microbial communities across space. The decay of the community similarity could be driven by the environmental heterogeneity, as observed by Horner-Devine *et al.*, (2004), or by the limited dispersion of microbes along a given geographic distance (Gumiere *et al.*, 2016).

Besides the results reported by the use of microbial ecological theories, the efficiency of microbial ecology prediction is poor, and the improvement of the use of microbial community rely on the better modeling of these systems (Wade *et al*., 2016). As previously stated, the soil microbial community is huge in size and vastly diverse (Torsvik and Øvreås, 2002). Thus, the ecological models aimed to classify the microbial community in differents groups increasing the resolution of biotic and abiotic interactions (Guisan and Zimmermann, 2000). Three important models are predominant in literature, the abundance model, differing the dominant and rare community, the occurrence frequency model differing the core and variable community, and the generalist-specialist model differing in niche system occupancy.

In the abundant model, the dominant community is defined as microorganism with high abundance in the system (Campbell *et al*., 2011). It has been suggested that the dominant microorganisms are more adapted to the current environmental conditions, or may also be selected by the hosts. For example, the plant rhizosphere may select the microbial community which became abundant, showing a distinction of the microbial community from the bulk soil (Mendes *et al*., 2014). The rare community is composed by microorganisms with low abundance, are described to act as microbial seed bank and also sustain the resilience of the system (Sogin *et al*., 2006). For example, it has been observed that the rare community plays important role in crop protection (Hol *et al*., 2010). It has also been suggested that members of the rare community may be widely distributed across the habitat and could be dominant depending on the environmental conditions (Shade *et al*., 2014), linking the abundant model with occurrence frequency model.

The occurrence frequency model classified the microorganism in the core community, which are microorganisms ubiquitous in the environment or host despite the environmental conditions (Turnbaugh, *et al.,* 2009), and variable community which are microorganisms varying across the environment conditions. This partition of microbial communities in the core and variable groups has been showed to possibly facilitate the study of complex microbial communities. For example, Saunders *et al*. (2015) identified microbial groups

shared across several activated sludge systems with important contributions to the turnover of carbon. It has been also observed that the core group found across sugarcane roots showed a little variation on its composition under distinct applications of nitrogen fertilizers (Yeoh *et al*., 2015). The different response of the core and variables communities may indicate that each of these groups inhabits different niches in the system, and possibly respond to distinct environmental variables. From this panorama, a parallel can be traced from this division with the generalist-specialist concept.

As presented by Lennon *et al*. (2012), generalist can be defined as a group of microorganisms which tolerated wide variations of conditions and not fill specific niches in the habitat, and the specialist is microorganisms with limited distribution by the habitat conditions filling specific niches in the system. The authors identified generalist microorganisms dry-adapted tolerating a broad range of water stress, and also specialists wet-adapted by biofilm production with restricted to less negative water potentials. The generalist-specialist concept could also be linked with models of species distribution abundance (SAD), classifying the microbial community distribution as stochastic (or neutral) and deterministic. The stochastic and generalist characteristics are similar because they represent microorganism with low influenced by the environmental conditions, and more influenced by historical events and dispersal limitations (Martiny *et al*., 2006; Dumbrell *et al*., 2010). The deterministic is more similar with the specialist where the group of microorganism is influenced by the environmental conditions and the existence niches (Martiny *et al*., 2006).

Then, we suggest that approaching the microbial communities in a partition system, such the core and variable microbial groups, may support an improved glance about the structuring a functioning of microbial communities in the light of ecological theories (Figure 2.1). For example, the diversity-stability theory may be better evidenced targeting the core microbial community. It corroborated previous statements, where variations in the structure of the core microbial community indicated the occurrence of system perturbation (Shade and Handelsman, 2012). Another important possible correlation between the diversity-stability and core microbial community is the suggested connection among these microbes with the environment or host, possibly indicating the occurrence of co-evolutionary processes (Pédron *et al*., 2012; Schlaeppi *et al*., 2014). The variable and core microbial communities may also show different dynamics along the environmental heterogeneity. It implies that each driver identified for the entire microbial community may be differentially related to the core and variable fractions of the microbial communities. One can also argue that if it is true, the

correlation derived from the intermediate disturbance or form taxa-relationship hypotheses may also fit particularly the core or the variable communities. It is possible to suggest that the taxa-relationship could be more clearly observed for the variable microbial community, linking its variation with environmental heterogeneity. Besides these hypotheses have to be further tested and validates, the partition of microbial communities in groups with differential behavior would certainly increase our knowledge on microbial interactions, supporting the advances on microbial predictions (*reviewed by* Widder *et al.*, 2016).

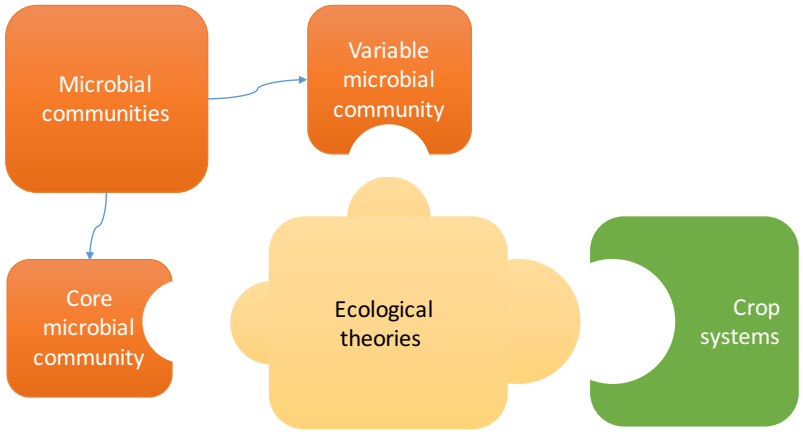

Figure 2.1 – The partition of microbial community improving the resolution of ecological theories applied for crop systems

However, the definition of the dominant, rare, core, variable, specialist, and generalist groups are mainly based on arbitrary thresholds. For example, the core microbial community is identified as microorganisms shared by the majority of samples given a habitat. The percentage of frequency cutoffs used have been variate from 100% (Turnbaugh *et al*., 2009) to 30% (Ainsworth *et al*., 2015) of occurrence frequency across samples. The same problem is observed for the abundant and rare community, where the rare community has been defined by arbitrarily cutoffs, such less than 0.01% of a sample (Galand *et al*., 2009). A possible solution to this problem is the used of probabilistic models to identify the different groups of microbes to properly fit ecological models. For example, Sato *et al*., (2012) identified that ectomycorrhizal fungi strongly constrained by the host specificity, and the saprotrophic fungi less constrained by climatic conditions, using the Bayesian zero-inflated models, which allowed them the to identify biogeographic pattern for these fungi.

2.5. Conclusions

Diversity-stability hypothesis, intermediate disturbance hypothesis (IDH), and taxa-area relationships are important ecological theories which have been shown important results in microbial communities. However, the poor efficiency of microbial ecology prediction relies on a better system modeling (Wade *et al*., 2016). Here, we argued that the link between the most significant environmental drivers with the microbial community across crop systems may be better achieved using ecological models, such as the division of communities in the core and variable communities. However, constant efforts to avoid the arbitrarily cutoffs in ecological models are required. Taken the ecological theories and modeling together, we can possibly advance on approaches to clarify the dynamics and the functioning of the microbial community in crop systems, implicating in a more sustainable agriculture.

References

Ainsworth TD, Krause L, Bridge T, Torda G, Raina J-B, Zakrzewski M, *et al*. (2015). The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. ISME J 1–14.

Bakker H. (1999). Sugar Cane Production Technology. In: Sugar Cane Cultivation and Management. Springer US: Boston, pp 167–218.

De Beenhouwer M, Van Geel M, Ceulemans T, Muleta D, Lievens B, Honnay O. (2015). Changing soil characteristics alter the arbuscular mycorrhizal fungi communities of Arabica coffee (*Coffea arabica*) in Ethiopia across a management intensity gradient. Soil Biol Biochem 91: 133–139.

Bell MJ, Stirling GR, Pankhurst CE. (2007). Management impacts on health of soils supporting Australian grain and sugarcane industries. Soil Tillage Res 97: 256–271.

Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. (2011). Activity of abundant and rare bacteria in a coastal ocean. Proc Natl Acad Sci 108: 12776–12781.

Ciccolini V, Bonari E, Pellegrino E. (2015). Land-use intensity and soil properties shape the composition of fungal communities in Mediterranean peaty soils drained for agricultural purposes. Biol Fertil Soils 51: 719–731.

Cohan FM. (2002). What are bacterial species? Annu Rev Microbiol 56: 457–487.

Connell JH. (1978). Diversity in tropical rain forests and coral reefs. Science (80) 199: 1302–10.

Conrad R, Klose M, Noll M. (2009). Functional and structural response of the methanogenic microbial community in rice field soil to temperature change. Environ Microbiol 11: 1844–1853.

Dick WA, Cheng L, Wang P. (2000). Soil acid and alkaline phosphatase activity as pH adjustment indicators. Soil Biol Biochem 32: 1915–1919.

Dong Z, Zhu B, Hua K, Jiang Y. (2015). Linkage of N_2O emissions to the abundance of soil ammonia oxidizers and denitrifiers in purple soil under long-term fertilization. Soil Sci Plant Nutr 61: 1–9.

Dumbrell AJ, Nelson M, Helgason T, Dytham C, Fitter AH. (2010). Relative roles of niche and neutral processes in structuring a soil microbial community. ISME J 4: 337–345.

Durrer A, Gumiere T, Taketani RG, Costa DP da, Pereira e Silva M de C, Andreote FD. (2016). The drivers underlying biogeographical patterns of bacterial communities in soils under sugarcane cultivation. Appl Soil Ecol. e-pub ahead of print, doi: 10.1016/j.apsoil.2016.11.005.

Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. Proc Natl Acad Sci U S A 106: 22427–22432.

Glick BR. (2010). Using soil bacteria to facilitate phytoremediation. Biotechnol Adv 28: 367–374.

Grant CA, Lafond GP. (1993). The effects of tillage systems and crop sequences on soil bulk density and penetration resistance on a clay soil in southern Saskatchewan. Can J Soil Sci 73: 223–232.

Griffiths BS, Ritz K, Bardgett RD, Cook R, Christensen S, Ekelund F, *et al*. (2000). Ecosystem response of pasture soil communities to fumigation-induced microbial diversity reductions: an examination of the biodiversity-ecosystem function relationship. Oikos 90: 279–294.

Griffiths RI, Bailey MJ, McNamara NP, Whiteley AS. (2006). The functions and components of the Sourhope soil microbiota. Appl Soil Ecol 33: 114–126.

Guisan A, Zimmermann NE. (2000). Predictive habitat distribution models in ecology. Ecol Modell 135: 147–186.

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

Gumiere T, Ribeiro CM, Vasconcellos RLF, Cardoso EJBN. (2014). Indole-3-acetic acid producing root-associated bacteria on growth of Brazil Pine (*Araucaria angustifolia*) and Slash Pine (*Pinus elliottii*). Antonie Van Leeuwenhoek 105: 1–7.

Herold MB, Baggs EM, Daniell TJ. (2012). Fungal and bacterial denitrification are differently affected by long-term pH amendment and cultivation of arable soil. Soil Biol Biochem 54: 25–35.

Hol WHG, de Boer W, Termorshuizen AJ, Meyer KM, Schneider JHM, van Dam NM, *et al*. (2010). Reduction of rare soil microbes modifies plant-herbivore interactions. Ecol Lett 13: 292–301.

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJM. (2004). A taxa-area relationship for bacteria. Nature 432: 750–3.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, *et al*. (2016). A new view of the tree of life. Nat Microbiol 1: 1–6.

Hussain Q, Liu Y, Jin Z, Zhang A, Pan G, Li L, *et al*. (2011). Temporal dynamics of ammonia oxidizer (*amoA*) and denitrifier (*nirK*) communities in the rhizosphere of a rice ecosystem from Tai Lake region, China. Appl Soil Ecol 48: 210–218.

İnceoğlu Ö, Sablayrolles C, van Elsas JD, Falcão Salles J. (2013). Shifts in soil bacterial communities associated with the potato rhizosphere in response to aromatic sulfonate amendments. Appl Soil Ecol 63: 78–87.

Lauber CCL, Hamady M, Knight R, Fierer N. (2009). Pyrosequencing-based assessment of soil ph as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75: 5111–5120.

Lennon JT, Aanderud ZT, Lehmkuhl BK, Schoolmaster DR. (2012). Mapping the niche space of soil microorganisms using taxonomy and traits. Ecology 93: 1867–1879.

Mäder P, Fliessbach A, Dubois D, Gunst L, Fried P, Niggli U. (2002). Soil fertility and biodiversity in organic farming. Science 296: 1694–1697.

Malavolta E, Vitti GC, Oliveira SA. (1997). Avaliação do estado nutricional das plantas: princípios e aplicações. 2nd. ed. Piracicaba: POTAFOS, 319p.

Manzoni S, Porporato A. (2009). Soil carbon and nitrogen mineralization: Theory and models across scales. Soil Biol Biochem 41: 1355–1379.

Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman J a, Green JL, *et al*. (2006). Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 4: 102–112.

Maul JE, Buyer JS, Lehman RM, Culman S, Blackwood CB, Roberts DP, *et al*. (2014). Microbial community structure and abundance in the rhizosphere and bulk soil of a tomato cropping system that includes cover crops. Appl Soil Ecol 77: 42–50.

Mayr E. (1970). Populations, species, and evolution. An abridgement of animal species and evolution. Cambridge: Mass. London Belknap P. Harvard Univ P XV, 453p.

McCann KS. (2000). The diversity-stability debate. Nature 405: 228–233.

McDaniel MD, Tiemann LK, Grandy AS. (2014). Does agricultural crop diversity enhance soil microbial biomass and organic matter dyanmics? A meta-analysis. Ecol Appl 24: 560–570.

Mendes LW, Kuramae EE, Navarrete A a, van Veen J a, Tsai SM. (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. ISME J 1–11.

Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM, *et al*. (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science 332: 1097–100.

Nelson MB, Martiny AC, Martiny JBH. (2016). Global biogeography of microbial nitrogen-cycling traits in soil. Proc Natl Acad Sci 201601070.

Odum E, Barrett G. (2005). Fundamentals of ecology. 5th ed. Thomson Brooks/Cole, 624p.

Ogilvie LA. (2012). Microbial Ecological Theory: Current Perspectives. Caister Academic Press. Norfolk, 120p.

Pédron T, Mulet C, Dauga C, Frangeul L, Chervaux C, Grompone G. (2012). A Crypt-Specific Core Microbiota Resides in the Mouse Colon. MBio 3: e00116-12.

Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, *et al*. (2007). The role of ecological theory in microbial ecology. Nat Rev Microbiol 5: 384–392.

Rodrigues JLM, Pellizari VH, Mueller R, Baek K, Jesus EDC, Paula FS, *et al*. (2013). Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. Proc Natl Acad Sci U S A 110: 988–93.

Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, *et al*. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1: 283–90.

Sato H, Tsujino R, Kurita K, Yokoyama K, Agata K. (2012). Modelling the global distribution of fungal species: new insights into microbial cosmopolitanism. Mol Ecol 21: 5599–612.

Saunders AM, Albertsen M, Vollesen J, Nielsen PH. (2015). The activated sludge ecosystem contains a core community of abundant organisms. ISME J 10: 1–10.

Schlaeppi K, Dombrowski N, Oter RG, Ver Loren van Themaat E, Schulze-Lefert P. (2014). Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives. Proc Natl Acad Sci U S A 111: 585–92.

Schultz P, Urban NR. (2008). Effects of bacterial dynamics on organic matter decomposition and nutrient release from sediments: A modeling study. Ecol Modell 210: 1–14.

Shade A, Handelsman J. (2012). Beyond the Venn diagram: The hunt for a core microbiome. Environ Microbiol 14: 4–12.

Shade A, Jones SE, Gregory Caporaso J, Handelsman J, Knight R, Fierer N, *et al*. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio (4) 5:e01371-14.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, *et al*. (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. Proc Natl Acad Sci 103: 12115–20.

Tang X, Placella SA, Daydé F, Bernard L, Robin A, Journet EP, *et al*. (2016). Phosphorus availability and microbial community in the rhizosphere of intercropped cereal and legume along a P-fertilizer gradient. Plant Soil 1–16.

Thilagar G, Bagyaraj DJ, Raoca MS. (2016). Selected microbial consortia developed for chilly reduces application of chemical fertilizers by 50% under field conditions. Sci Hortic (Amsterdam) 198: 27–35.

Torsvik V, Øvreås L. (2002). Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol 5: 240–245.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, *et al*. (2009). A core gut microbiome in obese and lean twins. Nature 457: 480–484.

van der Voort M, Kempenaar M, van Driel M, Raaijmakers JM, Mendes R. (2016). Impact of soil heat on reassembly of bacterial communities in the rhizosphere microbiome and plant disease suppression. Ecol Lett 19: 1–8.

Wade MJ, Harmand J, Benyahia B, Bouchez T, Chaillou S, Cloez B, *et al*. (2016). Perspectives in mathematical modelling for microbial ecology. Ecol Modell 321: 64–74.

Wakelin SA, Gupta VVSR, Harvey PR, Ryder MH. (2007). The effect of Penicillium fungi on plant growth and phosphorus mobilization in neutral to alkaline soils from southern Australia. Can J Microbiol 53: 106–115.

Wakelin SA, Macdonald LM, Rogers SL, Gregg AL, Bolger TP, Baldock JA. (2008). Habitat selective factors influencing the structural composition and functional capacity of microbial communities in agricultural soils. Soil Biol Biochem 40: 803–813.

Wickings K, Grandy A. S, Reed SC, Cleveland CC. (2012). The origin of litter chemical complexity during decomposition. Ecol Lett 15: 1180–8.

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, *et al*. (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. ISME J 1–12.

Yang C, Hamel C, Gan Y, Vujanovic V. (2013). Pyrosequencing reveals how pulses influence rhizobacterial communities with feedback on wheat growth in the semiarid Prairie. Plant Soil 367: 493–505.

Yeoh YK, Paungfoo-Lonhienne C, Dennis PG, Robinson N, Ragan M A., Schmidt S, *et al*. (2015). The core root microbiome of sugarcanes cultivated under varying nitrogen fertiliser application. Environ Microbiol 18: 1338–135.

Zarraonaindia I, Owens SSM, Weisenhorn P, West K, Hampton-Marcell J, Lax S, *et al*. (2015). The soil microbiome influences grapevine-associated microbiota. MBio 6: 1–10.

3. A PROBABILISTIC MODEL TO IDENTIFY THE CORE MICROBIAL COMMUNITY

Abstract

The core microbial community has been correlated with essential process ecosystem functioning, from maintaining the human guts homeostasis to protection against plant disease. However, the identification of the core microbial community is based on arbitrary cutoffs, what results in the preferential selection of the most abundant microorganisms. Here, we developed and tested an approach to identify the core community based on a probabilistic model. The Poisson distribution was used to identify OTUs with a probable occurrence in every sample of a given dataset. We have determined the core communities of four available microbial datasets by our probabilistic method comparing the results with arbitrary method. The datasets were composed by microbiomes of humans (tongue, gut, and palms), mice (gut), grapevine (plant organs and bulk soil) microbiome, and maize rhizosphere. We identified core microbial communities based on the Poisson distribution with R^2 varying between 0.46 (mice) and 0.91 (grape), and with p-values lower than 0.05. The probabilistic method revealed core microbial communities with higher frequencies and diversity than those previously described. This method also indicated a strong presence of the rare community in core, what has been neglected by the arbitrary method. The probabilistic model can extend our knowledge about the core microbial community, revealing new roles of this group in the functionality of ecosystems, from human health to increase the crop production. It constitutes a new tool to step forward in the microbial community investigation.

Keywords: Core microbial; Variable microbial; Poisson distribution; Probabilistic model

3.1. Introduction

The composition of microbial communities can vary spatially and temporally. To explain this variation, several terms are used to classify the dynamics and structures of microbial communities across environments. For example, the term “core” is defined as organisms that are ubiquitous in a given habitat, despite environmental variations (Hanski, 1982). In microbial ecology, the term core could refer to microbial gene or taxa occurring in a set of samples in a given ecosystem. For example, the Human Microbiome Project (HMP; Turnbaugh *et al*., 2007) defines the core microbiome as common microbial gene present in all or in the majority of human bodies. Other studies describe the core microbial community as common Operational Taxonomic Units (OTU) across space or time (Shade and Handelsman, 2012).

The origin of the core microbial community in a given habitat is associated with coevolutionary processes (Pédron *et al*., 2012; Schlaeppi *et al*., 2014). Consequently, the core microbial could be correlated with essential roles in ecosystem functioning, and also be used as indicators of system perturbation (Shade and Handelsman, 2012; Saunders *et al*., 2015).

For example, an abundant microbial core was identified across 210 human adult colons with widespread geographic origin, ethnic background and diet (Sekelja *et al*., 2011). The authors suggested that this core microbial has an important role in gut homeostasis and health. Similar attempts have defined roles for the microbial core in other ecosystems, including corals (Ainsworth *et al*., 2015), zebrafish (Roeselers *et al*., 2011), mouse (Pédron *et al*., 2012), ruminants (Henderson *et al*., 2015), *Arabidopsis thaliana* (Lundberg *et al*., 2012) and sugarcane plants (Yeoh *et al*., 2015). The core identified in *A. thaliana* roots are composed by plant growth promoting rhizobacteria (PGPR) and also bacterial families which promote plant health (Schlaeppi *et al*., 2014).

However, the identification of the core community is challenging in microbial ecology. As the core microbial is defined to be ubiquitous in a habitat, it is supposed that the microbial taxa or genes belonging to the core should be found in every sample collected given a habitat. Thus, Shade and Handelsman (2012) have suggested the Venn diagram as a method to determine the core microbial. In this approach, the core microbial is represented by taxa found in every sample analyzed (100% of frequency across samples). However, to date, no methodological approach can fully assess the extent of the microbial diversity of a given sample (Kanagawa, 2003; Feinstein *et al*., 2009; Prosser, 2015). It is known that in complex communities – commonly found in environmental samples – the methods available preferentially identify the most abundant groups of microorganisms (Caporaso *et al*., 2011). Consequently, the rare community is missed in core microbial community studies. The most commonly used approach to circumvent this problem is the definition of arbitrary cutoffs for the frequency of microbes or genes to be classified as a member of the core microbial. For instance, researchers have used cutoff values ranging from 30% to 99% frequency across samples (Li *et al*., 2013; Ainsworth *et al*., 2015) to define the core community in environmental samples. However, these arbitrary cutoffs still not include the rare community and also could result in false assignments to the core microbial, thus influencing inferences about its function and composition.

A possible solution for the proper identification of the core microbial is to use a probabilistic model to classify members of the microbial community belonging to the core community. Here, we developed and tested an approach to identify the core community based on the Poisson distribution. Given the occurrence distribution of an event, i.e., microorganism, in a period of samples, the model expresses the probability of this event to occur at all points of this same period of samples (Rao and Rubin, 1964). Among discrete

probability models, we selected the Poisson distribution because it is described as suitable for large datasets, e.g. high number of events, and occurrence of small or rare probabilities (Karlis, 2003). The Poisson distribution was selected because using the distribution of occurrence and abundance of each OTU, it is possible to identify which OTU has a probability to be found in every sample. Using this probabilistic model, the group of OTUs selected probabilistically estimates the composition of the core microbial community.

Based on this approach, we hypothesized that i) it is possible to identify a probable group of members shared by samples representing the core microbial community, ii) the probable core community could represent a high percentage of the whole community varying across different systems (environment or host), and iii) the group designated as the probable members of the core microbial community differs from the group selected by arbitrary methods, in particular for low abundance microorganisms. To test these hypotheses, we applied comparative methods to available microbial datasets, including animal hosts (human and mice) and environmental samples (soils and plants of maize and grapevine). The datasets were obtained from the Earth Microbiome Project (EMP; http://www.earthmicrobiome.org). We selected four biological datasets based on OTUs tables and verified the difference in the core microbial community identified by then.

3.2. Material and Methods

3.2.1. Microbial Databases

We selected four datasets composed by microbiomes of humans (tongue, gut, and palms), mice (gut), grapevine (plant organs and bulk soil) microbiome, and maize rhizosphere to study the core microbial community identified by arbitrary methods and by the probabilistic method based on the Poisson distribution (Table 3.1).

Table 3.1 – Databases selected from EMP on Qiita plataform

	Databases selected from EMP			
	Environmental systems		Hosts systems	
	Grape	**Maize**	**Human**	**Mice**
Study EMP-ID	2382	1792	550	77
Title	The Soil Microbiome Influences Grapevine-Associated Microbiota	Diversity and heritability of the maize rhizosphere microbiome under field conditions	Moving pictures of the human microbiome	A core gut microbiome in obese and lean twins
Number of samples	401	442	1,736	271
Data Type	16S - HiSeq	16S – 454 FLX	16S – 454 FLX	16S – 454 FLX
Number of reads / sample	1,000	2,080	5,000	1,000
OTUs	8,583	10,747	16,129	4,495
References	(Zarraonaindia *et al.*, 2015)	(Peiffer *et al.*, 2013)	(Caporaso *et al.*, 2011)	(Turnbaugh *et al.*, 2009)

The human microbiome database consists of 396 samples, collected along a temporal analysis of two individuals at four body sites, including gut, tongue, and left and right palm (Caporaso *et al*., 2011). In the original study, the authors aimed to evaluate the temporal variation in the human microbiome. The authors used the terms persistent (individuals with high occurrence across the samples), and transient (individuals with low occurrence across the samples) community, because it identified a very small temporal core across all samples. The core was obtained by a cutoff of 100% of samples.

The mice database was used to evaluate how the gut microbiome influences host adiposity (Turnbaugh *et al*., 2009). The data is composed of fecal samples from 154 individuals (mice) divided into adult females, monozygotic or dizygotic twin pairs, and their mothers. The identification of the core microbial community was established by the occurrence of OTUs in all samples (cutoff of 100% of samples). These authors included in their study the Phylotype Sampling Model, which is very similar to the model proposed in this work. However, in the previous study, the Poisson distribution was used to estimate the failures to observe the microbial groups possibly belonging to the core community, which it is different from our purpose with the Poisson distribution.

In the grapevine database, Zarraonaindia *et al*., (2015) identify shared OTUs across grapevine organs (flower, leaves, grapes, root), root zone, and bulk soil over two growing seasons. The authors reduced the cutoff to 75% of the samples to determine the core community. This decision was justified by the authors due to the absence of OTUs occurring across all samples.

The maize database is the only one not previously used to determine the core community. The authors aimed to determine the impact of genetic variance on the composition of bacterial communities that live in the maize rhizosphere (Peiffer *et al*., 2013).

The biological observation matrices (BIOM) derived from these data were obtained from the Earth Microbiome Project (EMP; http://www.earthmicrobiome.org), available on the Qiita platform (https://qiita.ucsd.edu). We used the BIOM files due to the similar treatment of data by bioinformatics, including quality filters and assignment of OTU taxonomy (Caporaso *et al*., 2011; Elli *et al*., 2010; Peiffer *et al*., 2013; Zarraonaindia *et al*., 2015). We used the software Qiime (Caporaso *et al*., 2010) to convert the BIOM files into text files, which were further imported into the R software (Team 2016), where we analyzed it using the packages '*RAM*' (Chen *et al*., 2016), '*vegan*' (Oksanen *et al*., 2016) and '*Hmisc*' (Harrell Jr *et al*., 2016).

3.2.2. Methods to identify the core microbial community

The identification of the core microbial community is conventionally obtained by defining limits of frequency across the samples, i.e., a core community could be defined as microorganisms occurring in all samples (100% of occurrence frequency) or in a part of the samples (varying from 30% to 90% of frequency). For example, Ainsworth *et al*., (2015) identified the ubiquitous endosymbiont bacterial community (or core community) associated with corals using the cut-off of 30% for the occurrence frequency. Similarly, the human, mice and grapevine studies were used determined the core community, respectively at levels of 100%, 100% and 75% of occurrence frequency across the samples. We used a range of limits - 30, 40, 50, 60,70, 80, 90 and 100% of occurrence frequency - based on the OTU tables across the samples to verify the difference in the core microbial community selected by these methods.

The method proposed here is based on the probability test for the distribution of each microorganism (indicated by OTU) among samples, what is directly linked to the probability of its occurrence in every sample of the given environment. This probability test is based on the Poisson distribution, which is a discrete random probability regression model. The Poisson distribution expresses the probability of events occurring in a period (Rao and Rubin, 1964). Here we treat events as OTUs across a series of collected samples. The Poisson distribution has been previously used in biogeographic studies to predict the abundance of

species in a targeted ecosystem (Vincent and Haworth, 1983; Guisan and Zimmermann, 2000) or to support the Phylotype Sampling Model (Turnbaugh *et al*., 2009), where the authors verify the sampling error expected given the sample size and the probability of observing the minimum abundance of a microorganism in any sample.

The probability (**P**) of Poisson distribution is obtained by equation (1):

(1) $$\mathbf{P(x) = \lambda^{x}\, e^{-\lambda}/x!}$$

Where, the lambda (λ) and x respectively represent the average of relative abundance and the occurrence frequency of each taxon across the communities. Using this formula, we have tested two hypotheses: H_0 – the individual (OTU) fits in the Poisson distribution and thus has a probability of occurring in every sample, indicating that it can not be excluded from the core microbial community; H_1 – the individual does not fit in the Poisson distribution, and thus does not have a probability to occurs in every sample, supporting its exclusion from the core microbial community.

The identification of the core microbial community by the Poisson distribution starts with the determination of the average of sequences per community source (***N*** in the equation 2), the average of relative abundance of each taxa across communities (***p*** in the equation 3) and the occurrence frequency of each taxa across communities (***f*** in the equation 4).

(2) $$\mathbf{N} = \frac{\textbf{Total number of sequences}}{\textbf{Number of OTUs}},$$

(3) $$\boldsymbol{p} = \frac{\sum_{i=1}^{n} \boldsymbol{ab}}{\boldsymbol{n} \times \boldsymbol{N}},$$

(4) $$\boldsymbol{f} = \frac{\sum_{i=1}^{n} \boldsymbol{rich}}{\boldsymbol{n} \times \boldsymbol{N}},$$

Where, the ***n*** is the number of samples, ab is the abundance of OTUs and rich is the richness of OTUs. The ***p*** and ***f*** are calculated with values of ***ab*** and ***rich*** > 0, and they are used in the Poisson distribution, where the $\boldsymbol{\lambda}$ is obtained per OTU by the equation (5):

(5) $$\boldsymbol{\lambda = N \times p}$$

The fitness of OTUs in the Poisson distribution is verified by the R^2 (adjusted) and p-value. The arbitrary (thresholds of 30, 40, 50, 60, 70, 80, 90 and 100%) and the probabilistic (Poisson distribution) methods resulted in OTU tables for the core microbial community and the "variable" community (made of those that do not belong to the core community). The statistical analyses comparing the results were performed using the R software version 3.2.2

(R Core Team, 2015), including the Shannon index. We also developed a function in R, which identifies a core microbial community by the method based on the Poisson distribution.

3.3. Results

3.3.1. The probabilistic method to identify the core microbial community

Despite the importance of the core microbial community, their identification has been based on arbitrary cutoffs. Thus, we developed a new method to identify a core microbial community based on the Poisson distribution. Using this probabilistic method, we identified core microbial communities for each dataset selected for analysis with R^2 varying between 0.46 (mice) and 0.91 (grape), and with p-values as lower than 0.05. The obtained curves indicated the occurrence of OTUs with distinct frequencies as components of the core microbial communities, what is not observed when other approaches are used (Figures 3.1, 3.2, 3.3 e 3.4).

Using the mice dataset as an example, the core microbial community identified by the probabilistic method (Figure 3.1A) differs from the core microbial community identified by the thresholds used conventionally (Figure 3.1B). The same differentiation was observed for the human, maize, and grape datasets (Figure 3.2, 3.3, and 3.4). In all analyzed datasets, we could not detect OTUs with 100% of occurrence frequency. However, when using other values for cut-offs, the numbers of OTUs belonging to the core community increased with decreasing threshold values (Figures 3.1B, 3.2B, 3.3B, and 3.4B). In addition, it was demonstrated by the new proposed method that most of the OTUs previously named as a component of the core community by the arbitrary frequency-based methods, did not fit the Poisson distribution (Figures 3.1A, 3.2A, 3.3A, and 3.4A), thus indicating a low probability thus these occur in every sample collected.

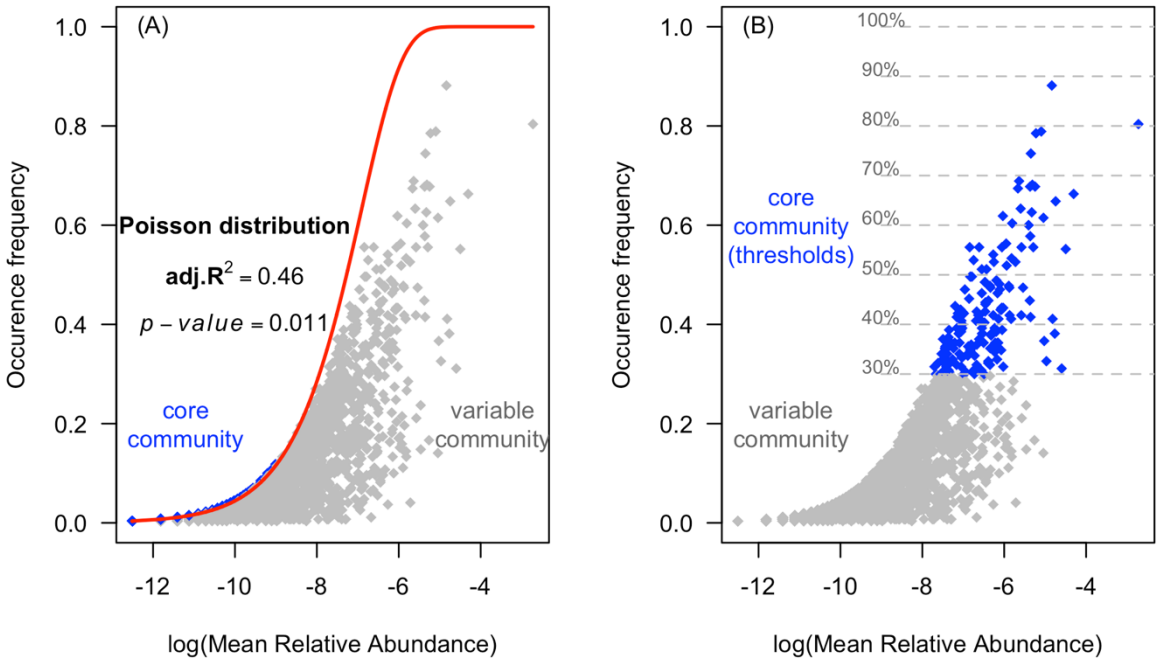

Figure 3.1 – Comparison of the core and variable communities from mice database identified by the probabilistic method based on the Poisson distribution (A) and by the arbitrary frequency cutoffs used in the literature (B)

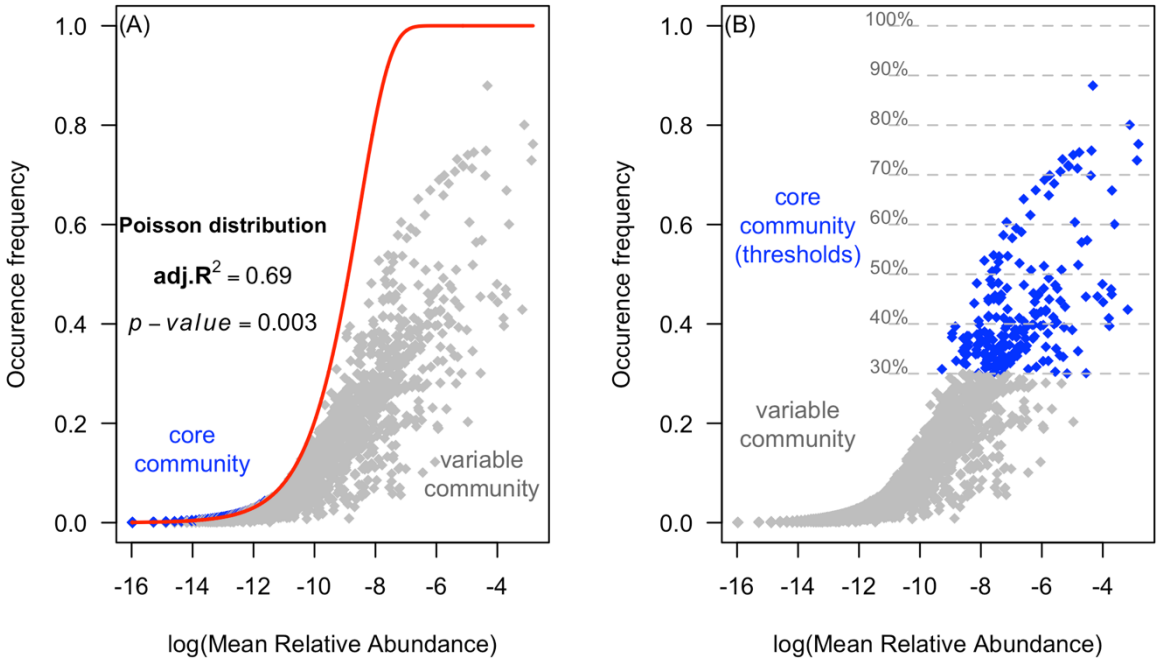

Figure 3.2 – Comparison of the core and variable communities from human database identified by the probabilistic method based on the Poisson distribution (A) and by the arbitrary frequency cutoffs used in the literature (B)

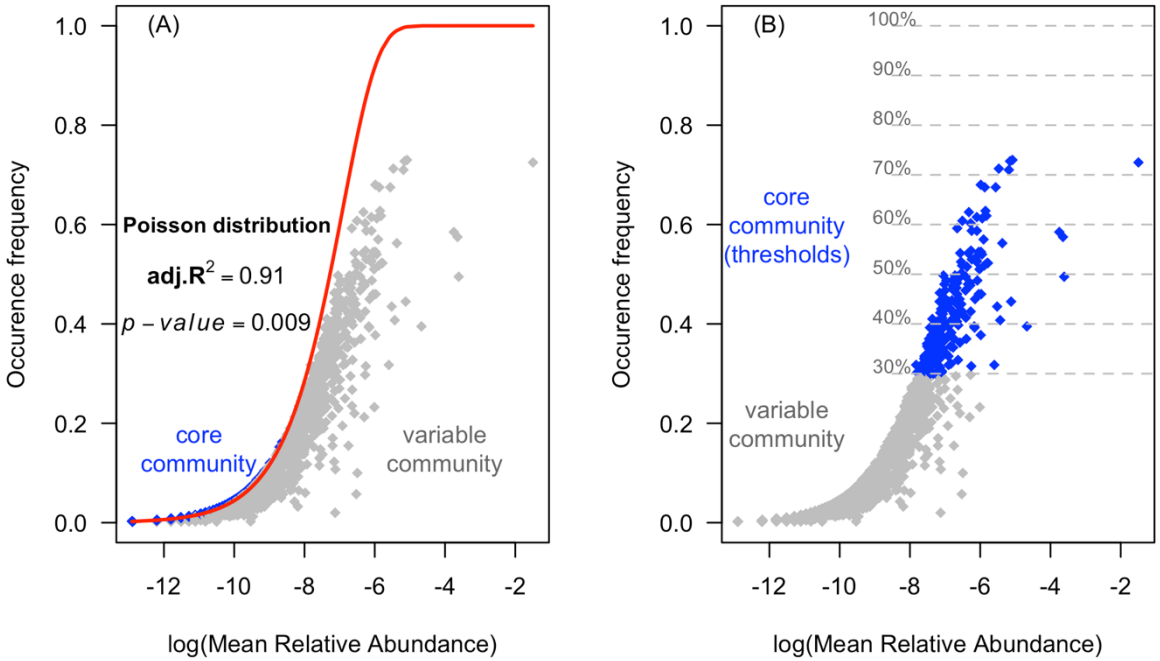

Figure 3.3 - Comparison of the core and variable communities from grape database identified by the probabilistic method based on the Poisson distribution (A) and by the arbitrary frequency cutoffs used in the literature (B)

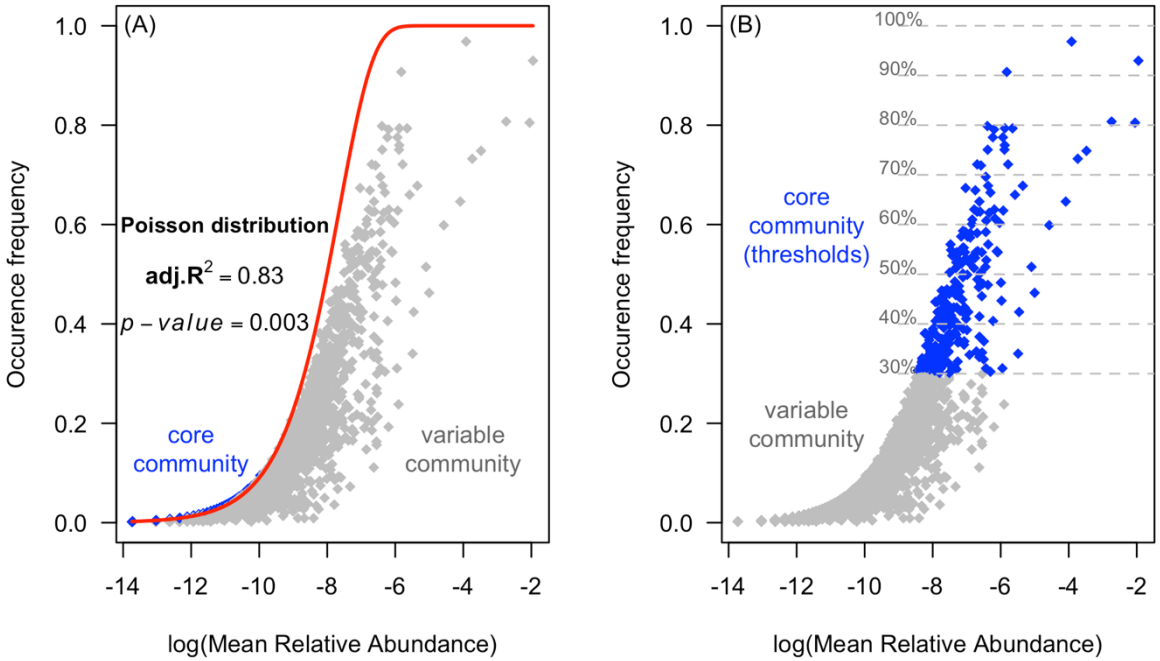

Figure 3.4 – Comparison of the core and variable communities from maize database identified by the probabilistic method based on the Poisson distribution (A) and by the arbitrary frequency cutoffs used in the literature (B)

Using the new method, we could find a greater number of OTUs classified as belonging to the core community (Table 3.2). For example, the core microbial community identified in

the mice database is composed of 170 OTUs using the threshold of 30% of frequency in the arbitrary method, and 1,717 OTUs using the method based on the Poisson distribution.

Table 3.2- Number of OTUs obtained by the arbitrary and probabilistic methods across the databases

		Databases							
		Grape		Maize		Human		Mice	
Methods		Core	Variable	Core	Variable	Core	Variable	Core	Variable
Arbitrary tresholds	30%	211	8,372	272	10,475	206	15,923	170	4,325
	40%	109	8,474	145	10,602	93	16,036	82	4,413
	50%	40	8,543	80	10,667	42	16,087	35	4,460
	60%	15	8,568	39	10,708	24	16,105	19	4,476
	70%	5	8,578	19	10,728	12	16,117	5	4,490
	80%	0	8,583	5	10,742	2	16,127	2	4,493
	90%	0	8,583	3	10,744	0	16,129	0	4,495
	100%	0	8,583	0	10,747	0	16,129	0	4,495
Poisson distribution		5,039	3,544	5,294	5,453	8,751	7,378	1,717	2,778

The percentage of sequences belonging to the core community as determined by the newly proposed method varied among datasets (Figure 3.5). For example, the core microbial community in grape and human represents 58.71% and 54.26% of total sequences, respectively. The core community associated with maize was nearly 50% of the sequences, and the mice data indicated the lowest percentage of sequences belonging to the core community (38.20%).

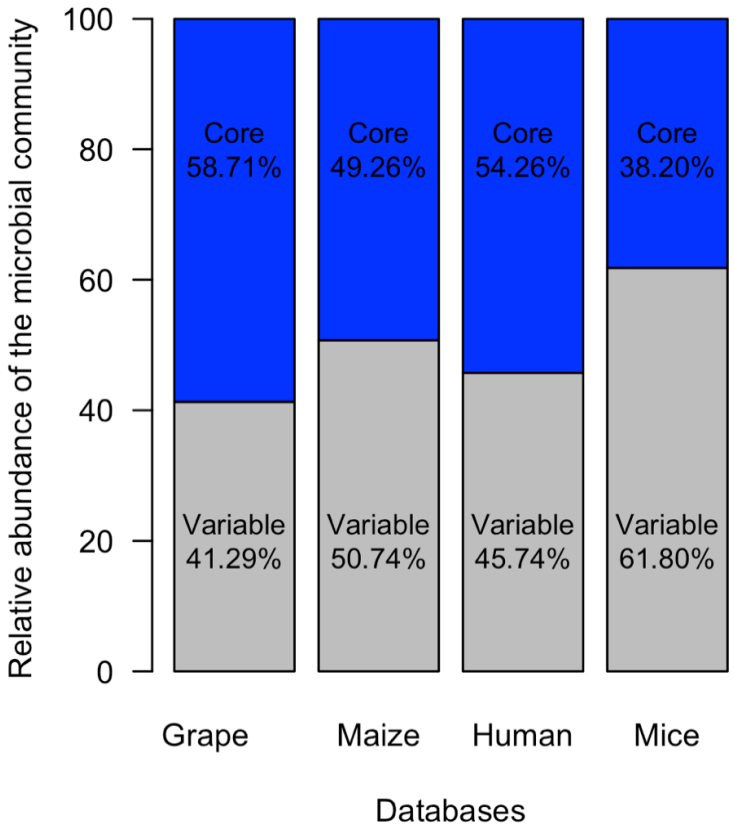

Figure 3.5 - Relative abundance of OTU classified as belonging to the core or variable communities in whole microbial community from environmental (grapevine and maize) and host (human and mice) databases, as determined by the distribution of Poisson

3.3.2. The composition of the core microbial community identified by the probabilistic and arbitrary methods

We found variation in the distribution of OTUs among taxonomic groups, even at the phylum level, when comparing results from distinct methodologies applied to all data. The number of phyla identified by the method proposed here is higher than those determined by the arbitrary methods.

In the mice analysis, the method based on the Poisson distribution identified the same three phyla as the arbitrary method (Actinobacteria, Bacterioidetes, and Firmicutes) and also identified eight more phyla (Cyanobacteria, Fusobacteria, Lentisphaerae, Proteobacteria, Synergistetes, Tenericutes, TM7, and Verrucomicrobia) as members of the core microbial community (Figure 3.6). Similar results were observed for the human, maize and grape data (Figures 3.7, 3.8, and 3.9).

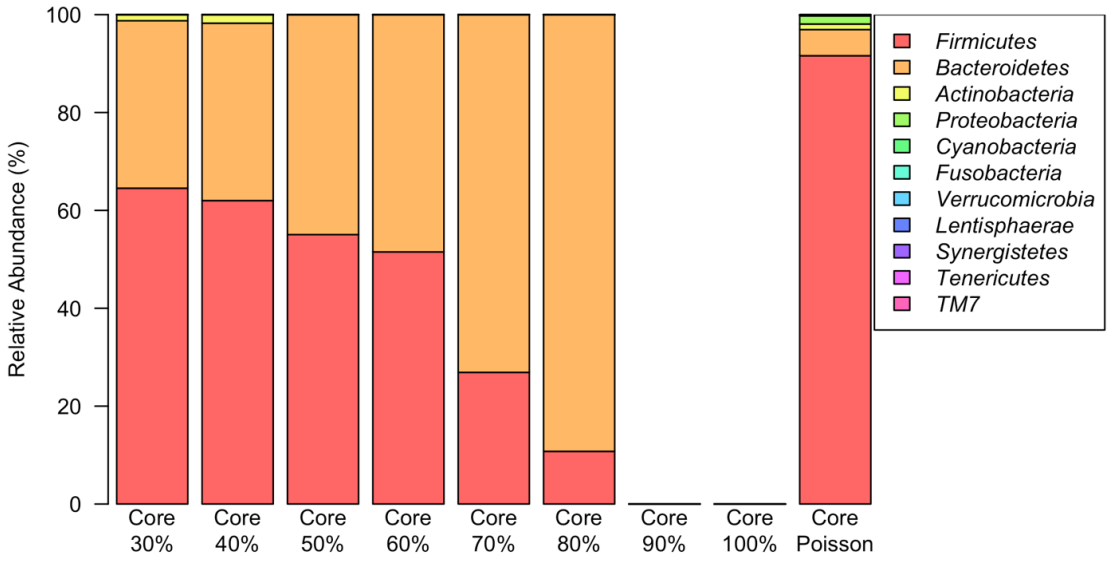

Figure 3.6 - Relative abundance of bacterial phyla in the core communities derived from the mice database, as determined by the arbitrary cutoffs or by the probabilistic model based on Poisson distribution

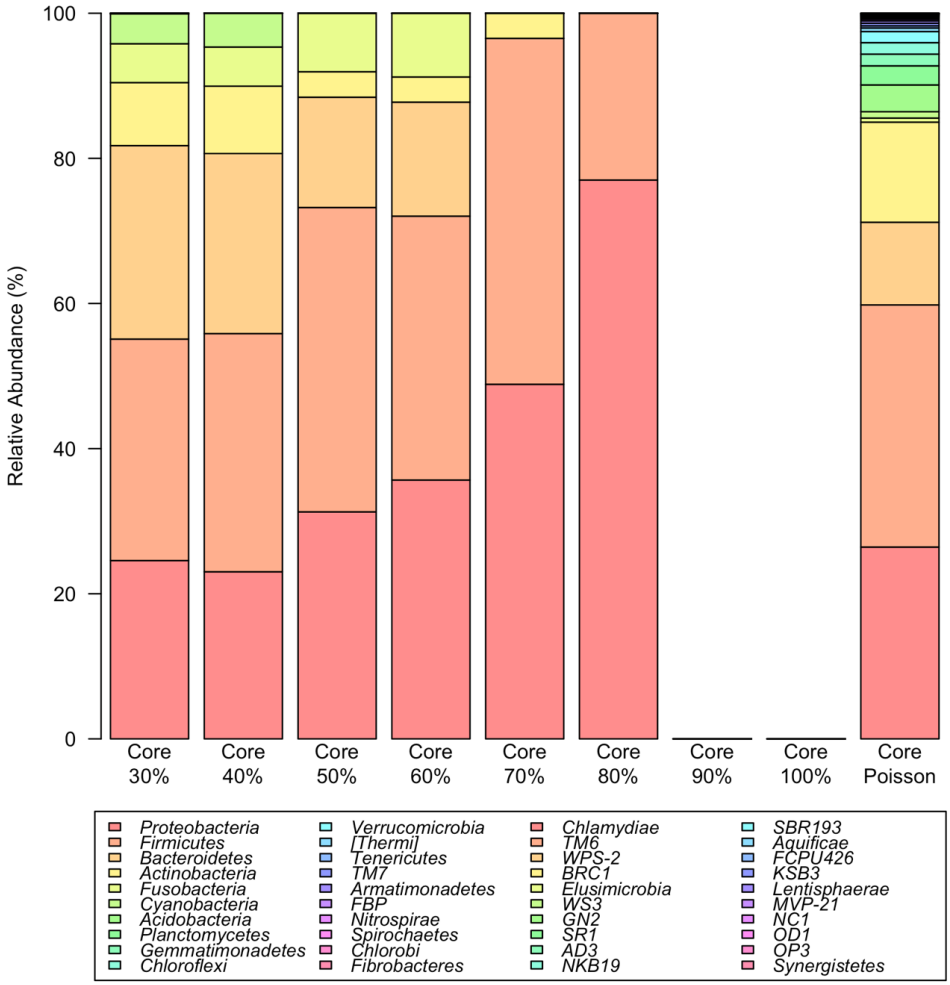

Figure 3.7 - Relative abundance of bacterial phyla in the core communities derived from the human database, as determined by the arbitrary cutoffs or by the probabilistic model based on Poisson distribution

We also observed for the arbitrary methods that the number of groups and the structure of the microbial core community identified are entirely dependent on the threshold level chose. For example, the *Cyanobacteria* and *Proteobacteria* are respectively the higher (70% of the relative abundance) and lower (5% of the relative abundance) phyla identified at the 90% of the occurrence frequency for the maize database (Figure 3.8). However, using the 30% of the occurrence frequency, the *Proteobacteria* phylum represent more than 50% of the abundance in the microbial core community.

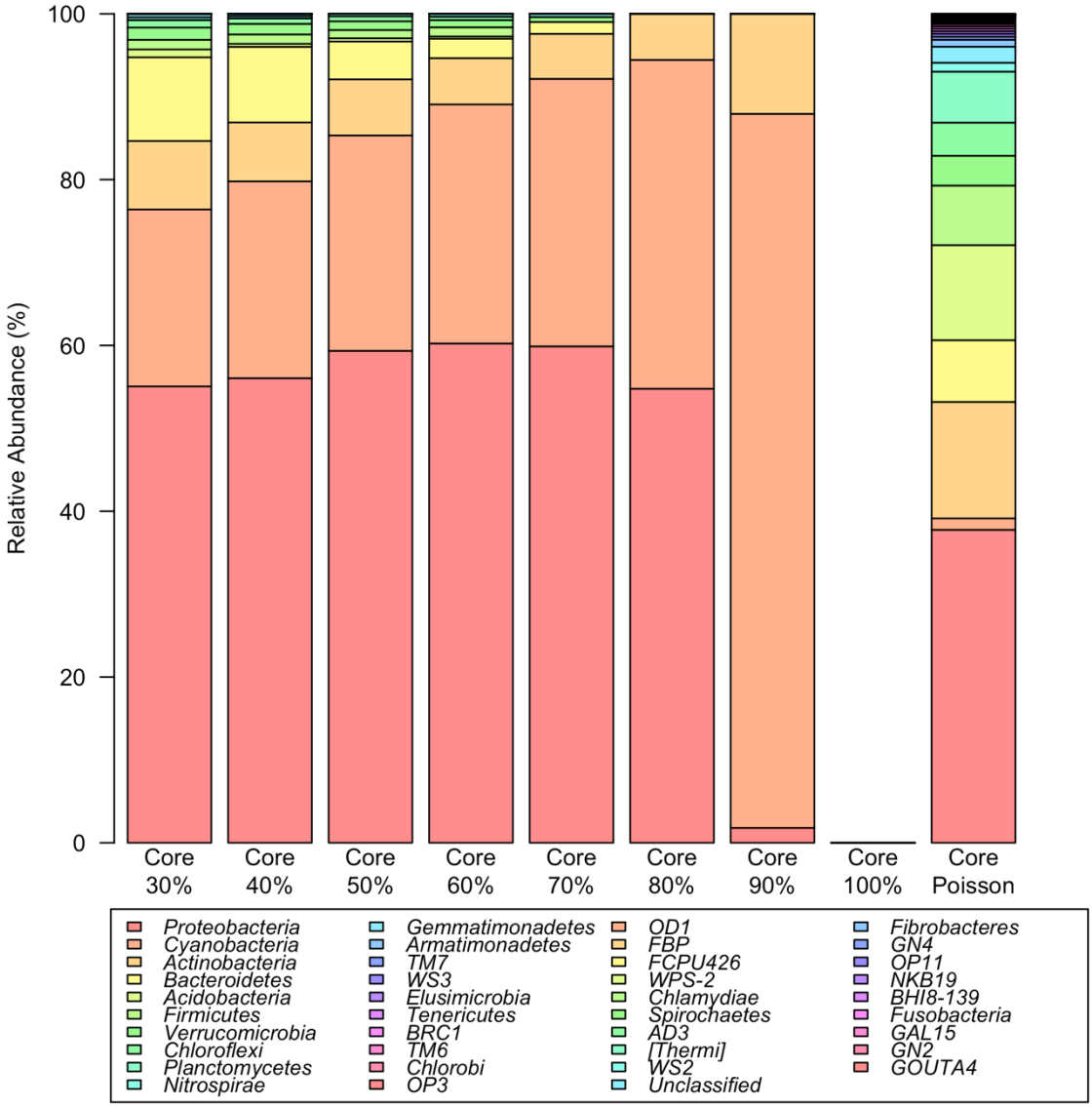

Figure 3.8 - Relative abundance of bacterial phyla in the core communities derived from the maize database, as determined by the arbitrary cutoffs or by the probabilistic model based on Poisson distribution

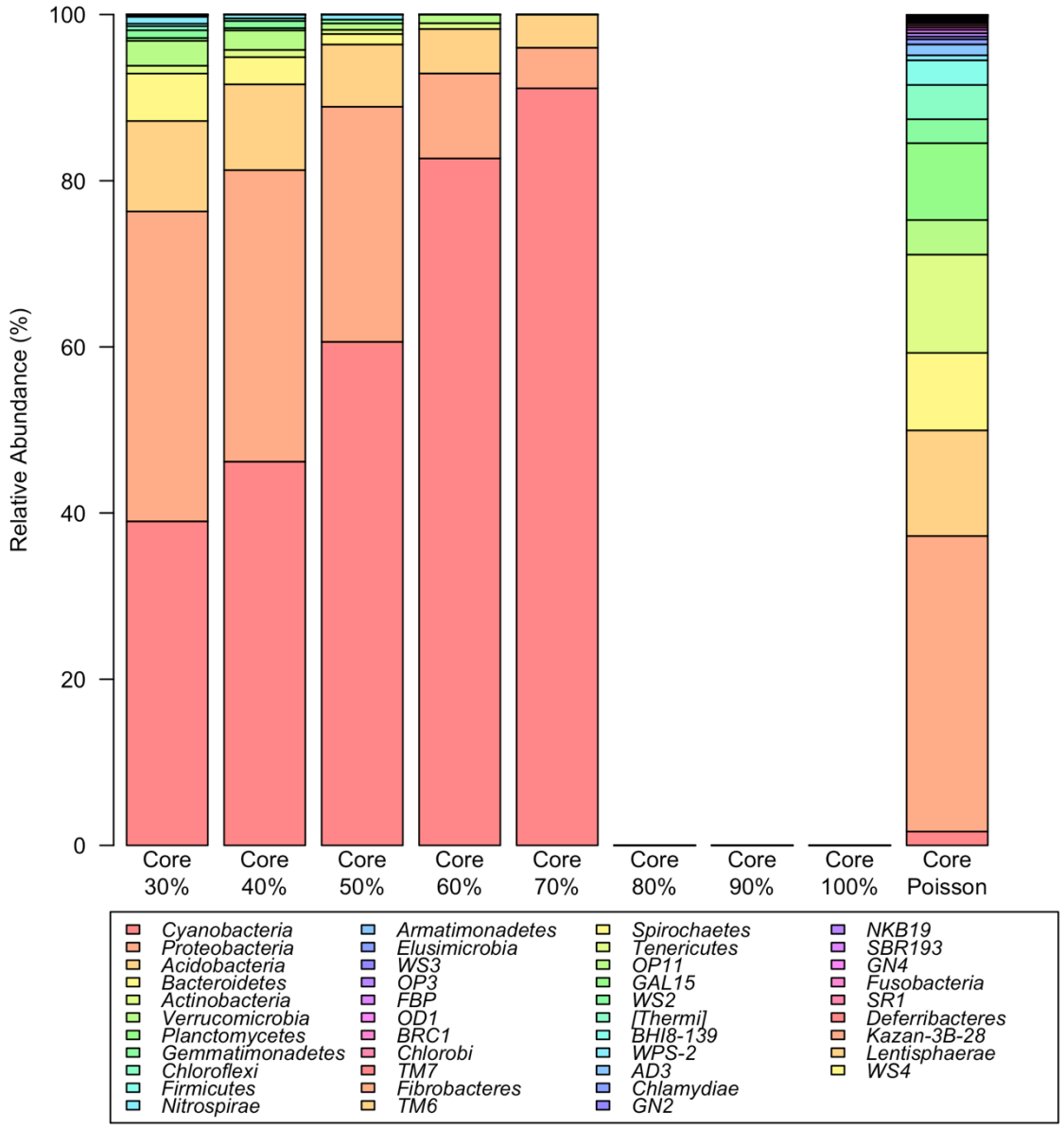

Figure 3.9 - Relative abundance of bacterial phyla in the core communities derived from the grape database, as determined by the arbitrary cutoffs or by the probabilistic model based on Poisson distribution

We also confirmed the higher diversity of the core community identified by the proposed method by comparing values of the Shannon index derived from communities obtained using each approach (Table 3.3).

Table 3.3 – Bacterial diversity, represented by the Shannon index, databases from environmental (grapevine and maize) and host (human and mice) databases, as determined by the the arbitrary methods or by the distribution of Poisson

		Databases			
Core methods		**Mice**	**Human**	**Maize**	**Grape**
Thresholds	**30%**	4.48	4.23	3.61	3.59
	40%	3.78	3.65	3.13	2.93
	50%	2.93	2.93	2.71	1.95
	60%	2.35	2.59	2.26	0.9
	70%	0.95	2.03	1.85	0.42
	80%	0.34	0.54	1.24	0
	90%	0	0	0.46	0
	100%	0	0	0	0
Poisson distribution		7.03	8.6	8.18	8.04

3.4. Discussion

The occurrence of a core microbial community given a habitat has been associated with a coevolutionary process. For example, a core microbial community was identified in ants with similar trophic niches, playing important roles as fixing, recycling, or upgrading nitrogen (Anderson *et al*., 2012). Thus, the dispersal of ant species may be correlated with their specific core microbial community. The importance of the core microbial community is also observed in human gut, maintaining the gut homeostasis (Sekelja *et al*., 2011); in insects promoting nutrition and defense (Douglas, 2011); in zebrafish, playing a role in vertebrate immunity (Star and Jentoft, 2012), and in rhizosphere of *A. thaliana*, promoting plant growth (Schlaeppi *et al*., 2014).

However, the arbitrary methods used to identify the core microbial community are based on arbitrary thresholds of occurrence frequency. The arbitrary methods can generate two main problems: the failure to identify taxa that belong to the core microbial community (false negatives), and the improper inclusion of taxa as a member of the core microbial community (false positives). These issues are mainly related to the limited capacity of methods to assess the full content of microbial complex communities. In combination, these failures might have blurred our view, limiting proper inferences on the functionality of the core communities in microbial communities and their associated hosts and environments. In this work, we aimed to solve this problem, proposing a new method to identify the core microbial community using a probabilistic method, based on the Poisson distribution. We compared the efficacy of this method to those based on arbitrary cut-offs in four available

databases, covering microbiomes of humans (tongue, gut, and palms), mice (gut), grapevine (plant organs and bulk soil) microbiome, and maize rhizosphere.

Despite the distinct characteristics of size, number of samples, or sequencing technologies of each dataset (Table 3.1), the Poisson distribution fit to each microbial community, and succeeded in identifying the possible members of the core community. As the results were based on a probabilistic method, we expected that our proposed method would identify a group closer to the real core community than the group identified by the arbitrary methods. We observed that the probabilistic method proposed here generated a probable core microbial community greater in frequency, richness and diversity (Table 3.3) than those obtained by the arbitrary methods. Especially, these differences were found for the occurrence of OTUs with low abundance, much more pronounced in the core community obtained by the method based on the Poisson distribution. In the literature, the microorganisms with low abundance are defined as components of the rare community or rare biosphere (Galand *et al*., 2009). The rare community was first described as microorganism with low growth rates, which could act as a "seed bank" of species or genes important to maintain the functional redundancy of a system (Pedrós-Alió, 2006). These taxa could become dominant (in high abundance) under different conditions (Shade *et al*., 2014). Under this view, the rare community is classified as conditionally rare taxa (CRTs), suggested to be ubiquitous in some systems (Shade and Gilbert, 2015). As the core microbial community, the CRTs is described as an important group to determine the stability and functional resilience of the system. Unfortunately, when arbitrary methods are used to determine the core community these CRTs are rarely identified as members of the core community due to their lower frequency (for example, *see* Figure 3.1B). These groups could be properly classified within the core community using the proposed method.

In some cases, the cut-offs used may fail to identify members of the core microbial community, which could be called "false negatives". For example, data from Turnbaugh *et al*., (2009) representing our mice database, did not identify a core microbial community across 100% of samples from the gut microbiome of mice. However, comparing the bacterial phyla found in obese and lean mice, the authors found a similar frequency of *Firmicutes* across the samples. Applying our approach to this dataset, most OTUs named as components of the core microbial community belong to the phylum *Firmicutes* (Figure 3.6). The authors also indicated the distinct proportions of the Bacteroidetes and Actinobacteria phyla

associated to obese and lean mice. Both phyla were also detected by the probabilistic method, with OTUs affiliated with these groups as components of the core microbial community.

In order to avoid problems related to the core concept, some researchers have used the term persistent which is applied for taxa with a high occurrence frequency, but below to 100%, and transient which is used for taxa occurring in a lower number of samples. For example, Caporaso *et al*., (2011) did not found a temporal core microbial community across the human-derived samples, but the authors identify a persistent and transient communities, arbitrary classified as those occurring in 60% or 20% of samples, respectively. Using this dataset (Caporaso *et al*. 2011), we identified a probable core community (based on OTUs) across all of the human site samples made of 8,751 OTUs (Figure 3.7). The authors identified classes belonging to the phyla Firmicutes, Proteobacteria, Bacteroidetes, and Tenericutes phyla communities in the human gut. Similar result was obtained by our approach, with the major affiliation of the OTUs to the phyla Firmicutes, Proteobacteria, and Bacteroidetes. We believe that our approach better succeeds to identify the core community for two reasons. First, our method identified core communities across assessments previously identified as not having a core community (as determined by the 100% of frequency occurrence). Second, our method could abolish the use of subjective and arbitrary concepts, such as “persistent” or “transient” communities.

The same efficiency was observed applying our proposed method to datasets derived from environmental studies. Peiffer *et al*., (2013) exposed a detailed experimental design to identifying a core microbial community in the bulk soil and maize rhizosphere cultivated at four different site. In this dataset, our proposed method identified 5,294 OTUs as components of the core microbial community (Figure 3.3), which was almost half of the total OTUs (Figure 3.5). Interestingly, with exception of the 90% threshold, Proteobacteria phylum was the dominant group identified by the proposed and arbitrary methods (Figure 3.8).

Zarraonaindia *et al*., (2015) suggested a bacterial core community identified by three OTUs across 75% of samples from grape (leaves, flowers, grapes, and roots) and soils, over two growing seasons. These OTUs were associated with the genera Bradyrhizobium, Steroidobacter and Acidobacteria. Using this dataset, the core community estimated by the arbitrary method (with 70% of occurrence frequency) identified five OTUs, all in accordance with taxonomic results previously described. By using our proposed method on the same dataset, 5,039 OTUs were identified as belonging to the core community (Figure 3.4A and 3.4B). In addition, the Cyanobacteria phylum, which was a dominant group identified by the

arbitrary methods (90% of relative abundance; Figure 3.9), was identified as a small group of the core microbial community using the probabilistic method.

Taking together, the results obtained in this study, clearly show that it is possible to identify a core microbial community, derived from multiple communities, by a probabilistic model. The probabilistic model revealed that the core microbial community could be higher in abundance and diversity than previously expected and described. This method also indicated a strong presence of the rare community to the core microbial community, what has been neglected by the arbitrary methods used for the determination of core communities. The probabilistic model can extend our knowledge about the core microbial community, revealing news roles of this group in the functionality systems. The better knowledge on core microbial functions could support more robust studies in several fields, from human health (Zaura *et al*., 2009) to increase the crop production. The microbial core community could also be used as indicators of system perturbation (Shade and Handelsman, 2012), informing with a high precision a disease occurrence, or being used to adjust the composition and quantity of fertilization applied in a crop field. Future study investigating the core microbial community, the drivers of its composition of the functionality of this microbial group, would be possibly provided of a more realistic strategy to define it. The probabilistic model is a new tool to step forward in the microbial community investigation.

References

Ainsworth TD, Krause L, Bridge T, Torda G, Raina J-B, Zakrzewski M, *et al*. (2015). The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. ISME J 1–14.

Anderson KE, Russell JA, Moreau CS, Kautz S, Sullam KE, Hu Y, *et al*. (2012). Highly similar microbial communities are shared among related and trophically similar ant species. Mol Ecol 21: 2282–2296.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al*. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7: 335–6.

Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, *et al*. (2011). Moving pictures of the human microbiome. Genome Biol 12: R50.

Chen W, Simpson J, Levesque CA. (2016). RAM: R for Amplicon-Sequencing-Based Microbial-Ecology. http://cran.r-project.org/package=RAM.

Douglas AE. (2011). Lessons from studying insect symbioses. Cell Host Microbe 10: 359–367.

Elli M, Colombo O, Tagliabue A. (2010). A common core microbiota between obese individuals and their lean relatives? Evaluation of the predisposition to obesity on the basis of the fecal microflora profile. Med Hypotheses 75: 350–352.

Feinstein LM, Sul WJ, Blackwood CB. (2009). Assessment of bias associated with incomplete extraction of microbial DNA from soil. Appl Environ Microbiol 75: 5428–5433.

Galand PE, Casamayor EO, Kirchman DL, Lovejoy C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. Proc Natl Acad Sci U S A 106: 22427–22432.

Guisan A, Zimmermann NE. (2000). Predictive habitat distribution models in ecology. Ecol Modell 135: 147–186.

Hanski I. (1982). Dynamics of regional distribution : the core and satellite species hypothesis. Oikos, 38:210–221.

Harrell Jr FE, with contributions from Charles Dupont, many others. (2016). Hmisc: Harrell Miscellaneous. https://cran.r-project.org/package=Hmisc.

Henderson G, Cox F, Ganesh S, Jonker A, Young W, Abecia L, *et al*. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. Sci Rep 5: 14567.

Kanagawa T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng 96: 317–323.

Karlis D. (2003). An EM algorithm for multivariate Poisson distribution and related models. J Appl Stat 30: 63–77.

Li K, Bihan M, Methé BA. (2013). Analyses of the stability and core taxonomic memberships of the human microbiome. PLoS 8(5): e63139.

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, *et al*. (2012). Defining the core Arabidopsis thaliana root microbiome. Nature 488: 86–90.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, *et al*. (2016). Vegan: Community Ecology Package. https://cran.r-project.org/package=vegan.

Pédron T, Mulet C, Dauga C, Frangeul L, Chervaux C, Grompone G. (2012). A Crypt-Specific Core Microbiota Resides in the Mouse Colon. MBio 3: e00116-12.

Pedrós-Alió C. (2006). Marine microbial diversity: can it be determined? Trends Microbiol 14: 257–263.

Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, *et al*. (2013). Diversity and

heritability of the maize rhizosphere microbiome under field conditions. Proc Natl Acad Sci U S A 110: 6548–6553.

Prosser JI. (2015). Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. Nat Rev 13: 439.

R Core Team. (2015). R: A Language and Environment for Statistical Computing. https://www.r-project.org/.

Rao CR, Rubin H. (1964). On a characterization of the poisson distribution. Indian J Stat 26: 295–298.

Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K, *et al*. (2011). Evidence for a core gut microbiota in the zebrafish. ISME J 5: 1595–608.

Saunders AM, Albertsen M, Vollesen J, Nielsen PH. (2015). The activated sludge ecosystem contains a core community of abundant organisms. ISME J 10: 1–10.

Schlaeppi K, Dombrowski N, Oter RG, Ver Loren van Themaat E, Schulze-Lefert P. (2014). Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives. Proc Natl Acad Sci U S A 111: 585–92.

Sekelja M, Berget I, Næs T, Rudi K. (2011). Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. ISME J 5: 519–531.

Shade A, Gilbert JA. (2015). Temporal patterns of rarity provide a more complete view of microbial diversity. Trends Microbiol 23: 335–340.

Shade A, Handelsman J. (2012). Beyond the venn diagram: the hunt for a core microbiome. Environ Microbiol 14: 4–12.

Shade A, Jones SE, Gregory Caporaso J, Handelsman J, Knight R, Fierer N, *et al*. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio (4) 5:e01371-14.

Star B, Jentoft S. (2012). Why does the immune system of Atlantic cod lack MHC II? BioEssays 34: 648–651.

Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, *et al*. (2009). A core gut microbiome in obese and lean twins. Nature 457: 480–484.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. (2007). The Human Microbiome Project. Nature 449: 804–810.

Vincent PJ, Haworth JM. (1983). Poisson regression models of species abundance. J Biogeogr 10: 153–160.

Yeoh YK, Paungfoo-Lonhienne C, Dennis PG, Robinson N, Ragan M A., Schmidt S, *et al*.

(2015). The core root microbiome of sugarcanes cultivated under varying nitrogen fertiliser application. Environ Microbiol 18: 1338–135.

Zarraonaindia I, Owens SSM, Weisenhorn P, West K, Hampton-Marcell J, Lax S, *et al*. (2015). The soil microbiome influences grapevine-associated microbiota. MBio 6: 1–10.

Zaura E, Keijser BJF, Huse SM, Crielaard W. (2009). Defining the healthy 'core microbiome' of oral microbial communities. BMC Microbiol 9: 259.

4. SPATIAL AND TEMPORAL DYNAMICS OF THE MICROBIAL COMMUNITIES IN SOILS CULTIVATED WITH SUGARCANE

Abstract

The identification of the core microbial community has been suggested as an important tool to understand the association of microorganisms and hosts or environmental systems. However, how abiotic components, such soil chemistry or physics, and crop management which could modulate the core and variable microbial communities remain unclear. Here, we identified the core and variable microbial communities across soils cultivated with sugarcane in four different soils, under distinct climate and crop managements, sampled along two years. Using a probabilistic model based on the Poisson distribution, we identified the core of bacterial and fungal communities which represented 70.35% and 57.55% of their respective communities. While the structure of the variable bacterial community showed a significant correlation with pH (12.37%), phosphorus (13.53%), and sulfur (8.28%), the difrerential structuring of the core bacterial community was correlated with temperature (1.67%). Similar results were observed for core fungal community. We also observed differences on the predictive functionalities of these fractions of communities, with genes involved in nitrogen fixation (*Nif*DKH) more predicted in the core bacterial community and genes involved in nitrification (*Amo*CAB) in the variable bacterial community. Overall, this study indicates that the core microbial community differs from the variable community in taxonomic composition, correlations with the abiotic components and also may have specific functionalities in soils. It certainly constitute a new view of the microbial systems in soils, supporting the progress for a more efficient management of microbial communities in agriculture.

Keywords: Sugarcane; Soil; Bacteria; Fungal; Core community; Variable community

4.1. Introduction

Understanding the dynamics of the microbial communities in cropping systems is one of the most important keys to develop a more sustainable crop system (Andreote *et al.*, 2014). The identification of the core microbial communities has emerged as an important tool revealing the microorganisms intimate associated with a host or an environmental system (Shade and Handelsman, 2012; Schlaeppi *et al.*, 2014; Nogueira *et al.*, 2015). The microbial core community is defined as ubiquitous microorganisms associated with habitat and persisting along time (Shade and Handelsman, 2012). Given a habitat, the occurrence of a core microbial community has been associated with a co-evolutionary process (Pédron *et al.*, 2012; Schlaeppi *et al.*, 2014), attributing to this group essential roles in the habitat. For example, it was identified an abundant core microbial community for the activated sludge ecosystem, which is made of microbial groups highly associated with carbon turnover (Saunders *et al.*, 2015).

In crop systems, a shared microbial community has been identified across the plants, such as grapevine (Zarraonaindia *et al.*, 2015) and sugarcane (Yeoh *et al.*, 2015; Souza *et al.*, 2016). Souza *et al.* (2016) described the occurrence of core bacterial and fungal communities inhabiting the endophytic and exophytic compartments of roots, shoots, and leaves of sugarcane. The authors also indicated that despite the core microbial community are composed by less the 20% of total microbial richness, they represent 90% of the microbial community in abundance. Also, for the microbiome of the sugarcane roots, it was verified that different quantities of nitrogen fertilizers showed little effect on the structure of the core microbial community (Yeoh *et al.*, 2015). Thus, the next step in the core microbial studies is to understand what abiotic components, as soil chemistry or physics, or climate variables, could drive the differential display of the core and variable microbial communities.

In this study, we identified the core and variable microbial communities across soils cultivated with sugarcane, sampled along two years. The soil samples were obtained from four distinct experimental fields cultivated with the same plant variety, under different soil attributes, climate and crop managements. The core microbial communities were identified by a probabilistic model based on the Poisson distribution, which tests the probability of each taxon to occurs in every sample collected (for details check chapter 2). Here, we hypothesized that *i)* there are bacterial and fungal core communities across the soils cultivated with sugarcane; *ii)* the core and the variable microbial communities are differentially structured by distinct abiotic components, and *iii)* the core and the variable microbial communities may host different functions in the soils cultivated with sugarcane.

4.2. Materials and Methods

4.2.1. Experimental location and soil management treatments

The experimental areas were established in soils previously cultivated with sugarcane. The experiments started in 2012 and were located in states of São Paulo (belonging to Iracema and Quata sugarcane mills) and Goias (Cerradinho and Boa Vista sugarcane mills), as observed in Figure 4.1A. The experiments were sampled during 2 years, which represents two annual cycles of the sugarcane crop. We used the same sugarcane plant variety for all of the areas (RB 96-6928), and harvesting the sugarcane crop in October of each year.

Each experiment area was composed by 24 sub-areas (plots) consisting of 15 lines of sugarcane plants (0.7 m between the sugarcane lines), with the 15 m in weight and 34 m in length. The soil management treatments were used with four replicates random distributed,

and are described as conventional tillage method (Ct) consisting in the subsoiling, plowing, harrowing, dig furrows and cultivation of sugarcane stems; no-tillage method (Nt) which consist only in digging furrows and cultivation of sugarcane stems; and no-tillage with reduced machinery traffic (Lt) which consist of the same soil management as Nt method, but there was no traffic of machinery under these sub-areas avoiding the soil compaction (Figure 4.1B). Half of the sub-areas were also submitted with a rotation of legume *Crotalaria spectabilis*, which was cultivated and incorporated in soils before the cultivation of sugarcane.

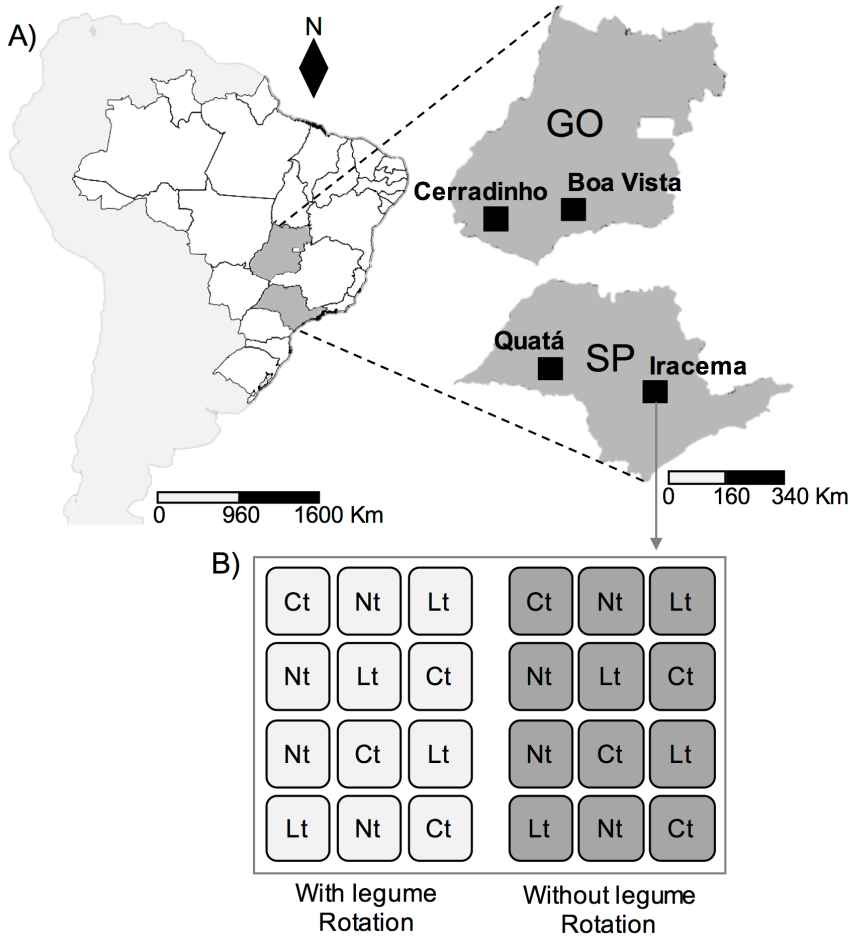

Figure 4.1 - Experimental location and soil management treatments

4.2.2. Soil samples and abiotic components evaluated

To represent each of soil treatments, we collected five subsoil samples (0-20 cm of deep) per plot and mixed resulting in one soil sample. We obtained four biological replicates of each soil management treatment with and without rotation with a legume (*Crotalaria spectabilis*) by experimental location and for two years, which results in a total of 96 soil samples.

For the purpose of correlation analyses, we evaluated the quantity of clay, silt and sand of each soil sample as physical soil factors, and also measured the chemistry factors as the pH (in CaCl2 0.01 M), and the concentrations of phosphorus (P; determined by resin), potassium (K), calcium (Ca), magnesium (Mg), Sulfur (S), and aluminum (Al). The methods used to evaluate the chemistry and physical soil factors are described by Raij *et al*. (2001). We also measured the pluvial precipitation (mm) and environment temperature (°C) as climate factors.

4.2.3. DNA extraction and phylogenetic diversity of the bacteria and fungi communities

Total DNA was extracted from 0.25g of soil using the PowerSoil DNA Isolation kit (MoBio, Carlsbad, USA) according to the manufacturer's instructions. The result of the DNA extraction was evaluated by electrophoresis in 1.2% agarose gel in TAE buffer (400 mM Tris, 20 mM glacial acetic acid, 1mM EDTA), in which DNA extracted product 5µl be applied and subsequently stained with GelRed TM and visualized in UV transilluminator. We performed phylogenetic description of the bacterial and fungi communities in soils cultivated with sugarcane. Firstly, we performed the amplification of target genes which was V6-V7 region of 16S rRNA gene and the ITS for bacterial and fungi community, respectively. The primers used in bacterial community description were 967F (Sogin *et al*., 2006) and 1193R (Wang and Qian, 2009), which were described by. To distinguish the samples, an additional tag of five nucleotides was synthesized together with the forward primer (http://vamps. mbl.edu/). The bacterial PCR reactions were adjusted to a final volume of 50 µl contained with final concentrations of 1X Platinium Taq PCR Buffer, 2 mM $MgCl_2$, 0.2 mM dNTP, 0.4 mM of each primer, 0.1 U.µL Platinium Taq DNA polymerase (Invitrogen, São Paulo, Brazil), 33 µl of optima water and finally, 1 µl of DNA template. The bacterial PCR reaction conditions were 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 seconds, 57 °C for 45 seconds and 72 °C for 1 minute, and a final extension of 72 °C for 10 min. For the fungi community, the primers used were ITS1f (Gardes and Bruns, 1993) and ITS2 (White *et al*., 1990), and it was also synthesized a tag of five nucleotides with the forward primer to distinguish the samples. The fungi PCR reactions were also adjusted to a final volume of 50 µl contained with final concentrations of 1X Platinium Taq PCR Buffer, 2.5 mM MgCl2, 0.2 mM dNTP, 0.4 mM of each primer, 0.05 $U.\mu L^{-1}$ Platinium Taq DNA polymerase (Invitrogen, São Paulo, Brazil), 32.4 µl of optima water and finally, 1 µl of DNA template. The fungi PCR reaction conditions used is described by Kemler *et al*. (2013). The PCR products of bacterial and fungi communities were purified with ChargeSwitch PCR Clean-Up Kit, quantified by Qubit 2.0

Fluorometer (Life Technologies), and finally sequenced on Ion Torrent® personal genome machine (PGM) system and using the Ion 316TM Chips (Life Technologies) according to Kavamura *et al*. (2013) and Kemler *et al*. (2013).

The Ion Torrent sequences were analyzed using QIIME 1.9.0 (Caporaso *et al*., 2010). Using the tutorials availables at QIIME website (http://qiime.org/tutorials/index.html), we first converted the fastq files of bacterial and fungi samples into fasta and quality file. The quality filters used were sequences equal or higher than 200 bp and had an average quality score equal or higher than 20. Similar sequences of were assign to operational taxonomic units (OTU) using the swarm method (Mahé *et al*., 2014). We performed the selection of the representative set of sequences of each OTU by the most abundant method and assigned the taxonomy using the Greengenes database for bacteria community and UNITE database for fungi community, both with uclust method and 97% of similarity. Finally, we assembled a table of OTU taxonomic identified for bacterial and fungi community with their respective abundance per sample. The OTU table of bacterial and fungi was used to verify the taxonomic composition across the treatments, identification of the core and variable communities, and also to correlated with abiotic components. However, before the statistical analyses, we rarified the OTU table of bacterial and fungal communities.

4.2.4. Identifying the core and variable microbial communities by a probabilistic method

Using the average of relative abundance of each taxa across communities and the occurrence frequency of each taxa across communities, we fitted the Poisson distribution on OTUs testing the hypotheses whether H0 – where the OTU fits in the Poisson distribution presenting a probability to occurs in every sample, belonging to the core community; or H1 – where the OTU does not fit in Poisson distribution indicating that it does not have a probability to occurs in every sample, belonging to the variable community.

The probability (P) of Poisson distribution is obtained by

$$\boldsymbol{P(x) = \lambda^x\, e^{-\lambda}/x!}$$

Where the lambda (λ) and x respectively represent the average of relative abundance and the occurrence frequency of each taxon across the communities. We designed a script in software R to identify the microbial core and variable communities.

4.2.5. Prediction functionality of the core and variable bacterial communities

Before the identification of the core and variable communities by the Poisson method, we used the 'Tax4Fun' R package (*see* Aßhauer *et al*., 2015) to predicted the functionality of each OTU resulting in the KEEG Orthology (KO) codes at the molecular level. The table containing the samples and KO codes were standardized in a range of -1 to 1 and used in 'Pathview' R package, selecting the nitrogen (KO00910) and sulfur (KO00920) metabolic paths.

4.2.6. Statistical analyses

In order to verify the variance of abiotic components, we transformed the chemical, physical and climate values by Hellinger method and performed the principal components analyses (PCA). The samples in PCA analyses were classified by the experimental location (Iracema, Quatá, Cerradinho, and Boa Vista), and also by the time of the experiment (first and second year). The variance of bacterial and fungi communities was evaluated by the non-metric multidimensional scaling (NMDS) using the Bray-Curtis similarity. The size of the core and variable communities and also the taxonomic compositions were verified using bar plot graphics. We evaluated the composition of the whole community and also of the core and variables portions of bacterial and fungi communities. Using the permutation MONOVA (PERMANOVA), we verified the correlation of each factor including time, crop managements, soil chemistry, soil physical, and climate, with the whole bacterial and fungi communities, and also with the portions of the core and variable communities. The abiotic components used in PERMANOVA analyses were also transformed by Hellinger method, and for the microbial tables, we also used the Bray-Curtis similarity. The statistical analyses and transformations were performed using the R software version 3.2.2 (R Core Team, 2015) and the packages 'RAM' (Chen *et al*., 2016), 'vegan' (Oksanen *et al*., 2016) and 'Hmisc' (Harrell Jr *et al*., 2016).

4.3. Results

4.3.1. Establishing the experimental field background

To identify the core and variable microbial communities in sugarcane soils, we firstly verified how different the experimental fields are. The variation of the soil attributes (chemical and physics) and climate factors (precipitation and temperature) was confirmed across the four distinct experimental fields. In overall, the precipitation was the main factor

that explains the variation between the samples from the first and second year for the four experimental fields (Figure 4.2). The soil characteristics from experimental fields of Quata and Iracema were clearly different in both periods. The Quata samples showed a high quantity of sandy, while Iracema samples showed higher amounts of clay and nutrients, including phosphorus, sulfur, and potassium. The Cerradinho and Boa Vista soil samples showed a high quantity of calcium, more pronounced in the second year (Figure 4.2).

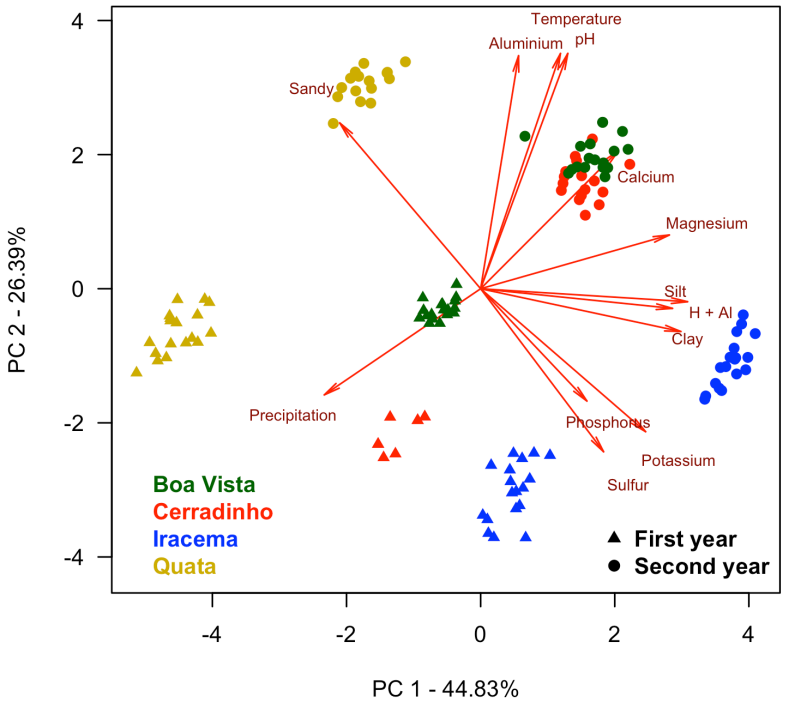

Figure 4.2 – Principal component analysis of soil attributes and climate factors obtained in each experimental field (indicated by different colors) during the first (triangules) and second (circles) years of sugarcane cultivation. Vallues in axes indicate the amount of variance explained in each axis

The second step aimed to verify the differential behavior of the bacterial and fungal communities across the areas, treatments and period (first and second year). We obtained a total of 5,579,534 and 7,159,659 of bacteria and fungi sequences, respectively, which we rarified to 7,300 sequences/sample for bacteria and 5,660 sequences/sample for fungi community. As it can be seen in Figure 4.3A, the bacterial community was distributed in four distinct groups (ANOSIM R= 0.95, *p-value* = 0.001) related to experimental fields. We observed similar results for the fungal community (Figure 4.3B), where the samples were also divided by experimental location (ANOSIM R= 0.85, *p-value* = 0.001). However, the fungal community from Boa Vista, Cerradinho, and Quata samples were more similar (ANOSIM R = 0.65, *p-value* = 0.001) than Iracema samples. There were no significant distinctions for the

communities of fungi and bacteria between the periods (ANOSIM R < 0.05, *p-value* > 0.05). In overall, we can state that distinctions were first placed due to locations, then between the two years analyzed, and lately by the treatments composing the experiment in each analyzed field.

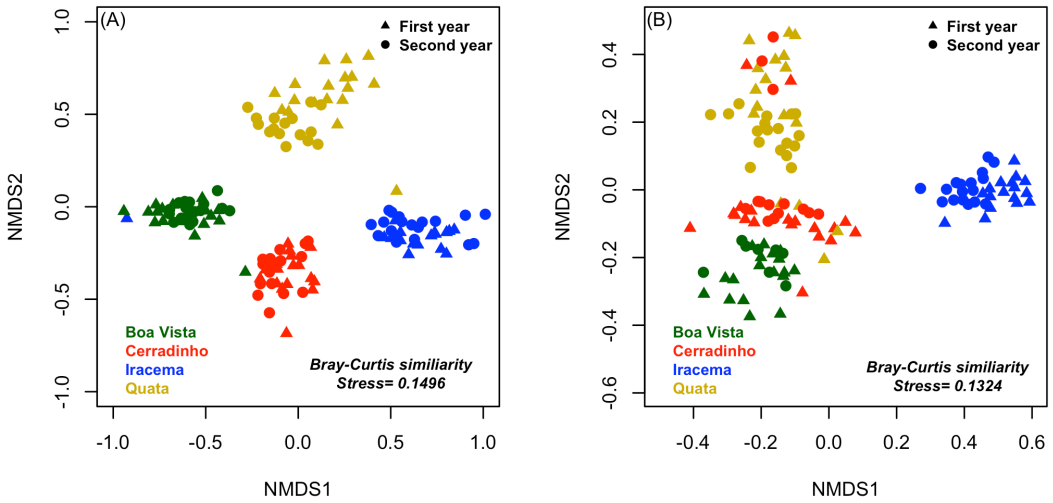

Figure 4.3 – Distribution of the whole bacterial (A) and fungal (B) communities across expeirmental fields (indicated by different colors) during the first (triangles) and second (circles) cultivating year. Vallues in axes indicate the amount of variance explained in each axis

4.3.2. Identifying the core and variable microbial communities of soils cultivated with sugarcane

We identified the core bacterial and fungal communities using the probabilistic method based on the Poisson distribution.The core bacterial presented the adjusted R^2 of 0.94 (*p-value* < 0.001) and the core fungal revealed the R^2 of 0.59 (*p-value* =0.0021), by the probabilistic method with 95% of confidence (Figure 4.4). Amounts of 70.35% (37,652 from 53,520 OTUs) and 57.55% (13,209 from 22,952 OTUs) of OTUs were named as members of the core community for bacterial and fungal analysis, respectively (Figure 4.5).

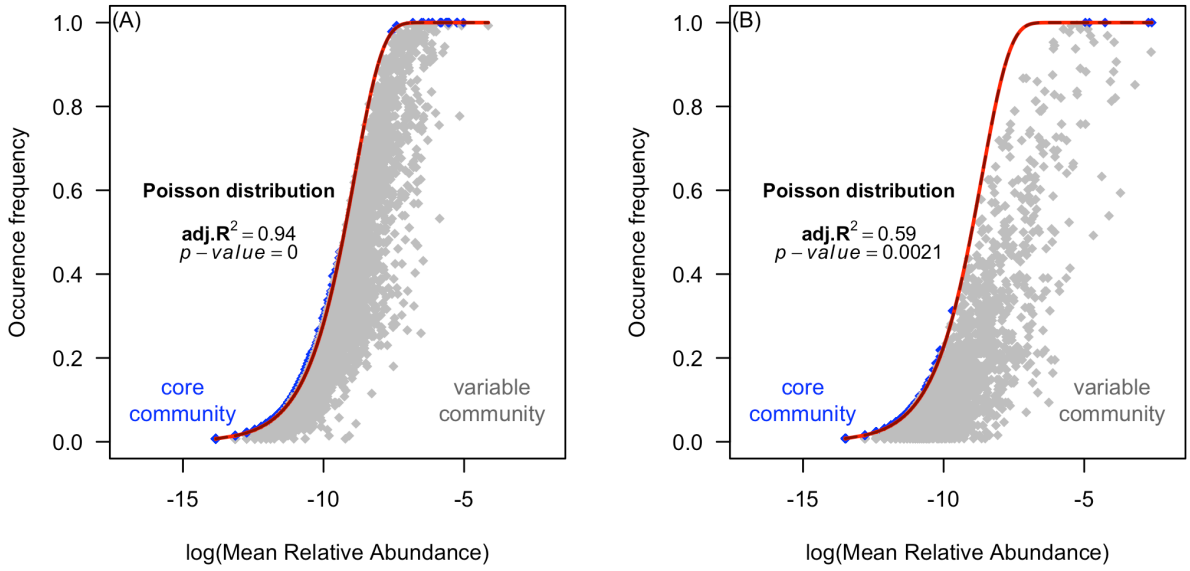

Figure 4.4 – Identification of the core bacterial (A) and fungal (B) community by probabilistic method based on the Poisson distribution. The OTUs identified as core are indicated by blue, and variable by gray

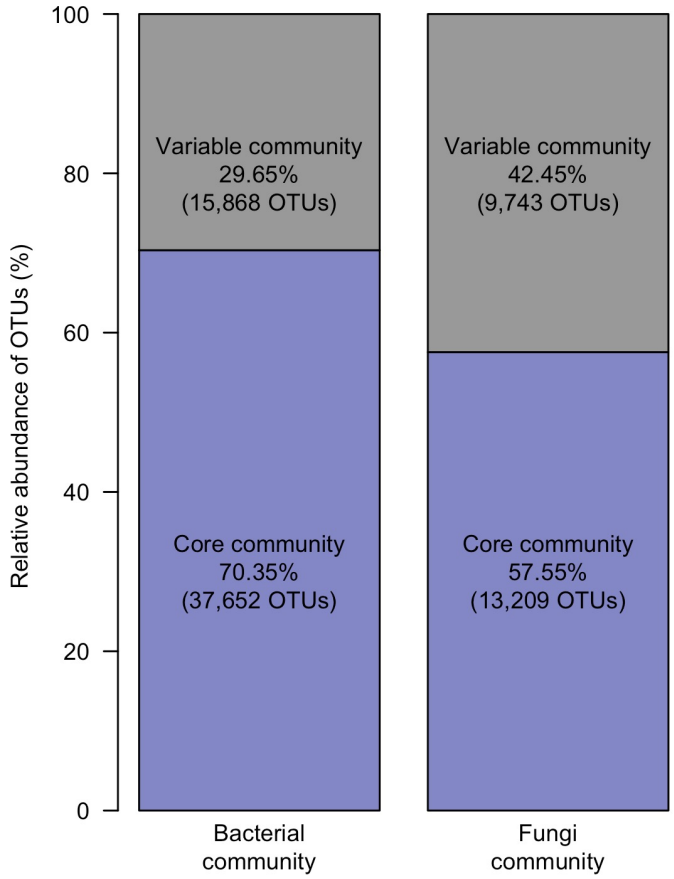

Figure 4.5 – Relative abundance of OTU classified as belonging to the core or variable communities determined by the distribution of Poisson

4.3.3. Taxonomic composition of the core and variable microbial communities is different

Although the most abundant phyla (*Proteobacteria* and *Actinobacteria*) were similarly found in each fraction of the bacterial community (Figure 4.6), we observed variations in the relative abundance of the phyla *Firmicutes* (core=15.1% and variable=5.9%), *Acidobacteria* (core=13.2% and variable=21.3%) and *Verrucomicrobia* (core=2.22% and variable=6.35%). We also observed the prevalent contribution of the families *Bacillaceae* (core=7.84% and variable=2.96%) and *Paenibacillaceae* (core=1.33% and variable=0.89%) for distinctions found in the phylum *Firmicutes*; and *Koribacteraceae* (core=2.97% and variable=4.86%) and *Solibacteraceae* (core=0.66% and variable=1.58%) for variations of *Acidobacteria*.

For the analysis of the fungal community (Figure 4.6), *Ascomycota* was the most abundant phylum identified, showing a high relative abundance for the core (96.9%) and variable (73.41%) communities. We also verified a strong variance for the phylum *Basidiomycota*, which was 2.17% in the core community and 24.96% in the variable community. Inside of the phylum *Ascomycota*, the most abundant families such as *Nectriaceae* (core=39.2% and variable=8.58%), *Hypocreaceae* (core=5.36% and variable=1.84%), and *Dipodascaceae* (core=0.33% and variable=6.85%), showed clear differences between the core and variable fungi communities. In the variable fungi community, the family *Filobasidiaceae* (core=0.15% and variable=2.48%) was the main divergence belonging to *Basidiomycota* phylum.

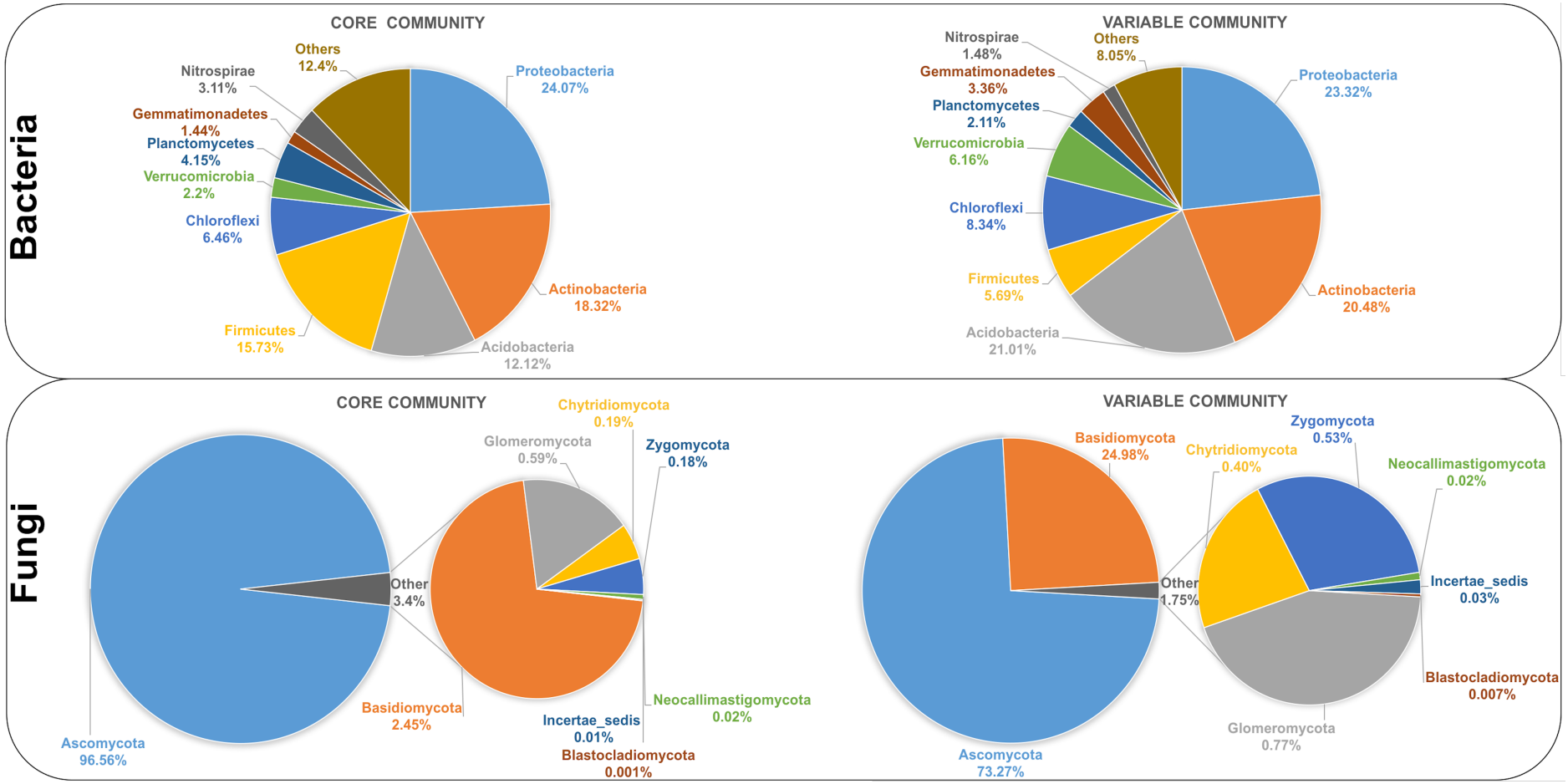

Figure 4.6 – Taxonomic composition of the core and variable microbial communities at the Phylum level

4.3.4. The distinctions of drivers modulating the core and variable communities

Besides the differences found in the taxonomic composition of the core and variable microbial communities, we also verified their correlation with several measured abiotic components. Using the PERMANOVA, we verify the correlation of time (first and second year of experiment), crop managements (soil treatments and rotation), soil chemistry (pH, phosphorus, sulfur, potassium, calcium, magnesium, aluminum, and H+Al), soil physics (sandy, clay, and silt), and climate (precipitation and temperature) with the whole community of bacterial and fungal communities, and also with the groups of the core and variable communities. In overall, almost 50% of the community variance was explained by the quantified abiotic components, except for the core bacterial community, where only 27.52% was explained (Table 4.1). We observed that the whole microbial community and their partitions in the core and the variable microbial communities showed a small but significant correlations with time.

We identified that percentages of correlations between the abiotic components and whole microbial communities were between the values observed for the core and variable microbial community correlations. Also, the whole bacterial and fungal community were correlated with distinct abiotic components. For example, while the fungi community was highly correlated with soils physics, the bacterial community was more correlated chemical characteristics of the soils.

However, the most important observation in this analysis is the clear variation between the abiotic components correlated with the core and variable microbial communities. The bacterial analysis showed that while the variable bacterial community was mainly correlated with phosphorus (13.53%), pH (12.37%), and sulfur (8.28%), these variables were not significantly correlated with the core microbial community (Table 4.1). Others biotic component, including potassium, calcium, aluminum, sand, and silt, showed a small difference (less than 2%) of correlation values for the core and variable bacteria community. We also observed that the highest significant correlation ($p\text{-value}<0.05$) of the core bacterial community with temperature (1.67%).

Table 4.1 - PERMANOVA analysis of abiotic componentes with the whole, core and variable microbial communities

Abiotic components	Bacterial community			Fungi community		
	Whole	**Core**	**Variable**	**Whole**	**Core**	**Variable**
Time [(1)]	1.68%***	1.26%***	1.73%***	2.27%***	2.62%**	2.09%**
Soil treatments [(2)]	(1.82%NS)	(1.46%NS)	(1.85%NS)	1.77%*	(0.70%NS)	1.92%*
Rotation [(3)]	0.52%*	(0.68%NS)	(0.49%NS)	0.88%*	1.21%*	0.87%*
pH	11.22%***	(4.25%NS)	12.37%***	0.80%*	1.21%*	(0.72%NS)
Phosphorus	12.22%***	(4.21%NS)	13.53%***	4.37%***	9.07%***	3.52%***
Sulfur	7.77%**	(4.12%NS)	8.28%**	(0.50%NS)	(0.54%NS)	(0.48%NS)
Potassium	3.71%**	2.37%**	3.87%**	1.32%**	2.21%**	1.10%*
Calcium	1.21%**	0.96%**	1.24%**	1.26%**	1.17%*	1.30%**
Magnesium	(0.58%NS)	(0.66%NS)	(0.57%NS)	(0.53%NS)	(0.60%NS)	(0.51%NS)
Aluminum	1.07%***	1.09%***	1.05%***	0.99%*	2.41%**	0.79%*
H+Al	(0.74%NS)	(0.82%NS)	(0.72%NS)	(0.43%NS)	(0.79%NS)	(0.36%NS)
Sandy	1.18%**	0.90%*	1.21%**	14.18%***	3.15%***	15.72%***
Clay	1.40%***	(0.99%NS)	1.45%***	16.76%***	30.17%***	14.66%***
Silt	1.38%*	1.10%**	1.41%*	9.10%***	(0.62%NS)	10.10%***
Precipitation	1.07%**	0.97%**	1.07%**	0.82%*	1.12%*	0.79%*
Temperature	(2.75%NS)	1.67%*	(2.90%NS)	(0.84%NS)	1.91%**	(0.76% NS)
Residuals	49.69%	72.48%	46.26%	43.21%	40.51%	44.32%

[(1)] First and second year of experiment; [(2)] No-tillage, No-tillage with reduced traffic and conventional methods; [(3)] With and without rotation with *Crotalaria spectabilis. ***p-value >0.001, **p-value >0.01, *p-value >0.05*, *NS* is no significant.

The core and variable fungal communities also showed distinctions in their correlation with abiotic components. For instance, while the core fungal community variance was explained by clay contents (30.17%) and availability of phosphorus (9.07%), the variable fungal community was correlated with sandy (15.72%), clay (14.66%), and silt (10.10%). Although the observed percentage of explanation was low, the temperature was significant correlated (*p-value<0.05*), only with the core fungi community (1.91%), as similar found for the core bacterial community. For the crop managements, the crop rotation showed correlation with core and variable fungal communities (core=1.21% and variable = 0.87%), and the soil treatments, such conventional, no-tillage and no-tillage with reduced traffic, showed a significant correlation (*p-value*<0.05) only with the variable fungal community (1.92%).

4.3.5. The core and variable bacterial communities display differential predictive functionality for nitrogen and sulfur cycling in soils

This approach was based in the predicted functional profiles of the core and variable components of the bacterial community. The comparison between the patterns suggested that the core and variable bacterial communities may be responsible for different metabolic paths in the soil functioning. For example, while the genes *Nif*DKH, involved in the nitrogen fixation, were more predicted for the core bacterial community and the nitrification, traced by the genes *Amo*CAB, was mainly predicted in the variable bacterial community (Figure 4.7). We also observed variations in the sulfur metabolism, where members of the core were prevalently predicted to host assimilatory sulfite reduction (genes *CysND*, *CysC*, and *CysJI*), dissimilatory sulfate reduction and oxidation (gene *AprAB*), and SOX system (*SoxD*).

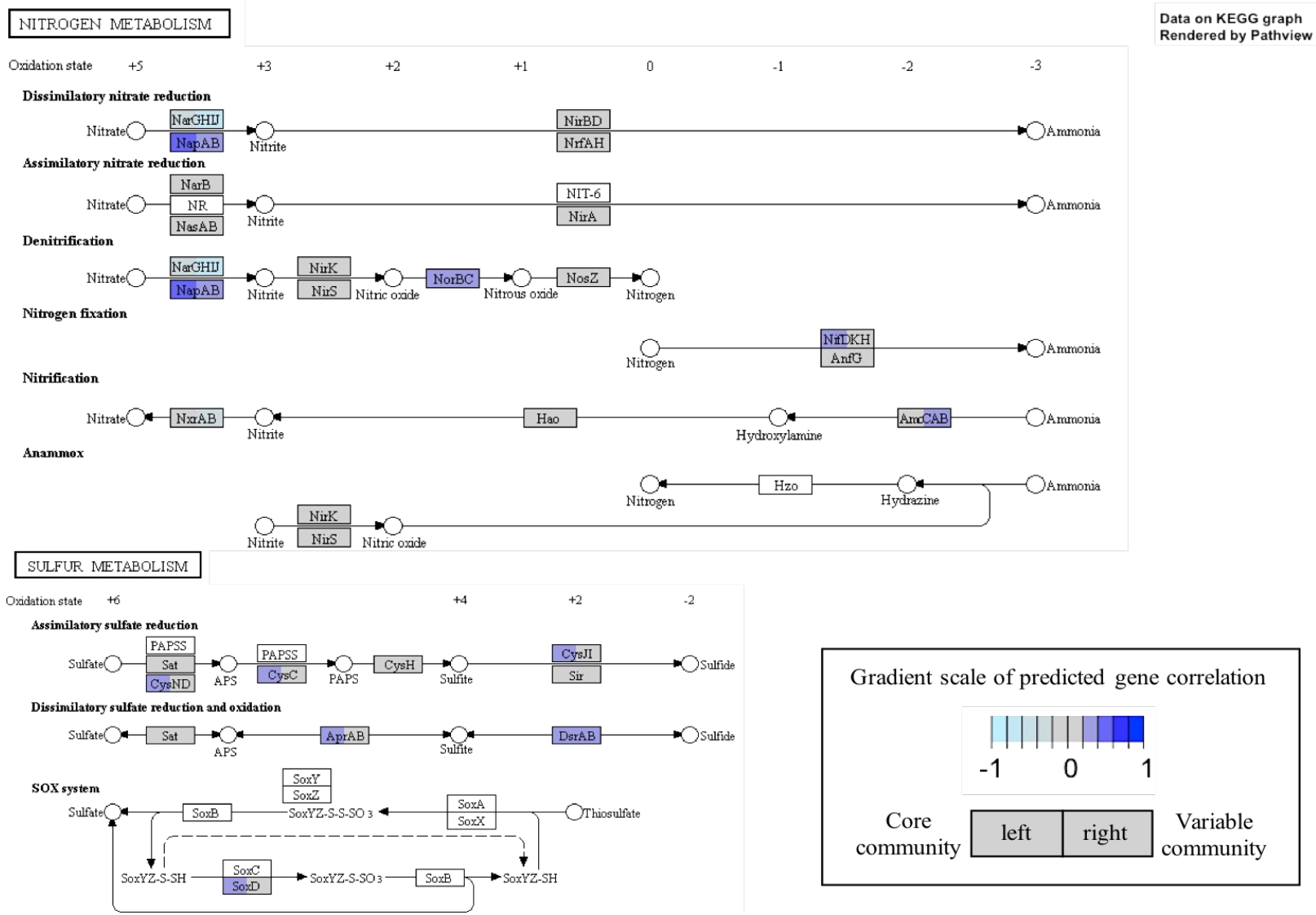

Figure 4.7 – Predicted functionality of the core and variable bacterial communities for nitrogen and sulfur cycling in soils

4.4. Discussion

Although several reports have identified a core microbial community across different hosts and environments, the influence of abiotic components on its structuring and functioning remains unclear. Here we used the soils cultivated with sugarcane, which showed high variance on the abiotic components across four locations to test it. The experimental design aimed to trace the role of soil chemistry, physics and management, together with climate variables, on the structuring of the whole, core and variable communities of bacteria and fungi.

Using a probabilistic method, we confirmed our first hypothesis, identifying the core microbial community across soils cultivated with sugarcane. The core bacterial and fungi communities identified were made of more than 50% of their whole respectively community, contradicting previous reports of the core communities, as estimated by subjective approaches. Considering the core as OTUs occurring in the least 90% of samples, Souza *et al*., (2016) identified the core bacterial and fungi community with respectively 11,349 and 1,711 OTUs, representing less than 20% of the whole microbial community present in the bulk soil, roots and shoots of sugarcane. We also observed that a great proportion of microorganisms with low abundance was identified as members of the core microbial community. Besides the possibility of the rare community composed the core microbial community (Shade *et al*., 2014), it was not previously reported. Thus, as we observed in chapter 2, the Poisson distribution allowed to identify OTUs as members of the core or the variable community despite their respective abundances.

Comparing the taxonomic composition of the core and variable microbial communities, we observed variations of phyla frequency, which also could be attributed to the phenotypic characteristics of its components. For example, bacteria belonging to the phylum *Firmicutes* were prevalently allocated in the core community. These bacterial phyla are known to produce endospore (Onyenwoke *et al*., 2004), especially the families *Bacillaceae* and *Paenibacillaceae*, what could support its high frequency in soils cultivated with sugarcane. The *Bacillaceae* family have been also reported as possible member associated with monocots (Schlaeppi *et al*., 2014; Yeoh *et al*., 2015). Thus, the constant occurrence across the soils cultivated with sugarcane may be linked with the resistance capacity of this microorganism, which could be compound the “microbial seed bank” (Lennon and Jones, 2011), playing an important role in the resilience of the system (Shade *et al*., 2012). In the other hand, members of the phylum *Acidobacteria* were a great proportion of the variable bacterial community.

These bacteria are described as strongly influenced by abiotic components, resulting in its differential spatial distribution along a gradient of pH (Shahnavaz and Geremia, 2012). For fungi, the phylum *Ascomycota* was more abundant in the core community. Possibly, its persistence lies in the capacity to produce ascospores (reviewed by Trail, 2007), or by the saprophytic behavior, found in the family *Nectriaceae* (Rodriguez *et al*., 1996), accounting for 39.2% of sequences of the core fungal community.

We have identified the main abiotic components driving the structure of the core and variable communities across soils cultivated with sugarcane. The most interesting finding was that the core bacterial and fungal community were modulated by different abiotic factors, confirming out the second hypothesis. The abiotic components which are adjusted for sugarcane cultivation such pH, phosphorus and sulfur contents better explains the dynamics of the variable bacterial community and presented no correlation with changes in the core bacterial community. Thus, some of these correlations could be linked with the taxonomic compositions of the group, i.e. the high abundance of *Acidobacteria* in the variable bacterial community could explain the higher correlation of pH with the variable community. As the origin of the core microbial community has been associated with the host by a coevolutionary process (Pédron *et al*., 2012), the low explanation of the core bacterial community by the abiotic components may suggest that this group are more correlated with plants or also microbial interactions.

The fungal community showed a distinct pattern, where the variations of the core and the variable communities were equally explained by connected abiotic components. In overall, the core and variable fungal communities were strongly correlated with soil physical components, connected to clay or sand contents. As the soil texture strongly influences the soil characteristics, such as water retention, soil structure, micro and macro pores (Dexter, 2004), the correlations identified for the core and variable fungal community may indicate an indirect effect from other soil components. Besides the soil physics, some chemical contents also affected differently the core and variable fungal communities, *i.e.* the connection between the core fungal and phosphorus were clear, while this nutrient poorly modulates the structuring of the variable fungal community.

In order to go further in our assumptions, we also approached the functionality of the core and variable communities. In this analysis, our third hypothesis was confirmed. We found the occurrence of specific steps of the nitrogen and sulfur transformations in soils, as promoted by members of the core or variable communities. These results, therefore, need to

be interpreted with caution, because there is no analysis of gene expression or protein functioning. The prediction functionality indicated the potential functionality that the microbial community may present (Aßhauer *et al*., 2015). Thus, suggesting a scenario where the microorganisms actually show the predicted functionality, as the variable bacterial community is more correlated with abiotic components, such pH, phosphorus and sulfur, this component may drive the nitrification process in soil. A similar scenario could be suggested for the core bacterial community, where the abiotic component does not influence directly the microbial group, as observed by Yeoh *et al*., (2015), but they affect the microbial function expression. For example, members of the core bacterial community under anaerobic condition could express the assimilatory sulfite reduction or dissimilatory sulfate reduction. In this situation, the core bacterial community could be used as biological indicators of system perturbation (Shade and Handelsman, 2012).

We can derive from our observations, that distinct components of microbial communities, such as core and variable communities, respond distinctly to environmental drivers. As it is described, the problematic of the prediction for the microbial communities structuring and functioning, a partially relies on the expectancy of a uniform response of the components of microbial communities, which is not observed. Possibly, this differential behavior is guided by the differential interactions of microbes with their niche. While some components of microbial communities are exogenous, others preserve a more intimate connection with the ecosystem, possibly resulted from a co-evolutionary process (Pédron *et al*., 2012; Schlaeppi *et al*., 2014).

The present study was designed to identify the core and variable microbial communities across soils cultivated with sugarcane and to name the major abiotic components driving the structuring of such communities. This research extends our knowledge on microbial communities of soils, indicating that the core microbial community differs from the variable community in taxonomic composition, correlation with the abiotic components and also may have specific functionalities. The next steps focusing on the active core and variable communities, using molecular techniques basing on RNA would be interesting to continue the exploration of the core and variables microbial communities. Continued efforts are needed to introduce the knowledge of microbial dynamics and functionality to crop systems managements, focusing on sustainable crop management.

References

Andreote, F.D.F.D., Gumiere, T., and Durrer, A. (2014) Exploring interactions of plant microbiomes. Sci. Agric. 71: 528–539.

Aßhauer, K.P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015) Tax4Fun: Predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31: 2882–2884.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., *et al*. (2010) QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7: 335–6.

Chen, W., Simpson, J., and Levesque, C.A. (2016) RAM: R for Amplicon-Sequencing-Based Microbial-Ecology. http://cran.r-project.org/package=RAM.

Dexter AR. (2004). Soil physical quality Part I. Theory , effects of soil texture , density , and organic matter , and effects on root growth. Geoderma 120: 201–214.

Harrell Jr, F.E., with contributions from Charles Dupont, and many others. (2016) Hmisc: Harrell Miscellaneous. https://cran.r-project.org/package=Hmisc.

Kavamura, V.N., Taketani, R.G., Lançoni, M.D., Andreote, F.D., Mendes, R., and Soares de Melo, I. (2013) Water Regime Influences Bulk Soil and Rhizosphere of Cereus jamacaru Bacterial Communities in the Brazilian Caatinga Biome. PLoS One (9)8:e73606.

Kemler, M., Garnas, J., Wingfield, M.J., Gryzenhout, M., Pillay, K.A., and Slippers, B. (2013) Ion Torrent PGM as tool for fungal community analysis: A case study of endophytes in eucalyptus grandis reveals high taxonomic diversity. PLoS One (12) 8:e81718.

Lennon, J.T. and Jones, S.E. (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. Nat Rev Microbiol 9: 119–130.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2014) Swarm: robust and fast clustering method for amplicon-based studies. PeerJ 2014; 2:e593.

Nogueira, V.L.R., Rocha, L.L., Colares, G.B., Angelim, A.L., Normando, L.R.O., Cantão, M.E., *et al*. (2015) Microbiomes and potential metabolic pathways of pristine and anthropized Brazilian mangroves. Reg. Stud. Mar. Sci. 2: 56–64.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al*. (2016) Vegan: Community Ecology Package. https://cran.r-project.org/package=vegan.

Onyenwoke, R.U., Brill, J.A., Farahi, K., and Wiegel, J. (2004) Sporulation genes in members of the low G+C Gram-type-positive phylogenetic branch (*Firmicutes*). Arch. Microbiol. 182: 182–192.

Pédron, T., Mulet, C., Dauga, C., Frangeul, L., Chervaux, C., and Grompone, G. (2012) A Crypt-Specific Core Microbiota Resides in the Mouse Colon. MBio 3: e00116-12.

R Core Team (2015) R: A Language and Environment for Statistical Computing.

Rodriguez, A., Perestelo, F., Carnicero, A., Regalado, V., Perez, R., De La Fuente, G., and Falcon, M.A. (1996) Degradation of natural lignins and lignocellulosic substrates by soil-inhabiting fungi imperfecti. FEMS Microbiol. Ecol. 21: 213–219.

Saunders, A.M., Albertsen, M., Vollesen, J., and Nielsen, P.H. (2015) The activated sludge ecosystem contains a core community of abundant organisms. ISME J. 10: 1–10.

Schlaeppi, K., Dombrowski, N., Oter, R.G., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2014) Quantitative divergence of the bacterial root microbiota in Arabidopsis thaliana relatives. Proc. Natl. Acad. Sci. U. S. A. 111: 585–92.

Shade, A. and Handelsman, J. (2012) Beyond the Venn diagram: The hunt for a core microbiome. Environ. Microbiol. 14: 4–12.

Shade, A., Jones, S.E., Gregory Caporaso, J., Handelsman, J., Knight, R., Fierer, N., and Gilbert, J.A. (2014) Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. MBio (4) 5:e01371-14.

Shade, A., Peter, H., and Allison, S. (2012) Fundamentals of microbial community resistance and resilience. Front. Microbiol. 3: 1–19.

Shahnavaz, B. and Geremia, R.A. (2012) The effect of pH and temperature on spatial variation of *Acidobacteria* / *Actinobacteria* communities from Alpine soil. J. Cell Mol. Res. 4: 28–33.

Souza, R.S.C. de, Okura, V.K., Armanhi, J.S.L., Jorrín, B., Lozano, N., Silva, M.J. da, *et al*. (2016) Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. Sci. Rep. 1–15.

Trail, F. (2007) Fungal cannons: Explosive spore discharge in the Ascomycota. FEMS Microbiol. Lett. 276: 12–18.

Wang Y, Qian PY. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PloS One 2009;4:e7401. DOI: 10.1371/journal.pone.0007401.

Yeoh YK, Paungfoo-Lonhienne C, Dennis PG, Robinson N, Ragan M A., Schmidt S, *et al*. (2015). The core root microbiome of sugarcanes cultivated under varying nitrogen fertiliser application. Environ Microbiol 18: 1338–135.

Zarraonaindia, I., Owens, S.S.M., Weisenhorn, P., West, K., Hampton-Marcell, J., Lax, S., *et al*. (2015) The soil microbiome influences grapevine-associated microbiota. MBio 6: 1–10.

5. FINAL CONSIDERATIONS

This study contributed to elucidate that ecological models are important tools to reveal the microbial community distribution in crop systems, and future research should focus on reducing the arbitrary of methods used to feed ecological models in microbiological studies.

The fisrt landmark of this study was the development of a probabilistic method to assess the core microbial community, validates in several microbiologicla systems. Unlike the arbitrary models previously used, the probabilistic model developed revealed a bigger core microbial community. It also includes on the core community memebers of the rare community, which it has been neglected in previous approaches. Applying it for differents microbial systems, the probabilistic model may clarify the correlation between the core and the variable microbial communities with the abiotic components.

The second important contribution of this study is the identification of the core and the variable microbial community in soils cultivated with sugarcane. We verified that the core and variable microbial communities showed distinct correlations with abiotic components, including the soil management, chemistry, physical, and also climate. We also observed a distintic predictive functionality harbored by the core and the variable microbial community, such nitrogen fixation and nitrification processes.

The inclusion of the knowledge of microbial dynamics and functionality to crop systems managements could help the sustainability of crop system, even reducing mineral fertilization, or the pesticides application. Therefore, continued efforts are needed to unlock the cropping system microbiome, providing scientific suport for a better management of microbial communities in agriculture.