Using Metagenomics to Connect Microbial Community Biodiversity and Functions

Lucas W. Mendes[1]*, Lucas Palma Perez Braga[1], Acacio A. Navarrete[1], Dennis Goss de Souza[1], Genivaldo G. Z. Silva[2] and Siu M. Tsai[1]

[1]Cell and Molecular Biology Laboratory, Center for Nuclear Energy in Agriculture CENA, University of São Paulo, Piracicaba, SP, Brazil.
[2]Computational Science Research Center, San Diego State University, San Diego, CA, USA.

*Correspondence: lucaswmendes@gmail.com

https://doi.org/10.21775/cimb.024.103

Abstract

Microbes constitute about a third of the Earth's biomass and are composed by an enormous genetic diversity. In a majority of environments the microbial communities play crucial roles for the ecosystem functioning, where a drastic biodiversity alteration or loss could lead to negative effects on the environment and sustainability. A central goal in microbiome studies is to elucidate the relation between microbial diversity to functions. A better understanding of the relation diversity-function would increase the ability to manipulate that diversity to improve plant and animal health and also setting conservation priorities. The recent advances in genomic methodologies in microbial ecology have provided means to assess highly complex communities in detail, making possible the link between diversity and the functions performed by the microbes. In this work we first explore some advances in bioinformatics tools to connect the microbial community biodiversity to their potential metabolism and after present some examples of how this information can be useful for a better understanding of the microbial role in the environment.

Introduction

Biodiversity has been recognized as the primary factor that affects the functioning of the ecosystem (Wagg *et al.*, 2014). Diversity has been linked to nutrient cycling, biogeochemical process, ecosystem stability and productivity. A better understanding of the ecosystem functioning is of paramount importance in the pursuit for sustainable practices, and the microorganisms that inhabit the soil should be taken into consideration. Soil-borne microorganisms represent the largest biodiversity pool on Earth, with more than 10^{30} microbial

cells and estimates of 10^4 to 10^6 different species per gram of soil (Whitman *et al.*, 1998; Torsvik *et al.*, 2002). Soil microbial communities hold a central place in terrestrial ecosystems due to their enormous number, large biomass and involvement in numerous key biogeochemical functions. These communities carry out essential ecosystem functions, such as nutrient cycling, plant nutrition, and disease suppression (Bardgett *et al.*, 2008). However, soil microorganisms are sensitive to environmental disturbances, with consequences for microbial diversity and functions. In this sense, increased attention has recently been paid to microbial communities living in soils from different environments, but despite the increased appreciation of belowground microbial diversity, little is known about the functional responses of the microorganisms to alterations in soil management and chemical properties.

Microbial diversity studies are important in order to understand the microbial ecology in soils and other ecosystems (Atlas, 1984). In the past two decades, molecular tools have been allied to classical microbiology methods to provide new insights into microbial ecology of soils. The rapid increase of molecular ecology issues has emerged as a result of the advancement of molecular biology. The molecular ecology applied to environmental microbiology should include not only diversity studies, but should address studies of functional traits related to microbial process. The development of tools in molecular biology has helped researchers to answer important questions in microbial ecology. Fingerprinting technologies, such as T-RFLP (Thies, 2007), ARISA (Fisher and Triplett, 1999), and DGGE (Muyzer and Smalla, 1998), have helped answer diversity issues, such as the richness and abundance of species. Sequencing technologies have provided consistent information about 'who' are the species in a given environment. The challenge now is to respond what these species are doing there, in other words, what are the functions performed by each species. In this sense, the assessment of microbial diversity will be advanced by the development of new technologies that answer some key questions about the 'who, what, when, where, why, and how' of microbial communities (Knight *et al.*, 2012).

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

In this work we discuss the use of metagenomics to link the microbial community diversity to functions. Here we show some tools and few examples on how to bring taxonomic–functional dichotomy in microbial ecology into one perspective by analysing metagenomic datasets.

Metagenomics in microbial ecology and bioinformatics tools

In view of the recent technological advances in the genomic area, studies based on metagenomics, defined here as the sequencing of total DNA extracted from environmental samples, can be of great value to investigate microbial community responses when environmental factors are being experimentally tested (Prosser, 2015). In many environments, 99% of the

lineages in the resident microbial community cannot be cultured (Handelsman, 2004). Metagenomics is a powerful tool that allows the study of environmental communities without this culture-bias and understand the real diversity present in microorganisms just by using rapidly developing DNA sequencing approaches (Handelsman, 2004). Next-generation DNA sequencing technologies have accelerated the sequencing process, decreased the cost, and opened new horizons in the environmental understanding (Zhang *et al.*, 2011). In order to facilitate better-educated comparative analysis and biological interpretations of the sequence data the integration of metagenomics data with their associated metadata is strongly recommended (Pagani *et al.*, 2011). If applied to test microbial community response under different environmental conditions, metagenomics could contribute to the discovery of new metabolic pathways as well as novelty in the way that microorganisms interact with each other in the environment (Prosser, 2015). Understanding microbial communities is important in many areas of biology. For example, it discriminates taxonomic and functional profiles of microbes associated with human (T.H.M.P. Consortium, 2012), global ocean microbiomes (Sunagawa *et al.*, 2015) and soil ecosystems (Leff, 2015). Particularly, soil is the most challenging environment due to the diversity of microbial communities living in such environment. It has been estimated that one gram of soil may contain more than 10^7 prokaryotic cells in forest ecosystems, while in cultivated soils this number can increase by 100 times (Daniel, 2005). Microbial diversity can be translated to genetic, taxonomic, phylogenetic and functional diversity. Depending on the environmental conditions, the same microorganisms may play different functions. However, decreasing taxonomical diversity might implicate in loss of functions (Philippot, 2013). On the other hand, increasing the taxonomical diversity could lead to a higher functional redundancy and, consequently, the maintenance of essential ecosystem services (Mendes *et al.*, 2015a). Thus, the assessment of microbial diversity and their link to functional activities would increase our comprehension of the microbial communities' role in the environment and the consequences of diversity losses.

Computational tools for functional profiling of metagenomics datasets

Next generation sequencing (NGS) technologies can generate gigabase (Gb)-sized sequences. However, raw datasets require preliminary analysis before properly assignment of the genetic information encoded in the reads. This preliminary analysis in general includes normally the filtering of low quality and low length sequences. In general analysis, it is recommended to remove reads below quality Q20 and below length 50 bp. After the filtering step, the tools implemented to decode the information from the reads will depend basically on the scientific question driving the experiment (Prosser, 2015). Reads can be assembled into contigs (contiguous sequences), which enhances the chances of retrieving complete genes but reduces the analytical precision regarding the conclusions inferred for the abundance of these genes. Additionally, avoiding assembly of chimeric contigs is also a question of concern. Here in this work we will focus on the direct analysis of metagenomics reads, since we believe that by using appropriate methods, as discussed below, this approach provides a powerful comprehension of the abundance of functions and pathways/subsystems across the datasets.

Functional profiling of metagenomics sequences normally requires the alignment of sequences to a curated large database of annotated sequences to identify similar matches

(Mendoza *et al.*, 2015). There are many databases for functional annotations, including the SEED (Overbeek *et al.*, 2005), which contains subsystems (protein families with a similar function), and the large metabolic pathway databases KEGG (Kanehisa and Goto, 2000) and MetaCyc (Caspi *et al.*, 2010). These databases overlap at some point. The SEED database, mostly derived from KEGG, was pointed to harbour more compounds than MetaCyc. In turn, MetaCyc was found to harbour more diversity of pathways and reactions (Altman *et al.*, 2013).

MEGAN (Mitra *et al.*, 2014) and MG-RAST (Meyer *et al.*, 2008) align the query sequences to a reference database to profile the metagenomics data. MEGAN uses as input the tabular results files created from programs such as BLAST, DIAMOND, or RAPSearch2 to the non-redundant database, and creates taxonomic and/or functional profiles. MG-RAST predicts the open reading frames (ORFs) on the metagenomics data and aligns them to the M5 non-redundant database using BLAST. Alignment of sequences is normally the most expensive computational process, and it can be the most time-consuming step.

Most of the available tools for alignment of sequences use homology (evolutionary hypothesis based on similarity measurements between alignments), i.e. the metagenomics sequencing reads are aligned against a curated reference database. Homology-based approaches normally use BLAST (Altschul *et al.*, 1997) or BLAT (Kent, 2002) to find the best hit in a reference database; however, those algorithms are slow because they were developed when metagenomics datasets inputs were small. New homology search methods such RAPSearch2 (Zhao *et al.*, 2012) and DIAMOND (Buchfink *et al.*, 2014) have recently been developed to reduce the run time. Both tools use basically the same strategy to reduce time of analysis and they are recommended for alignments of short reads.

As an alternative to alignment based approaches, metagenomics reads can also be assigned to database sequences via *k*-mer based approaches, where *k* means the size of the sub-set of a sequence, use short sequences to find exact matches in a metagenomics query. These approaches assume that the size (k) of the sub-set encodes the genomic signature of the sequence, which is sufficient enough to assign any other sequence with the same genomic signature by exact match. For instance, real time metagenomics (RTMg) (Edwards *et al.*, 2012) identifies all words (i.e. sub-sets of the sequence) of length k (between $k=7$ and $k=12$ amino acids) for a set of functionally related proteins, and uses them to find exact matches in the query to identify the functions present in the metagenomics sample. These methods based on *k*-mer approaches, as alignment-free approaches, tend to be more reliable due to the lower computational burden required.

Recently, there has been released a new approach – SUPER-FOCUS SUbsystems Profile by databasE Reduction – combining *K*-mer matches with fast alignment tools such as DIAMOND, RAPSearch2 and BLAST for taxonomical and function annotation, respectively (Silva *et al.*, 2016). SUPER-FOCUS classifies each sequence in the metagenome into a subsystem by aligning all the input data against a reduced database. The speed-up compared to other available tools comes from three improvements compared to the standard metagenome annotation pipelines. First, a clustered version of the SEED database; second, identify the genera present in the metagenomics sample using FOCUS (Silva *et al.*, 2014); and finally, align input data against the reduced database using RAPSearch2 which is 100 times faster than blastx with no reduction in sensitivity or specificity (Berendzen *et al.*, 2012). SUPER-FOCUS is up to 1000 times faster than other tools with little loss of sensitivity and still computationally efficient. In turn, FOCUS produces in seconds the taxonomic

profile for metagenomes. It is much faster than homology approaches because it does not align all the query sequences against the database. Instead, it reduces the computational time using *k*-mers and non-negative least squares to find the optimal set of genomes present in the metagenome. We believe this approach is useful especially for trivial users, because the database reduction performed by FOCUS demand less computational resources to the complete (taxonomical and functional) annotation of metagenomics datasets. Even though, other computational tools might provide the same results.

The advances of bioinformatics tools for data analysis follow the current advances of sequencing techniques. However, the next challenge will be to create and/or improve the databases used as reference for a more robust and reliable analysis and interpretation of the information generated. Nevertheless, the current methods to access the microbial diversity has contributed to a better understanding of the dynamics and functional role of communities in wide range of environments, as exemplified in the following section.

Metagenomics and microbial ecology in tropical soils: linking diversity to functions

The development of techniques for sequencing DNA directly from environmental samples was a crucial factor for the discovery of the enormous degree of prokaryotic diversity in almost every environment. Along with that, the development of appropriate bioinformatics tools has helped researchers to assess and interpret the results obtained by the high-throughput sequencing methodologies. This approach has increased our understanding of the microbial community dynamics in changing environments. In the last year, much attention has been paid to the effects of land use changes on microbial communities, due to the importance of microorganisms for soil environments (Nesme *et al*., 2016). The growing demand for agricultural and livestock production has led to the opening of new areas of native forest. However, when conservation practices are not adopted, the intensive land use affects negatively both the environment and agricultural productivity (Cerri *et al*., 2004; Foley *et al*., 2005). One assumption often made is that biodiversity loss is happening more rapidly in the tropics due to agricultural activities, and these transformations in the environment due to land use changes have a direct effect on the soil microbial community. The advances in agriculture in these regions have transformed natural areas into agricultural fields, where studies on microbial ecology are needed to increase our knowledge about the effects of land use changes over the microbial community structure, composition, and function. Several studies have linked alterations in biodiversity to soil functioning, showing that soil stability is very dependent on microbial diversity (Tardy *et al*., 2014), and higher microbial diversity increases resistance and resilience of microbial processes (Girvan *et al*., 2005). In this sense, a deeper comprehension of the composition and role of microorganisms in soil systems is of paramount importance for the development of sustainable practices in agriculture.

In this section, we discuss the effects of land use changes on soil microbial diversity, composition and distribution in tropical soils. Through studies using a metagenomics approach we examine the effects of agriculture practices on aspects of the soil ecosystem from both a taxonomic and functional perspective, in order to illustrate how the assessment of microbial communities and their functional profiles can contribute to our comprehension of soil management and its sustainable use.

Effect of land use changes in Amazon soils on microbial communities

The Amazon forest harbours 40% of all remaining tropical rainforest, playing a fundamental role in biodiversity conservation, biogeochemical cycles, and climate regulation. The Amazon is the Earth's largest reservoir of plant and animal diversity, including the microorganisms (Rodrigues *et al.*, 2013). Despite its global importance, the Amazon's 'arc of deforestation' has been the world's most active deforestation frontier in recent decades. Deforestation in the Brazilian Amazon can be defined as clear cutting and conversion of the native forest to other land uses, mainly agriculture and cattle pasture. The increasing expansion of the agriculture has become the main agent of disturbance in the Amazon region, and such land use change has consequences on soil microorganisms (Mendes *et al.*, 2015a). For this reason, increased attention has recently been paid to the belowground microbial communities in Amazon soils.

The first description of the microbial diversity in Amazon soils was made by Borneman and Triplett (1997) who used a Sanger sequencing approach to analyse sequences from forest and pastures. In this first glimpse, they showed an immense microbial diversity and differences in composition between forest and pasture. Sequences affiliated to *Clostridia* were related to forest samples while *Bacillus* were related to pasture. More recently, other studies in the Brazilian Amazon have also contributed to a better comprehension of the effects of land use changes on the soil microbial communities (Cenciani *et al.*, 2009; Jesus *et al.*, 2009; O'Neill *et al.*, 2009; Grossman *et al.*, 2010; Pazinato *et al.*, 2010; Taketani and Tsai, 2010; Navarrete *et al.*, 2010, 2011, 2013a, 2015b,c; Germano *et al.*, 2012; Rodrigues *et al.*, 2013; Taketani *et al.*, 2013; Brossi *et al.*, 2014; Mendes *et al.*, 2014, 2015a). These studies have shown that Amazon soils harbour high diverse microbial communities and have demonstrated that the land use changes alter the structure, diversity and composition of the communities.

For a better comprehension of the effects of land use change in Amazon soils, some recent studies explored the microbial communities using soil DNA shotgun metagenomics approach to assess the microbiome in a chronosequence from a native tropical forest, followed by deforestation and cultivation of soybean croplands and pasture. They found that the land use system had a primary effect on the soil bacterial communities, whereas parameters such as pH, C, N, NO_3^- and K content significantly correlated to overall community structure (Mendes *et al.*, 2015b). Soils of forest were related to the phyla Proteobacteria, Acidobacteria and Verrucomicrobia. Deforested sites presented high abundance of Chloroflexi and Actinobacteria. Agriculture and pasture soils related to Bacteroidetes and Firmicutes, respectively (Fig. 6.1a). The differential distribution of these phyla in each land use system can be explained by the life-style and functions performed by these bacterial groups. Members of the phylum Proteobacteria are important to global carbon, nitrogen, and sulfur cycling, functions that were abundant in forest soils (Mendes *et al.*, 2015a). Acidobacterial responses arise at least in part through alterations on acidity- and nutrient-related properties of the Amazon soils (Navarrete *et al.*, 2013a, 2015c). Also, high abundance of Verrucomicrobia in forest may be explained due to its specialization on the degradation of more recalcitrant carbon compounds and high soil moisture, that are characteristics parameters of forest soils (Fierer *et al.*, 2013; Navarrete *et al.*, 2015b). As a consequence of slash-and-burn, the deforested site presents an increased soil temperature and a deposit of a high amount of carbon from the ashes. This process can explain the abundance of Chloroflexi and Actinobacteria in this area. Some members of the Chloroflexi phylum are aerobic thermophiles,

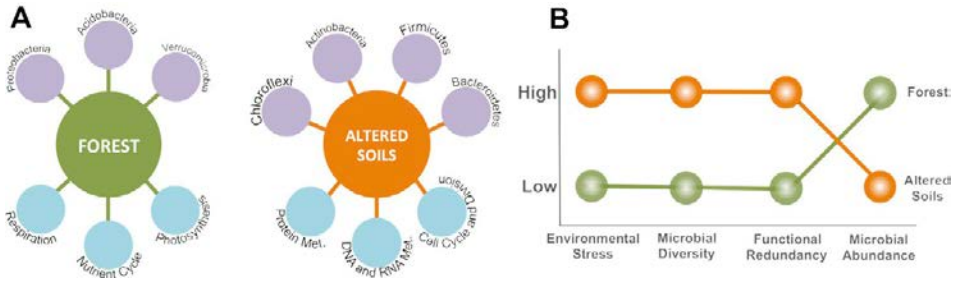

Figure 6.1 (a) Most abundant microbial groups (upper circles) and functional categories (bottom circles) present in forest and altered soils. (b) Graph showing the comparison between forest and altered soils for environmental and microbiological parameters. This figure is based on the results of Mendes *et al*. (2015a).

growing well in high temperatures; Actinobacteria are important decomposers and also capable to produce spores, which allow them to resist perturbation (Ventura *et al*., 2007). Bacteroidetes phylum includes plant-growth promoting bacteria and cellulose decomposers, which might be related to the soybean cultivation in the agricultural area (Mendes *et al*., 2015b). The high abundance of Firmicutes in pasture soils may be due to their resistance to desiccation and extremes of environmental variation, scenario commonly found in pastures (Battistuzzi and Hedges, 2009; Rodrigues *et al*., 2013).

In this context, gathering information about functional traits of the microbial groups is fundamental to a better interpretation of the diversity and composition of the soil communities. An analysis of the microbial functional profile in these Amazon soils revealed that the most prevalent core of functions in native forest soils was related to respiration, photosynthesis and nutrient cycle (phosphorus, potassium, sulfur and nitrogen). In the other hand, the altered soils (deforested, agriculture and pasture) presented a core of genes associated to DNA and RNA metabolism, protein metabolism, cell division and cell cycle (Mendes *et al*., 2015a) (Fig. 6.1a). These results are an indication of soil stress after the deforestation process, since these enriched functions in the altered soils are linked to adaptations of the organisms to thrive under disturbed conditions.

Interestingly, several studies with different molecular approaches showed an increased taxonomic diversity after the conversion of Amazon forest (Cenciani *et al*., 2009; Jesus *et al*., 2009; Rodrigues *et al*., 2013). Mendes *et al*. (2015a) also found that agriculture and pasture soils were among the most diverse in comparison to natural forest soils. In order to better understand the increased microbial diversity in the altered soils, they linked the diversity with functions by analysing metagenomics data. To predict the functional patterns of the microbial communities in each site, they selected functional categories related to nutrient cycle metabolisms (nitrogen, phosphorus, potassium, and sulfur), respiration and response to stress (DNA and RNA metabolism, stress response, and virulence/disease/defence). They evaluated the number of genera and the abundance of genes related to these functions and found that altered sites presented higher taxonomical and functional diversities and higher richness of genera playing specific functions in comparison to forest. On the other hand, although the forest soils presented lower diversity of genera playing these functions, these functional genes were 2- to to 4-fold more abundant than in the disturbed sites (Fig. 6.1b). In this sense, they suggest that the higher diversity found in altered soils led to higher

functional redundancy, i.e. different groups playing same function, which are important to maintain the ecosystem functioning after a stress event. But then, in an environment in equilibrium, such as forest, the ecosystem functioning is maintained based on a high abundance of microorganisms. In this example, the access of the functional profile of the soil microbial communities allowed a better interpretation of the role of diversity in these soils.

Ecological process governing soil microbial assembly and functions during land use change in Amazon soils

The last few years have brought a new landmark for microbial ecology. Due the increased efforts in sequencing technologies (e.g. new-generation sequencing) and data mining, microbial ecologists may now go deeply into the 'roots' of microbial interactions in soil and plants (Metzker, 2010). The main challenges now concern about: how to analyse and interpret microbial patterns of diversity generated by millions of taxonomic and functional data? Which are the ecological effects of the forest-to-agriculture conversion? And, how to integrate microbial taxonomic and functional biodiversity with the environmental factors that it can interact with? Microbial metagenomics is trending in community ecology, and allows us to elucidate microbial black box in several agroecosystems (Daniel, 2005; Mendes *et al.*, 2015a; Souza *et al.*, 2015).

Here we summarize the results of two studies, emphasizing the relationship among taxonomical and functional diversity and environmental factors that modulate them. The first study was performed in the Brazilian Atlantic Rainforest Biome, with the aim to evaluate soil microbial taxonomic and functional alterations imposed by niche and neutral model assemblies, in order to gain insights into the relationship between community assembly and ecosystem functions (Souza, 2015). Soil samples were collected in a land use gradient, from forest remnants to grassland and cropping systems. The results showed that land use was more important to drive microbial assembly than the forest conversion. Factors like restriction in range of C and energy sources in grassland may subject microbial communities to a strong process of homogenizing selection (Wawrik *et al.*, 2005). Grass root system increases the soil connectivity by removing potential barriers to migration (e.g. aggregate-to-aggregate continuity) and may lead to higher microbial dispersal in space and time (Murphy and Foster, 2014). As a consequence, the overall microbial diversity is impacted. No-till cropping increases carbon storage, maintains carbon residues and supplies microbial communities with resource heterogeneity. Ecological theory supports that species coexistence increases as a result of niche-partitioning due to resource availability (Carrol *et al.*, 2011) and empirical support has been provided for plant and animal communities (Finke and Snyder, 2008; Mori *et al.*, 2013). It was found that resource heterogeneity in no-till increased niche specialization and reduced the importance of the process of homogenizing selection on microbial assembly.

By analysing a large range of taxa, functional and environmental data, it was shown that soil microbial community assembly is not dependent on conversion, but on the land use after conversion, and influenced by the net biodiversity outcome. Furthermore, the relative weight of divergent ecological processes, depending on land use, determined a niche- or neutral-based model of assembly, with implications to function. The long-term grassland leaded to biotic homogenization and functional stress, through environmental selection. In the other hand, long-term no-till cropping has shown no differences, when compared to forest. Future investigations stand to gain valuable information from assessing ecological

processes combined effect in different ecosystems and across multiple scales of space and time. Understanding microbial community assembly as well as characterizing fundamental processes that guide dynamic responses in community organization have the potential to provide important insights into ecosystem modelling of biodiversity and microbial conservation.

The second study was performed in the last border of Amazon Rainforest (Souza, 2015). The aim of this work was to evaluate the microbial community dynamics in long-term no-till cropping and the power of selection of soybean along a chronosequence, using agriculture sites arisen by the deforestation of the Amazon Rainforest. The outcome was a framework of the microbial dynamics along soil fractions and time. Soil samples were collected in a land use gradient, from forest remnants to grassland and cropping systems. Microbial communities have evolved in long-term no-till cropping management, shifting taxa along the time. The microbial community fluctuations were constant in bulk soil and gradually increased in the rhizosphere, the extent the chronosequence advanced. Despite of the trade-offs, the microbial communities of bulk soil–rhizosphere interface became more similar along time. In the bulk soil, both taxa and functions shifted along the chronosequence, meaning that stochasticity is a major driver of the community dynamics, which explains the higher number of samples assembled by neutral model and the tighter connections between taxa, functions and environmental factors. In the rhizosphere, the fight for niche occupancy increased the extent the chronosequence advances, characterized by the high taxa trade-off, high homogenizing selection and reducing stochasticity. Despite of that, the functional resilience acted in order to guarantee the maintenance of the niches, with high functional redundancy and niche stability. This explains why, despite of the increased number of connections between taxa, functions and environment, these connections were sparser. In conclusion, the long-term no-till leads to homogenization of the microbial community in both bulk soil and rhizosphere. The rhizosphere of the plant increases its power of taxa homogenizing selection along the chronosequence, but the functions are kept.

Microbial indicators in sugarcane systems

Sugarcane (*Saccharum* sp.) is one of the most important croplands for tropical agriculture. This plant is a high-sucrose accumulator and one of the crops that more efficiently convert solar energy to chemical energy. The production of sugarcane biomass serves to supply the sugar and ethanol market. Additionally, more recently, the residues of the harvesting are being destined to the generation of bioelectricity. Brazil is currently the world's major producer of sugarcane, where a total area of cropland occupies more than 9 million hectares (Della-Bianca *et al.*, 2013).

Considering the economical importance of sugarcane fields in Brazil, metagenomics data have been integrated with environmental metadata to reveal microbial indicators in soil for sugarcane model systems. A multianalytical approach based on measurements of carbon dioxide and nitrous oxide emissions from soil, chemical factor analysis of the soil samples, and a survey of the soil microbial community using methods to determine the microbial biomass (conventional microbiological parameter for soil quality), and abundance of taxonomic groups of bacteria (57 soil metagenomics datasets and real-time quantitative PCR) was used. This approach revealed Acidobacteria, Actinobacteria and Verrucomicrobia, and their subgroups as potential indicators of effects of organic and inorganic amendments and straw retention in sugarcane-cultivated soils, which can alter the soil chemical factors

(Navarrete *et al.*, 2015b). Functional microbial profiling of metagenomics datasets can also compose this multianalytical approach as shown in Fig. 6.2.

Based on the integration of soil metagenomics with their associated environment metadata biologically relevant methods can be developed to assess soil quality. In this sense, high-throughput soil DNA sequencing applied to screening microbial community diversity and functions in soils under sugarcane biomass production systems can be associated to soil characteristics and processes occurring in soil, such as, fertility, nutrients availability, microbial biomass, enzyme activity and gas emissions from soil. This combinatorial approach has been initially applied to evaluate the potential impacts of bio-based sugarcane production processes on microbial functions and soil quality.

Sugarcane soils receive high amounts of fertilizers. Every year around to 80–150 kg/ha of nitrogen is amended to the soil to maintain the production of biomass (van Raij *et al.*, 1997). However, a great proportion of the N input is lost to the atmosphere due to denitrification processes. Loss of N is caused by volatilization and by the action of microbes capable of producing denitrification enzymes. The incomplete denitrification seems to be prevalent over the complete process, which is terminated by the release of N_2 through the reduction of N_2O, thereby enhancing the impact of agriculture by increasing emissions of greenhouse gases (GHG). Nitrogen inputs can naturally happen as a consequence of the metabolism of N_2-fixing microbes. Even though N_2 reduction by nitrogenases is an exergonic process, the

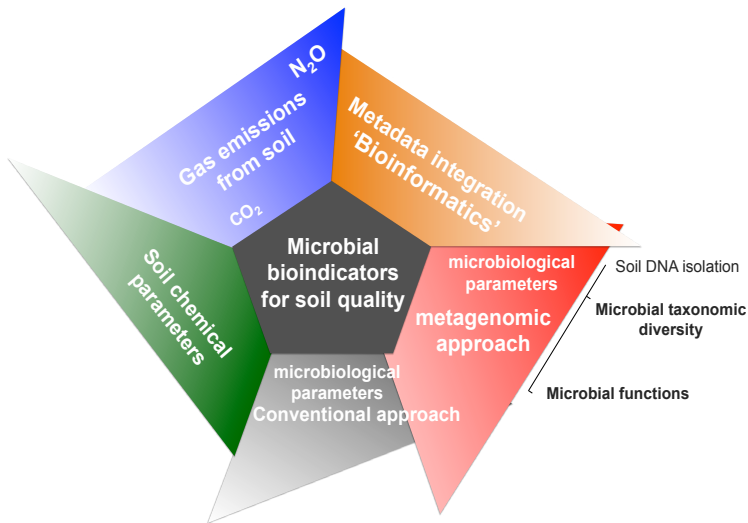

Figure 6.2 Representative scheme of the multianalytical approach used for integration of metagenomics with their associated environment metadata to reveal microbial indicators in soil for sugarcane model systems. Gas emissions from soil: carbon dioxide (CO_2) and nitrous oxide (N_2O); soil chemical parameters: soil fertility factors including pH, potential acidity, Ca, Mg, P, K, S, available micronutrients, exchangeable bases, cation exchange capacity, and base saturation; conventional microbiological parameters: microbial biomass and soil enzyme; soil metagenomics: which comprises isolation of soil DNA, and taxonomic and functional analysis of the soil microbial community using high-throughput DNA sequencing; Bioinformatics and statistics: analysis of metagenomics datasets and univariate and multivariate statistical analysis. Adapted from Navarrete *et al*. (2013b).

flow of energy generated is very expensive requiring much ATP; for this region, nitrogenases are inhibited by NH_3 (Boyd and Peters, 2013). However, in sugarcane, endophytes symbiosis with N_2-fixing microbes is known to happen, and they have been reported for more than 25 years (Cavalcante and Dobereiner, 1988).

Soil microbial community plays a crucial role not only on nutrients cycling, such as N as mentioned, but many other beneficial functions in rhizosphere are maintained by soil microbes. Pathogen suppression, probably by antibiotic production, and the supply of growth hormones are additional consequences of plant–microbe interactions in the roots (Bruto *et al*., 2014). Revealing the mechanisms behind these processes is of great concern to relieve the exaggerated impacts of fertilizers inputs and GHG emissions. In this direction, it is of considerable relevance to apply metagenomics approaches towards the understanding of microbial functions and diversity in the rhizosphere.

Interestingly, in a similar approach as described in Navarrete *et al.* (2015a), it is shown that sugarcane rhizosphere can enrich the genetic potential of specific microbial subsystems which are not detected in bulk soils. As demonstrated in Fig. 6.3, the enrichment of pathways

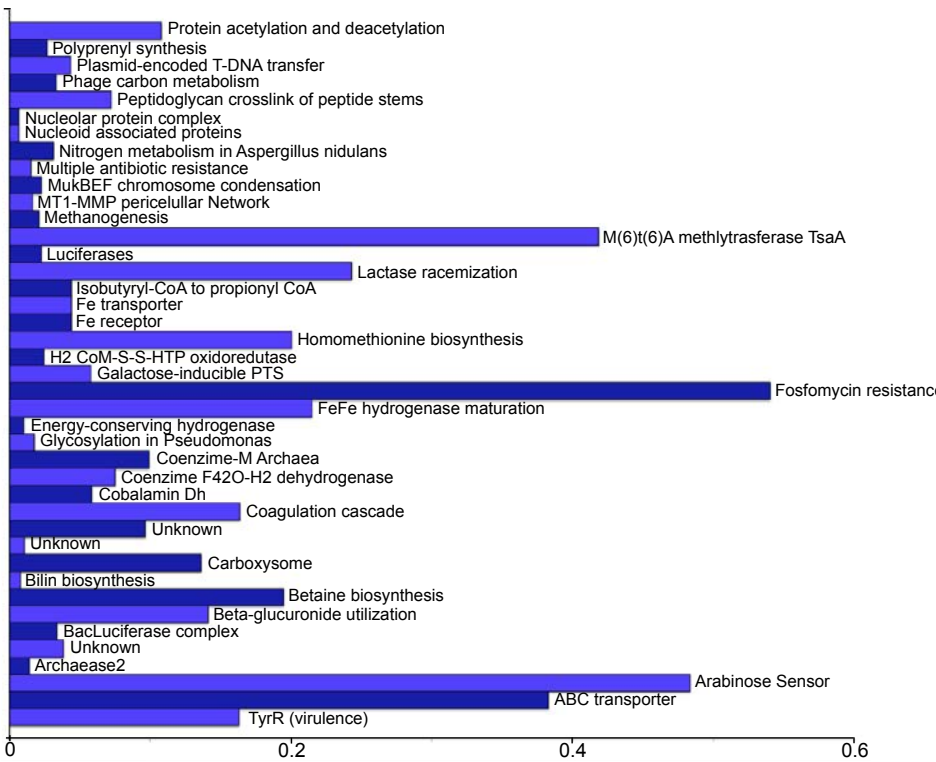

Figure 6.3 Average relative abundance (%) of microbial functional subsystems detected on sugarcane rhizosphere. Only microbial subsystems detected uniquely in rhizosphere are represented. Metagenomics profiling was performed using datasets extracted from sugarcane macrocosms in a similar approach as described in Navarrete *et al*. (2015a). Datasets from bulk soil were compared to datasets from rhizosphere (n = 3). Functional annotation of subsystems was performed using SUPERFOCUS (Silva *et al*., 2016).

associated to carbon cycling such as sugars and proteins was detected by metagenomics functional profiling, including also the process of methanogenesis. N cycling pathways enriched in the roots were probably qualitatively the same as enriched in the bulk soils. Additionally, some mechanisms of resistance to antibiotics were also reported. This analysis illustrates the importance of applying metagenomics approaches and its potential of reveling microbial functions in a global perspective putting all the microbial community in one scale.

Final remarks

Metagenomics has become an indispensable tool for studying the diversity and metabolic potential of environmental microbes, whose majority is still non-cultivable (Teeling and Glöckner, 2012). The analysis of metagenomics sequences has enabled researchers to study genes and functions from previously inaccessible microbes, opening the black box of microbial diversity in several environments. Despite recent advances in sequencing technologies and bioinformatics tools, the obtaining data are still biased by methodological process, such as soil sampling, DNA extraction, adsorption of nucleic acids to soil particles, sample preparation, sequencing protocols, sequence analysis and taxonomical and functional annotations (Nesme *et al.*, 2016). The use of different methodologies influences both the qualitative and quantitative results of molecular surveys and metagenomics (Delmont *et al.*, 2013). Another recurring problem is that about half of the genes in a metagenome have functions as yet unknown (Teeling and Glöckner, 2016).

However, metagenomics studies have allowed the access to the enormous microbial diversity, accelerating the discovery of novel phyla, classes, genera, and providing means for correlation between taxonomy and potential metabolic functions. From now on, the microbiologists are challenged to find less biased methodologies, which also should include a better access to the rare biosphere (minimizing the effect of dominant organisms), differentiating between active microorganisms and dormant cells, and increase the ability to link microorganisms to their metabolic roles within a community. Also, the reference databanks would be improved with information obtained through the combination of metagenomics and classical microbiology, which could lead to more accurate ecological inferences and interpretations.

Acknowledgements

LWM, LPPB, and AAN were supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP 2010/20353-7, 2014/03217-3, 2015/00251-9, 2013/22845-2, 2012/13321-7 and 2014/06077-8). DGS was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (DS-CAPES PDSE 99999.003968/2014-06), and Conselho Nacional de Desenvolvimento Científico (PRONEX-CNPq 140317/2014-7). GGZS was supported by NSF Grants (CNS-1305112, MCB-1330800, and DUE-132809). SMT thanks FAPESP (2011/51749-6) and CNPq for productivity scholarship.

References

Altman, T., Travers, M., Kothari, A., Caspi, R., and Karp, P.D. (2013). A systematic comparison of the MetaCyc and KEGG pathway databases. BMC Bioinf. *14*, 112. http://dx.doi.org/10.1186/1471-2105-14-112

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Atlas, R.M. (1984). Diversity of microbial communities. In Advances in Microbial Ecology, K.C. Marshall, ed. (New York City, US: Springer), v. 7, pp. 1-47.

Bardgett, R.D., Freeman, C., and Ostle, N.J. (2008). Microbial contributions to climate change through carbon cycle feedbacks. ISME. J. *2*, 805–814. http://dx.doi.org/10.1038/ismej.2008.58

Battistuzzi, F.U., and Hedges, S.B. (2009). A major clade of prokaryotes with ancient adaptations to life on land. Mol. Biol. Evol. *26*, 335–343. http://dx.doi.org/10.1093/molbev/msn247

Berendzen, J., Bruno, W.J., Cohn, J.D., Hengartner, N.W., Kuske, C.R., McMahon, B.H., Wolinsky, M.A., and Xie, G. (2012). Rapid phylogenetic and functional classification of short genomic fragments with signature peptides. BMC. Res. Notes. *5*, 460. http://dx.doi.org/10.1186/1756-0500-5-460

Borneman, J., and Triplett, E.W. (1997). Molecular microbial diversity in soils from Eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. Appl. Environ. Microbiol. *63*, 2647-2653.

Boyd, E.S., and Peters, J.W. (2013). New insights into the evolutionary history of biological nitrogen fixation. Front. Microbiol. *4*, 201. http://dx.doi.org/10.3389/fmicb.2013.00201

Brossi, M.J., Mendes, L.W., Germano, M.G., Lima, A.B., and Tsai, S.M. (2014). Assessment of bacterial bph gene in Amazonian dark earth and their adjacent soils. PLOS ONE *9*, e99597. http://dx.doi.org/10.1371/journal.pone.0099597

Bruto, M., Prigent-Combaret, C., Muller, D., and Moënne-Loccoz, Y. (2014). Analysis of genes contributing to plant-beneficial functions in Plant Growth-Promoting Rhizobacteria and related Proteobacteria. Sci. Rep. *4*, 6261. http://dx.doi.org/10.1038/srep06261

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. http://dx.doi.org/10.1038/nmeth.3176

Carroll, I.T., Cardinale, B.J., and Nisbet, R.M. (2011). Niche and fitness differences relate the maintenance of diversity to ecosystem function. Ecology *92*, 1157–1165.

Caspi, R., Altman, T., Dale, J.M., Dreher, K., Fulcher, C.A., Gilham, F., Kaipa, P., Karthikeyan, A.S., Kothari, A., Krummenacker, M., *et al.* (2010). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. *38*, D473–479. http://dx.doi.org/10.1093/nar/gkp875

Cavalcante, V.A., and Dobereiner, J. (1988). A new acid-tolerant nitrogen fixing bacterium associated with sugarcane. Plant Soil *108*, 23-31.

Cenciani, K., Lambais, M.R., Cerri, C.C., De Azevedo, L.C.B., and Feigl, B.J. (2009). Bacteria diversity and microbial biomass in forest, pasture and fallow soils in the southwestern Amazon Basin. Rev. Bras. Cienc. Solo*33*, 907-916.

Cerri, C.C., Bernoux, M., Cerri, C.E.P., and Feller, C. (2004). Carbon cycling and sequestration opportunities in South America: the case of Brazil. Soil Use Manage.*20*, 248-254.

Daniel, R. (2005). The metagenomics of soil. Nat. Rev. Microbiol. *3*, 470–478.

Della-Bianca, B.E., Basso, T.O., Stambuk, B.U., Basso, L.C., and Gombert, A.K. (2013). What do we know about the yeast strains from the Brazilian fuel ethanol industry? Appl. Microbiol. Biotechnol. *97*, 979–991. http://dx.doi.org/10.1007/s00253-012-4631-x

Delmont, T.O., Simonet, P., and Vogel, T.M. (2013). Mastering methodological pitfalls for surviving the metagenomic jungle. Bioessays *35*, 744–754. http://dx.doi.org/10.1002/bies.201200155

Edwards, R.A., Olson, R., Disz, T., Pusch, G.D., Vonstein, V., Stevens, R., and Overbeek, R. (2012). Real time metagenomics: using k-mers to annotate metagenomes. Bioinformatics *28*, 3316–3317. http://dx.doi.org/10.1093/bioinformatics/bts599

Ekkers, D.M., Cretoiu, M.S., Kielak, A.M., and Elsas, J.D. (2012). The great screen anomaly – a new frontier in product discovery through functional metagenomics. Appl. Microbiol. Biotechnol. *93*, 1005–1020. http://dx.doi.org/10.1007/s00253-011-3804-3

Fierer, N., Ladau, J., Clemente, J.C., Leff, J.W., Owens, S.M., Pollard, K.S., Knight, R., Gilbert, J.A., and McCulley, R.L. (2013). Reconstructing the microbial diversity and function of pre-agricultural tallgrass prairie soils in the United States. Science *342*, 621–624. http://dx.doi.org/10.1126/science.1243768

Finke, D.L.; Snyder, W.E. (2008). Niche partitioning increases resource exploitation by diverse communities. Science*321*, 1488–1490.

Fischer, M.M., and Triplett, E. (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. App. Environ. Microbiol.*65*, 4630-4636, 1999.

Foley, J.A., Defries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., *et al*. (2005). Global consequences of land use. Science *309*, 570–574.

Germano, M.G., Cannavan, F.S., Mendes, L.W., Lima, A.B., Teixeira, W.G., Pellizari, V.H., and Tsai, S.M. (2012). Functional diversity of bacterial genes associated with aromatic hydrocarbon degradation in anthropogenic dark earth of Amazonia. Pesqui. Agropecu. Bras. *47*, 654-664.

Girvan, M.S., Campbell, C.D., Kilham, K., Prosser, J.I., and Glover, L.A. (2005). Bacterial diversity promotes community stability and functional resilience after perturbation. Environ. Microbiol. *7*, 301–313.

Grossman, J.M., Sheaffer, C., Wyse, D., Bucciarelli, B., Vance, C., and Graha, P.H. (2010). An assessment of nodulation and nitrogen fixation in inoculated *Inga oestediana*, a nitrogen-fixing tree shading organically grown coffee in Chiapas, Mexico. Soil Biol. Biochem. *38*, 769-784.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. *68*, 669–685.

da C Jesus, E., Marsh, T.L., Tiedje, J.M., and de S Moreira, F.M. (2009). Changes in land use alter the structure of bacterial communities in Western Amazon soils. ISME. J. *3*, 1004–1011. http://dx.doi.org/10.1038/ismej.2009.47

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30.

Kent, W.J. (2002). BLAT – the BLAST-like alignment tool. Genome. Res. *12*, 656–664. http://dx.doi.org/10.1101/gr.229202. Article published online before March 2002

Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J.A., Hugenholtz, P., van der Lelie, D., Meyer, F., Stevens, R., *et al*. (2012). Unlocking the potential of metagenomics through replicated experimental design. Nat. Biotechnol. *30*, 513–520. http://dx.doi.org/10.1038/nbt.2235

Leff, J.W., Jones, S.E., Prober, S.M., Barberán, A., Borer, E.T., Firn, J.L., Harpole, W.S., Hobbie, S.E., Hofmocker, K.S., Knops, J.M.H., *et al.* (2015). Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. Proc. Natl. Acad. Sc. USA *112*, 10976-72.

Mendes, L.W., Brossi, M.J.L., Kuramae, E.E., and Tsai, S.M. (2015a). Land-use system shapes soil bacterial communities in Southeastern Amazon region. App. Soil Ecol. *95*, 151-160.

Mendes, L.W., Tsai, S.M., Navarrete, A.A., de Hollander, M., van Veen, J.A., and Kuramae, E.E. (2015b). Soil-borne microbiome: linking diversity to function. Microb. Ecol. *70*, 255–265. http://dx.doi.org/10.1007/s00248-014-0559-2

Mendes, L.W., Kuramae, E.E., Navarrete, A.A., van Veen, J.A., and Tsai, S.M. (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. ISME. J. *8*, 1577–1587. http://dx.doi.org/10.1038/ismej.2014.17

Mendoza, M.L.Z., Sicheritz-Pontén, T., and Gilbert, M.T. (2015). Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. Briefings Bioinf. *16*, 745–758. http://dx.doi.org/10.1093/bib/bbv001

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., *et al.* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinf. *9*, 386. http://dx.doi.org/10.1186/1471-2105-9-386

Metzker, M.L. (2010). Sequencing technologies – the next generation. Nat. Rev. Genet. *11*, 31–46. http://dx.doi.org/10.1038/nrg2626

Mitra, S., Rupek, P., Richter, D.C., Urich, T., Gilbert, J.A., Meyer, F., Wilke, A., and Huson, D.H. (2011). Functional analysis of metagenomes and metatranscriptomes using SEED and KEGG. BMC Bioinf. *12* (Suppl. 1), S21. http://dx.doi.org/10.1186/1471-2105-12-S1-S21

Mori, A.S.; Shiono, T.; Koide, D.; Kitagawa, R.; Ota, A.T.; Mizumachi, E.(2013). Community assembly processes shape an altitudinal gradient of forest biodiversity. Global Ecol. Biogeog.*22*, 878–888.

Murphy, C.A.; Foster, B.L. (2014). Soil properties and spatial processes influence bacterial metacommunities within a grassland restoration experiment. Restor. Ecol. *22*, 685–691.

Muyzer, G., and Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. Antonie. Van. Leeuwenhoek. *73*, 127–141.

Navarrete, A.A., Cannavan, F.S., Taketani, R.G., and Tsai, S.M. (2010). A molecular survey of the diversity of microbial communities in different Amazonian agricultural model systems. Diversity *2*, 787-809.

Navarrete, A.A., Kuramae, E.E., de Hollander, M., Pijl, A.S., van Veen, J.A., and Tsai, S.M. (2013a). Acidobacterial community responses to agricultural management of soybean in Amazon forest soils. FEMS Microbiol. Ecol. *83*, 607–621. http://dx.doi.org/10.1111/1574-6941.12018

Navarrete, A.A.; Perez Braga, L.P.; Diniz, T.R.; Rossetto, R.; Tsai, S.M. (2013b). Perspectives on combinatorial biological, chemical and physical parameters in soil quality: an approach for sugarcane production systems. In: 12th Symposium on Bacterial Genetics and Ecology (BAGECO), Ljubljana, Slovenia. Conference abstracts, p. 185.

Navarrete, A.A., Taketani, R.G., Mendes, L.W., Cannavan, F.S., Moreira, F.M.S., and Tsai, S.M. (2011). Land-use systems affect archaeal community structure and functional diversity in Western Amazon soils. Rev. Bras. Cienc. Solo *35*, 1527-1540.

Navarrete, A.A.; Diniz, T.R.; Perez Braga, L.P.; Silva, G.G.Z.; Franchini, J.C.; Rossetto, R.; Edwards, R.A.; Tsai, S.M. (2015a). Multi-analytical approach reveals potential microbial indicators in soil for sugarcane model systems. PLOS ONE *10*, e0129765.

Navarrete, A.A., Tsai, S.M., Mendes, L.W., Faust, K., de Hollander, M., Cassman, N.A., Raes, J., van Veen, J.A., and Kuramae, E.E. (2015b). Soil microbiome responses to the short-term effects of Amazonian deforestation. Mol. Ecol. *24*, 2433–2448. http://dx.doi.org/10.1111/mec.13172

Navarrete, A.A., Venturini, A.V.; Meyer, K.M.; Klein, A.M.; Tiedje, J.M.; Bohannan, B.J.M.; Nüsslein, K.; Tsai, S.M.; Rodrigues, J.L.M. (2015c). Differential response of *Acidobacteria* subgroups to forest-to-pasture conversion and their biogeographic patterns in the Western Brazilian Amazon. Front. Microbiol. *6*, 1443.

Nesme, J., Achouak, W., Agathos, S.N., Bailey, M., Baldrian, P., Brunel, D., Frostegård, Å., Heulin, T., Jansson, J.K., Jurkevitch, E., *et al.* (2016). Back to the Future of Soil Metagenomics. Front. Microbiol. *7*, 73. http://dx.doi.org/10.3389/fmicb.2016.00073

O'Neill, B., Grossman, J., Tsai, S.M., Gomes, J.E., Lehmann, J., Peterson, J., Neves, E., and Thies, J.E. (2009). Bacterial community composition in Brazilian Anthrosols and adjacent soils characterized using culture and molecular identification. Microb. Ecol. *58*, 23-35.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., *et al.* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res. *33*, 5691–5702.

Pagani, I., Liolios, K., Jansson, J., Chen, I.M., Smirnova, T., Nosrat, B., Markowitz, V.M., and Kyrpides, N.C. (2012). The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. *40*, D571–579. http://dx.doi.org/10.1093/nar/gkr1100

Pazinato, J.M., Mendes, L.W., Paulo, E.N., Vazoller, R.F., and Tsai, S.M. (2010). Molecular characterization of the archaeal community in an Amazonian Wetland Soil and culture-dependent isolation of methanogenic Archaea. Diversity*2*, 1026- 1047.

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C.M., Sarr, A., and Maron, P.A. (2013). Loss in microbial diversity affects nitrogen cycling in soil. ISME. J. *7*, 1609–1619. http://dx.doi.org/10.1038/ismej.2013.34

Prosser, J.I. (2015). Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. Nat. Rev. Microbiol. *13*, 439–446. http://dx.doi.org/10.1038/nrmicro3468

Rodrigues, J.L., Pellizari, V.H., Mueller, R., Baek, K., Jesus, E.d.a.C., Paula, F.S., Mirza, B., Hamaoui, G.S., Tsai, S.M., Feigl, B., *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*, 988–993. http://dx.doi.org/10.1073/pnas.1220608110

Silva, G.G., Cuevas, D.A., Dutilh, B.E., and Edwards, R.A. (2014). FOCUS: an alignment-free model to identify organisms in metagenomes using non-negative least squares. PeerJ *2*, e425. http://dx.doi.org/10.7717/peerj.425

Silva, G.G., Green, K.T., Dutilh, B.E., and Edwards, R.A. (2016). SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. Bioinformatics *32*, 354–361. http://dx.doi.org/10.1093/bioinformatics/btv584

Souza, D.G.; Mendes, L.W.; Navarrete, A.A.; Tsai, S.M. (2015). Microbial assembly in agroecosystems: from the small arise the big. In: Blanco, J.A.; Lo, Y.-H. (Eds.). Biodiversity in ecosystems - Linking structure and function. 1 ed. Rijeka: Intech, 187–201.

Souza, D.G. (2015). Comparative analyses of microbial phylogenetic and functional processes following long-term land-use change. Doctoral Thesis. University of Sao Paulo, Piracicaba, Brazil.

Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., *et al.* (2015). Ocean plankton. Structure and function of the global ocean microbiome. Science *348*, 1261359. http://dx.doi.org/10.1126/science.1261359

Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214. http://dx.doi.org/10.1038/nature11234

Taketani, R.G., and Tsai, S.M. (2010). The influence of different land uses on the structure of archaeal communities in Amazonian anthrosols based on 16S rRNA and amoA genes. Microb. Ecol. *59*, 734–743. http://dx.doi.org/10.1007/s00248-010-9638-1

Taketani, R.G., Lima, A.B., da Conceição Jesus, E., Teixeira, W.G., Tiedje, J.M., and Tsai, S.M. (2013). Bacterial community composition of anthropogenic biochar and Amazonian anthrosols assessed by 16S rRNA gene 454 pyrosequencing. Antonie. Van. Leeuwenhoek. *104*, 233–242. http://dx.doi.org/10.1007/s10482-013-9942-0

Tardy, V., Mathieu, O., Lévêque, J., Terrat, S., Chabbi, A., Lemanceau, P., Ranjard, L., and Maron, P.A. (2014). Stability of soil microbial structure and activity depends on microbial diversity. Environ. Microbiol. Rep. *6*, 173–183. http://dx.doi.org/10.1111/1758-2229.12126

Teeling, H., and Glöckner, F.O. (2012). Current opportunities and challenges in microbial metagenome analysis – a bioinformatics perspective. Briefings Bioinf. *13*, 728-742.

Thies, J.E. (2007). Soil microbial community analysis using terminal restriction fragment length polymorphism. Soil Sci. Soc. Am. J. *71*, 579-591.

Torsvik, V., Øvreås, L., and Thingstad, T.F. (2002). Prokaryotic diversity – magnitude, dynamics, and controlling factors. Science *296*, 1064–1066. http://dx.doi.org/10.1126/science.1071698

vanRaij, B., Cantarella, H., Quaggio, J.A., and Furlani, A.M.C. (1997). Fertilization and limin recommendations for Sao Paulo state. BoletimTécnico 100-Instituto Agronômico, Campinas, Brazil.

Vellend, M. (2010). Conceptual synthesis in community ecology. Q. Rev. Biol. *85*, 183–206.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G.F., Chater, K.F., and van Sinderen, D. (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. Microbiol. Mol. Biol. Rev. *71*, 495–548.

Wagg, C., Bender, S.F., Widmer, F., and van der Heijden, M.G. (2014). Soil biodiversity and soil community composition determine ecosystem multifunctionality. Proc. Natl. Acad. Sci. U.S.A. *111*, 5266–5270. http://dx.doi.org/10.1073/pnas.1320054111

Wawrik, B., Kerkhof, L., Kukor, J., and Zylstra, G. (2005). Effect of different carbon sources on community composition of bacterial enrichments from soil. Appl. Environ. Microbiol. *71*, 6776–6783.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. U.S.A. *95*, 6578–6583.

Zhang, J., Chiodini, R., Badr, A., and Zhang, G. (2011). The impact of next-generation sequencing on genomics. J. Genet. Genomics. *38*, 95–109. http://dx.doi.org/10.1016/j.jgg.2011.02.003

Zhao, Y., Tang, H., and Ye, Y. (2012). RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics *28*, 125–126. http://dx.doi.org/10.1093/bioinformatics/btr595