RESPONSE OF MICROBIAL NETWORKS AND MICROBIOMES TO THE FOREST-TO-PASTURE CONVERSION IN AMAZON SOILS

By

Md Abdul Wadud Khan

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

November 2016

Acknowledgements

During my PhD years at UTA, I have received tremendous support from a number of people, especially my PhD advisors: Dr. Jorge L. M. Rodrigues (currently, Associate Professor in the Department of Land, Air and Water Resources at University of California-Davis), and Dr. James P. Grover, Professor in Biology. I could not have asked better advisors. Besides them, my committee members have always been very supportive: Dr. Thomas Chrzanowski, Dr. Jeffery Demuth, and Dr. Julian Hurdle. Without their proper guidance, it would not have been possible to accomplish my research goals on time. The rest of the faculty in Biology, especially those I had taken courses with, and office ladies (Paulette Batten, Linda Taylor, Gloria Burlingham, and Sherri Echols) at the Life Science building have also been a blessing to me. Dr. Christensen was very helpful as graduate student advisor. I am very grateful to them.

The support from project collaborators across country expedited my work. Collaborators include Dr. Brendan Bohannan (University of Oregon), Dr. Klaus Nüsslein (University of Massachusetts, Amherst), and Dr. James Tiedje (Michigan State University). In addition, I thank the owners and staff of Agropecuaria Nova Vida for logistical support and permission to work on their property. I am grateful to Wagner Piccinini and Gunther Brucha for field sampling, Mark Charbonneau for DNA amplification and Dr. Ann Klein (Bohannan lab) for providing me with the metagenome datasets. This project was supported by the Agriculture and Food Research Initiative Competitive Grant from the US Department of Agriculture National Institute of Food, and Agriculture and Fundação de Amparo à Pesquisa do Estado de São Paulo Grant.

I am also thankful to my lab mates and graduate students of the department who helped me not only in the laboratory experiments and analysis but also provided me with helpful suggestions through scientific discussions. Especially, Malini Kotak (PhD candidate), Dr. Babur Mirza (postdoc), Luis Velazquez and Michael Fanning (undergrad), Dr. Acacio A. Navarrete and Dennis Goss-Souza (visiting scholars) were very supportive.

My parents- Md Joynul Abedin Khan and Rekhanur Begum, siblings, especially my brother- Abu Hamim Khan had successfully instilled a dream in me to pursue PhD from a reputed university in the United States of America, and supported accordingly. My wife Rezbin Karim has been a great encouragement during my PhD years. This dissertation is dedicated to them.

November 09, 2016

Abstract

RESPONSE OF MICROBIAL NETWORKS AND MICROBIOMES
TO THE FOREST-TO-PASTURE CONVERSION IN AMAZON SOILS

Md Abdul Wadud Khan, PhD

The University of Texas at Arlington, 2016

Supervising Professors: Jorge L. M. Rodrigues and James P. Grover

The Amazon rainforest is the largest tropical rainforest ecosystem in the world, and an extremely important region ecologically. Due to the growing human needs mainly for cattle ranching and cultivation, the rainforest has been in the process of clearing since 1970s losing an average of 25,000 km annually. This ecosystem conversion has a substantially adverse impact on biological diversities, in which the response of microbial diversity has not been studied until recently. Several studies showed that the conversion severely altered microbial diversities, with an increased alpha and decreased beta diversities. Given that microorganisms are closely associated with soil-associated functions, where microbial interactions are important in addition to diversity, I attempted to explore the spatial pattern of microbial co-occurrence networks based on the relative abundance of microbial species. Moreover, I used metagenomes to analyze differential abundances of metabolic processes related to ecosystem functioning. The taxonomic (16S rRNA) and functional gene datasets were obtained from Amazon forest and pasture soils. The analyses showed that forest-to-pasture conversion altered network topological properties, with decreases in the number of nodes and edges, average connectivity, betweenness centrality, and in-degree. My results also revealed that network topological

patterns across ecosystem, and topological roles of major bacteria taxa differed significantly between forest and pasture, indicating substantial alteration of relationship pattern between different bacterial taxa. I estimated this pattern, which may discern life-style strategies of species, and syntrophy between species, critical for adaptation. For example, higher co-occurrence of *Alphaproteobacteria* and *Nitrospirae*, known to have members involved in nitrification and denitrification processes, in forest soils may indicate higher reliance on nitrogen metabolism for energy harvest. Consistent with the taxonomic shifts, functional genes of forest soils showed a similar pattern of metabolic pathways involved in nitrogen cycle, whereas the pasture microbiome had increased abundance of genes involved in methanogenesis and fermentation. Conversely, forest microbiomes had higher relative abundance of genes involved in the TCA cycle and oxidative phosphorylation. With the survey of taxonomic and functional diversities, this study sheds light on how the anthropogenic activities impact on co-occurrence and ecosystem functioning. Our approach, however, requires further validation, and approaches such as multi-omics, cultivation and lab experiments should be integrated for better understanding of soil-associated functions in these ecosystems.

Table of Contents

List of Figures

Chapter 1

Overview on the response of microbial diversity to forest-to-pasture conversion in the Amazon and significance of this research project

1.1 Introduction

The Amazon is the largest reservoir of plant and animal diversities (Dirzo et al. 2003), and presumably microbial diversity as well. Due to the growing demand of humans, the Amazon ecosystem has been facing a huge environmental pressure over the last few decades. One of the primary threats is deforestation. Anthropogenic activities such as pollution and deforestation lead to adverse consequences to the biological composition and thereby ecosystem functions mediated by the entire biological system (Laurance et al. 2006). While there are many comprehensive studies on the plant and animal communities in the Amazon and effects of deforestation on them (Bierregaard et al. 2001, Soares-Filho et al. 2009), few studies were conducted on microbial communities, which were restricted to only microbial composition and diversity (Borneman et al. 1997, Cenciani et al. 2009, Jesus et al. 2009, Rodrigues et al. 2013, Mirza and Rodrigues 2012), and functional diversity (Navarrete et al. 2011, Paula et al. 2014). Microorganisms comprise of a major portion of terrestrial ecosystems, and are closely associated with ecosystem functions such as carbon and nitrogen sequestration and biogeochemical cycling of elements (Madsen 2011). Any significant changes in microbial diversity therefore might result in alterations of large-scale ecosystem processes. To obtain an in-depth knowledge of microbial diversity and to measure the impact of rapid deforestation on its diversity, the Amazon Rainforest Microbial Observatory (ARMO) was created in 2009.

Previous studies have shown that bacterial communities are different between Amazon forest soils and pasture soils, and ecosystem conversion from forest to pasture in Amazon rainforest alters biogeochemical cycling processes (Herpin et al. 2002, Neill et al. 1997a and b, Neill et al. 1995, Neill et al. 1999, Fernandes et al. 2002). However the link between the microbial diversity and function remains unknown. One of the most comprehensive studies on microbial composition and diversity in Amazon revealed that the conversion of primary forest to pasture in Amazon increases microbial alpha diversity, while decreasing beta diversity (Rodrigues et al. 2013). This causes spatial homogenization of bacterial communities at taxonomic and phylogenetic levels. However, it remains unknown whether bacterial homogenization results in functional losses. To obtain more insights into it, it is now imperative to study the microbial taxa and the ecosystem functions associated with them are impacted by deforestation in the Amazon soil.

Soil is a heterogeneous environment that harbors enormously diverse microbial communities, where they dwell by interacting with each other and affect growth and metabolism of others (Faust and Raes 2012, Papke and Ward 2004, Horner-Devine et al. 2004, Fierer and Jackson 2006). This complex ensemble of microbes plays essential roles in biogeochemical processes and to maintain a healthy ecosystem. The inter-taxa relationships may help reveal the niche spaces shared by community members of microbes (Williams et al. 2014), which is particularly valuable in evaluating the impacts of ecosystem conversion. However, based on our current knowledge of biology, it is currently impossible to map out direct interactions in complex microbial communities, which are mostly uncultivable and poorly studied. Therefore microbial network studies

are widely used to explore potential inter-taxa relationships for better understanding of microbial community structure, where a variety of methods have been developed for network construction (Faust et al. 2012, Zhou et al. 2011, Williams et al. 2014, Widder et al. 2014).

I have analyzed taxonomic network and metagenomes of Amazon forest and pasture soils to explore the potential consequences on ecosystem processes due to forest-to-pasture conversion. The network analysis determines the co-occurrence relationships of microbes based on relative abundances of different microbial species present across and within samples. Topology based analysis allows us to obtain insights into relative significance of individual species in the co-occurrence networks. The implementation of metagenomic datasets, on the other hand, aids in examining the differential abundances of biochemical pathways between ecosystems based on the relative abundances of functional genes. Therefore comparative examination of metagenomes demonstrates the potential variations of important ecosystem processes that may result from forest-to-pasture conversion.

1.2 Background and significance

A microbial ecosystem is a complex system where microorganisms interact with each other and the environment surrounding them, which dictates their contributions to ecosystem functions (Fuhrman et al. 2008, Raes and Bork 2008). Therefore network analysis has been applied widely to understand the potential ecological relationships within microbial communities in different ecosystems (Faust et al. 2012, Zhou et al. 2011, Williams et al. 2014, Widder et al. 2014, Ding et al. 2015, Barberan et al. 2012).

While cooperative metabolic interactions potentially increase the abundance of interacting microbes, competition for same resources may lead to an opposite abundance pattern. Thus network analysis may identify potential microbial interactions, posing hypotheses for further work on the mechanisms behind community patterns. My study aimed to demonstrate these co-occurrence relationships based on relative abundances of microorganisms. However, abundance patterns of microbial pairs may reflect their response to a common environmental factor rather than their direct interactions.

Each network may be divided into sub-networks, termed modules, which contain a set of members i.e., microbial species, that have a higher number links among them than with other members of other modules (Bascompte et al. 2009). It is assumed that the microbial species that belong to a module have similar ecological niches (Olesen et al. 2007). According to Hartwell et al. 1999, a module in a biological system is "a discrete unit whose function is separable from those of other modules". Therefore modeling the entire ecosystem with networks reveals the properties of ecosystem, where different modules would capture the different functionalities of a complex ecosystem.

Following the construction of networks, a myriad of topological features are used to describe network properties. Topologies, especially some of the centrality measures, are used to determine the role of network members such as degrees, betweenness centrality etc. In a network, degree of a node is the number of connections it has to other nodes. It is an important measure while it provides a local characteristic of nodes and does not provide the information about the importance of nodes in the network. Betweenness centrality, on the other hand, reflects the number of times a node (i.e., an item in the network) plays a role as a connector along the shortest path between two other

nodes (González et al. 2010; Vick-Majors et al. 2014). While nodes with high betweenness values are likely to be situated in the core of the network, nodes with low values are expected to have a more peripheral location.

Over the last few years, the advances in second generation sequencing technology, together with the development of analysis tools, provide new insight into microbial ecology by deciphering the hidden diversity of uncultivable microorganisms and their functional potentialities. We now have to take the challenge to expand microbial community analysis beyond the estimation of alpha and beta diversities. Taxonomic information alone cannot provide a complete picture of microbial contributions in ecosystem processes. Therefore we have to study their roles in specific ecosystem functions and possible changes in functions due to anthropogenic activities. In my research projects, I have analyzed 16S rRNA genes and metagenomes of Amazon forest and pasture soils, and my results are expected to contribute significantly to the understanding of the impact of ecosystem conversion on microbial co-occurrence patterns, and soil-associated ecosystem functions conducted by microorganisms.

1.3 Project aims

The specific research aims are:

Aim 1: To explore the effect of diversity loss on microbial co-occurrence network structure and composition.

Aim 2: To examine the spatial pattern of bacterial co-occurrence networks across ecosystems, and ecological relationships between bacterial taxa.

Aim 3: To analyze the impact of ecosystem conversion on the functional diversity and differential abundances of biochemical pathways with ecological significance.

1.4 Hypotheses

Based on the previously published results, my study revolves around the following three hypotheses.

Hypothesis 1: Forest-to-pasture conversion will lead to less connected microbial communities in co-occurrence networks.

Hypothesis 2: The alteration of microbial diversity will result in the alteration of ecological relationship patterns between microbial taxa, and the spatial distribution pattern of co-occurrence networks.

Hypothesis 3: Microbial diversity loss will cause functional diversity loss, and differential abundances of biochemical pathways.

Chapter 2

Network analysis reveals substantial alteration of microbial co-occurrence patterns due to deforestation in the Amazon rainforest

2.1 Abstract

In complex communities, organisms interact with the surrounding environment and among themselves, driving the success or failure of each species. Historically, ecologists emphasized competitive interactions as an important factor for microbial assemblies, but later, a more pluralistic view of interactions suggests that non-random co-occurrence of specific taxa may be the result of either different species responding to similar environmental conditions and/or interacting positively. Co-occurrence networks analysis allows us for the identification of potential interactions between species. Here, we performed Pearson's correlation calculations for all possible pairwise combinations of 16S rRNA gene sequence abundances of forest and pasture datasets from the Amazon to identify network links ($r \geq 0.95$). The forest-to-pasture conversion altered network topological properties, with decreases in the number of nodes and edges, average connectivity, betweenness centrality, and in-degree. Network membership abundance changed with land use and similar results were confirmed for selected taxa using quantitative PCR. A predictive functional profiling based on high throughput 16S rRNA gene sequences indicated that the removal of hub nodes could have a profound impact on ecosystem functions, including nitrification and denitrification. The number of 16S rRNA gene sequences identified as belonging to the ammonia-oxidizing *Candidatus* Nitrososphaera and the nitrite-oxidizing *Nitrospira* were greatly reduced in pasture. This study shows that microbial taxa may occur together despite their metabolic dissimilarity.

We argue that such relationships require us to reinterpret microbial community assembly models taking positive interactions into consideration.

2.2 Importance

Advances in next-generation sequencing technology coupled with novel analytical tools can provide new insights by deciphering the hidden organization of yet-to-be-cultured microorganisms. Traditionally, biodiversity studies have relied on species richness and turnover, but ignore co-occurrence patterns, which are important for understanding how communities assemble and respond to changes. We report that ecosystem conversion substantially alters microbial co-occurrence patterns in soil. We identified microbial groups serving as network hubs, the equivalent of keystone species, and predicted their functional profiles, providing novel information regarding changes to ecosystem biogeochemical processes following forest-to-pasture conversion. This work provides the foundation to identify species interactions occurring in natural ecosystems.

2.3 Introduction

The Amazon is the largest continuous rainforest ecosystem in the world and home to thousands of species (Dirzo et al. 2003). Due to the growing demand for resources, mainly fuel and food, the Amazon rainforest has been facing continuous pressure over the last few decades, leaving its biological composition in jeopardy. One of the primary threats is forest-to-pasture conversion, with a substantial impact on communities of plants and animals (Bierregaard et al. 2001, Soares-Filho et al. 2009), as well as microorganisms (Rodrigues et al. 2013, Ranjan et al. 2015, Navarrete et al. 2015).

Microorganisms, being closely associated with ecosystem functions such as the biogeochemical cycling of elements (Madsen 2011), make a substantial contribution to ecosystem services and stability. Therefore, any significant changes in microbial communities could result in alterations of large-scale processes.

Soil is a heterogeneous environment that harbors an enormously diverse microbial community (Torsvik et al. 1990, Curtis et al. 2002). Members of this community interact ecologically in ways that affect their growth and metabolism. Such interactions can result in patterns of species abundance across space and time. Cooperative metabolic interactions can lead to increased growth of interacting microbes and ultimately to positive co-occurrence patterns in abundance, while competition for the same resources may lead to an inverse pattern (Greenblum et al. 2013). Abundance patterns may also reflect the response of different species to a common environmental factor rather than their direct interactions (Zhou et al. 2011). A co-occurring microbial pair therefore indicates they are either interacting synergistically or they have similar responses to environmental factors. This inter-taxa relationship may help reveal the niche spaces shared by members of a microbial community (Williams et al. 2014), which is particularly valuable in evaluating the impacts of environmental changes such as deforestation on microbial communities. To analyze co-occurrence patterns, it can be helpful to organize community data into networks, where each node represents a species and the edges represent correlations in abundance (Zhou et al. 2011). Each network may be divided into sub-networks, termed modules, which contain a set of members i.e., microbial species, that have a higher number links among them than with other members of other modules (Bascompte et al. 2009). Groups belonging to a module may have

similar ecological niches (Olesen et al. 2007). Although it is currently not possible to map out direct interactions in complex microbial communities, an empirical study reported that phylogenetic markers (e.g., 16S rRNA genes) properly predict the niche-defining properties (Fuhrman et al. 2006). Hence, microbial network studies may be a way forward to understand and test potential inter-taxa relationships in microbial communities (Faust et al. 2012, Zhou et al. 2011, Williams et al. 2014, Widder et al. 2014), and to enable questions of higher focus.

In this study, we hypothesize that alteration of microbial composition and diversity due to deforestation will lead to the substantial alteration of co-occurrence patterns across microbial taxa. By constructing correlation networks using 16S rRNA gene data, and estimating network defining topologies, we aimed: (*i*) to quantify the impact of microbial diversity loss on the network structure and composition; (*ii*) to obtain insights into the relative importance of individual microbial species in the networks; and (*iii*) to detect the potential variation in important ecosystem processes that may result from forest-to-pasture conversion.

2.4 Materials and methods

2.4.1 Site description and sampling

The Amazon Rainforest Microbial Observatory (ARMO) study site is located at Fazenda Nova Vida in the State of Rondonia, Brazil (10°10′18.71″S, 62°47′15.67″W), representing one of the highest rates of deforestation of the Brazilian Amazonia in the last two decades (INPE 2011). Our study was conducted on ten soil samples collected from a primary forest and pasture at the end of the rainy season, March 2010. The primary forest

is a typical wet Terra Firme forest with the majority of the trees yet to be identified, while the pasture was established in 1972 after a slash-and-burn procedure followed by aerial seeding of the fast growing grasses *Urochloa brizantha* and *Panicum maximum*. Sampling plots are 5.5 km apart and on a red-yellow podzolic latosol with physicochemical characteristics described in detail elsewhere (Neill et al. 1997a).

A sampling design based on a 100m^2 quadrat with 10-m^2, 1-m^2, 0.1-m^2, and 0.01-m^2 quadrats within was established for primary forest and pasture). For this study, samples were collected from a transect along the cardinal direction North in both forest and pasture. Following the removal of litter, soil sampling was performed with a 10 cm-depth by 5 cm-diameter corer with samples being transported on ice to the laboratory and stored at -80^oC until soil DNA extractions. Soil physicochemical properties were analyzed at the Laboratorio de Fertilidade do Solo, Department of Soil Sciences, University of Sao Paulo. At the time of sampling, pH measured *in situ* averaged 3.96 in forest and 4.24 in pasture, while average soil temperature for both sites was 25.5^oC.

2.4.2 DNA extraction, amplification and sequencing

Ten grams of soil from each sample were used for total genomic DNA using the PowerLyzer PowerSoil DNA isolation kit (MoBio Inc, Carlsbad, CA, USA). The concentration and purity of soil DNA were determined spectrophotometrically (NanoDrop Technologies Inc., Wilmington, DE). Prokaryotic primers 515F and 806R, which amplify the V4 hyper-variable region of 16S rRNA gene, were used for PCR amplification as described elsewhere (Caporaso et al. 2012). A unique 12-bp long barcode was incorporated to the forward primer for specific identification of each sample. The resulting bar-coded amplicons were pooled in an equimolar concentration to carry

out sequencing using the MiSeq platform (Illumina, Inc., San Diego, California) at the DOE Joint Genome Institute (Walnut Creek, CA).

2.4.3 Sequence processing and taxonomic assignment

All the upstream and downstream analyses of raw Illumina sequences were carried out in the QIIME 1.8.0 environment as described in Caporaso et al. (2010) and Kuczynski et al. (2012). Raw sequences belonging to a specific sample were sorted based on the bar code sequences, followed by quality filtering to discard anonymous bases and removal of primer sequences. Sequences were then assigned to Operational Taxonomic Units (OTUs) at a minimum of 97% sequence identity using *de novo* OTU-picking protocol with QIIME 1.8.0 (Caporaso et al. 2010) with the following controls before analysis: (1) singleton OTUs were removed from downstream analysis to reduce the possibility of sequencing error and to differentiate unique OTUs from potential data noise, (2) rarefaction was performed ten times and subsequent results were based on standard means of rarefactions. The algorithm Uclust (Edgar, 2010) was used to cluster the quality-filtered reads against the GreenGenes database (DeSantis et al. 2006), which was followed by the assignment of taxonomy using the RDP classifier (Wang et al. 2007).

2.4.4 Taxonomic network analysis

Networks were constructed to identify correlations among members of soil bacterial communities using the open-access pipeline Molecular Ecological Network Analysis (MENA, Zhou et al. 2011). The correlation coefficient cutoff was $r \geq 0.95$ with significance of $P \leq 0.01$ for pairwise comparisons. When constructing networks, relative abundance data for OTUs were transformed to logarithmic scale, followed by Pearson

correlation assessment of all possible pairwise combination of OTUs across samples. Only significant pairwise relationships were combined for network construction with each node representing an OTU and each edge representing a significant pairwise association between them: a positive correlation between two OTUs denoted similar abundance patterns, while a negative correlation was characterized as opposite abundance patterns. Interacting nodes within networks represented co-occurrence across samples. Networks were visualized with Cytoscape 2.6.0 using group attributes layout, where it separates OTUs nodes into modules (Cline et al. 2007). Only modules with at least five nodes were used in the subsequent analysis.

Ecological networks exhibit several properties, such as scale-free, small world, and modularity (Barabási et al. 2004). After the construction of networks, the following topological features were used to describe network properties: (1) connectivity, which is the average number of edges originating from a node; (2) degree, which is the average number of edges connecting nodes; (3) path length, which is the average number of edges connecting any two nodes; (4) modularity, which refers to the ability of a network to become organized into individually separated and distinct units (i.e., modules); and (5) betweenness centrality, which reflects the number of times a node plays a role as a connector along the shortest path between two other nodes. While average values of the indexes are generally used to describe the overall features of the network, the betweenness centrality value of each node can indicate its relative importance in the network. Nodes with higher betweenness values are likely to be situated in the core of the network and those with lower values are expected to have a more peripheral location.

2.4.5 Community evenness

Because deforestation may have a direct effect on soil microbial diversity through inequality of species composition, we tested whether network structures were altered between forest and pasture. Microbial community evenness was compared using an equitability index (Shannon 1948). The index value ranges between 0 to 1, where 0 indicates completely uneven community and 1 indicates complete evenness. Statistical analyses for both indexes were carried out with a two-sample *t*-test between land uses.

2.4.6 Bipartite network analysis

A bipartite network study was conducted to determine how taxonomic network members were shared between land use types (Navas-Molina et al. 2013). The analysis provided two node types: OTU-nodes and sample-nodes. If an OTU is present in a sample, then the OTU-node is connected to sample-node(s) by edge(s). The node and edge files were generated with a QIIME script using an OTU table as input file and imported into Cytoscape 2.6.0 for descriptive visualization (Cline et al. 2007). In the diagram, the number of shared members between samples and their abundances shape the clustering pattern of samples. While OTUs shared by multiple samples tend to cluster together, unique OTUs restricted to a single sample remain isolated. Likewise, samples sharing more OTUs tend to remain close together. In order to simulate the output of the bipartite network analyses, taxonomic network similarity was evaluated by using the Jaccard index. The value of the index was constrained between 0 and 1, with a higher value representing higher similarity of microbial communities between two network members. This analysis was performed in the QIIME 1.8.0 environment.

2.4.7 Predictive functional profiling

To determine the functional potential of the network members, we used the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) approach (Langille et al. 2013). To test whether our observed network-derived microbial communities were in agreement with PICRUSt predictions, two Procrustes analyses were carried out: First, a PICRUSt compatible OTU-picking method was employed for selection of representative sequences from the 16S rRNA gene dataset to obtain a closed-reference OTU table, which was compared against a *de novo* OTU table. Second, the closed-reference table was compared against KEGG Orthologies (KO), level 4, which represents specific functional genes, including those for enzymes with Enzyme Commission (EC) numbers (Kanehisa 2000). Pairwise distances between samples were calculated using the Bray-Curtis similarity matrix (Muegge et al. 2011). Statistical significance of the goodness of fit, M^2, was estimated by a Monte Carlo permutation strategy with 1,000 repetitions for both comparisons.

2.4.8 Quantitative PCR targeting bacterial phyla

Five dominant bacterial phyla (the phylum *Proteobacteria* was analyzed for three of its six classes) were quantified using quantitative PCR using phylum/class specific primer sets (Supplementary Table 2-1). A total of 15 replicates (5 biological X 3 technical) were used for each land use and each taxon. Each 20µl-reaction mixture contained 10µl of SYBR Green Supermix (2X, Bio-Rad, Hercules, CA), 0.3 µM of each primer and 5 ng of DNA sample. Reactions were performed with the Applied Biosystems 7300 real-time PCR system and conditions were experimentally determined for each primer set (Supplementary Table 2-1). Negative controls were included with each

reaction using PCR-grade water. Dissociation curve analysis of each post-reaction confirmed the specificity of the products. Standard curves (10^3 to 10^7 copies per reaction) were generated with the pCR2.1-TOPO vector containing a PCR-amplified fragment for each phylum or class (Invitrogen Corp., Carlsbad, CA). Reaction efficiency (E) was determined with the equation $E = 10^{(-1/slope)}$. Differences in abundance between forest and pasture samples were assessed using the two-sample *t* test.

2.5 Results

2.5.1 OTU filtering and data standardization

To reduce the noise caused by potential PCR or sequencing errors, singleton OTUs from each sample were filtered out prior to downstream analysis. Following filtering, the number of sequences obtained from different samples ranged from 27,014 to 34,659 with the average length of 250 bases. Rarefaction was performed through random sampling for ten times at an even depth of 10,000 sequences per sample and data were merged for downstream analysis, resulting in a total number of 11,673 OTUs that belong to forest and 12,227 OTUs to pasture. A total of 16,660 unique OTUs were observed in both land uses.

2.5.2 Topological features of taxonomic networks

Correlation networks, which included only significant OTU-OTU relationships based on their relative abundances, were constructed for forest (Figure 2-1A) and pasture (Figure 2-1B). The pasture network had a lower number of nodes (544) and edges (1089) in comparison to those observed in the forest (624 nodes and 1351 edges). The average connectivity value was found to be higher for forest (4.33) in comparison to pasture

(4.00), where the connectivity followed a power-law model with coefficients of 0.93 and 0.91, respectively. We estimated a higher average path length in forest (2.84) compared to pasture (2.4). Network modularities were estimated to be 0.77 and 0.80 for forest and pasture, respectively.

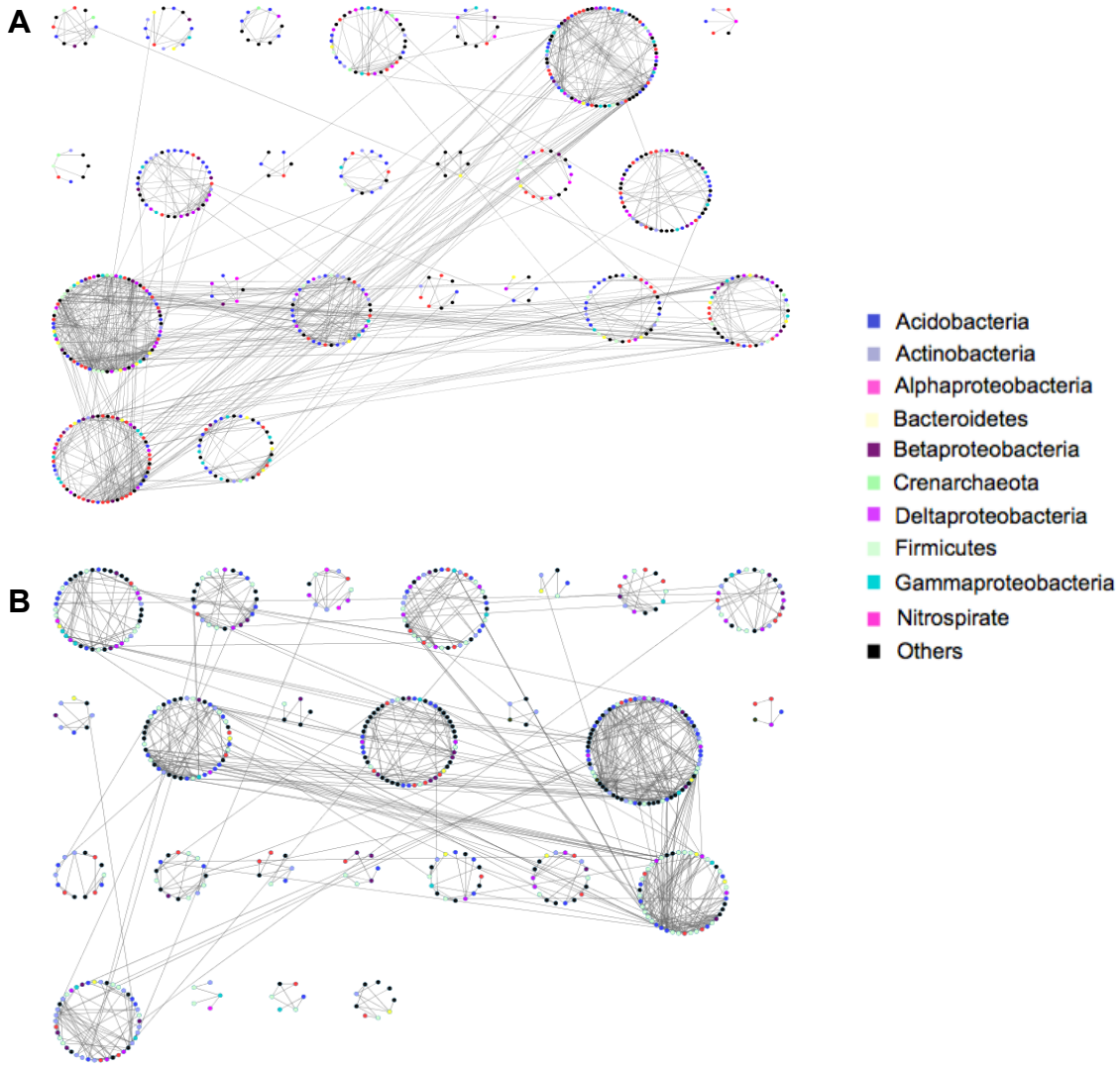


Figure 2-1. Networks of co-occurring OTUs for (*A*) forest and (*B*) pasture based on a Pearson correlation ($r \geq 0.95$) with significance of $P \leq 0.01$. An interaction (edge) between nodes (OTUs) implies a significant correlation. Only modules with at least five members are shown.

Next, we measured two important centrality measures of networks: betweenness centrality and degree, where forest network had significantly higher values than those observed for pasture ($P < 0.01$ and $P < 0.05$, respectively). While the forest network had an average betweenness centrality value of 4.79 x 10^{-5} and degree value of 2.17, the pasture network had values of 3.11 x 10^{-5} and 2.00, respectively (Figure 2-2).

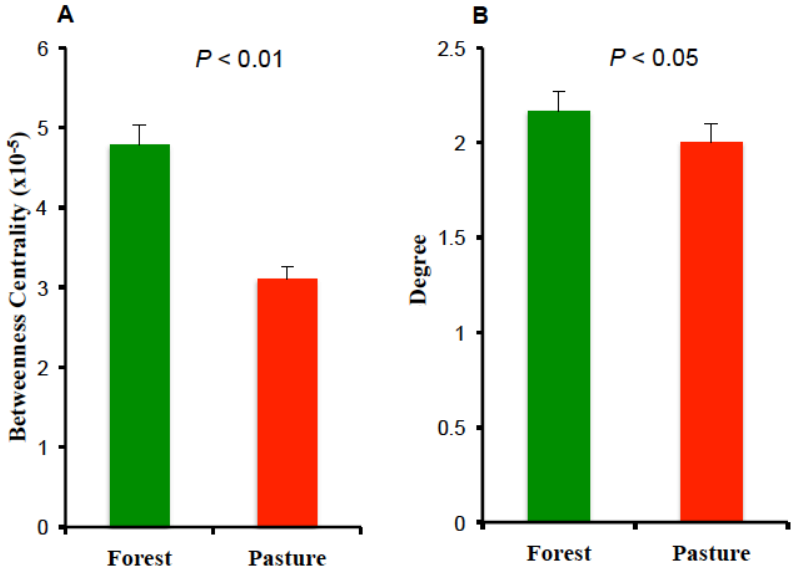


Figure 2-2. Network properties change with ecosystem conversion: (*A*) betweenness centrality and (*B*) degree of taxonomic networks. Error bars represent standard error.

2.5.3 Network memberships

In addition to changes in network structure, the distribution of major microbial phyla between the two networks was not preserved. We estimated that microbial communities comprising the networks were significantly different between forest and pasture. These differences were true at the phylum - [analysis of similarity (ANOSIM): $R = 0.81$, $P < 0.01$], genus – [ANOSIM: $R = 1.00$, $P < 0.02$, using Bray-Curtis dissimilarity values], and OTU-level [ANOSIM: $R = 1.00$, $P < 0.0$, using Bray-Curtis dissimilarity

values; R = 0.97, P < 0.02, using weighted-Unifrac distance]. The largest percentage variations of network phyla/class in response to forest-to-pasture conversion were observed for *Alphaproteobacteria* and *Firmicutes* (Figure 2-3A). The former taxa decreased from a mean value of 25.41% [±1.19; α=0.05 confidence intervals (CI)] to 11.69% (±2.51; α=0.05 CI), and latter taxa increased from 4.16% (±2.33; α=0.05 CI) to 29.88% (±14.36; α=0.05 CI).

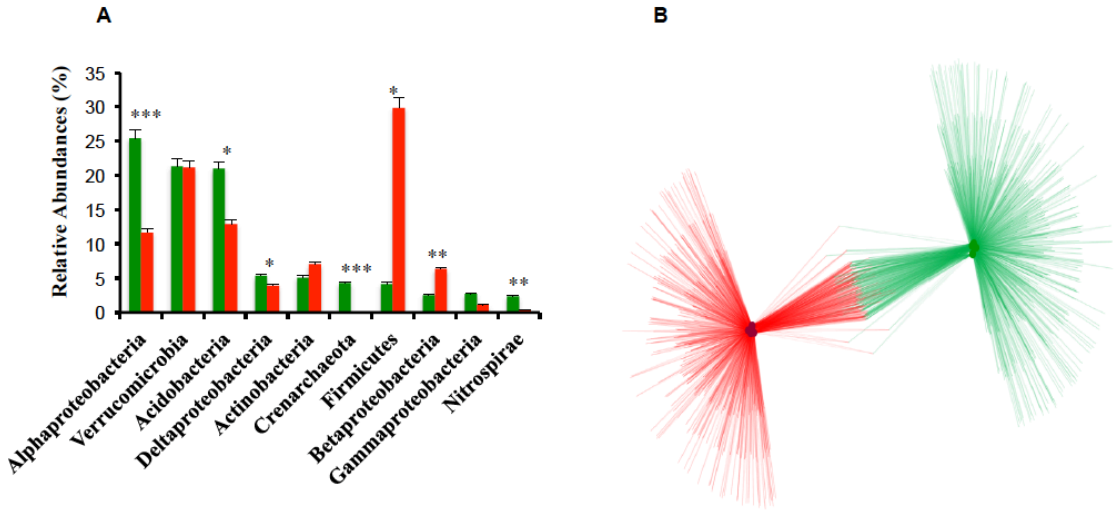


Figure 2-3. (*A*) Relative abundance of major microbial phyla and classes in forest (green) and (red) pasture networks. Error bars represent standard error. Symbols (*), (**), and (***) indicate significance values of P < 0.05, P < 0.01, P < 0.001, respectively. (*B*) Bipartite network of OTUs depicting inter-relationships between members comprising forest (green) and pasture (red) networks.

Three other major groups, *Acidobacteria* (P < 0.05), *Deltaproteobacteria* (P < 0.05), and *Nitrospirae* (P < 0.01) showed significant decreases with forest-to-pasture conversion from 21.01% (±4.46; α=0.05 CI) to 12.90% (±5.53; α=0.05 CI), from 5.31%

(±0.54; α=0.05 CI) to 3.88% (±1.02; α=0.05 CI), and 2.33% (±0.89; α=0.05 CI) to 0.36% (±0.2; α=0.05 CI), respectively, while the class *Betaproteobacteria* showed a significant increase ($P < 0.01$) from 2.55% (±1.66; α=0.05 CI) to 6.33% (±0.98; α=0.05 CI). Intriguingly, *Crenarchaeota* comprised 4.3% (±0.99; α=0.05 CI) of the forest network, but none of the OTUs in the pasture network were assigned to this archaeal phylum (Figure 2-3A). We constructed a bipartite network by combining individual forest and pasture networks to identify similarity patterns between community members detected in both datasets (Figure 2-3B). Microbial communities were clustered by land use with a Jaccard similarity index of 0.15. We estimated that only 145 OTU-nodes were shared between these two taxonomic networks, of which 55 were statistically significant between forest and pasture ($P < 0.05$).

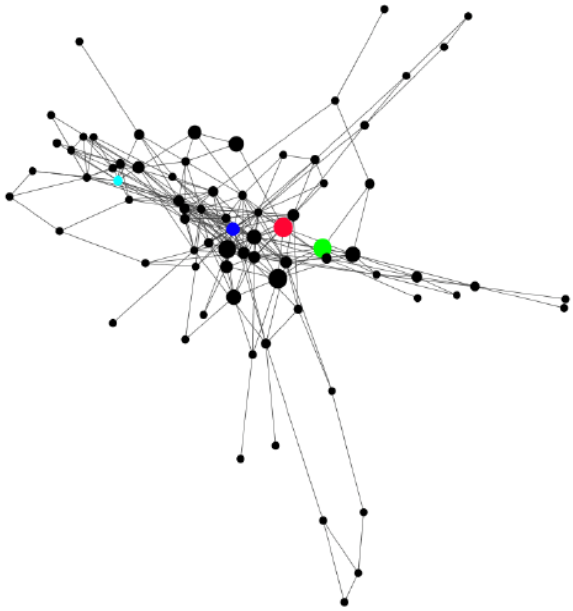


Figure 2-4. Network representation of the largest forest module. A total of 87 nodes are scaled according to their betweenness centrality values, indicating their importance in the network. Red node = *Candidatus* Nitrososphaera, green node = *Opitutus,* blue node = *Paenibacillus*, and cyan node = *Rhodoplanes.*

The network evenness for the forest and pasture were 0.71 and 0.70 respectively (Supplementary Figure 2-1). This represents skewed microbial communities in which effectively only a few members predominate. After binning OTUs into genera, 49.11%

and 38.4% of all forest and pasture sequences, respectively, were assigned to the top 10 genera in each system (Supplementary Figure 2-2). Among these abundant genera, the largest percentage decrease for genera in response to the ecosystem conversion was observed with *Rhodoplanes* [18.07% (±0.56; α=0.05 CI) in forest and 7.27% (±1.78; α=0.05 CI) in pasture, $P < 0.001$], which was counterbalanced by two endospore-forming genera, *Sporosarcina* [1.16% (±0.66; α=0.05 CI) in forest and 11.28% (±4.95; α=0.05 CI) in pasture, $P < 0.01$] and *Bacillus* [0.30% (±0.17; α=0.05 CI) in forest and 2.73% (±1.30; α=0.05 CI) in pasture, $P < 0.05$]. More noticeably, *Candidatus* Nitrososphaera ($P < 0.001$) and *Nitrospira* ($P < 0.01$) contributed 4.11% (±1.08; α=0.05 CI) and 1.38% (±0.44; α=0.05 CI) in the forest network, respectively, whereas none of the OTUs in pasture network belong to these genera. *Rhodoplanes* and *Candidatus* Nitrososphaera have not only high frequencies in forest network, but also high relative importance. Betweenness centrality values for these genera were estimated to be 7.08×10^{-4} and 1.05×10^{-4}, respectively, in the largest module of forest network (Figure 2-4), which were much higher than the average value (Figure 2-2A). While most of the OTUs in this module were not identified to the genus-level, members of two other genera, *Opitutus* and *Paenibacillus*, had high betweenness centrality values, 6.7×10^{-4} and 3.35×10^{-4}, respectively.

2.5.4 Procrustes analysis

The analyses showed that two different OTU-picking algorithms, *de novo* and closed-reference OTU lineages, produced similar clustering patterns (Supplementary Figure 2-3A). Clustering based on PICRUSt predicted KO gene profiles also resulted in patterns similar to closed-reference OTU clustering (Supplementary Figure 2-3B). The

coordinates in both cases explained most of the variances, estimated to be 78.9% and 90%, respectively. Associations between taxonomic and predictive functional profiles were robust with goodness of fit values of M^2 = 0.307 and 0.009, respectively, and statistically significant ($P < 0.01$).

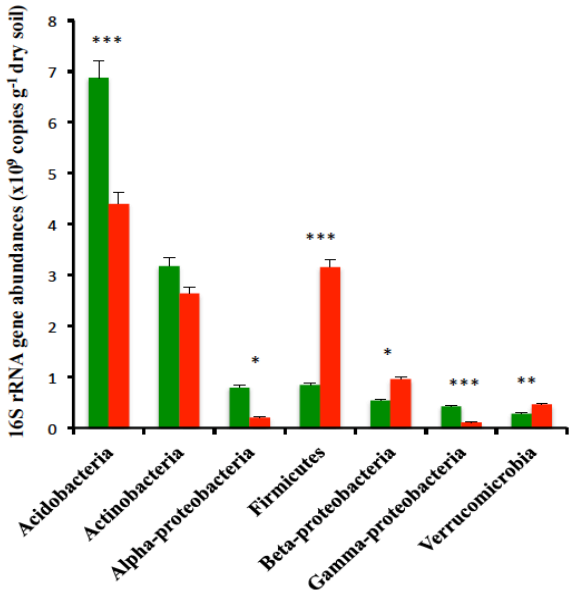


Figure 2-5. Copy numbers of seven major bacterial taxa per gram of dry soil determined by quantitative PCR in two different land uses in the Amazon: forest (green) and pasture (red). Error bars represent the standard error. Symbols (*), (**), and (***) indicate significance values of $P < 0.05$, $P < 0.01$, $P < 0.001$, respectively.

2.5.5 Quantitative PCR (qPCR)

We selected five dominant phyla to be quantified by qPCR in Amazon soils. Cycle threshold (C_t) values were obtained for each specific primer set using a representative 16S rRNA gene sequence per phylum or class. The C_t values were used to calculate the actual copy numbers of each bacteria phylum or class from standard curves (Supplementary Table 2-2). All seven qPCR assays showed R^2 values above 0.97 with reaction efficiencies ranging between 1.79 and 2.35. The 16S rRNA gene copy numbers of *Acidobacteria* ($P > 0.001$), *Alphaproteobacteria* ($P > 0.05$), and *Gammaproteobacteria* ($P > 0.001$) were significantly higher for forest samples, while those numbers observed for *Firmicutes* ($P > 0.001$), *Betaproteobacteria* ($P > 0.05$), and

Verrucomicrobia ($P > 0.01$) were found to be higher for pasture samples. The 16S rRNA gene copy numbers for members of the phylum *Actinobacteria* were not significantly different in both land uses (Figure 2-5).

2.6 Discussion

Soil microbial communities are known to have diverse compositions and varied abundances, even when analyzed at the small spatial scales that matter to microorganisms (Martiny et al. 2006, Fierer and Jackson 2006). These complex microbial patterns are not entirely random distributions, but the outcome of multiple ecological interactions (Lidicker 1979). In this study, we used a co-occurrence-based mathematical analysis to represent interactions/coupling across multiple microbial species, and identify alterations in potential relationships with forest-to-pasture conversion in the Amazon rainforest, the largest tropical ecosystem in the world.

In order to compare our resulting networks, we first confirmed that they were of biological origin (as opposed to randomly self-organizing ones) and evaluated three main properties of microbial network associations (Faust and Raes 2012). Both forest and pasture networks were found to be: (*i*) modular in structure with modularity values of at least 0.4 (Newman 2006); (*ii*) scale-free as their connectivity followed a power-law model; and (*iii*) small-world as their average path lengths were short and nodes highly clustered (Watts and Strogatz 1998). All of the above properties suggest these co-occurrence patterns are of non-random organization.

Our results indicate that key features of forest and pasture networks were altered with conversion. The significant decrease ($P \leq 0.01$) in number of nodes (OTUs) and

edges (OTU-OTU relationships) signals microbial diversity losses in response to ecosystem conversion. This finding is in agreement with previous studies, in which land use conversion resulted in biotic homogenization (Rodrigues et al. 2013, Ranjan et al. 2015). In comparison to forest, the lower connectivity of pasture network has two non-mutually exclusive explanations. First, it implies that land use conversion increased the potential for negative interactions (e.g. competitive exclusion and parasitism/predation) with loss of taxa. Resource heterogeneity and spatial isolation have been hypothesized as key mechanisms of control on soil microbial diversity (Zhou et al. 2002). It is reasonable to assume that conversion of tropical forests to pastures containing only two grass species will reduce the variety of resources, even if the total amount of resources, specifically C-based compounds, increases. There is evidence that this is occurring in our study system as we detected decreases in community dissimilarity at taxonomic, phylogenetic, and functional gene levels (Hamaoui et al. 2016, Mirza et al. 2014, Navarrete et al. 2015, Ranjan et al. 2015) with increased values of total C in pastures (Cenciani et al. 2009). A second and alternative explanation is the effect of positive interactions (e.g. mutualism and commensalism) on driving microbial co-occurrences in the forest. It is well established that certain microorganisms are functionally complementary to each other, such as microbial consortia for anaerobic methane oxidation, thermodynamically interdependent degradation, and nutritional exchange, among others (Morris et al 2013). Positive interactions through interspecific facilitation have been reported for plants (Valiente-Banuet et al. 2007, Elias and Dias 2009) and animals (Cardinale et al. 2002), but have yet to be empirically tested in soil microbial communities. Our experimental approach was not designed to provide direct evidence of beneficial interactions,

nevertheless, the co-occurrence patterns described in our study provide guidance for isolation efforts for poorly characterized microbial species that share the same or complementary physiological traits with known species, and increase our limited understanding of interactions along processes of community assembly involving competition and habitat filtering.

Given the potential taxonomic and phylogenetic changes with forest-to-pasture conversion, we asked which groups were more susceptible to alterations. Our results indicate that the relative abundances of *Alphaproteobacteria*, *Deltaproteobacteria*, *Acidobacteria*, and *Nitrospirae* were significantly decreased, while *Firmicutes* and *Betaproteobacteria* increased. We found direct support for these results with the use of qPCR (Figure 2-5). In a previous study, a co-occurrence network analysis for 151 soil samples has been used to classify microbial taxa into two ecological strategies, namely generalists and specialists (Barberan et al. 2012). Our findings agree with these two very general categories as some groups were found in both land uses and others were limited to one land use. For example, the phyla *Crenarchaeota* and *Nitrospirae* were found only in forest soils. OTUs associated with members of these phyla have been linked to specialized functions of ammonia and nitrite oxidations in soils (Leininger et al. 2006, Spang et al. 2002, Stahl et al. 2012), which are in line with our recent studies (Paula et al. 2014, Hamaoui et al. 2016). The local abundances of these groups may be directly associated with their resource/functional specialization as theoretical (Buchi and Builleumier 2014) and experimental (MacLean et al. 2004) studies have shown for other groups. There is, however, often a fitness cost associated with specialization, which is the diminished range of potential niches for colonization as the primary environment

becomes disturbed. Consistent with this idea, we observed that land use change in the Amazon rainforest increased the relative abundance of generalist fungi (Mueller et al. 2016). We argue that similar responses occur for *Bacteria* and *Archaea*, but these patterns have yet to be identified on a group-by-group basis, as not all phylogenetic groups are expected to respond the same. There is evidence that even within groups previously characterized as generalists, *Acidobacteria* and *Verrucomicrobia* (Barberan et al. 2012), the biogeographical patterns and relative abundances change among subgroups (Navarrete et al. 2015; Ranjan et al. 2015), albeit the overall response of phylum variation in percent remains the same.

Co-occurrence patterns do not allow mapping of microbial interactions directly, but provide information on particular groups sharing habitats or performing similar ecological functions (Freilich et al. 2010). We asked whether taxa observed in our study would be associated with specific functions. In order to validate our results, we performed two Procrustes analyses (*de novo* OTU table vs. closed-reference OTU table and closed-reference OTU table vs. KO annotation using PICRUSt) using the nearest-neighbor model (Muegge et al. 2011), and confirmed significant relationships between taxonomic and predictive functional profiling ($P < 0.01$, Supplementary Figure 2-3). Our results indicate that dominant members of the forest network include potential contributors to the processes of nitrification and denitrification (Supplementary Figure 2-2). Nitrification, the metabolic process by which ammonia (NH_3) is oxidized to nitrate (NO_3^-), is mediated jointly by an ammonia oxidizer and a nitrite oxidizer (Martens-Habbena et al. 2009). Forest networks were comprised of a substantial portion of these microorganisms with *Candidatus* Nitrososphaera being characterized as an ammonia

oxidizer, while *Nitrospira* is known to be a nitrite oxidizer but sometimes is capable of ammonia oxidation as well (Daims et al. 2015, Spang et al. 2002, Stahl et al. 2012, van Kessel et al. 2015). This is in agreement with our recently published findings that only thaumarchaeal sequences were retrieved for the gene *amoA*, encoding the α subunit of the ammonia monooxygenase enzyme and consistent with our failed attempts to amplify *amoA* gene sequences associated with bacterial ammonia oxidizing microorganisms, despite using seven different primer combinations and ten amplification conditions (Hamaoui et al. 2016). Moreover, a network node identified in the largest forest module and estimated to have a very high betweenness centrality value was classified as *Candidatus* Nitrososphaera (Figure 2-4).

There are two important implications associated with our findings. At the compositional level, the network theory predicts that the nodes with high betweenness centrality values should be more vulnerable to disturbance (Montoya et al. 2006, Saavedra et al. 2011, Solé et al. 2001) and removal of these nodes results in network fragmentation (Widder et al. 2014). There is evidence for both loss of this archaeal genus (from 4.11% to 0%, $P < 0.001$, Supplementary Figure 2-2) and decrease in the number of connections (loss of 80 nodes and 262 edges) with forest-to-pasture conversion. At the functional level, ecological theory predicts that losses of keystone species (identified as hub nodes in the network theory) can have a large impact on ecosystem functioning (Chapin et al. 2000). There is also evidence this is happening in our study site as previous biogeochemical studies at Fazenda Nova Vida observed higher net nitrification rates in forest soils (1.32 to 3.51 μg N g^{-1} dry soil day^{-1}) in comparison to pastures (0.02 to 0.77 μg N g^{-1} dry soil day^{-1}) (Neill et al. 1995, Neill et al. 1997b). These studies have

hypothesized that a direct consequence of the change in net N mineralization is the reduction of NO and N_2O emissions from pastures (Melillo et al. 2001), but the reasons for these alterations have not been established. We asked if any nodes (taxa) present in our networks were associated with denitrification, the process by which NO_3^- is reduced to N_2 (or NO and N_2O). We detected the presence of nodes with high betweeness centrality associated for the genera *Rhodoplanes* and *Opitutus* in the largest forest network module. Members of these groups are characterized for their ability to use NO_3^- as an electron acceptor (Hiraishi et al. 1994, Chin et al. 2001) and while only the former has been experimentally shown to be a denitrifier, the latter is known to possess a complete denitrification pathway in its genome (Sanford et al. 2012). Nodes associated with the above groups were not depicted in pasture networks, suggesting their memberships and associated traits were not preserved with land use change. The lower denitrification rates measured by others in our study site (Melillo et al. 2001) can be explained, at least partially, by the loss of inter-taxa relationships represented by specific nodes and edges in co-occurrence networks.

2.7 Final remarks

The astonishing diversity of soil microorganisms is generally viewed as a consequence of their spatial isolation and ecological strategies. Lately, co-occurrence association studies draw a different picture, one where certain phylogenetic and functional groups are non-randomly associated through resource sharing and metabolic exchange. We found that forest-to-pasture conversion strongly impacts microbial co-occurrence patterns, implying long lasting effects to biogeochemical processes such as

the N transformation dynamics in the Amazon rainforest. We identify microbial taxa that may occur together despite their metabolic dissimilarity. These novel relationships require us to reinterpret microbial community assembly models taking potential positive interactions into consideration.

Chapter 3

Forest-to-pasture conversion results in the shifts of spatial co-occurrence patterns of bacterial communities in the Amazon soils

3.1 Abstract

A central goal in ecology is to explore the interactions between organisms and their response to any treatment, such as land use change. Yet, this is poorly understood in microbial ecology. Here we report a new strategy to construct co-occurrence networks, and describe a suite of approaches for analyzing topological measures to infer the ecological relationships of bacterial communities in two landscape-scale ecosystems, Amazon forest and pasture. Our results revealed that ecosystem-wide topological patterns, and topological roles of major bacteria taxa differed significantly between forest and pasture, indicating substantial alteration of interaction relationships between different taxa. We estimated relationship pattern, which may discern the life-style strategies and metabolic synergisms of microorganism critical for adaptation. For example, higher co-occurrence of *Alphaproteobacteria* and *Nitrospirae*, known to have members involved in nitrogen metabolism, in forest soils may indicate their higher degree of association in the nitrogen cycle. Identifying broadly distributed species, namely generalists, further increases our understanding of microbial ecology in terrestrial ecosystems. Forest-to-pasture conversion decreased generalist OTUs that are known to be involved in carbon and nitrogen metabolism, which may explain, at least partially, the higher content of these compounds in the pasture soils. We conducted a multivariate analysis to correlate network topological measures with environmental factors, which showed that C/N content and factors related to soil pH primarily drive the variation of network features.

Given that the microbial community is closely associated with ecosystem functioning, our novel approach demonstrates microbial interaction associations and allows incorporating multiple variables, which is particularly useful to infer ecological traits of unstudied microorganisms.

3.2 Importance

The interactions between microbial species are critical in important ecosystem functions. Yet microbial co-occurrences and interaction strategies between them are poorly understood, which challenges us in exploring the mechanism of microbial adaptation in their natural habitat, and the alteration of microbial life-style strategy in response to an anthropogenic activity. Here we report a novel strategy to construct co-occurrence networks, and analyze the shifts in topological patterns of networks at spatial scale due to land use change in the Amazon. This analytical strategy allows us to examine co-occurrence pattern of microbial consortium that is selected (or adapted) by environmental factors, and subsets of that consortium that are selected (or separated) by habitat properties in each sampling site. Therefore an ecosystem-wide study allows us to examine the variation of co-occurrence pattern of microbes across different sites, and a comparative study allows us to examine the impact of a treatment on co-occurrence patterns.

3.3 Introduction

Soil microbial communities are known to have diverse compositions and varied abundances, even when analyzed at small spatial scales (Martiny et al. 2006). Several

deterministic processes such as habitat filtering, and niche differentiation drive the co-occurrence of microorganisms and help maintaining microbial diversity in nature (Costello et al. 2009, Ofiţeru et al. 2010, Langenheder and Székely 2011, Stegen et al. 2012). The abundance patterns of microbial species in any ecosystem are driven by a complex web of interactions between microbial and macrobial components, and abiotic components of the ecosystem (Faust and Raes 2012, Papke and Ward 2004, Horner-Devine et al. 2004, Fierer and Jackson 2006). Therefore a similar abundance pattern of two microbial species does not necessarily mean they have metabolic interactions; it may be a result of their similar response to environmental properties. Also, a similar abundance pattern of two species may be a result of interaction(s) with other species, or both biotic and abiotic components of the ecosystem. Since the abundance profile of microbial species is the only variable used in constituting the co-occurrence networks, we may decipher neither direct interactions between microbial species nor the factors shaping the microbial frequencies from network survey. An integrated approach synthesizing additional information (e.g. spatial variation, taxonomy, metabolism), however, can link the abiotic factors associated with the network co-occurrence pattern (Zhou et al. 2011, Ding et al. 2014) or network topological features (Ma et al. 2016) in order to discern the impact of environmental factors on the interactions between microorganisms. The integrated approach might also allow inferences about biotic factors, such as likely symbioses or cross-feeding interactions.

Microbial interactions may contribute to ecosystem functioning more than their diversity (Zhou et al. 2011). Therefore exploring the interactions provides insight into soil-associated function and may help reveal the role of specific microorganisms in the

ecosystem. Due to the scarcity of known biology of diverse microbial communities in soil, we rely on abundance patterns of microbial species to comprehend the complex partnerships of microorganisms. Based on the relative abundance of microbial species across ecosystem, a correlation test is performed to calculate strongly related microbial pairs in community, which constitutes a co-occurrence network (Faust et al. 2012, Zhou et al. 2011, Barberan et al. 2012, Widder et al. 2014). Since spatial variability has a profound influence on the occurrence and abundance pattern of each microbial species (Fierer and Jackson 2006, Krave et al. 2002, Wallenstein et al. 2007, Girvan et al. 2003), the traditional approach may be incapable of capturing valid pairs of microorganisms, which have interactions and/or similar response to environmental factors, in the following situations: (*i*) if a pair of microbes have metabolic interactions but have differential responses to other factors; (*ii*) if a pair of microbes have metabolic interactions when they co-occur but are detected sporadically across different soil samples. Both situations would cause the dissimilar abundance patterns of this pair across ecosystem, especially with high beta diversity, which the traditional approach may neglect to calculate as a valid pair. Likewise, a pair of microbes can be calculated as valid that are neither interacting nor have a similar response to environmental factors. Hence, this approach may be less conducive in calculating valid pairs of microorganisms. Here we introduce a novel approach that takes these situations into consideration. Unlike the traditional approach, site-wise co-occurrence networks are constructed based on the relative abundances of species detected in each site.

In this study, we constructed site-wise networks to demonstrate relationships of multiple microbial species within sites of two contrasting ecosystems, Amazon forest and

pasture. We used over 13,000 bacterial operational taxonomic units from a total of 69 sampling sites of these two ecosystems to deduce the variations of microbial assemblage patterns in response to the ecosystem conversions. Using topology-based analyses, we aimed (*i*) to calculate the impact of spatial variability on the co-occurrence patterns within and between ecosystems; (*ii*) to identify the network constituents (and co-occurring pairs) whose occurrences are consistent to the spatial variability and ecosystem conversion; (*iii*) to identify the environmental factors that are associated with the variation of network properties.

3.4 Materials and methods

3.4.1 Sampling strategy

Soil samples were collected from three one-hectare plots spaced at 1-km and 10-km distances of Amazon forest and pasture in 2009. Each of 100-m^2 quadrat with 10-m^2, 1-m^2, 0.1-m^2, and 0.01-m^2 quadrats nested within was established, which provided a total of 36 replications per soil type. This sampling scheme was designed to offer sufficient replications such that statistically rigorous techniques could be used to profile microbial diversity and determine its response to ecosystem conversion. Site location, soil abiotic factors, and sequencing procedures were described in details elsewhere (Rodrigues et al. 2013).

3.4.2 16S rRNA gene sequencing

DNA was extracted from soil samples using PowerSoil DNA Isolation kit (MoBio Laboratories) as per manufacturer's instructions. Prior to the sequencing, the purity of the extracted DNA samples was checked with the Nanodrop ND-1000 spectrophotometer,

which quantified the ratio of A260 and A280. The average ratios were 1.78 and 1.84 for forest and pasture, respectively. For PCR, a primer set, 577F (5′-AYTGGGYDTAAAGNG-3′) and 926R (5′-CCGTCAATTCMTTTRAGT-3′), was used targeting the hypervariable V4 region of the 16S rRNA gene. This hypervariable region therefore offers comprehensive coverage and evolutionary relatedness across microbial lineages (Sul et al. 2011). The amplicons were subjected to high-throughput sequencing with the 454 GS FLX Sequencer (454 Life Sciences) at the Michigan State University Research Technology Support Facility.

3.4.3 Sequence processing and taxonomy assignment

All upstream processing of raw sequence datasets, including quality filtering was performed using the Mothur software (Schloss et al. 2009), as suggested by Huse et al. (2007). Results of three pasture samples were discarded due to low number of sequencing reads (<1,000). Remaining sample reads were assigned to Operational Taxonomic Units (OTUs) at a minimum of 97% sequence identity using a *de novo* OTU-picking protocol with QIIME 1.8.0 (Caporaso et al. 2010). The algorithm Uclust (Edgar, 2010) was used to cluster quality-filtered reads against the GreenGenes database (DeSantis et al. 2006), and assignment of taxonomy was performed with the RDP classifier (Wang et al. 2007). For all downstream processes, samples were rarefied to 1,000 randomly selected sequence reads for 10 times with replacement per sample to avoid sample heterogeneity and site-wise networks were constructed using these subsets of 1,000 sequences as replicates.

3.4.4 Network construction

To construct co-occurrence networks, we calculated Spearman correlation coefficients between all possible OTU pairs within each site. The analysis was performed in R 3.2.1 (R Core Team, 2013) using "multtest" package (Pollard et al. 2005). We adjusted the *P*-values using Benjamini and Hochberg false discovery rate (FDR; Benjamini et al. 2006), where the cutoff was $P < 0.01$ for statistical significance. Non-random co-occurrence pattern of networks were tested with degree distribution following the power-law model, where most of the network members (i.e., nodes) have lower number of significant correlations to other nodes (i.e., links) and few nodes will have high links. The threshold of correlation coefficient, which ranged from 0.79 to 0.83, was therefore estimated when degree distribution follows a power-law. Ten networks of each ecosystem were removed from the analysis as they failed to follow the power-law degree distribution for any of the correlation coefficient values.

3.4.5 Topological measures and analysis

The following topological features were used to describe network properties: (*i*) degree, which is the average number of edges connecting nodes; (*ii*) average path length, which is the average number of edges connecting any two nodes; (*iii*) modularity, which refers to the ability of a network to become organized into individually separated and distinct units (i.e., modules); and (*iv*) betweenness centrality, which reflects the number of times a node plays a role as a connector along the shortest path between two other nodes. Degree is an important measure and provides a local quantity of nodes, however fails to provide information about the relative importance of nodes in network structure. The relative importance of a node can be measured by its betweenness centrality value.

Nodes with higher betweenness values are likely to be situated in the core of the network and are considered having higher influences to the relationships with other nodes, while those with lower values are expected to have a more peripheral location and lower influences in the network. These topological features describing individual network nodes such as betweenness centrality and degrees, and overall network structures such as average path length and modularity were calculated with Gephi (Bastian et al. 2009).

In addition to the power-law distribution of networks, ecological networks also exhibit several other properties, such as low average path length, also known as small world, meaning there tends to be a path between any two nodes that involves only a few links (Barabási et al. 2004), and modularity, where >0.4 is considered as modular structure (Newman 2006).

3.4.6 Statistical analysis

The Pearson correlation test was used to examine the correlation between OTU counts and their degrees in a network to check the non-random pattern of networks. Two principle coordinate analyses (PCoA) were performed to calculate the variations of network features across ecosystems: one addressing topology as measured by betweenness centrality of each node (i.e. OTU), and another addressing presence/absence of each node-node pair (i.e., OTU-OTU pair), where a value of one was used for the presence of each pair in a network, while zero was used for the absence of it. The former index was used to examine the variation of spatial topological pattern and the latter index was used to evaluate the variation of potentially interacting taxa pairs across ecosystems. In both cases, the datasets were normalized, and Bray-Curtis distance was used as distance metric. Analyses of similarities (ANOSIM) tests of PCoA's were performed to

assess whether forest networks are significantly different from those of pasture. Nonparametric two-sample *t*-tests with 999 Monte Carlo permutations were carried out to examine whether the relative importance of different bacterial phyla, as measured by their betweenness values, varied between ecosystems. All of these tests were performed in the QIIME 1.8.0 platform.

A Redundancy Analysis (RDA) was performed to analyze the impact of spatial variability on the network topology. RDA is a regression-based method that outlines the variation within a set of response variables that is explained by a set of explanatory variables. In this study, environmental factors were used as explanatory variables, and betweenness centrality value of each node in networks was used as quantitative response variable that was then categorized into soil types and bacteria taxa. Thus RDA ordination plot represents the impact of environmental parameters on the network topological feature and relative influence as measured by betweenness of individual bacterial taxa. A permutation test under the reduced model was implemented to check whether the pattern that is shown by RDA appears random or not. This analysis was performed in R 3.2.0 using package "vegan" (Dixon 2003).

3.5 Results

3.5.1 Datasets

We obtained 454 GS FLX Sequencer-derived 16S rRNA gene sequence datasets from 69 soil samples, of which 36 belonged to Amazon forest and 33 belonged to 38-year old pasture soils. Upstream processing removed those sequences whose lengths were outside of the main distribution, and those with any ambiguous base calls and not having

identical primer sequences. The depths of quality-trimmed sequence reads ranged from 1,242 to 15,415 in forest and 1,386 to 11,814 in pasture soil samples with an average length of 350 bases. Following taxonomy assignment, we observed 7,547 OTUs in forest and 9,021 in pasture soils, totaling 13,359 unique OTUs in both ecosystems. A rarefaction procedure was conducted through random sampling for ten times at the depth of 1,000 sequences per sample, which were then used for the construction of site-wise networks.

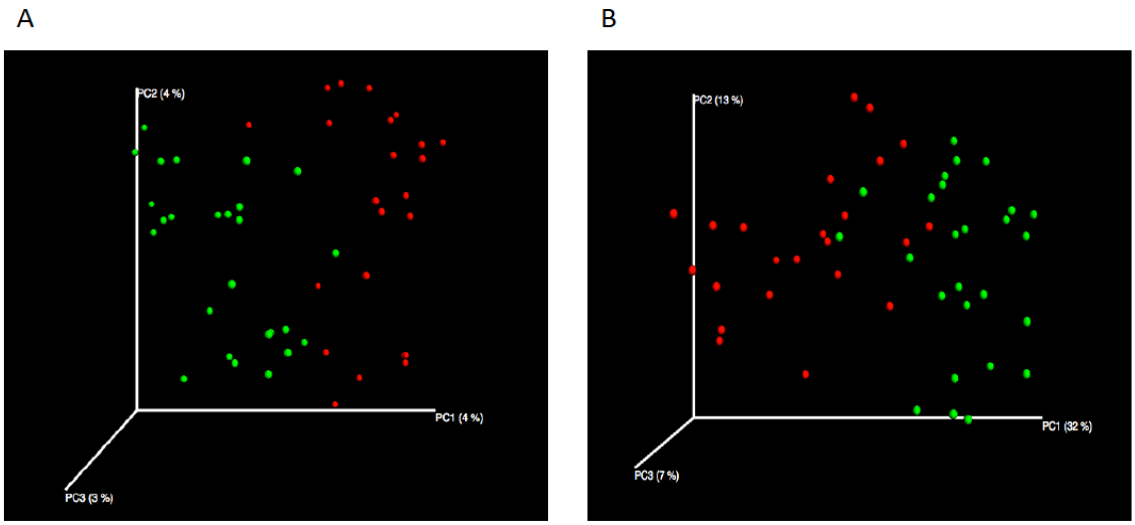


Figure 3-1. Spatial distribution patterns of forest and pasture networks. (*A*) Principle coordinate analysis showing the response of topological variations to ecosystem conversion (betweenness centrality). (*B*) Principle coordinate analysis showing the response of occurrence of bacterial pairs (OTU-OTU) at phylum-level to ecosystem conversion (betweenness centrality). Bray-Curtis distance was used to measure the ecological distance in both cases. Green dots, forest networks; red dots, pasture networks.

3.5.2 Network description and topological pattern across ecosystems

Prior to describing individual networks, first we confirmed the non-random nature of networks, which were identified by the power-law distribution of degrees (Supplementary Table 3-1). Following the exclusion of the OTUs that did not have a strong correlation to any OTU, we observed that the number of nodes ranged from 253 to 1,119 in forest, and 455 to 1,158 in pasture networks. Two other properties calculated were modularity and average path length. The range of former index values were estimated from 0.79 to 0.98 in forest and 0.74 to 0.96 in pasture, while for latter index, forest had a range between 1.56 and 16.54 and pasture had 5.97 and 14.15.

We analyzed the pattern of co-occurrence networks using PCoA plots in order to comprehend variation in network features across forest and pasture ecosystems. Clusters specific to these ecosystems were evident for betweenness centrality (Figure 3-1A; ANOSIM: $R = 0.27$, $P < 0.01$), and for proportional co-occurrence of microbial pairs (Figure 3-1B; ANOSIM: $R = 0.48$, $P < 0.01$). We also evaluated the impact of land use change on the relative influences of major bacterial taxa as measured by their relative betweenness centrality values (Figure 3-2). The largest proportional decrease in response to forest-to-pasture conversion was by *Acidobacteria* ($P < 0.001$), which decreased from an average of 19.07% (±0.009; α=0.05 confidence interval (CI) to 12.77% (±0.022; α=0.05 CI) followed by the phyla *Alphaproteobacteria* ($P < 0.01$) and *Gemmatimonadetes* ($P < 0.001$). In contrast, the largest proportional increase was by *Firmicutes* ($P < 0.001$), which increased from an average of 2.50% (±0.022; α=0.05 CI) to 10.57% (±0.014; α=0.05 CI) followed by the phyla *Planctomycetes* ($P > 0.05$) and *Chloroflexi* ($P < 0.05$).

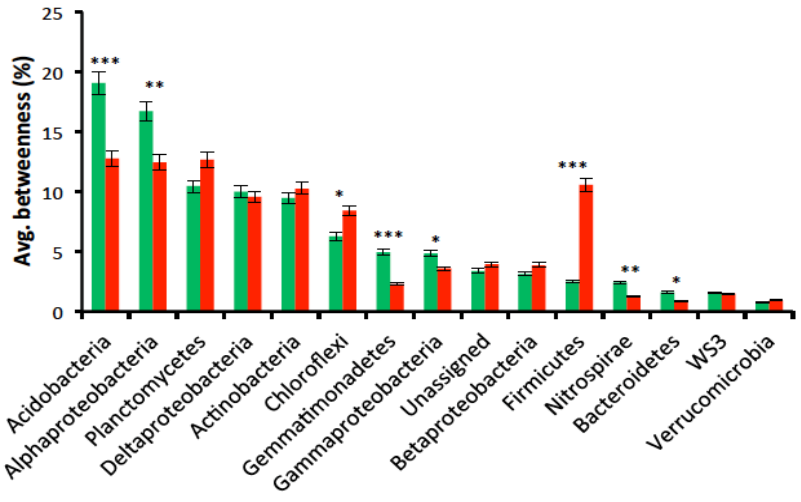


Figure 3-2. Relative influences (measured as betweenness) of different bacterial taxa comprising forest (green) and pasture (red) networks. Error bars represent standard errors. Symbols (*), (**), and (***) indicate significance values of $P < 0.05$, $P < 0.01$, $P < 0.001$, respectively, which were calculated using two-sample *t*-test.

3.5.3 Relationship between network-level topological features and environmental properties

Several important environmental factors varied between forest and pasture soil (Supplementary Table 3-2). We calculated that forest soil has significantly higher base saturation ($P < 0.05$), and lower total carbon ($P < 0.001$), moisture ($P < 0.01$), nitrogen ($P < 0.001$) and positive hydrogen ion ($P < 0.01$) contents in comparison to pasture soils. We correlated the soil properties with network topology as measured by betweenness centrality by redundancy analysis (RDA; Figure 3-3). The significance test of the overall RDA model showed $P < 0.01$ (number of permutation: 999). While the distribution of forest and pasture networks had no obvious pattern, networks on the RDA plot were well separated by soil type. On the other hand, the distribution pattern of environmental

factors seems to distinguish soils that are more acidic and dominated by Al^{+3} minerals from those that are more basic and have more K^{+}, Ca^{+2} and Mg^{+2}. Along with these factors, the analysis showed that soil H^{+} and C/N were strongly associated with spatial network patterns. While forest networks were negatively correlated with C/N and H^{+}, pasture networks had positive correlations with them. RDA was also conducted to demonstrate the link between relative influences of individual bacterial taxa and environmental factors. Our result showed that *Acidobacteria* is correlated with decreasing H^{+}, while *Firmicutes* showed the reverse pattern. Two proteobacterial classes *Alpha- and Gammaproteobacteria* were negatively correlated with moisture, C and N content of soils, which on the other hand, are in positive correlation with the phyla *Actinobacteria* and *Planctomycetes.* There is a trend for *Deltaproteobacteria* to be negatively correlated with the concentration of different metal ions (K^{+}, Mg^{+2} and Ca^{+2}), base saturation (V), S, and P, and to be positively correlated with the concentration of Al^{+3}.

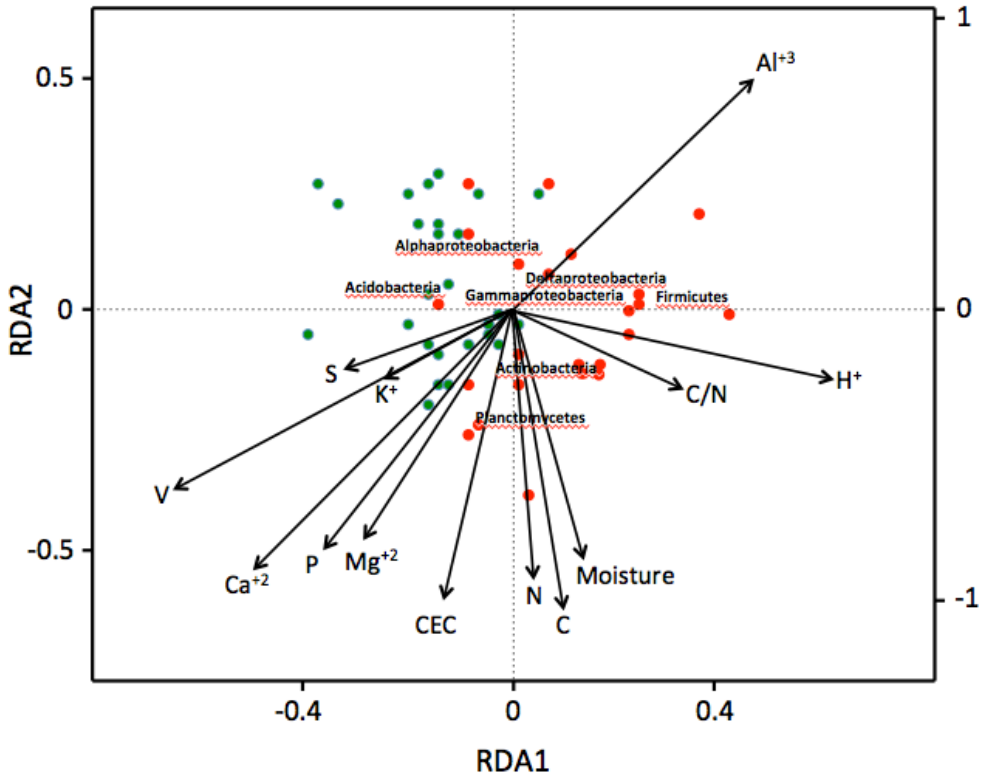


Figure 3-3. Redundancy analysis shows the influence of environmental factors, known as explanatory variables, on the network topology and relative influence of individual bacterial taxa, known as response variables, as measured by betweenness centrality. The taxa that were correlated with the environmental factors were only shown on the plot. Environmental factors are indicated with arrows. The length of arrows corresponds to the variance that can be explained by the environmental factors and the direction of arrows points to an increasing magnitude of these factors. The perpendicular distance between explanatory and response variables indicates their correlations, where a smaller distance reflects stronger correlation, and vice versa. Al^{+3}, aluminum; C, carbon; Ca^{+2}, calcium; K^{+}, potassium; CEC, cation exchange capacity; H^{+}, hydrogen; Mg^{+2}, magnesium; N, nitrogen; P, phosphorus; S, sulfur; V, base saturation. Green circles, forest networks; red circles, pasture networks.

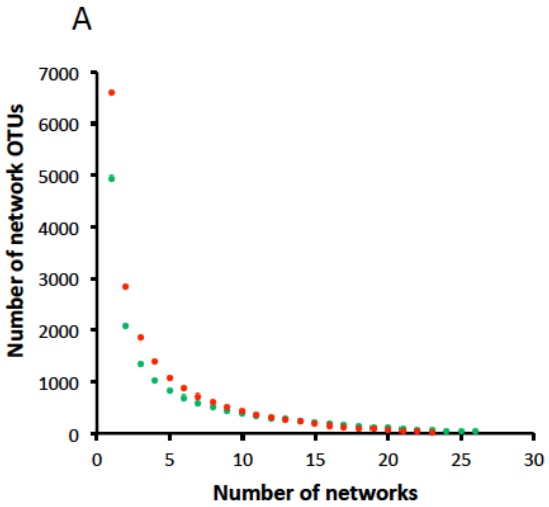


Figure 3-4. The distribution of network node counts was plotted as a function of the number of networks. Exponential decay of nodes identifies a core suite OTU-nodes, known as network generalists. In this analysis, generalists were arbitrarily defined as taxa occurring in at least 18 or 20 networks, in forest and pasture, respectively.

3.5.4 Network generalists

We attempted to explore core microbial communities shared across networks of both soil types. Using the OTUs comprising networks, the distribution of counts is plotted as a function of the number of samples in which OTUs are detected (Figure 3-4). An exponential decay was observed within each type of networks, with only few OTUs consistently comprising the networks. This analysis was used to identify network generalists, which are broadly distributed OTUs, in both soil types (Supplementary Tables 3-3 and 3-4), and across soil types (Supplementary Table 3-5). In this study, the network OTUs (i.e., nodes) were considered generalists if their presences were detected in around 75% of networks or higher, which gave 20 networks in case of forest, and 18 networks in case of pasture. Under these arbitrarily defined criteria, 2.1% of the nodes were identified as forest generalists, and 1.49% as pasture generalists.

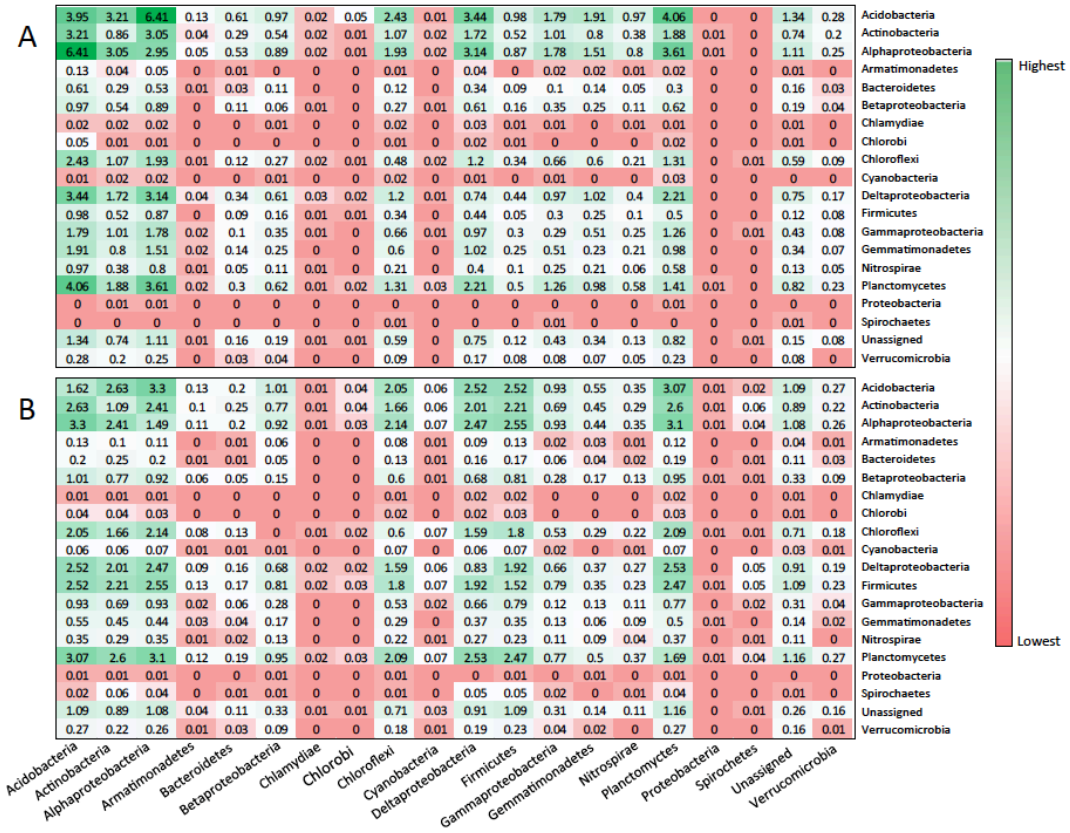


Figure 3-5. Ecological community association. (*A*) Forest networks exhibiting percentage occurrences of bacterial taxa pairs. (*B*) Pasture networks exhibiting percentage occurrences of bacterial taxa pairs.

3.5.5 Ecological community association

Using the same approach described above, we further explored the distribution pattern of OTU-OTU pairs, where we observed a very low occurrence of pairs more than once. Of more than 20,000 pairs observed in forest, we detected only 102 pairs twice and 3 pairs thrice. Of more than 20,000 pairs observed in pasture, 126 pairs were detected twice in pasture networks, with no pair detected more than this. The correlation pattern of each microbial taxon was then explored by estimating the proportional co-occurrence of taxa pairs at higher taxonomic level (Figure 3-5). We estimated that the largest percentage variations of co-occurrence frequencies between forest and pasture involved *Acidobacteria*, *Alphaproteobacteria* and *Firmicutes*. Ecosystem conversion decreased the co-occurrence frequencies of *Acidobacteria* with *Alphaproteobacteria* from a mean value of 6.41% (±0.397; α=0.05 CI) to 3.3% (±0.286; α=0.05 CI), with *Acidobacteria* itself from 3.95% (±0.821; α=0.05 CI) to 1.62% (±0.25; α=0.05 CI), and with *Gemmatimonadetes* from 1.91% (±0.195; α=0.05 CI) to 0.55 % (±0.074; α=0.05 CI), whereas it increased with *Firmicutes* from 0.98% (±0.129; α=0.05 CI) to 2.52% (±242; α=0.05 CI). In addition, we estimated a decrease of *Alphaproteobacteria* with itself from 2.95% (±0.549; α=0.05 CI) to 1.49% (±0.334; α=0.05 CI), and with *Gemmatimonadetes* from 1.51% (±0.15; α=0.05 CI) to 0.44 % (±0.058; α=0.05 CI). On the other hand, pasture ecosystem was estimated to have an increase of *Firmicutes* with *Actinobacteria*

from a mean value of 0.52% (±0.096; α=0.05 CI) to 2.21% (±0.252; α=0.05 CI), with *Alphaproteobacteria* from 0.87% (±0.145; α=0.05 CI) to 2.55% (±0.269; α=0.05 CI), with *Chloroflexi* from 0.34% (±0.063; α=0.05 CI) to 1.80% (±0.249; α=0.05 CI), with *Deltaproteobacteria* from 0.44% (±0.074; α=0.05 CI) to 1.92% (±0.225; α=0.05 CI), with itself from 0.05% (±0.038; α=0.05 CI) to 1.52% (±0.639; α=0.05 CI), and with *Planctomycetes* from 0.50% (±0.082; α=0.05 CI) to 2.47% (±0.296; α=0.05 CI). None of the other pairs has over 1% of variation in co-occurrence frequencies between forest and pasture.

3.6 Discussion

Terrestrial ecosystems harbor a wide variety of microbial species, where they adapt by interacting with a variety of biotic and abiotic variables (Faust and Raes 2012, Papke and Ward 2004, Horner-Devine et al. 2004). The environmental conditions are associated with the diversity of microbial communities (Martiny et al. 2006, Fierer and Jackson 2006, Rodrigues et al. 2013), and therefore the complex interactions between microorganisms. In this study, we used bacterial datasets to construct co-occurrence network for each sample, and analyze the spatial pattern of topological features and ecological relationships between microbial taxa.

Complex microbial networks are not entirely random distributions, but the outcome of multiple ecological interactions (Lidicker 1979). In order to compare our resulting networks, we first confirmed that they were of biological origin by evaluating three main properties of microbial network associations (Supplementary Table 3-1). Networks were retained for the subsequent analysis if they were found to be: (*i*) modular

in structure with modularity values of at least 0.4; (*ii*) scale-free as their degree distribution followed power-law model; and (*iii*) small-world as their average path lengths were short and nodes highly clustered (Watts and Strogatz 1998), however no threshold value is proposed. Strong negative correlation between degree of nodes and count of nodes for each degree indicated non-random nature of networks. Under this criterion, several networks were identified random as indicated by binomial degree distribution and were discarded. A recent study on co-occurrence networks of soil microbiota showed that archaeal degree follows a binomial distribution, while bacterial and fungal degrees follow power-law distributions (Ma et al. 2016). The assembly pattern of archaea is mainly influenced by the stochastic forces (Zheng et al. 2013), which possibly explains the random pattern of archaeal degree distribution. Although stochastic processes may play a partial role, the assembly pattern of soil bacterial community is primarily influenced by the deterministic forces (Langenheder and Székely, 2011, Faust and Raes 2012, Horner-Devine et al. 2007, Costello et al. 2009), even following the burning of forest (Ferrenberg et al., 2013), which probably explains non-random pattern of bacterial degree distribution (Horner-Devine et al. 2007).

Soils are known to possess diverse array of microbial niches, and spatial patterns, which largely depend on the aboveground floristic diversity, drives the spatial pattern of microbial community (Prober et al. 2015). Therefore ecosystem conversion in the Amazon alters the bacterial beta diversity as a function of soil environments (Rodrigues et al. 2013, Jesus et al. 2013). Using a novel strategy, we report the impact of spatial pattern of bacterial community on the spatial pattern of network topological features (Figure 3-1), which illustrates substantial variations not only between ecosystems but also

within ecosystems. While the soil pH was identified as the predictor of soil bacterial composition at the continental scale (Lauber et al. 2009), other factors such as vegetation type, carbon and nutrient contents, soil moisture dictate the composition at local or landscape scales (Fierer and Jackson 2006). The landscape scale study showed that forest bacterial communities were associated with, in addition to pH, increasing base saturation and calcium, whereas pasture communities were associated with increasing carbon, nitrogen, and moisture concentrations (Rodrigues et al. 2013). The abundance and diversity of individual taxa can also be explained by environmental factors (Navarette et al. 2015). Likewise, impact of different abiotic factors can be linked to the co-occurrence relationships (Gilbert et al. 2012, Steele et al. 2011, Fuhrman 2009), which was proven successful for studying uncultured microorganism (Duran-Pinedo et al. 2011, Faust and Raes 2012). Similar to the community composition pattern, topological features of co-occurrence networks across ecosystems are governed by environmental factors, including other biotic factors (Gilbert et al. 2012), to optimize the adaptation of community in soil habitats with minimal interspecific competition. A study with forest soils in China showed that soil variables affect the topological features of co-occurrence networks (Ma et al. 2016). In our analysis, we noticed substantial variations between relative abundances of taxa that were described previously (Rodrigues et al. 2013) and relative influences of them as measured by their betweenness centrality values (Figure 3-3). The largest percentage variations between relative abundances and relative influences were observed for *Alphaproteobacteria, Actinobacteria, Planctomycetes* and *Chloroflexi.* Forest-to-pasture conversion decreased the relative abundance of *Alphaproteobacteria* from a mean value of 25.41% to 11.69%, whereas the relative influence decreased from

16.72% to 10.43%. In contrast to the relative abundances of latter three taxa, which remained a mean of around 4%, 3% and 2%, respectively in both soil types, ecosystem conversion increased the relative influences of them from 10.43% to 12.63%, 9.47% to 10.27%, and 6.26% to 8.4%, respectively. Intriguingly, some of the taxa including *Chloroflexi, Planctomycetes, Bacteroidetes* and *Gemmatimonadetes* that were estimated to have considerable influences in both network types, while they were not detected in top 10 taxa of higher relative abundances in our previous study (Chapter 2). Conversely, *Verrucomicrobia* that was detected in higher abundance (over 21% in each of the forest and pasture networks), contributed only about 1% in each of network types (Figure 3-2). We reasoned that factors affecting the relative abundance of taxa are different to those affecting the relative influences of them in co-occurrence networks. Nonetheless, the depth of sequencing effort, and choice of universal primer sets, which may preferentially target specific bacterial taxa might play a role for these variations. This is true for, at least, *Verrucomicrobia*, which typically accounts approximately one-fourth of all bacterial sequences in soils (Bergmann et al. 2011). While this is evident in previous studies (Chapter 2, Ranjan et al. 2015), the relative abundance of *Verrucomicrobia* in the datasets of current study was estimated only 0.28% in forest and 0.79% in pasture soils.

Amazon forest soil is characterized by lower pH, moisture, nitrogen and carbon content (Supplementary Table 3-2). Although statistical significance was not observed, we estimated that forest soil has slightly lower average pH (4.3) compared to pasture (4.5), which supports previous reports (Chapter 2, Kauffman et al. 1995, Fernandes et al. 2002). Redundancy analysis in our study revealed that topological pattern of forest networks is related to decreasing C/N, while pasture networks followed the opposite

(Figure 3-3), which might have important insights into ecological traits of microorganisms (Cordovil et al. 2005, Seneviratne 2000, Manojlovic et al. 2010). Decreasing C/N may indicate higher mineralization rate due to the fact that there is an extra supply of nitrogen compounds as a result of rapid decomposition of litters in the forest soil probably to satisfy plant growth. In contrast, increasing C/N may indicate higher immobilization rate due to the fact that there is a higher carbon content, and higher ammonium as a result of rapid nitrogen fixation in the pasture soil probably to support the growth of fast growing grass species (Mirza and Rodrigues 2012). The association of individual bacterial taxa with the environmental factors may also demonstrate their ecological traits. For example, *Actinobacteria* and *Planctomyctes* are positively correlated with carbon content of the soil, indicating that these two bacteria phyla may be involved in carbon metabolism. Recently, it is known that *Actinobacteria* decompose polysaccharides or phenolic compounds in dead plant biomass (Větrovský et al. 2014), and *Planctomycetes* degradea number of plant cell wall sugars, namely L-fucose and L-rhamnose (Erbilgin et al. 2014). The RDA plot also shows that *Alphaproteobacteria* are correlated with decreasing nitrogen content of the soil, which may indicate that nitrogen deficiency is in proportion to the increasing influence of this taxon. The taxonomic order *Rhizobiales*, which comprises a major portion of the known soil-dwelling nitrogen fixers (Brown et al., 2012), belongs to *Alphaproteobacteria*. Intriguingly, the carbon content is negatively correlated with the relative influence of *Alphaproteobacteria*, known as copiotrophs (Eilers et al. 2010), further supporting to the conclusion that relative abundance and relative influence may be driven by different factors.

Our analysis approach has identified network generalist nodes (Figure 3-4). Our results showed an essentially exponential decay of nodes as successive samples are added, with an initial slower decay of pasture nodes. This observation is consistent with the pervious observation of higher taxonomic similarity in pasture community (Rodrigues et al. 2013). Most of the forest generalists in our study belonged to the phylum *Acidobacteria,* of which a known representative genus was *Candidatus Koribacter;* and class *Alphaproteobacteria*, of which a known representative was *Rhodoplanes* (Supplementary Table 3-3). Genomic evidence suggests that *Candidatus Koribacter* plays an important role in the global carbon cycle, in which they are capable of oxidizing carbon monoxide (CO) and degrading complex plant polymers, and the nitrogen cycle, in which they are capable of reducing nitrogen compounds (Ward et al. 2009). This alphaproteobacterial genus, which was detected with high centrality value in forest network in our previous study (Chapter 2), is known for carrying out the denitrification process (Hiraishi et. al. 1994). While the ecological significance of most of the other generalists are understudied, a few alphaproteobacterial families such as *Beijerinckiaceae, Hyphomicrobiaceae, Methylocystaceae* and *Bradyrhizobiaceae,* belonging to the order *Rhizobiales*, are known for their nitrogen-fixing ability (Brown et al., 2012). Moreover, *Methylocystaceae* is known for its ability to obtain carbon and energy from methane (Bowman 2006). Hence these generalists are speculated to make significant contributions to carbon and nitrogen cycles, where scavenging single-carbon compounds and atmospheric nitrogen may be important to optimize life of these microbes in the Amazon forest soils. Our result aligns with the higher methanotrophic ability of forest soil (Paula et al. 2014) though higher nitrogen fixation is evident in pasture soil

(Mirza and Rodrigues et al. 2012). Conversely, pasture generalists are spread out more evenly to several major phyla, where two OTUs are close relatives of forest generalists: *Candidatus Koribacter* and *Rhodoplanes* (Supplementary Table 3-4). While the majority of these are understudied to decipher their ecological significance, some of the generalists are known to synthesize antimicrobial agents, such as *Bacillus cereus* (Naclerio et al. 1993), others are known for their ability to sporulate, such as members of *Firmicutes*. Given the lower floristic diversity in pasture soil, which may not contribute to a wide variety of carbon compounds in rhizospheric regions, this observation may reflect increased competition between bacterial taxa. Another important change observed by forest-to-pasture conversion was the detection of lower acidobacterial and actinobacterial OTUs as generalists; a similar pattern was observed previously due to anthropogenic activity (Zhou et al. 2011). In addition to ecosystem-wise study for generalists, we identified shared generalists residing across ecosystems, whose presence was not influenced by land use change (Supplementary Table 3-5). Most of the shared generalists in our study belonged to *Acidobacteria* and *Alphaproteobacteria*. Within the former taxa, two known representatives were *Edaphobacter modestus*, which can utilize plant-derived sugars and sugar alcohols (Koch et al. 2008), and *Candidatus Solibacter*, which can produce biofilms (Ude et al. 2006). On the other hand, two alphaproteobacteria nodes, *Phenylobacterium* and *Pedomicrobium* were involved, respectively, in xenobiotic degradation (Eberspächer and Lingens 2006) and biofilm production (Sly et al. 1988). These shared generalists are therefore versatile at utilizing plant derived compounds as carbon and energy sources, and in dealing with extreme variations in environmental conditions by producing biofilms. Within the shared generalists, forest nodes mostly had

higher centrality values compared to pasture, including the ones, whose closest relatives were similar to generalists of forest and pasture networks, meaning that these forest nodes are more influential over the interactions with other nodes. This result therefore sheds light in understanding the ecology of microorganisms in terrestrial ecosystems. Using similar approaches, this ecologically important category was proved useful in microbial (Zhou et al. 2011, Barberan et al. 2012) and macrobial (Pandit et al. 2009) ecology studies.

With the same methodology, we also attempted to discern the conservation of the ecological community relationships and functional architecture of OTUs. Co-occurring pairs of microbes in networks are assumed to share similar niches (Fuhrman et al. 2008, Leibold and McPeek 2006, Raes and Bork 2008), which can provide valuable insights about ecological interactions and functional distribution of microbes (Williams et al. 2014). Our results demonstrated that, the co-occurring pairs of nodes, including of generalists nodes, were neither preserved across both of the soil types nor within each of the soil types. Since most of the network nodes are specialists, meaning they were not cosmopolitan across networks (Figure 3-4), it is plausible to observe this pattern for the microbial pairs. We investigated bacterial co-occurrence patterns at higher taxonomic level (Figure 3-5). The quantitative framework of our approach may suggest metabolic interactions of microorganisms through co-occurrence with other microorganism based on the assumption that statistically significant correlation pairs are ecologically similar (Barberan et al. 2012, Williams et al. 2014, Leibold and McPeek, 2006). It may also explain the mechanisms of ecological processes, and alteration of them due to land use change (or, any other treatment). For example, *Alphaproteobacteria*, which comprises

both nitrogen-fixers and denitrifiers, is in higher relative proportion with *Nitrospirae*, which comprises nitrifiers (Spang et al. 2002, Stahl et al. 2012) in forest soils (0.8%) compared to pasture (0.35%). This higher mathematical correlation between these groups of microbes aligns with the experimental observations of higher nitrification and denitrification processes in forest soils (Neill et al. 1995, Neill et al. 1997b, Neill et al. 1999, Melillo et al. 2001). The relationships between these microbial taxa may therefore represent testable hypotheses in understanding ecological significance and metabolic mechanism of their co-occurrence. This approach would be useful to study microbial ecology, especially to microbial groups and ecosystem that are poorly studied. Although these hypotheses on metabolic interactions require verification using approaches such as co-culture in microcosms along with biochemical tests, our novel approaches provide us with a starting point in understanding the physiology and ecology of microorganisms.

3.7 Final remarks

With this work we demonstrated a novel co-occurrence framework, where topology-based analysis approach enhances our knowledge of bacterial biogeography. The use of site-wise co-occurrence networks from bacterial datasets of multiple terrestrial locations along with the incorporation of environmental factors constitutes a more comprehensive approach in revealing ecological traits of microorganisms. We identified bacteria taxa that were correlated to soil carbon and nitrogen contents, implying their associations of biogeochemical processes and alteration of these processes by ecosystem conversion. More importantly, our approach attempted to explore potential metabolic interactions between bacterial taxa, which may provide insights in life style strategies of

soil bacterial communities. To the best of our knowledge, this is the first study to document the spatial assemblage pattern of bacteria, and its alteration in response to the anthropogenic activity.

Chapter 4

Forest-to-pasture conversion shifts microbiome as a function of environmental challenges in the Amazon soils

4.1 Abstract

The Amazon rainforest plays a crucial role in global ecosystem processes, and it has been subjected to high rates of forest-to-pasture conversion. This conversion substantially alters the biological and chemical composition of the soil ecosystem. In this study, we used a shotgun metagenomic approach to compare the microbial communities and their functional attributes between these two contrasting ecosystems. Here we report that taxonomic alpha and beta diversities increased due to ecosystem conversion, which mirror functional diversities. Across soils, functional beta diversity had a strong concordance with taxonomic beta diversity, where each soil type showed distinct clustering pattern. We estimated that ecosystem conversion significantly increased copiotrophic microbial taxa such as *Actinobacteria, Firmicutes, Bacteroidetes*. Consistent with the taxonomic shifts, pasture microbial community had overrepresentation of genes associated with the utilization of carbohydrates. In contrast, the forest microbiome is enriched in genes involved in the gluconeogenesis and utilization of xenobiotic compounds. The functional genes observed might have links to the quality and/or quantity of plant derived carbon inputs in the soils. For energy harvest, the pasture community had lower relative abundance of genes involved in the TCA cycle, denitrification, and higher relative abundance of genes involved in fermentation, and methanogenesis. Comparative analyses of biochemical pathways revealed that land use change from forest to pasture resulted the alteration of the microbiome in relation to the

alteration of biological and chemical composition of the soil. Therefore the identification of genes, which were differentially abundant between forest and pasture, offers insights for the lifestyle strategies of microbial communities in their ecosystems. To the best of my knowledge, this is the most comprehensive study on microbiomes to examine the effect of deforestation in the Amazon rainforest.

4.2 Importance

In the Amazon, forest-to-pasture conversion has a strong negative impact on biological composition that results in alteration of important ecosystem processes. Most of the previous studies were related to understanding the microbial diversity. Though there were some studies targeting the specific functional genes, a comprehensive understanding of functional attributes of microbial communities remained understudied. Employing a shotgun metagenomics approach, we report here that forest and pasture communities are functionally distinct. As the most comprehensive survey known to date, this study demonstrates predictive understanding of how anthropogenic activity shapes microbial and functional diversities across ecosystems, and the link between them. This study demonstrates the energy harvest strategies to specific environmental challenges, providing insights to understand the mechanism of ecosystem processes.

4.3 Introduction

The Amazon is the largest continuous rainforest ecosystem in the world and provides essential ecosystem services in global scale. It harbors the largest collection of plants and animal species in the world (Dirzo et al. 2003), and balances the flux of

atmospheric gases, both at regional and global scale (Betts et al. 2008). Despite this importance, forest clearing has been alarming over the last few decades. The process of deforestation and land use change concurrently alters the chemical and biological composition of the soils (Fearnside 1999, Feeley and Silman 2009, Rodrigues et al. 2013, Herpin et al. 2002, Neill et al. 1997a, Fernandes et al. 2002), which has important consequences to the ecosystem processes.

Land use change is predicted to be the most important factor in altering biodiversity in tropical areas for the twenty first century (Sala et al. 2000). Ecologists have long been studying the impact of deforestation on the plant and animal diversity in tropical forests (Fearnside 1999, Feeley and Silman 2009), and their contribution in ecosystem functions (Andrade et al. 2015, Eva et al. 2004, Feigl et al. 2006, Foley et al. 2007). In contrast, microbial ecologists have just started to explore similar set of questions in the Amazon (Rodrigues et al. 2013, Mirza and Rodrigues 2012, Navarrete et al. 2011 and 2015, Ranjan et al. 2015). Despite being closely related to important biogeochemical processes such as carbon and nitrogen cycles, microbial diversity has not been studied until recently. These microbiological studies, which largely relied on ribosomal genes, showed that ecosystem conversion alters bacterial (Borneman and Triplett 1997, Cenciani et al. 2009, Jesus et al. 2009, Rodrigues et al. 2013), archaeal (Chapter 2, Navarrete et al. 2011) and fungal (Mueller et al. 2014) community composition in various Amazonian soil types. These studies showed that microbial diversity loss essentially followed the patterns of plant and animal diversity. Plant diversity not only increases microbial diversity (Garbeva et al. 2004, Prober et al. 2015) but also their activity in the soil (Lange et al. 2015). While other studies demonstrated the

consequences of deforestation on microbial activity, these were only restricted to specific functional groups of microbes or specific functional genes of interests such as those associated with carbon and nitrogen cycles (Chapters 2 and 3, Mirza et al. 2014, Navarrete et al. 2011, Paula et al. 2014, Taketani and Tsai 2010). To date, no work has been published to understand the change of the microbiome in response to the land use change in the Amazon.

In this study, we attempted to investigate the impact of forest-to-pasture conversion on the functional gene abundance of shotgun metagenomic datasets and functional processes carried out by microbes. In particular, we aimed (*i*) to estimate the functional composition of the microbiome in primary forest and established pasture; (*ii*) to understand the impact of microbial diversity loss on the status of functional diversity; (*iii*) to demonstrate differential abundances of key functional genes involved in ecologically relevant metabolic processes. This study therefore sheds light in a comparative understanding of metabolic activities, which are related to specific environmental challenges and important biogeochemical processes as a result of alteration of soil biological and chemical composition. The slash-and-burn process of deforestation is largely responsible for the alteration of the chemical composition of soil nutrients, which may result in the alteration of soil microbial structure and function (Mäder et al. 2002). Following deforestation, the increased availability of nutrients through the incorporation of ashes from forest vegetation and combustion of soil organic matter initially increases soil fertility, which is usually maintained by the incorporation of cow dung in established pastureland. We therefore hypothesized that forest-to-pasture conversion will increase the genes that are involved in nutrient utilization. The strategies

for adaptation and energy generation of the microbiome will parallel the availability of nutrients in ecosystems and environmental condition.

4.4 Methods and materials

4.4.1 Site description and sampling

The sampling sites were located at Fazenda Nova Vida in the State of Rondonia, Brazil (10°10′18.71″S, 62°47′15.67″W), representing one of the highest rates of deforestation of the Brazilian Amazonia in the last two decades (INPE 2011). The pasture was established in 1972 after a slash-and-burn procedure followed by aerial seeding of the fast growing grasses *Urochloa brizantha* and *Panicum maximum*. The sampling and sample processing were describes in detail elsewhere (Chapter 2, Rodrigues et al. 2013). In brief, a sampling design based on a 100m^2 quadrat with 10-m^2, 1-m^2, 0.1-m^2, and 0.01-m^2 quadrats within was established for primary forest and pasture. Sampling plots are 5.5 km apart between forest and pasture. For this study, samples were collected in 2010 following the rainy season along a transect with the cardinal direction North in both forest and pasture. Following the removal of litter, soil sampling was performed with a 10 cm-depth by 5 cm-diameter corer with samples being transported on ice to the laboratory and stored at -80^{o}C until soil DNA extractions.

4.4.2 Total DNA extraction and shotgun sequencing

Ten grams of soil from each sample were used for total genomic DNA extraction using the PowerLyzer PowerSoil DNA isolation kit (MoBio Inc, Carlsbad, CA, USA) according to the manufacturer's instruction. The concentration and purity of soil DNA were determined spectrophotometrically (NanoDrop Technologies Inc., Wilmington,

DE). To obtain taxonomic and functional information from the extracted DNA samples, shotgun metagenome sequencing was carried out on total DNA samples extracted from forest and pasture soils using Illumina HiSeq 2000 Paired-End Prep kit protocol (Illumina, Inc., San Diego, California).

4.4.3 Sequence processing and annotation of shotgun metagenomes

Raw sequences were uploaded to Rapid Annotation using Subsystems Technology for Metagenomes (MG-RAST; Meyer et al. 2008) for quality filtering and annotation. Paired end reads were joined using fastq-join as part of the MG-RAST pipeline version 3.2. Single end reads that could not be joined were retained. Quality filtered sequence reads were annotated using M5rna database for taxonomy and Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa et al. 2000) database for mapping metabolic pathways. The annotations were conducted at an e-value cutoff of 1e-5, minimum identity cutoff of 60% and minimum alignment length cutoff of 15.

4.4.4 Data analysis

Following the annotations, taxonomic and functional (KO) profiles were downloaded from the MG-RAST server in biome file format. For the downstream processes, we discarded taxonomic and KO genes that were observed in less than 3 samples in each ecosystem. They were considered as less represented genes in their soil type. Then all samples were rarefied to their lowest sequencing depth to bring all the samples to equal sequencing depth in order to remove sequencing effort bias. The threshold numbers of sequences for taxonomic and functional genes were 804,055 and 60,373,709, respectively. Rarefaction plots were generated to check whether sampling effort was sufficient.

The relative abundances of major taxa and functional categories of genes (KEGG level 2) in the shotgun metagenomes obtained from forest and pasture ecosystems were compared. Alpha diversity of functional genes was calculated using Shannon index, *H*' (Shannon 1948). To test the hypothesis that forest and pasture have differential metabolic profiles, principal coordinate analysis (PCoA) tests were performed using the pairwise Hellinger distance from the relative abundances of individual annotated genes (KEGG level 4).

To estimate the magnitude of impact of ecosystem conversion on the relative frequencies of KO genes, fold difference of forest genes was calculated compared to that in pasture and represented as $\log_2$foldchange. A score of zero indicates that the KO gene has statistically same proportional abundance in metagenomes of both soil types. A positive value for a given gene indicates overrepresentation in forest metagenomes compared to pasture, while a negative value indicates the reverse pattern. We used these KO genes, especially genes encoding enzymes, to conduct ecosystem-level analyses in understanding the major differences in metabolic pathways between forest and pasture, especially those that are related to important ecosystem processes.

4.4.5 Statistical analysis

To determine whether the different relative abundances of functional gene categories between the ecosystems are statistically significant, a nonparametric two-sample *t*-test with 999 Monte Carlo permutations was used. Alpha diversity values between soil samples were compared using the same *t*-test. Analyses of similarities (ANOSIM) tests were performed to assess whether metagenomic composition are significantly different across the terrestrial ecosystems. Procrustes analyses were carried

out to test whether functional datasets produce concordance with taxonomic datasets. We used DESeq2 statistic to identify individual KO genes, whose relative abundances between forest and pasture differ significantly (Love et al. 2014). Therefore this approach identified KO genes that are differentially abundance between forest and pasture. Data analyses were performed in the QIIME platform (Caporaso et al. 2010).

4.5 Results

4.5.1 Metagenome profiles

Rarefaction of functional genes to even depth yielded a total number of 5,292 KOs that belong to forest and 5,432 KOs to pasture. A total of 5,607 KOs were estimated in both land uses. In both ecosystems, genes associated with central metabolic processes comprised the majority of functional gene pools (Figure 4-1). Highest proportional abundances of genes were estimated for amino acid metabolism, which comprised 21.54% [±0.12; α=0.05 confidence interval (CI)] in forest and 21.48% (±0.15; α=0.05 CI) in pasture. Then carbohydrate metabolism, membrane transport, translation and energy metabolism followed. While we have not observed large differences in the relative frequencies of major functional categories at KEGG level 2 between forest and pasture, the difference in following categories, however, were statistically significant: carbohydrate metabolism ($P < 0.001$), energy metabolism ($P < 0.01$), replication and repair ($P < 0.05$), nucleotide metabolism ($P < 0.05$), cell growth and death ($P < 0.001$), and xenobiotic biodegradation and metabolism ($P < 0.01$).

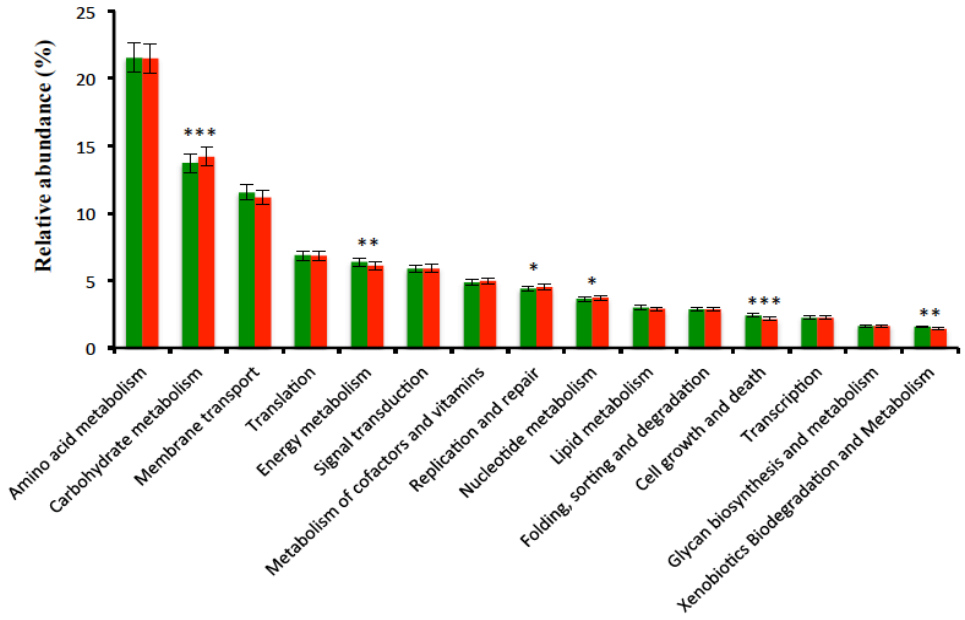


Figure 4-1. Relative abundances of functional categories at KEGG level 2 in the shotgun metagenomes obtained from the Amazon forest (green) and pasture (red). Error bars represent standard error. Symbols (*), (**), and (***) indicate significance values of $P < 0.05$, $P < 0.01$, $P < 0.001$, respectively, which were calculated using two-sample *t*-test.

Analysis of taxonomic profiles of metagenomes results in a total of 20,635 OTUs in forest and 21,066 OTUs in pasture, totaling 21,214 OTUs in both soil types. With over 25% [±1.62; α=0.05 CI] in forest and 18% (±0.48; α=0.05 CI) in pasture, *Alphaproteobacteria* represents the dominant microbial taxon in both ecosystems and the largest percentage variation in response to forest-to-pasture conversion (Supplementary Figure 4-1). Other major microbial groups that differed substantially between forest and pasture: *Actinobacteria* ($P < 0.001$) and *Firmicutes* ($P < 0.01$), which increased from 12.61% (±0.48; α=0.05 CI) to 16.2% (±0.83; α=0.05 CI) and 10.96% (±1.22; α=0.05 CI) to 14.02% (±0.4; α=0.05 CI), respectively. A slight increase of *Cyanobacteria* ($P<0.05$)

was observed as abundance of this phylum changed from an average of 4.06% [±0.14; α=0.05 CI] to 4.38% (±0.13; α=0.05 CI). Importantly, *Thaumarchaeota,* an ecologically important archaeal phylum, comprised only 1.02% (±0.19; α=0.95 CI) in forest but decreased drastically to 0.21% (±0.03; α=0.95 CI) in pasture.

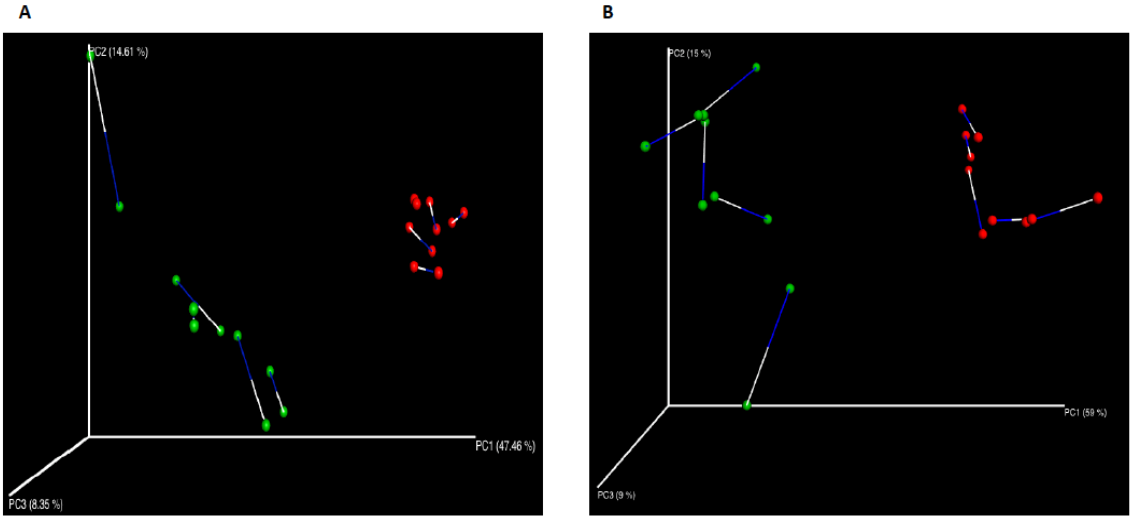


Figure 4-2. Procrustes analyses of functional genes with operational taxonomic units (OTUs) observed in (*A*) shotgun metagenomes and (*B*) 16S rRNA gene amplicons (Chapter 2). The metagenomes and 16S rRNA gene amplicons datasets were obtained from the same soil samples. Clustering patterns were statistically significant ($P < 0.01$) with goodness of fit (M^2) values of 0.14 and 0.29, respectively. Forest samples are represented by green dots and pasture samples are represented by red dots in both panels. The white ends of each connector line between two dots connect to the 16S rRNA data for the sample, whereas the blue ends connect to the metagenomic data. The length of the connector lines is inversely proportional to the overall concordance between taxonomic and functional gene datasets.

4.5.2 Metagenome diversity

To test whether the metagenomic datasets have reached the full breadth of genomic diversity, we generated rarefaction plots. These plots showed that both the taxonomic and functional genes (KO) have reached plateaus (Supplementary Figure 4-2). However rarefaction plots of recently studied 16S rRNA amplicons (Chapter 2), which were obtained from the same soil samples of metagenomes, have not leveled off (Supplementary Figure 4-3). This observation reflects on the procrustes analysis, where functional genes have better concordance with taxonomic genes from metagenomes than with 16S rRNA amplicons (Figure 4-2; M^2=0.007 and 0.307, respectively; $P < 0.01$ in both cases)

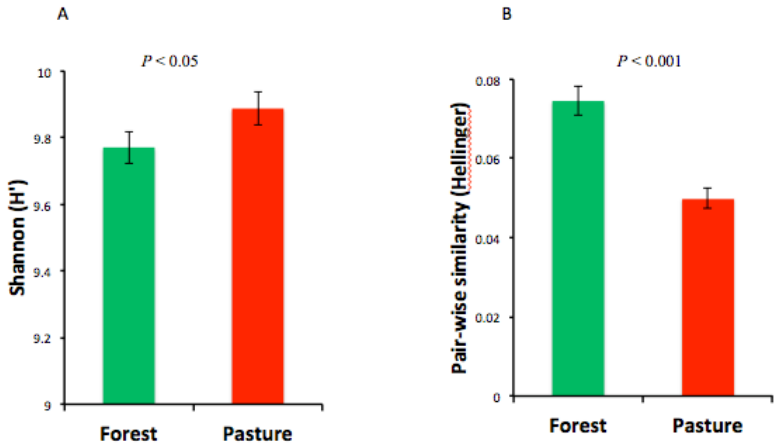


Figure 4-3. Response of soil ecosystem conversion to functional diversities in shotgun metagenomes. (*A*) Alpha diversity (Shannon, *H'*), $P < 0.05$; (*B*) Beta diversity (Hellinger), $P < 0.001$. Means (n = 5 for alpha diversity, n = 10 for pair-wise similarity as beta diversity) are depicted. Error bars represent standard error.

We calculated alpha and beta diversity of metagenomes, where Shannon index (*H'*) and Hollinger distance measures were used, respectively. Alpha diversity reflects the

mean of sample diversities, whereas beta diversity implies the mean diversity between any two samples. We observed that both alpha [two sample *t*-test: t=-4.17 and $P < 0.05$ for functional (Figure 4-3A); t=-3.24 and $P < 0.01$ for taxonomic (Supplementary Figure 4-4A)], and beta [two-sample *t*-test: t=-10.35 and $P < 0.001$ for functional (Figure 4-3B); t=-3.88 and $P < 0.01$ for taxonomic (Supplementary Figure 4-4B)] diversities of metagenomes increased following the land use change. We also report beta diversity pattern of functional genes using a PCoA plot (Figure 4-4), where most of the variance (89.05%) between forest and pasture was estimated. The ordination plots clearly showed that forest and pasture metagenomes were functionally distinct (ANOSIM: $R = 0.98$, $P < 0.02$).

4.5.3 Differential abundance of functional genes and pathways

Next, we focused on determining what functional genes and pathways were responsible for the disparities. We used the DESeq2 method, which is based on the negative binomial distribution, to identify functional genes whose proportional representation differed significantly between forest and pasture. This analysis yielded 2,070 KEGG orthology (KO) genes that are differentially abundant between forest and pasture, of which 1,254 belong to KEGG EC categories ($P < 0.01$, Benjamini-Hochberg adjusted). DESeq2 analysis showed that the major difference of forest and pasture metagenomes was associated with genes involved in carbohydrate and energy metabolism.

4.*5.3.1 Carbohydrate metabolism*- Forest-to-pasture conversion results in the enrichment of EC genes associated with the metabolism of starch and sucrose, and galactose, in which starch, glycogen, cellulose, sucrose and galactose were potential sugar sources

(Supplementary Figure 4-5, Supplementary Table 4-1). In addition, genes involved in the transport of mono- and di-saccharides were far more abundant in pasture, of which log_2foldchange values ranged from -1.09 to -4.79. The forest microbiome, on the other hand, was enriched in genes involved in gluconeogenesis (Supplementary Figure 4-6). A gene encoding fructose-1, 6-bisphosphatase (EC3.1.3.11) involved in a key step of gluconeogenesis increased over 2.5 fold (log_2foldchange=1.19) in the forest samples (Figure 4-5). In addition, genes associated with the pentose phosphate pathway (PPP), the conversion of pyruvate to oxaloacetate [pyruvate carboxylase (EC6.4.1.1)], the conversion of oxaloacetate to phosphoenolpyruvate [PEP; phosphoenolpyruvate carboxykinase (EC4.1.1.49]) have higher relative abundance in the forest soils.

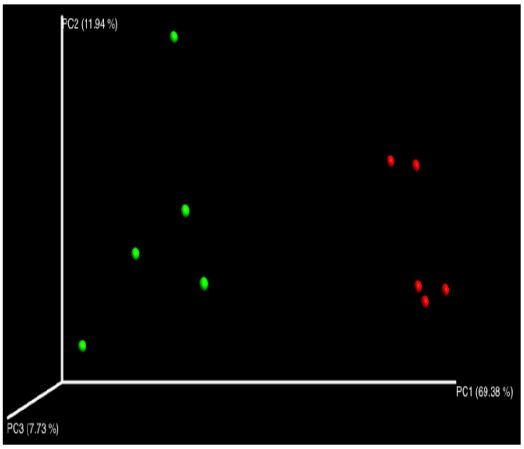


Figure 4-4. Principal coordinates analyses of Hellinger distances between functional metagenomes across soil samples. Forest samples, green dots; pasture samples, red dots.

We observed considerable variation in gene frequencies that are involved in subsequent metabolism of these sugars between forest and pasture. Several genes involved in fermentation such as metabolism of pyruvate [acetyl-CoA hydrolase (EC3.1.2.1) and phosphate acetyltransferase (EC2.3.1.8)], butanoate [phosphate

butyryltransferase (EC2.3.1.19)] and propanoate [methylmalonyl-CoA carboxyltransferase (EC2.1.3.1) and methylmalonyl-CoA decarboxylase (EC4.1.1.41)] are overrepresented in pasture datasets (Figure 4-5, Supplementary Table 4-1). Unlike the abundance of different fermentation genes, which increased through ecosystem conversion, the forest microbiome demonstrated higher proportional representation of genes involved in the utilization of fermentation products, such as L-lactate dehydrogenase (EC1.1.2.3) (Supplementary Figure 4-6, Supplementary Table 4-1).

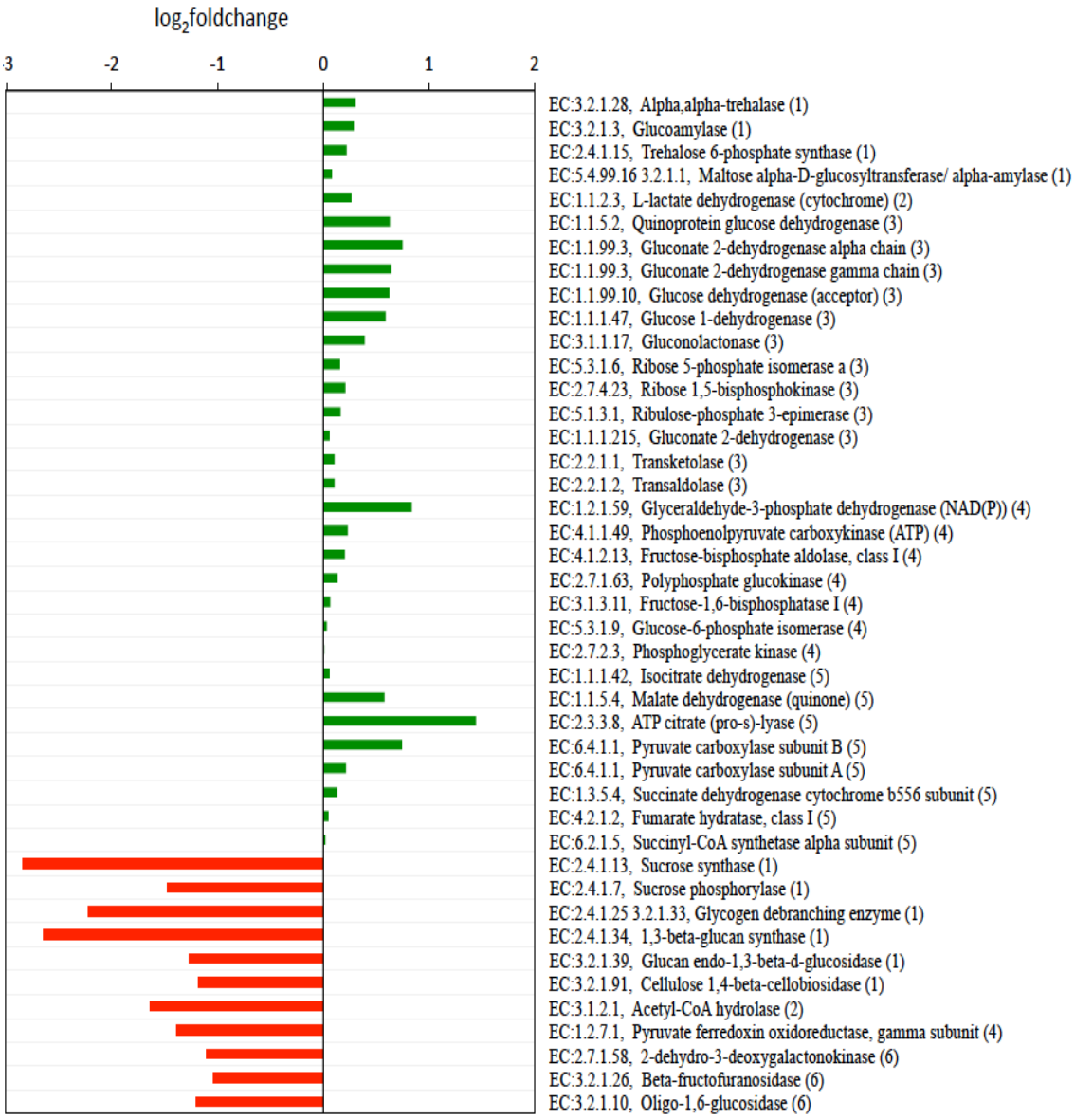


Figure 4-5. Differentially abundant genes encoding ECs involved in the carbohydrate metabolism ($P < 0.01$, Bonferroni corrected); x axis shows the values of log_2foldchange, where positive values indicate the genes enriched in forest (green) and negative values indicate the genes enriched in pasture (red). 1= Starch and sucrose metabolism; 2= Pyruvate metabolism; 3= Pentose phosphate pathway; 4= Glycolysis/Gluconeogenesis; 5= TCA cycle; 6= Galactose metabolism.

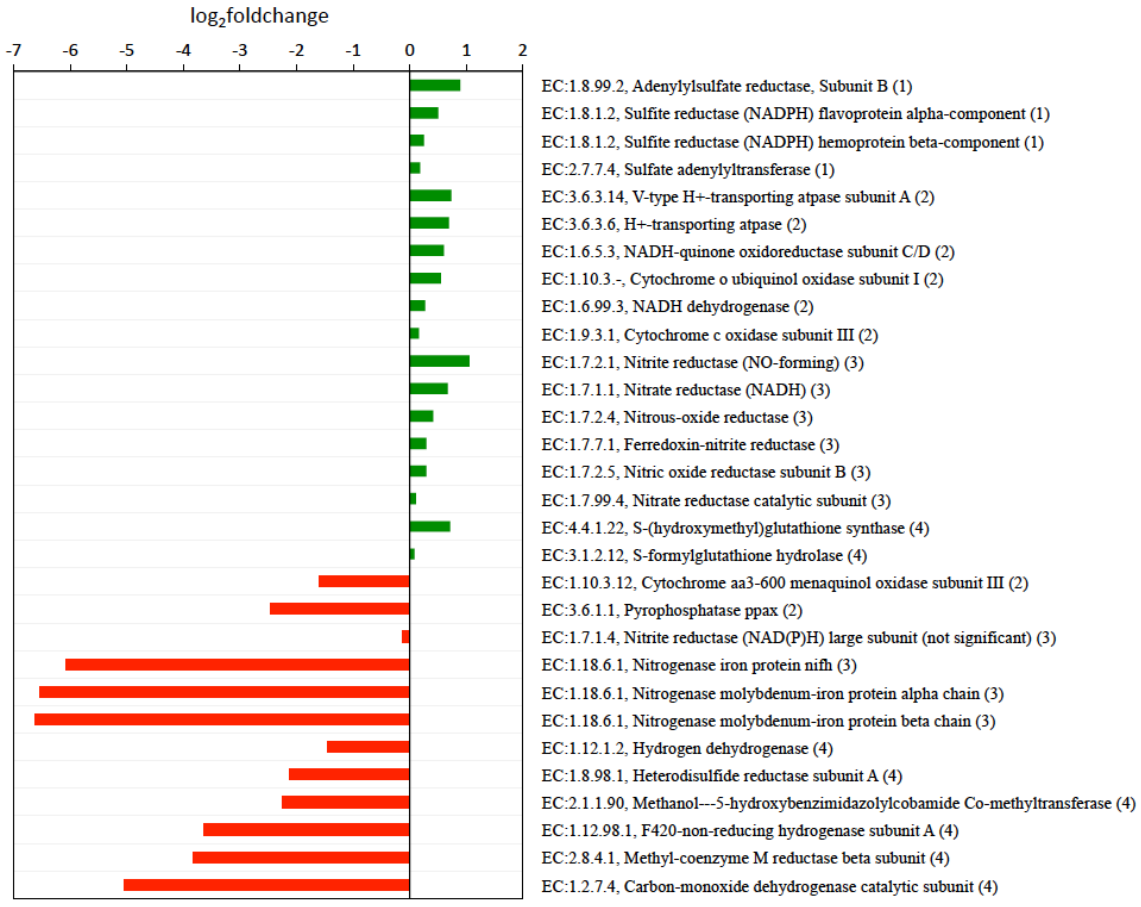


Figure 4-6. Differentially abundant genes encoding ECs involved in energy metabolism ($P < 0.01$, Bonferroni corrected); x axis shows the values of log_2foldchange, where positive values indicate the genes enriched in forest (green) and negative values indicate the genes enriched in pasture (red). 1= Sulfur metabolism; 2= Oxidative phosphorylation; 3= Nitrogen metabolism; 4= Methane metabolism.

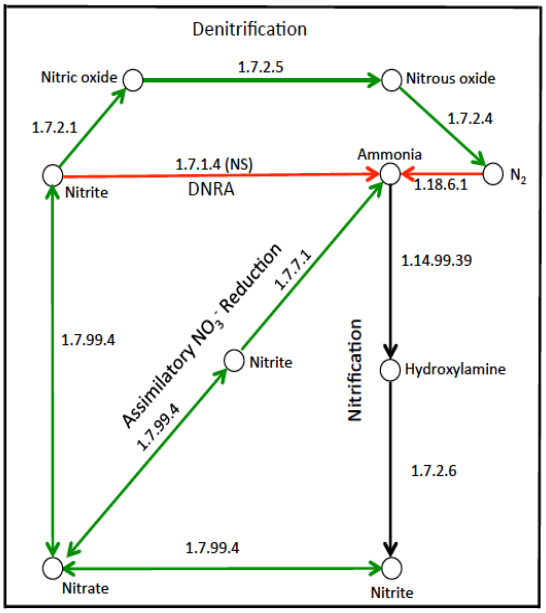


Figure 4-7. Response of soil ecosystem conversion in the differential representation of genes encoding ECs involved in nitrogen metabolism. Arrows indicate the enzyme-mediated steps in the pathway, where green and red arrows indicate genes enriched in forest and pasture, respectively, and black arrows indicate genes that were not estimated to be differentially abundant between forest and pasture. NS; non-significant

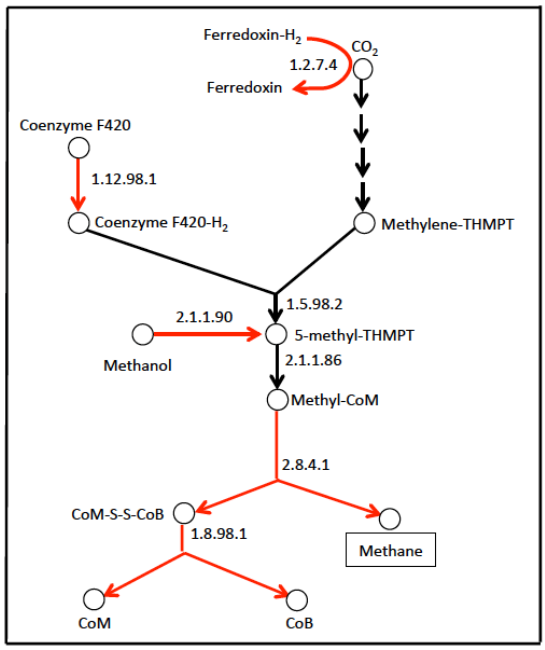


Figure 4-8. Response of soil ecosystem conversion in the differential representation of genes encoding ECs involved in methanogenesis. Arrows indicate the enzyme-mediated steps in the pathway, where green and red arrows indicate genes enriched in forest and pasture, respectively, and black arrows indicate genes that were not estimated differentially abundant between forest and pasture.

4.5.3.2 Energy metabolism- The forest microbiome had more KO genes that are components of the TCA cycle (Supplementary Figure 4-7, Supplementary Table 4-1),

whereas the relative number of specific genes involved in oxidative phosphorylation, typical of respiration, varied considerably between soil types (Supplementary Table 4-2). In contrast to the forest microbiome, where different V-type H+-transporting ATPase subunits predominated (log_2foldchange: 0.44 to 1.08), the pasture microbiome was enriched in genes for bidirectional Ni-Fe hydrogenase diaphorase subunits (K05586, K05587, K05588; log_2foldchange: -1.32 to -1.56; Figure 4-6). Genes involved in anaerobic respirations also varied considerably between forest and pasture. This variation involved nitrogen metabolism (Figure 4-7) and methane metabolism (Figure 4-8). The forest microbiome had overrepresentation of EC genes involved in nitrification (EC1.7.99.4), and likewise, denitrification (EC1.7.2.1, EC1.7.2.5 and EC1.7.2.4) and assimilatory nitrate reduction (EC1.7.99.4 and EC1.7.7.1). On the other hand, the proportions of genes involved in nitrogen fixation (EC1.18.6.1) and dissimilatory nitrate reduction to ammonia (DNRA) (EC1.7.1.4) were estimated higher in the pasture microbiome. The fold differences of different subunits of nitrogen fixing gene ranged from 67.85 to 99.37 (log_2foldchange: -6.08 to -6.63; Figure 4-6). The differences in abundance for the gene associated with the DNRA, however, was not statistically significant. In methane metabolism, several genes (EC1.2.7.4, EC1.12.98.1 and EC2.8.4.1; log_2foldchange: -5.05, -3.63 and -3.83, respectively) were enriched in the pasture microbiome, while others (EC3.1.2.12 and EC4.4.1.22; log_2foldchange: 0.08, and 0.71, respectively) were enriched in the forest. The former set of genes is involved in the synthesis of methane (methanogenesis) and latter set involved in the utilization of it (methanotrophy).

4.5.3.3 Amino acids metabolism- While we have not observed a statistically significant difference in amino acid metabolism, we observed overrepresentation of genes for the biosynthesis of several amino acids and intermediate compounds including alanine (EC4.1.1.12), asparagine (EC6.3.5.4), homoserine (EC1.1.1.3), threonine (EC2.7.1.39), homocysteine (EC2.3.1.31 and EC4.4.1.8), arginine (EC6.3.4.5), methionine (EC2.1.1.13 and EC2.1.1.14), glutamine (EC3.5.1.2) and putrescine (EC4.1.1.17) in the forest microbiome (Figure 4-9). Among the ABC transporter genes for amino acids, the pasture microbiome had higher representation for the transport of lysine, arginine and ornithine (K10013), and D-methionine (K02071-K02073), while forest microbiome had higher representation for the transport of branched-chain amino acids (K01995-K01999), glutamate and aspartate (K10001-K10003), General L-amino acids (K09969-K09972) and putrescine (K11073-K11076) (Supplementary Table 4-3).

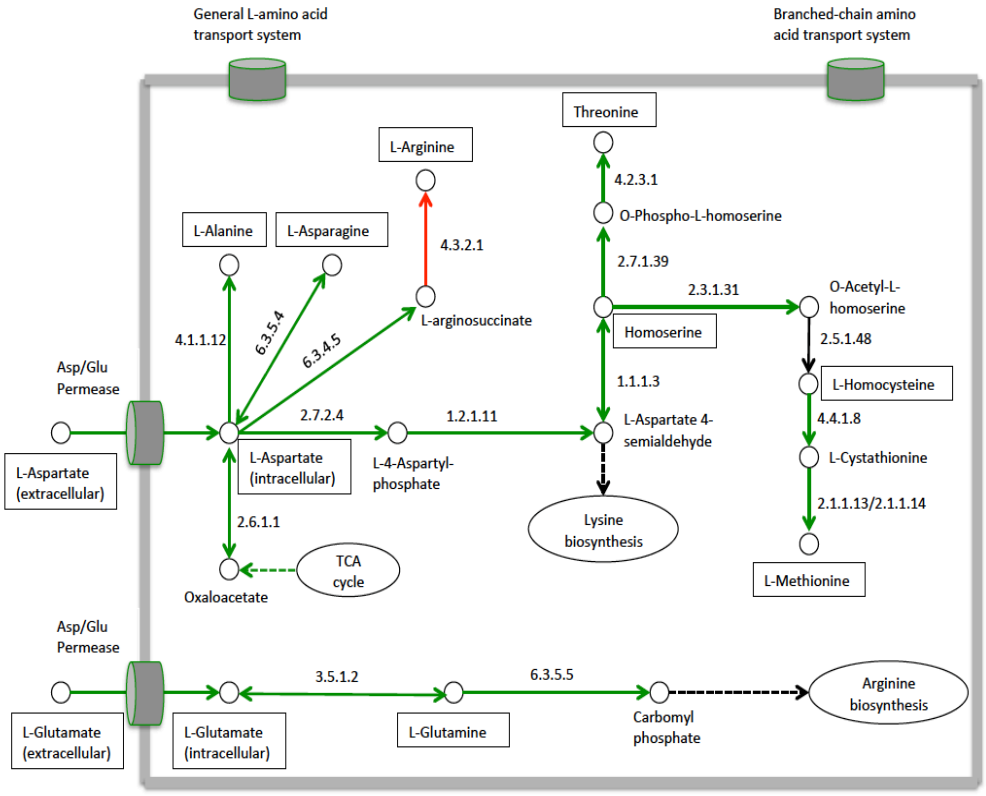


Figure 4-9. Response of soil ecosystem conversion in the differential representation of genes encoding ECs involved in the amino acid metabolism. Arrows indicate the enzyme-mediated steps in the pathway, where green and red arrows indicate genes enriched in forest and pasture, respectively.

4.5.3.4 Metabolism of xenobiotic compounds- According to the results of the present study, the forest microbiomes had higher representation of genes associated with the degradation of xenobiotic compounds, including benzoate, polycyclic aromatic hydrocarbon, aminobenzoate (Supplementary Table 4-4). Breakdown of benzoate may lead to the production of the intermediates of TCA cycle including succinyl-CoA and oxaloacetate (Supplementary Figure 4-8).

4.5.3.5 Sporulation and cell motility- Endospores and flagella are two special structures of microorganisms. Forest-to-pasture conversion increased the number of genes affiliated with both of these structures. Our analyses demonstrated that the following sporulation associated genes were far more abundant in the pasture microbiome: K07697, K07699 and K13533, where the values of $\log_2$foldchange were -2.99, -2.54 and -1.96 (Supplementary Table 4-5). In cell motility, genes involved in the flagellar assembly that were more abundant in the pasture were K02396, K02402, K02406, K02422 and K13820, where $\log_2$foldchange values were -1.09, -1.44, -1.28, -1.82 and -1.93, respectively; as were the chemotaxis genes K03406, K03408, K03411, K05874 and K05875, where $\log_2$foldchange values were -1.33, -1.07, -1.58, -1.33 and -1.53, respectively (Supplementary Table 4-6).

4.5.3.6 Biosynthesis of other secondary metabolites- We observed varied abundances of genes involved in secondary metabolites, which were antibiotics and bacitracin, an antimicrobial agent that inhibits the biosynthesis of cell wall. We estimated a slight variation of the mean relative abundances of genes involved in the biosynthesis of antibiotics penicillin and cephalosporin (EC3.5.1.11), and streptomycin (EC5.1.3.13) (Supplementary Table 4-7). While we have not observed a significant variation of biosynthetic enzymes of bacitracin, the two-component signal transduction system (K11629-K11630) and the ABC transporter genes for it (K11631-K11632) were far more enriched in the pasture microbiome, where $\log_2$foldchange values ranged from -2.29 to -3.57.

4.6 Discussion

This study represents the first metagenome analysis compared between Amazon forest and pasture ecosystems. We obtained soil samples from top layer (0-10-cm) of both soil-types, which generally represents the most biologically active zone. By comparing the functional metagenome profiles, we evaluated the consequences of forest-to-pasture conversion in central metabolic processes, especially those that have important ecological significance.

4.6.1 Coverage of metagenomic diversity

We first checked whether the sequencing effort of our samples has reached the full breadth of diversity. The rarefaction curves of both the taxonomic and functional genes (KOs) of metagenomes confirmed that we obtained the full coverage of genomic diversity (Supplementary Figure 4-2). While a similar approach with operational

taxonomic units (OTUs) of 16S rRNA amplicon datasets of our study site was not conducted in a previous study (Rodrigues et al. 2013), we have not obtained a full coverage of taxonomic datasets used in our previous study (Chapter 2, Supplementary Figure 4-3). Along with the Procrustes analysis (Figure 4-2), this observation suggests that microbial taxa inferred from metagenomes better predict the functional potentiality of metagenomes. While the relative proportion of major taxonomic groups showed similar abundance patterns between current and previous studies, intriguingly we estimated that both alpha and beta diversities of functional genes increased following land use change (Figure 4-3). This pattern indicates that ecosystem conversion increases mean sample diversity and diversity between any two samples, which is in agreement with previous report of taxonomic and phylogenetic alpha diversity but not with beta diversity (Rodrigues et al. 2013). We argue that this observation may have two mutually exclusive implications: (*i*) the lower number of samples (five, for each soil type) of the current study may limit us to fully explore the spatial diversity of functional genes; (*ii*) the lower sequencing depth of each sample in previous study may not have reached the full breadth of diversity coverage.

4.6.2 A trade-off between glycolysis and gluconeogenesis in forest microbiome

Plants typically contribute major sources of organic carbon to the soils and these are often enriched in aromatic compounds, and microbial communities rely on these external carbon compounds (Grandy and Neff 2008). The extraordinary physiological diversity of microorganism for the utilization of these carbon compounds implies their ability to adapt in diverse habitats, which is often associated with energy metabolism. They employ different strategies for energy conversion and their primary challenge is

how to maintain adequate energy reserves despite challenges in their environment (Gianoulis et al. 2009). The differences in functional genes for energy conversion pathways between the two soil types presumably relate to the differences in the quantity and/or quality of organic carbon compounds present in the soils and environmental features such as edaphic factors. For example, pasture soil is rich in cellulosic litter from grasses, and one of the two grass species that dominates in pasture is *Urochloa brizantha,* which stores starch, proteins etc. in their rhizomes (Deinum et al. 1996). Therefore microbial communities in this soil type are likely to receive these carbohydrates from the grass species along with other nutrients. This pattern is supported by the increased number of genes involved in the transport of mono- and di-saccharides in pasture (Supplementary Table 4-1), which suggests that the pasture microbiome may rely more on exogenous saccharides. In contrast to the pasture microbiome, it is apparent that the forest microbiome has higher capacity for the utilization of xenobiotic compounds, which presumably are more abundant in the floristically richer forest habitat. The breakdown products of these compounds such as benzoate may supply the intermediates of the TCA cycle that would provide flexibility to energy harvest strategies and anaplerosis in the forest microbiome. Based on the results of KEGG pathway mapping, the conversion of anaplerotic compound oxaloacetate into phosphoenolpyruvate (PEP) might take place when gluconeogenesis is required and the conversion of pyruvate into oxaloacetate might take place when glycolysis is required (Supplementary Figure 4-6). Moreover, higher pentose phosphate pathway (PPP) associated genes in forest microbiomes may provide added flexibility to carbohydrate metabolism because they have higher potentiality to funnel glyceraldehyde-3-P, an intermediate of PPP, into gluconeogenesis or glycolysis

(Supplementary Figure 4-6). This observation suggests that the forest microbiome is better able to maintain a trade-off between glycolysis and gluconeogenesis, a strategy that may be critical in maintaining a healthy microbial community.

The genes affiliated with the production of antibiotics and resistances to them are widespread within microbial communities (Allen et al. 2010, Dantas et al. 2008), and probably are indicative of competition within microbial communities (Fierer el al. 2012). Likewise, bacitracin, which interferes with the cell wall synthesis (Stone et al. 1971), may have links to competitive interactions. In contrast to the genes affiliated with the synthesis of penicillin and cephalosporin (EC3.5.1.11), and streptomycin (EC5.1.3.13), which do not show substantial variations in frequency between forest and pasture, genes involved in the transport of bacitracin and resistance response to it were estimated to be elevated in the pasture microbiome (Supplementary Table 4-7). In pasture soil, which is dominated by just two grass species, the microbial community is expected to have a less diverse physiology for the utilization of organic compounds from grasses compared to that in forest. This probably explains the possibility of having more interspecific competition within the pasture microbial community for the utilization of more similar resources, which also supports our previous results (Chapters 2 and 3).

4.6.3 Forest-to-pasture conversion shifts energy harvest strategy in the Amazon

The most prominent differences imposed by ecosystem conversion in Amazon involved pathways relate to the energy metabolism, which is linked functionally to the carbohydrate metabolism. Gene abundance data suggest that the strategy of oxidative phosphorylation varied considerably between soil types (Supplementary Table 4-2). This might have special implications to understanding the ecosystem ecology of

microorganisms in both soil types. For example, bidirectional Ni-Fe hydrogenase diaphorase subunits, enriched in pasture, is oxygen sensitive and was purified under anaerobic condition (Schmitz et al. 2002). On the other hand, genes for superoxide dismutase (EC1.15.1.1) and biosynthesis of glutathione (EC6.3.2.2 and EC1.8.1.7) were abundant in the forest microbiome, indicating more aerobic processes for energy generation.

Another striking difference involved the metabolism of nitrogen and methane, which are related to anaerobic respiration. Deforestation by slash-and-burn process causes an initial loss of terrestrial nutrients through emission of gases and hydrologic leaching, especially nitrate (Davidson et al. 2007). Therefore it is not surprising to report that KEGG EC genes involved in the nitrification and likewise, assimilatory nitrate reduction were higher in the forest microbiome, whereas genes involved in nitrogen fixation were elevated in the pasture microbiome. These observations support previous reports (Neill et al. 1995, Paula et al. 2014). In addition, increased representation of *Thaumarchaea* in forest may be related to the observed abundance of nitrification genes (Hamaoui et al. 2016). It is noteworthy that nitrogen deficiency in the pasture habitat could increase the frequency and activity of free-living diazotrophs, which was evident in the previous study (Mirza et al. 2012). This probably explains why the pasture microbiome has higher capacity for nitrogen fixation, which may contribute to the fitness of the microbial community following deforestation. We detected an interesting pattern in nitrogen reduction processes between forest and pasture microbiomes. The forest sample is enriched in genes engaged in energy-yielding denitrification, a pattern that predominates in oligotrophic habitats and leads to the production of nitrous oxide and

nitrogen gases (Tiedje 1988, Yoon et al. 2015). In contrast, the pasture sample has higher representation in genes engaged in dissimilatory nitrate reduction to ammonia (DNRA), a pattern that relates to a copiotrophic and anaerobic habitat such as the bovine rumen. The difference in genes for DNRA in this study, which did not achieve statistical significance, does not support the previous report (Paula et al. 2014). DNRA is not an energy generation process but it probably helps reduce the further loss of the nitrogen from pasture soil in gaseous forms. On the other hand, forest-to-pasture conversion shifts the predominance of genes from methane oxidation to methane generation. KEGG pathway analysis therefore shows that ecosystem conversion potentially turns the soil ecosystem from a sink of methane to a source of it. It is noteworthy that methanogenesis is entirely carried out by strictly anaerobic archaea belonging to the *Euryarchaeota* (Dridi et al. 2012, Liu et al. 2008, Paul et al. 2012), although we did not notice a substantial increase of this group of archaea by forest-to-pasture conversion (Supplementary Figure 4-1). Methanogens may increase the efficiency of fermentation by removing one of its products, acetate. Livestock, which contributes about 37% of global methane emission (Steinfeld et al. 2006), may represent a potential source of this archaeal group in pastureland. This study therefore showed that strategies for anaerobic respiration in both soil types have the potential to produce important greenhouse gases (methane, nitrous oxide) and our findings are in line with previous studies conducted in Amazon (Davidson et al. 2012, Melillo et al. 2001, Neill et al. 2005, Steudler et al. 1996).

Although soil samples were collected after the rainy season, the better draining capacity of sandy loam forest soil prevents persistent waterlogging, which contributes to better aeration in the active layer of soil. Therefore forest habitats potentially support the

growth of microbial communities of diverse physiologies. In contrast, pasture soil of our study site is classified as sandy clay loam (Neill et al. 1997a). A better water-retaining capacity of this type of soil would eventually make the soil habitat more anaerobic, which is aided by the soil compaction during cattle grazing. Therefore cattle may not only disperse microbes to the land, but also maintain reduced state suitable for anaerobic processes to take place such as DNRA, nitrogen fixation, fermentation, and methanogenesis. Thus we hypothesize that forest-to-pasture conversion shifts the microbial energy harvest strategy towards more anaerobic.

4.6.4 Possible mechanisms for subsisting other environmental challenges in the Amazon

The differences in genes involved in amino acid metabolism between forest and pasture microbiomes can also be related to the observed differences in environmental features, which can play important role in the adaptation of microbiome in their habitats (Fierer et al. 2012, Mary et al. 2008, Greenblum et al. 2012). For instance, amino acid metabolism produces putrescine, which offers resistance to acidic pH (Konings et al. 1997, Griswold et al. 2006). The genes encoding enzymes for putrescine production are overrepresented in forest microbiome, hinting that it may contribute substantially in the adaptation of microbial community to the acidic soil habitats. More importantly, higher biosynthetic capacity of amino acids (Supplementary Table 4-3) from acidic amino acids might be advantageous for the forest microbiome in their oligotrophic habitat. Another forest-enriched enzyme, homospermidine synthase (EC2.5.1.44), which is probably involved in protecting plant community from herbivores (Ober et al. 1999).

It is conceivable that pasture microbial communities had to endure the slash-and-burn process and then adapt a lifestyle that is different from pristine forest. Despite the

increased carbon content, the deforestation process and subsequent loss of essential elements including nitrogen might have led to the selective enrichment of small group of microbial communities with special structures, such as flagella and endospores, which help adapt to the altered ecosystem. Our finding is in line with previous studies, where forest-to-pasture conversion increased relative abundance of known endospore-formers such as *Firmicutes* (Rodrigues et al. 2013, Chapters 2 and 3). Several authors have demonstrated that the ability of motile microbial species to move towards the gradients of chemoattractants enhances their ability to establish colonization in the rhizosphere (Ames and Bergman 1981, Bais et al. 2006, Caetano-Anollés et al. 1988), which enables bacteria to respond quickly to ecosystem alteration, especially those that are not part of native soil microbiota.

4.7 Final remarks

Microbial diversity in soils has a great influence on ecosystem functions. Anthropogenic activities such as deforestation alter microbial communities, which thereafter lead to the alteration of functional attributes. Since the pristine Amazon soil harbors a major portion of microbial community, many of the microbial species remain undefined based on our current knowledge. While defined, biology of many of them is not fully understood, which challenges us to explore and compare their complete ecological roles. This study attempted to overcome this limitation. Metagenome datasets used in this study are of sufficient quality to allow the identification of KEGG orthology genes and quantity to obtain the full extent of genomic diversity. The identification of genes, which were differentially abundant between forest and pasture, offers insights for

the strategies of microbial communities in their ecosystems, which may contribute to of adaptation. Such information is essentially important for environments like Amazon soils, where the basic ecology and microbiology are mostly inconspicuous. We found that ecosystem conversion altered microbial lifestyle strategies, which in turn alters ecologically relevant biochemical pathways.

Chapter 5

Conclusions, caveats, and future directions

5.1 Conclusion

Microbial diversity in soils has a great influence on ecosystem functions. Biodiversity study in microbial ecology traditionally demonstrates mainly abundances and species richness and ignores the microbial interactions, which could be more important to understand ecosystem processes with central importance. I used co-occurrence network approaches to analyze the microbial communities of Amazon forest and pasture. My studies showed how altered microbial communities lead to the alteration of co-occurrence patterns and topological properties of networks. In addition to the network properties, the association of soil properties, especially carbon and nitrogen content with the topological properties of co-occurrence networks provided valuable insights about microbial associations with ecosystem functions, which are consistent with the results that I observed with metagenome analyses. In addition, the analyses in my research projects attest to the previously observed alteration of functions caused by forest-to-pasture conversion. Transcriptomic and proteomic approaches are also widely used in studying the ecosystem functions. However, they are more expensive and may not always provide better comprehension about the microbial roles in ecosystem functions. Many microbial genes are only expressed in a certain situation, for example nitrogenase genes are only expressed during nitrogen fixation. Our approach, however, may require further validation using other approaches and an integration of metagenomics, cultivation and lab experiments should be conducted in order to link the functional genes with microbial taxa (Fuhrman 2009). Nonetheless, network analysis

provides us with the starting point in understanding the ecology of these microorganisms. Such information is essentially important for environments like Amazon soils, where the basic ecology and microbiology are mostly inconspicuous. Therefore, this network strategy can guide us to identify habitat affinities of microbial groups, and potentially their shared physiologies that could govern more focused studies and design experimental settings accordingly.

5.2 Caveats

Despite the usefulness of co-occurrence networks, we have to be very careful in inferring microbial interactions from network analysis as they only demonstrate the association based on their relative abundance pattern. Since the abundance patterns of microorganism are a result of complex biotic and abiotic interactions, co-occurrence patterns therefore do not prove true ecological interactions. Nevertheless, the novel strategy that I described in my study may increase the probability of identifying co-occurring microorganisms with physical and physiological interactions. In addition, networks described here provide a snapshot of the co-occurring microbial communities at a given time, which may not explain some important phenomena, for example, responses to perturbation, succession etc. On the other hand, we have to be very cautious in coupling ecological attributes and taxonomy for several reasons. First, lateral gene transfer allows microbes to gain and lose genes rapidly, and it is common in microbial lineages that share similar ecologies (Smillie et al. 2011). The high degree of genomic plasticity in microbial communities therefore promotes functional convergence within distantly related microbial species (Tettelin et al. 2008). Ecosystem conversion probably

triggers this process to adapt microbial communities to new habitats. Second, since the pristine Amazon soil harbors novel microorganisms, many of the operational taxonomic units (OTUs) remain undefined based on our current knowledge. When defined, the biology of many of them is not fully understood, which challenges us to explore and compare the complete ecological roles. Therefore we cannot predict if any contribution played by a forest OTU, which remains undetected in pasture can be rescued by currently undefined member in pasture.

5.3 Future directions

5.3.1 Study of microbial dynamics

My study has broadened our understanding of microbial co-occurrence patterns and functional attributes, and impact on them by ecosystem conversion in the Amazon. However, it focused on the effect of spatial variability and ignored that of temporal variability, which could be more important for understanding molecular mechanisms in successional shifts and the stability of microbial communities, and the projection of ecological consequences due to ecosystem conversion. In the future, I propose to collect samples from Amazon forest and pasture soils at both spatial and temporal scales. While we can follow the same sampling strategy across ecosystems, I propose to collect samples at four different seasons of a year for at least three consecutive years. This sampling scheme is designed to offer sufficient replications at both spatial and temporal scales such that statistically rigorous techniques could be used to determine the response of microbial taxonomic and functional diversities to seasonal variability in both ecosystems. Microbial networks can be constructed with datasets of multiple time points

that would contribute to understand a dynamic structure of microbial communities, and may reveal true ecological interactions between microbial species. The incorporation of environmental factors to the co-occurrence networks would guide us to unearth the impact of these factors in shaping the co-occurrence patterns. In addition, this approach will identify a core set of microbial species and their co-occurrents that are consistent with temporal variation, which may guide us to identify the microbial species and conservation of the ecological community relationships required for adaptations and homeostasis. Later, supervised classification of KEGG pathways and individual KEGG ECs can be performed, for example using a randomForest classifier (Knights et al. 2011). This approach will identify a core set of functional genes that are not differentially abundant and significantly discriminatory across different time points. The integration of functional genes in the analysis can consolidate our understanding of molecular mechanisms of syntrophy between microbial species for adaptation in terrestrial ecosystem, which therefore exhibit their ecological significance, and response of their adaptation strategy to anthropogenic activities.

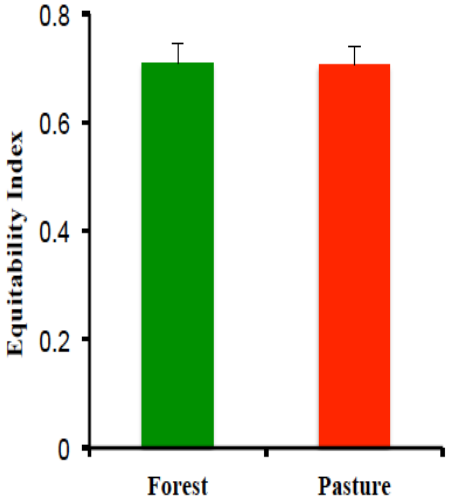
5.3.2 Secondary metabolites and bioremediation

Since the majority of Amazon soil microbial species is currently undefined, it is reasonable to expect the presence of novel genes and pathways, which are involved in the degradation of recalcitrant chemical compounds, and in the synthesis of commercially and medically important compounds. My metagenome analyses attest to the possibility to discover novel secondary metabolites and microbial species involved in bioremediation of recalcitrant compounds. Previously, the Amazon Rainforest Microbial Observatory (ARMO) team has successfully isolated 260 species belonging to *Burkholderia*, a

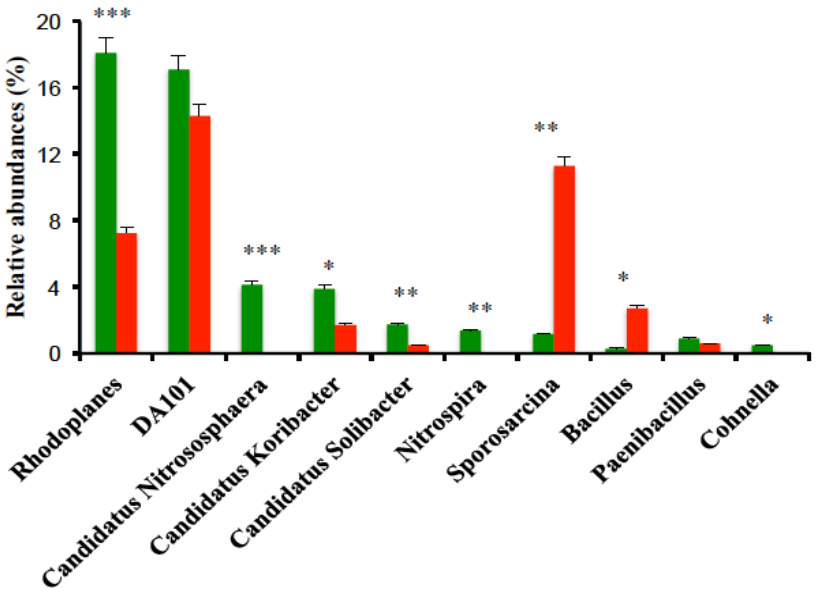
betaproteobacterial genus, with the ability of pollutant degradation. On the other hand, it is imperative to discover new antibiotic class as medically important microbial species have already become resistant to most of the major antibiotic classes discovered to date. The last novel class of antibiotics was discovered in 1987 (Silver 2011). In my study, I have estimated high frequencies of *Actinobacteria, Firmicutes* and *Proteobacteria* in Amazon soil samples. While the former two phyla are known as major antibiotic producers (de Lima et al. 2012, Silo-Suh et al. 1994, Pichard et al. 1995, Laland and Zimmer 1972), the latter taxon is known for having species, especially belonging to *Betaproteobacteria* and *Gammaproteobacteria*, for bioremediation of recalcitrant compounds (Bell et al. 2013, Lors et al. 2010). Therefore Amazon soil samples provide an ideal opportunity to explore microbial species involved in antibiotic production and bioremediations. Also, since the biology of members belonging to *Verrucomicrobia* and *Acidobacteria* are mostly unstudied, and they comprise major fractions of microbial communities in Amazon soils, they may also have candidate species of interest.

Appendix A

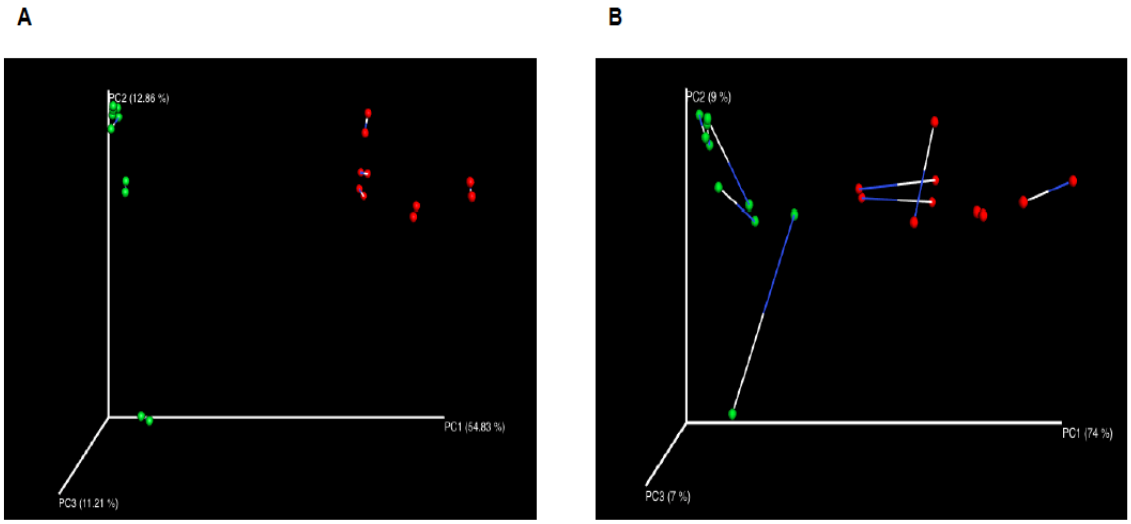
Supplementary Material for Chapter 2



Supplementary Figure 2-1. Evenness values for members comprising taxonomic networks ($P > 0.05$) in forest (green) and pasture (red). Error bars represent standard error.



Supplementary Figure 2-2. Relative abundance of 10 major genera observed in forest (green) and (red) pasture networks. Error bars represent standard error. Symbols (*), (**), and (***) indicate significance values of $P < 0.05$, $P < 0.01$, $P < 0.001$, respectively, which were calculated using two-sample *t*-test.



Supplementary Figure 2-3. Procrustes analyses of two different OTU-picking algorithms: (*A*) *de novo* vs. closed-reference, and (*B*) closed-reference OTUs vs. PICRUSt predicted KEGG Orthology gene profiles. Clustering patterns were statistically significant ($P < 0.01$) with goodness of fit (M^2) values of = 0.009 and 0.307, respectively. Forest samples are represented by green dots and pasture samples are represented by red dots in both panels. Forest samples are represented by green dots and pasture samples are represented by red dots in both panels. The white ends of each connector line connect to the OTU data derived by *de novo* OUT-picking method (*A*), and functional annotation data (*B*), whereas the blue ends connect to the OTU data derived by closed-reference OTU-picking method. The length of the connector lines is inversely proportional to the overall concordance between taxonomic and functional gene datasets.

Supplementary Table 2-1. Quantitative PCR conditions for each primer set used in this study and their approximate amplicon length.

Taxa	Primer set	Primer specificity (RDP database, Feb 2014)	Initial denaturation	Temperature -Time cycle	No. of cycle	Amplicon length (bp)	Reference
Acidobacteria	Acid31+ Eub518	96.84%	95ºC 5'	*95ºC-30", 53ºC-30", 72ºC-60"	40	500	Barns et al. 1999
Actinobacteria	Act920F3+Act1200R	91.58%	95ºC 3'	95ºC-15", 61.5ºC-30", 72ºC-27"	30	170	Gregoris et al. 2011
α-proteobacteria	α682F+ 908αR	89.8%	95ºC 3'	95ºC-15", 61.5ºC-30", 72ºC-27"	30	180	Gregoris et al.2011
β-proteobacteria	Eub338+ Bet680	99.87%	95ºC 5'	*95ºC-15", 60ºC-15", 72ºC-45"	30	360	Overmann et al. 1999
γ-proteobacteria	1080γF+ γ1202R	96.57%	95ºC 3'	95ºC-15", 61.5ºC-30", 72ºC-27"	30	170	Gregoris et al. 2011
Firmicutes	Lgc353+ Eub518	99.82%	95ºC 5'	*95ºC-15", 60ºC-15", 72ºC-45"	30	180	Meier et al. 1999
Verrucomicrobia	VER_673R+ VER_37F	99.89%	95ºC 3'	*95ºC-30", 50ºC-30", 72ºC-60"	40	650	Ranjan et al. 2015

Symbol (*) indicates that reaction conditions were experimentally determined in this study.

Supplementary Table 2-2. Standard curve properties for each taxon determined by quantitative PCR.

Taxa	Vector containing 16S fragment of	Efficiency	Standard curve equation	R^2
Acidobacteria	*Unknown	2.03	Y = -3.25X + 40.27	0.985
Actinobacteria	*Rhodococcus* sp. 4-8 (FJ905295)	1.87	Y = -3.66X + 40.29	0.990
α-proteobacteria	*Rhizobium tropici* (FN178365)	1.79	Y = -3.97X + 41.97	0.973
β-proteobacteria	*Burkholderia tropica* (EF622219)	1.98	Y = -3.37X + 41.23	0.980
γ-proteobacteria	*Xanthomonas campestris* (HQ256868)	2.14	Y = -3.03X + 34.37	0.977
Firmicutes	*Bacillus cereus* GMX6 (AM422128)	2.13	Y = -3.37X + 35.28	0.992
Verrucomicrobia	*Unknown	2.35	Y = -2.70X + 33.63	0.998

Symbol (*) Indicates that forest soil PCR amplified products were used in cloning as representatives for *Acidobacteria* and *Verrucomicrobia*.

Appendix B

Supplementary Material for Chapter 3

Supplementary Table 3-1. Major topological properties of 49 site-wise networks

Site	ρ- cutoff	Degree distribution (r-values, P<0.05)	Node counts	Edge counts	Modularity	Average path length
F1A001	0.8	-0.89	669	3387	0.903	7.926
F1A01	0.8	-0.85	560	2294	0.897	8.782
F1A1	0.8	-0.87	650	2904	0.834	7.227
F1A10	0.8	-0.87	646	3016	0.787	6.568
F1A100	0.83	-0.95	656	1566	0.983	1.563
F1B1	0.8	-0.82	656	3064	0.86	7.181
F1B10	0.8	-0.88	632	2958	0.801	6.61
F1B100	0.8	-0.85	713	4061	0.9	7.647
F1C01	0.82	-0.92	578	1620	0.969	2.632
F1C1	0.82	-0.87	1119	4831	0.969	12.191
F1C10	0.82	-0.81	1015	4411	0.972	11.179
F1C100	0.8	-0.94	483	1753	0.826	7.728
F2A01	0.8	-0.96	427	1495	0.835	9.06
F2A1	0.8	-0.94	547	2151	0.79	7.1
F2A100	0.8	-0.96	422	1368	0.834	8.318
F2B1	0.8	-0.98	435	1445	0.815	8.247
F2B10	0.8	-0.88	623	2623	0.883	7.863
F2C1	0.8	-0.88	546	2130	0.832	7.441
F3A001	0.8	-0.94	253	623	0.893	9.476
F3A01	0.8	-0.95	382	1208	0.913	11.121
F3A1	0.8	-0.94	440	1460	0.811	7.854
F3A100	0.8	-0.78	666	4250	0.921	9.128
F3B1	0.81	-0.8	768	3398	0.951	16.54
F3B10	0.8	-0.84	699	3113	0.841	7.305
F3C01	0.8	-0.92	271	731	0.871	5.787
F3C10	0.79	-0.88	312	1084	0.949	4.914
P1A1	0.81	-0.9	770	2948	0.921	9.78
P1A10	0.8	-0.77	788	4944	0.917	8.353
P1A100	0.8	-0.75	889	5569	0.868	6.684
P1B1	0.8	-0.78	1011	7181	0.855	6.132
P1B10	0.82	-0.92	1158	5182	0.964	13.355
P1B100	0.81	-0.84	910	3606	0.926	10.848
P1C01	0.81	-0.85	1038	5280	0.934	8.87
P1C1	0.8	-0.91	699	3363	0.762	6.653
P1C10	0.8	-0.9	803	4365	0.745	6.102
P1C100	0.8	-0.93	455	1791	0.887	7.509
P2A001	0.8	-0.77	1012	7976	0.868	6.224
P2A1	0.8	-0.93	596	2558	0.806	7.026
P2A10	0.81	-0.87	859	3343	0.915	9.753
P2A100	0.8	-0.95	574	2270	0.785	7.026
P2B1	0.8	-0.72	1028	8080	0.886	6.665
P2B10	0.82	-0.9	1103	4519	0.962	14.146
P2C1	0.8	-0.91	941	5937	0.738	5.97
P3A01	0.8	-0.96	684	3312	0.767	6.183
P3A1	0.8	-0.95	572	2218	0.843	7.853
P3A10	0.8	-0.92	651	2703	0.807	7.102
P3A100	0.8	-0.96	482	1832	0.792	7.477
P3B10	0.8	-0.72	996	7230	0.849	6.29
P3C01	0.8	-0.87	781	4075	0.845	6.724

Supplementary Table 3-2. Comparison of environmental variables between forest and pasture soils. Symbols (*), (**), and (***) indicate significance values of $P < 0.05$, $P < 0.01$, $P < 0.001$, respectively; NS, non-significant.

Soil variable	**Unit**	**Forest (CI, 95%)**	**Pasture (CI, 95%)**	***P*-value (two-sample *t*-test)**
Aluminum (Al)	mmolc dm^{-3}	0.88±0.56	1.00±0.62	NS
Base saturation (V)	%	50.35±6.77	40.00±6.41	*
Calcium (Ca)	mmolc dm^{-3}	22.27±6.25	19.04±7.62	NS
Carbon (C)	-	1.20±0.16	1.91±0.33	***
C/N	N/A	12.31±0.43	12.74±0.35	NS
Cation Exchange Capacity (CEC)	mmolc dm^{-3}	58.89±6.03	64.03±9.49	NS
Magnesium (mg)	mmolc dm^{-3}	6.77±0.79	7.74±1.35	NS
Moisture	**%**	14.03±0.89	17.62±2.03	**
Nitrogen (N)	-	0.10±0.01	0.15±0.03	***
pH (in $CaCl_2$)	N/A	4.3±0.24	4.5±0.19	NS
Phosphorus (P)	mg dm^{-3}	8.42±1.03	9.13±2.68	NS
Potassium (K)	mmolc dm^{-3}	2.54±0.96	1.77±0.51	NS
Proton (H^+)	mmolc dm^{-3}	27.31±2.88	35.48±2.75	**
Sulfur (S)	mg dm^{-3}	4.08±0.43	4.04±0.44	NS

Supplementary Table 3-3. Taxonomy of generalist nodes, and their average topological values in forest networks.

Forest Generalist	Closest known representative	Avg. betweenness	Avg. degree	Number of co-occurring networks
Acidobacteria	*Chloracidobacteria*	0.0022	8.09	23
	Acidobacteria-5	0.0014	5.91	23
	Acidobacteria-5	0.0014	6.09	22
	Acidobacteria-5	0.0013	7.2	20
	Acidobacteria-6	0.0035	8.27	22
	Acidobacteria-6	0.0007	5.71	21
	Acidobacteria-6	0.0006	5.83	23
	Acidobacteria-6	0.0009	6.55	22
	Koribacteraceae	0.0005	5.39	23
	Koribacteraceae	0.0018	7.4	20
	Candidatus Koribacter	0.0014	6.08	24
	Candidatus Koribacter	0.0008	5	26
	Koribacteraceae	0.0004	6.43	23
	Ellin6513	0.0028	7.9	20
	Ellin6513	0.001	5.64	22
	Candidatus Solibacter	0.0013	5.6	25
	Candidatus Solibacter	0.0012	6.92	24
	Candidatus Solibacter	0.0012	6.82	22
	Candidatus Solibacter	0.001	7.62	21
	Candidatus Solibacter	0.0012	4.8	25
Actinobacteria	*Actinomycetales*	0.0003	5.64	22
	MB-A2-108	0.0002	5.09	22
Alphaproteobacteria	*Beijerinckiaceae*	0.0006	6.69	26
	Hyphomicrobiaceae	0.0003	5.64	22
	Rhodospirillaceae	0.0008	5.62	26
	Rhodoplanes	0.0013	7.14	21
	Rhodospirillaceae	0.0008	7.14	21
	Methylocystaceae	0.0006	7.3	20
	Rhodospirillaceae	0.0003	5.12	25
	Rhodospirillaceae	0.0013	6.78	23
	Rhodospirillaceae	0.0013	6.35	23
	Caulobacteraceae	0.001	6.4	20
	Bradyrhizobiaceae	0.0012	6.7	23
	Rhodoplanes	0.001	7.27	22
Betaproteobacteria	*Betaproteobacteria*	0.0051	7	20
Chloroflexi	*Ktedonobacteria*	0.0018	5.71	21
Deltaproteobacteria	*Syntrophobacteraceae*	0.0009	5.9	20
	Syntrophobacteraceae	0.0014	6.86	21
	Entotheonellaceae	0.0002	5.6	20
	Syntrophobacteraceae	0.0013	5.31	26
GAL15	*GAL15*	0.0006	5.64	22
Gemmatimonadetes	*Gemm-1*	0.0015	6.57	21
	Gemm-1	0.0009	6.3	20
	Gemm-1	0.001	6.38	26
	Ellin5290	0.0006	4.83	24
	Gemmatimonadetes	0.0029	5.82	22
	Ellin5290	0.0009	7.05	21
	Ellin5290	0.0006	5.17	24
Nitrospirae	*Nitrospirales*	0.0008	5.69	26
	Nitrospirales	0.0002	4.91	22
	Nitrospirales	0.0017	5.52	25
Planctomycetes	*Pirellulaceae*	0.0003	5.13	23
	Gemmataceae	0.0008	6	21

Supplementary Table 3-4. Taxonomy of generalist nodes, and their average topological values in pasture networks.

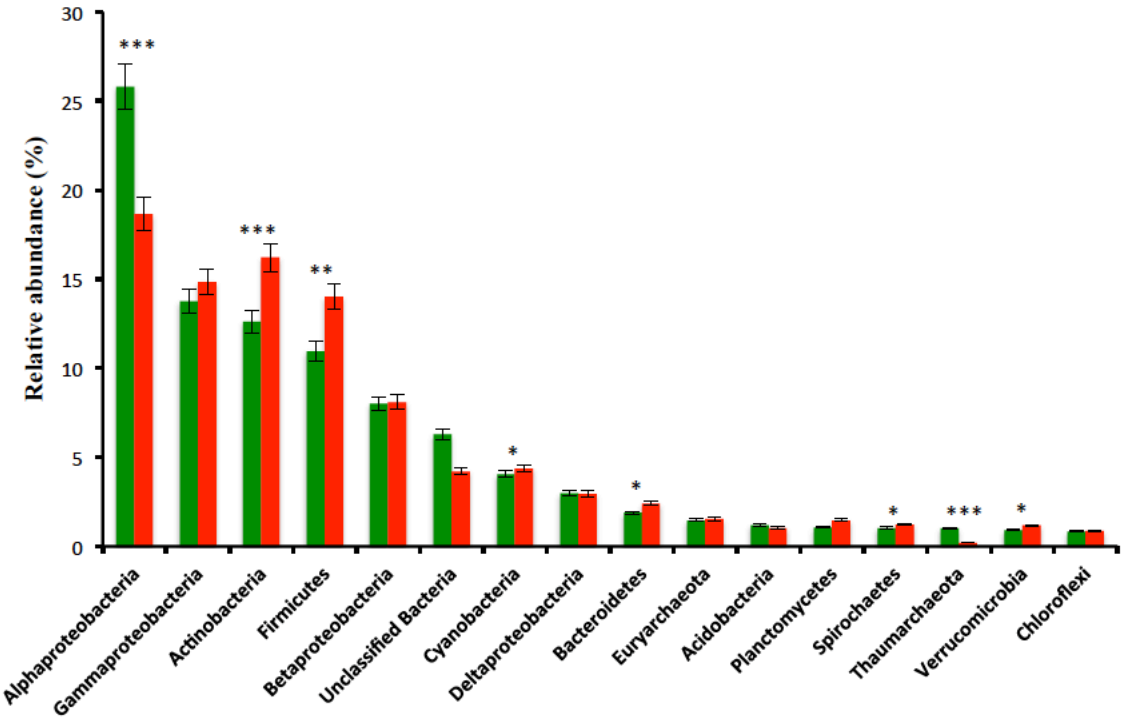
Pasture Generalist	Closest known representative	Avg. betweenness	Avg. degree	Number of co-occurring networks
Acidobacteria	*RB41*	0.0009	6.36	22
	Chloracidobacteria	0.0009	7.71	21
	Acidobacteria-6	0.0014	11.00	18
	Acidobacteria-6	0.0004	6.00	19
	Candidatus Koribacter	0.0012	7.81	21
	Koribacteraceae	0.0016	8.38	21
	Koribacteraceae	0.0011	9.16	19
	Micromonosporaceae	0.0008	9.00	18
	Nocardioides	0.0012	8.78	18
	Nocardioidaceae	0.0016	9.81	21
	Gaiellaceae	0.0004	6.10	21
	Solirubrobacteraceae	0.0008	8.00	20
	Gaiellaceae	0.0012	8.17	23
	Solirubrobacterales	0.0010	8.00	20
Alphaproteobacteria	*Rhodospirillaceae*	0.0012	10.63	19
	Labrys	0.0008	9.00	20
	Rickettsiales	0.0014	10.22	18
	Alphaproteobacteria	0.0008	7.67	18
	Rhodospirillaceae	0.0007	9.68	19
	Rhodospirillaceae	0.0010	8.00	18
	Rhodoplanes	0.0011	8.67	18
Betaproteobacteria	*Betaproteobacteria*	0.0008	9.44	18
	Betaproteobacteria	0.0009	7.68	19
	Betaproteobacteria	0.0019	9.05	19
	Comamonadaceae	0.0006	12.22	18
	Betaproteobacteria	0.0006	8.21	19
	Betaproteobacteria	0.0009	6.26	23
Deltaproteobacteria	*Myxococcales*	0.0010	8.29	21
	Myxococcales	0.0010	10.60	20
	Syntrophobacteraceae	0.0018	10.30	20
	Syntrophobacteraceae	0.0009	6.70	20
	Deltaproteobacteria	0.0009	8.82	22
	Deltaproteobacteria	0.0009	13.00	18
	Anaeromyxobacter	0.0004	7.16	19
Firmicutes	*Bacillales*	0.0010	6.55	22
	Bacillales	0.0003	8.70	20
	Bacillus cereus	0.0013	8.19	21
	Alicyclobacillus	0.0004	6.10	21
	Bacillus	0.0009	9.44	18
	Solibacillus	0.0010	7.58	19
	Planococcaceae	0.0008	7.55	22
	Clostridium	0.0006	6.87	23
	Veillonellaceae	0.0007	8.11	18
	Clostridiales	0.0007	8.00	20
Gammaproteobacteria	*Piscirickettsiaceae*	0.0007	9.33	18
Gemmatimonadetes	*Gemmatimonadetes*	0.0005	7.44	18
WS3	*PRR-12*	0.0016	7.48	23
	PRR-12	0.0005	6.36	22

Supplementary Table 3-5. Taxonomy of shared generalist nodes, and their average topological values in forest and pasture networks.

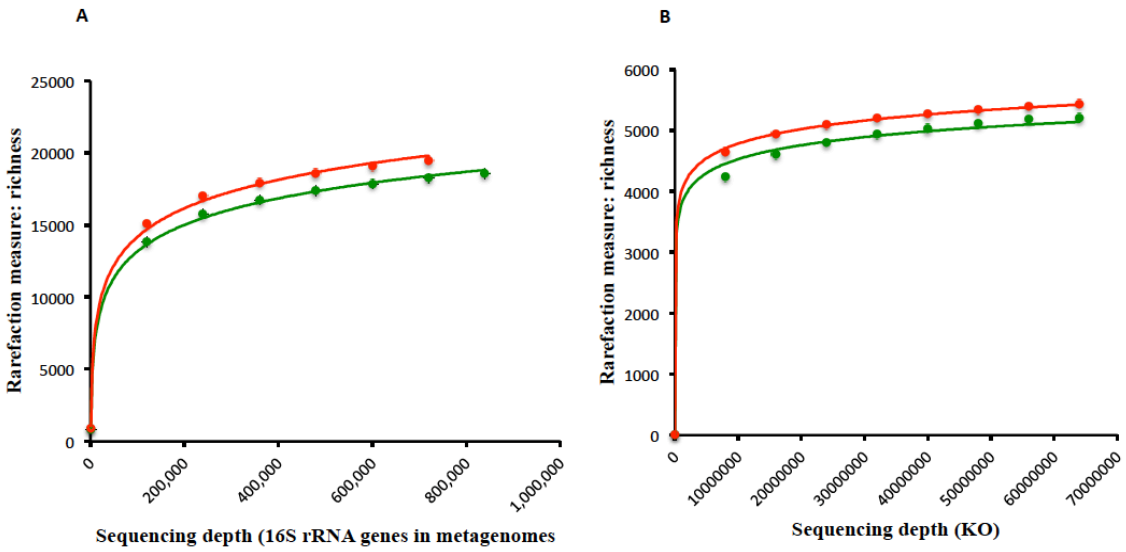
Shared generalist	Closed known representative	Avg. betweenness		Avg. degree		No of co-occurring networks	
		Forest	Pasture	Forest	Pasture	Forest	Pasture
Acidobacteria	*Acidobacteria-5*	0.0023	0.0006	4.77	5.5	26	20
	Acidobacteria-5	0.0007	0.0004	5.54	7.39	26	23
	Acidobacteria-5	0.0021	0.0006	6.72	8.44	25	18
	Acidobacteria-6	0.0014	0.0009	7.42	7.55	24	22
	Acidobacteria-6	0.0015	0.0008	5.54	11.11	26	18
	Acidobacteria-6	0.0009	0.0006	7	6.8	24	20
	Acidobacteria-6	0.0006	0.0016	5.23	9.04	26	23
	Acidobacteria-6	0.0014	0.0007	6.24	6.64	25	22
	Edaphobacter modestus	0.001	0.0006	5.92	7.89	24	18
	Koribacteraceae	0.0023	0.0007	5.84	9.78	25	18
	Koribacteraceae	0.0006	0.0013	5.46	7.52	26	21
	iii1-8	0.0002	0.0012	5.2	7.18	25	22
	Candidatus Solibacter	0.0012	0.0004	5.69	6.17	26	23
	Candidatus Solibacter	0.0029	0.0007	5.92	6.35	26	23
Actinobacteria	*Acidimicrobiia*	0.0012	0.0007	6.61	7.2	23	20
	Actinomycetales	0.0032	0.001	5.52	9.89	21	18
	Gaiellaceae	0.0017	0.0011	6	7.48	23	23
	Conexibacteraceae	0.0004	0.0012	5.62	7.91	26	22
	Gaiellaceae	0.0007	0.0006	7	8.11	20	19
	Gaiellaceae	0.0007	0.0013	6.6	9.1	20	20
Alphaproteobacteria	*Rhizobiales*	0.0012	0.0004	7	6.55	20	22
	Rhodoplanes	0.0031	0.0008	6.46	8.44	26	18
	Phenylobacterium	0.0007	0.0008	7	7.18	26	22
	Rhodoplanes	0.0013	0.0008	6.08	7.39	26	23
	Bradyrhizobiaceae	0.0002	0.0005	5.62	9.22	26	23
	Alphaproteobacteria	0.0007	0.0003	5.85	5.57	26	23
	Rhodoplanes	0.0006	0.0009	5.92	6.35	25	23
	Hyphomicrobiaceae	0.0015	0.0004	7.13	6.26	23	23
	Rhodospirillaceae	0.0005	0.0012	5.9	9.26	21	19
	Rhodoplanes	0.0009	0.001	5.31	8.17	26	23
	Pedomicrobium	0.0026	0.0004	6	5.64	26	22
	Rhodoplanes	0.0024	0.0005	5.84	8.21	25	19
	Hyphomicrobiaceae	0.0034	0.0006	7.39	7.79	23	19
	Rhodoplanes	0.0004	0.0006	5.54	5.22	26	23
Chloroflexi	*Ellin6529*	0.0008	0.0013	6.67	7.57	21	23
	Ktedonobacteria	0.002	0.0005	5.08	6.09	26	22
	B07_WMSP1	0.0037	0.0019	6.73	8.1	22	21
Deltaproteobacteria	*Myxococcales*	0.0002	0.0006	6.4	6.61	20	23
	Syntrophobacteraceae	0.0046	0.0008	6	7.09	26	22
	Syntrophobacteraceae	0.0005	0.0006	5.85	7.3	26	23
	Syntrophobacteraceae	0.0008	0.0005	5.91	9.2	22	20
	Syntrophobacteraceae	0.0009	0.0006	5.54	6.78	26	23
Firmicutes	*Bacillus*	0.0021	0.0007	7.81	7.83	21	23
	Bacillus	0.0015	0.0006	7.14	6.17	21	23
Gammaproteobacteria	*Sinobacteraceae*	0.0016	0.0009	6.08	8.84	26	19
Gemmatimonadetes	*Ellin5290*	0.0025	0.0011	6.6	8.22	20	18
Nitrospirae	*Nitrospirales*	0.0013	0.0006	5.77	6.17	26	23
	Nitrospirales	0.0007	0.0005	6.6	8.63	20	19
	Nitrospirales	0.0016	0.0007	6.85	7.43	26	21
Planctomycetes	*Pla4*	0.001	0.0007	6.24	6.09	25	22

Appendix C

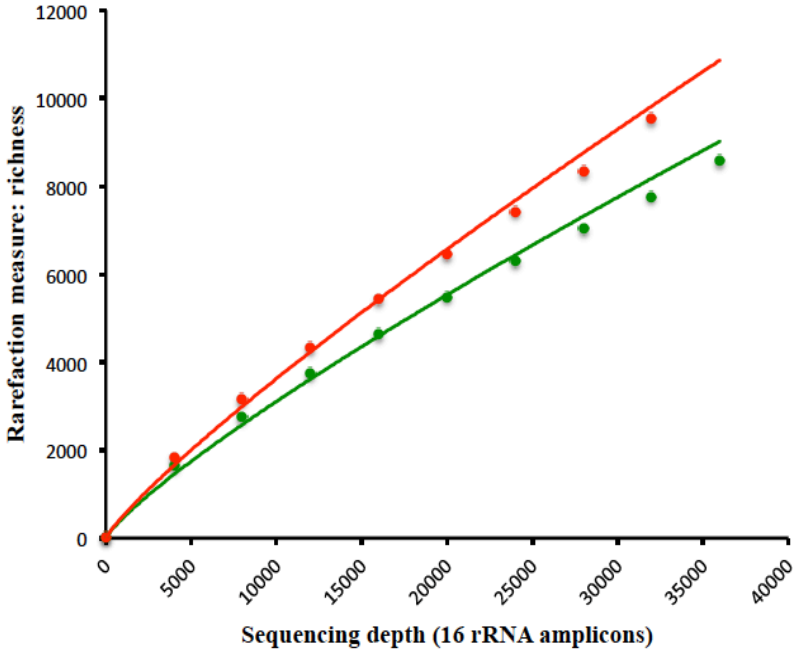
Supplementary Material for Chapter 4



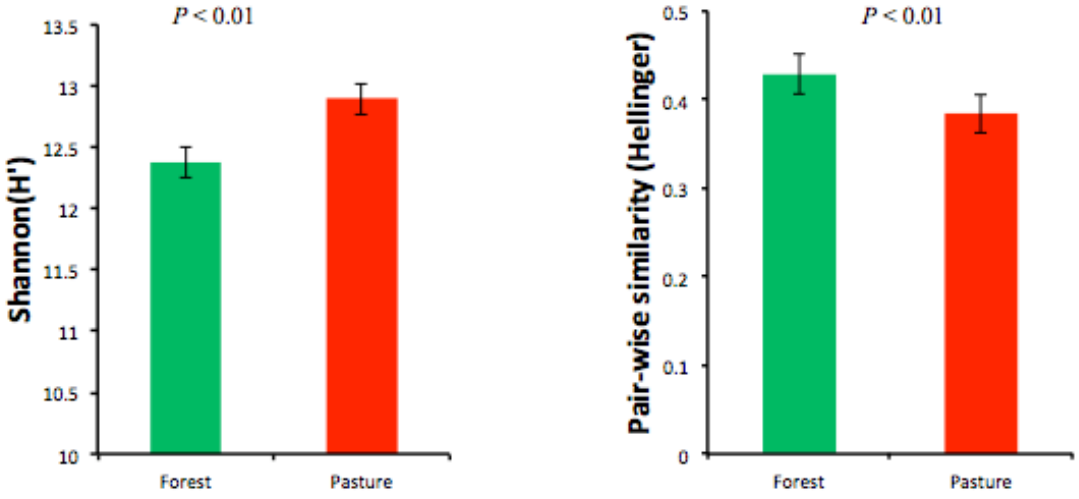
Supplementary Figure 4-1. Relative abundances of taxonomic distribution at phylum and class levels in the shotgun metagenomes obtained from the Amazon forest (green) and pasture (red). Error bars represent standard error. Symbols (*), (**), and (***) indicate significance values of $P < 0.05$, $P < 0.01$, $P < 0.001$, respectively, which were calculated using two-sample *t*-test.



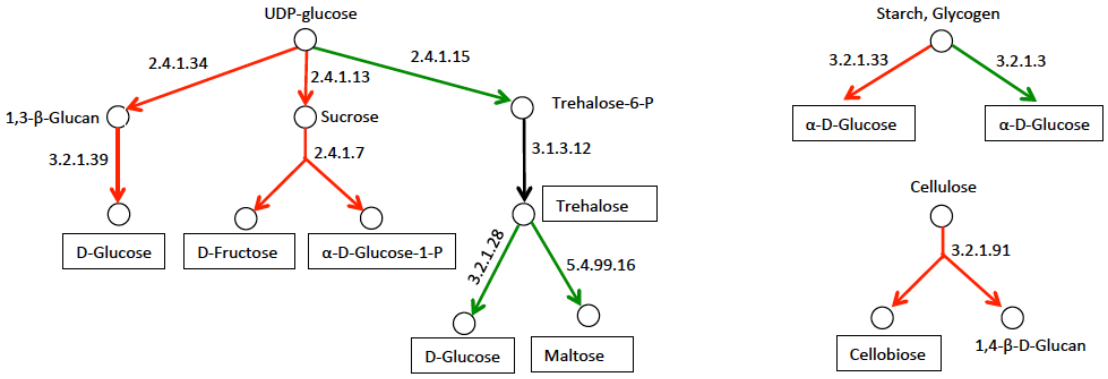
Supplementary Figure 4-2. Rarefaction curves of taxonomic (A) and functional (B) distributions of metagenomes obtained from Amazon forest (green) and pasture (red) soil samples, where species richness was used as alpha-diversity index. Each line connects an average number of observed OTUs (A) and KOs (B) at each rarefaction depth.



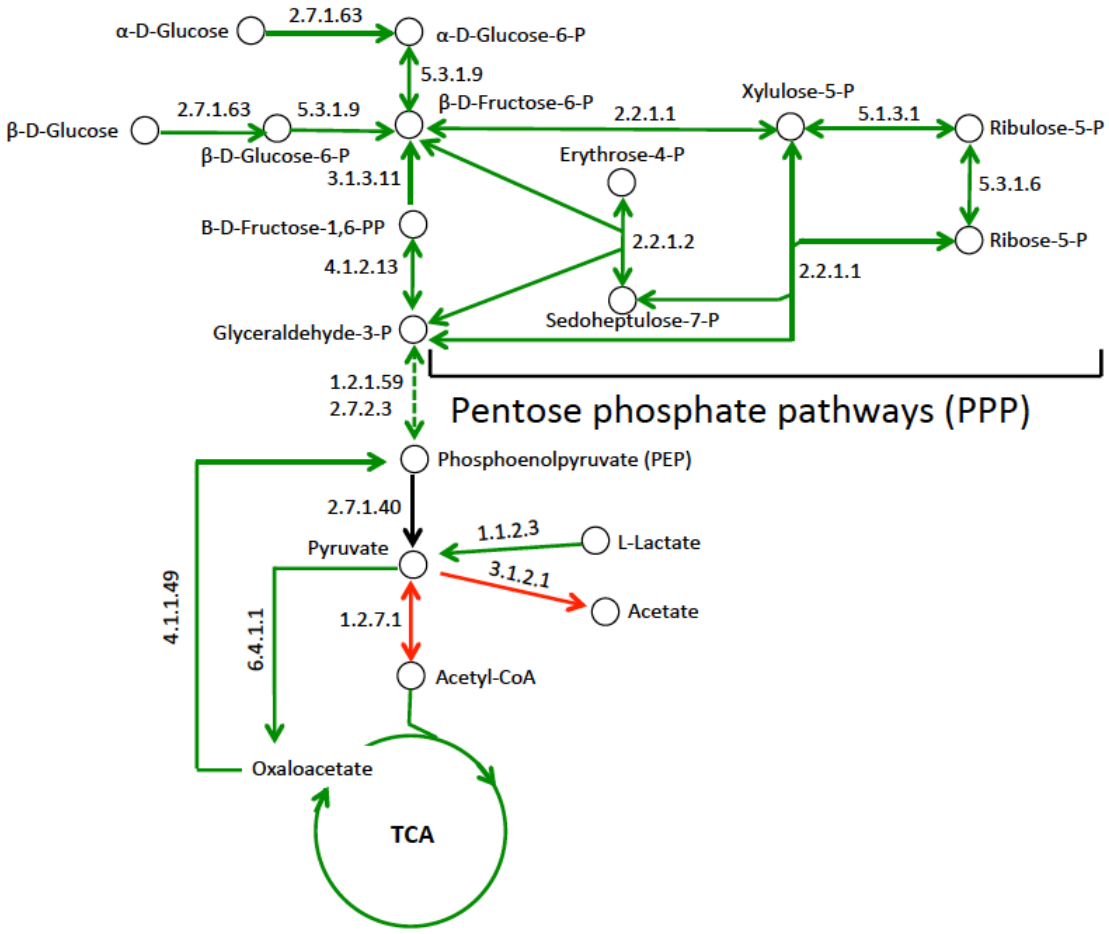
Supplementary Figure 4-3. Rarefaction curves of 16S rRNA gene amplicons obtained from a recently published Amazon forest (green) and pasture (red) soil samples (**Chapter 2**), where species richness was used as alpha-diversity index. Each line connects an average number of observed OTUs at each rarefaction depth.



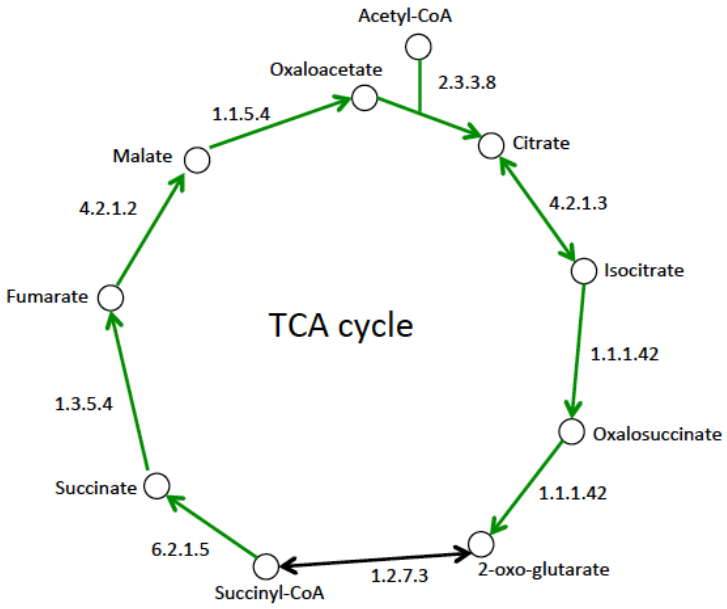
Supplementary Figure 4-4. Response of soil ecosystem conversion to taxonomic diversities in shotgun metagenomes. (A) Alpha diversity (Shannon, *H'*), $P < 0.01$; (B) Beta diversity (Hellinger), $P < 0.01$. Means (n = 5 for alpha diversity, n = 10 for pair-wise similarity as beta diversity) are depicted, standard error. Error bars represent standard error.



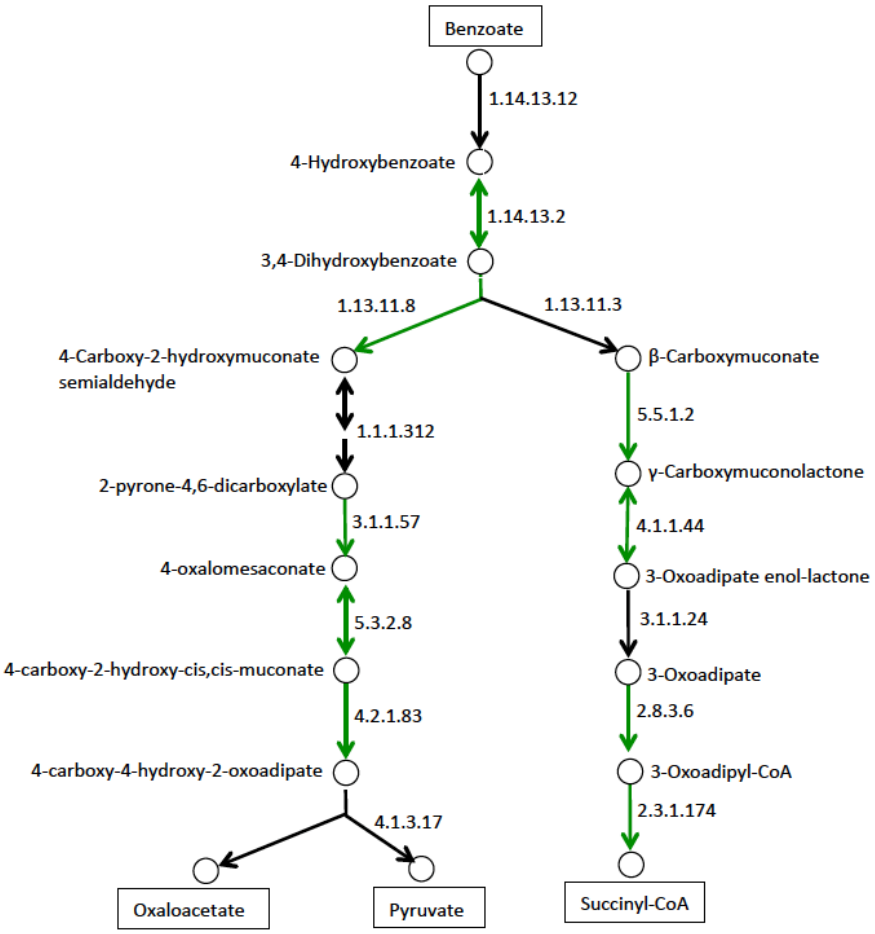
Supplementary Figure 4-5. Response of soil ecosystem conversion in the differential representation of genes encoding ECs involved in the starch and sucrose metabolism (carbohydrate metabolism). Arrows indicate the enzyme-mediated steps in the pathway, where green arrows indicate genes enriched in forest, red arrows indicate genes enriched in pasture, and black arrows indicate genes that were not estimated differentially abundant between forest and pasture.



Supplementary Figure 4-6. Response of soil ecosystem conversion in the differential representation of genes encoding ECs involved in monosaccharide metabolism (carbohydrate metabolism). Arrows indicate the enzyme-mediated steps in the pathway, where green arrows indicate genes enriched in forest, red arrows indicate genes enriched in pasture, and black arrows indicate genes that were not estimated differentially abundant between forest and pasture.



Supplementary Figure 4-7. Response of soil ecosystem conversion in the differential representation of genes encoding ECs involved in the TCA cycle (carbohydrate metabolism). Arrows indicate the enzyme-mediated steps in the pathway, where green arrows indicate genes enriched in forest, and black arrows indicate genes that were not estimated differentially abundant between forest and pasture.



Supplementary Figure 4-8. Response of soil ecosystem conversion in the differential representation of genes encoding ECs involved in benzoate metabolism. Arrows indicate the enzyme-mediated steps in the pathway, where green arrows indicate genes enriched in forest, and black arrows indicate genes that were not estimated differentially abundant between forest and pasture.

Supplementary Table 4-1. Frequencies of genes involved in carbohydrate metabolism between forest and pasture. The positive values of log_2foldchange indicate the extent of increased genes frequencies in forest and negative values of log_2foldchange indicate the extent of increased genes frequencies in pasture.

KO ID	EC	Forest	Pasture	log_2foldchange	KEGG Level 3	EC Annotation
K01903	EC:6.2.1.5	0.1481 (±0.0042)	0.1478 (±0.0022)	0.003	TCA cycle	Succinyl-coa synthetase beta subunit
K01902	EC:6.2.1.5	0.1173 (±0.003)	0.1157 (±0.0042)	0.02	TCA cycle	Succinyl-coa synthetase alpha subunit
K01676	EC:4.2.1.2	0.0869 (±0.007)	0.084 (±0.0051)	0.049	TCA cycle	Fumarate hydratase, class I
K00241	unknown	0.0211 (±0.001)	0.0193 (±0.0011)	0.129	TCA cycle	Succinate dehydrogenase cytochrome b556 subunit
K01959	EC:6.4.1.1	0.0134 (±0.001)	0.0115 (±0.0006)	0.215	TCA cycle	Pyruvate carboxylase subunit A
K00242	unknown	0.0082 (±0.0005)	0.0065 (±0.0003)	0.344	TCA cycle	Succinate dehydrogenase membrane anchor subunit
K01960	EC:6.4.1.1	0.0092 (±0.001)	0.0055 (±0.0008)	0.746	TCA cycle	Pyruvate carboxylase subunit B
K01648	EC:2.3.3.8	0.0028 (±0.0004)	0.001 (±0.0002)	1.447	TCA cycle	ATP citrate (pro-S)-lyase
K01182	EC:3.2.1.10	0.0034 (±0.0003)	0.0078 (±0.0007)	-1.207	Galactose metabolism	Oligo-1,6-glucosidase
K01193	EC:3.2.1.26	0.0022 (±0.0005)	0.0044 (±0.0003)	-1.042	Galactose metabolism	Beta-fructofuranosidase
K00883	EC:2.7.1.58	0.0011 (±0.0004)	0.0024 (±0.0003)	-1.101	Galactose metabolism	2-dehydro-3-deoxygalactonokinase
K00927	EC:2.7.2.3	0.1479 (±0.0037)	0.1472 (±0.0023)	0.007	Glycolysis / Gluconeogenesis	Phosphoglycerate kinase
K01810	EC:5.3.1.9	0.1862 (±0.0092)	0.1818 (±0.0056)	0.035	Glycolysis / Gluconeogenesis	Glucose-6-phosphate isomerase
K00172	EC:1.2.7.1	0.0021 (±0.0003)	0.0054 (±0.0008)	-1.391	Glycolysis / Gluconeogenesis	Pyruvate ferredoxin oxidoreductase, gamma subunit
K03841	EC:3.1.3.11	0.0206 (±0.003)	0.0197 (±0.0014)	0.065	Glycolysis / Gluconeogenesis	Fructose-1,6-bisphosphatase I
K00886	EC:2.7.1.63	0.0285 (±0.0022)	0.0259 (±0.0016)	0.136	Glycolysis / Gluconeogenesis	Polyphosphate glucokinase
K01623	EC:4.1.2.13	0.1264 (±0.0044)	0.1099 (±0.0027)	0.201	Glycolysis / Gluconeogenesis	Fructose-bisphosphate aldolase, class I
K01610	EC:4.1.1.49	0.1087 (±0.0037)	0.0927 (±0.0017)	0.23	Glycolysis / Gluconeogenesis	Phosphoenolpyruvate carboxykinase (ATP)
K00150	EC:1.2.1.59	0.0022 (±0.0003)	0.0012 (±0.0001)	0.835	Glycolysis / Gluconeogenesis	Glyceraldehyde-3-phosphate dehydrogenase (NAD(P))
K00616	EC:2.2.1.2	0.166 (±0.0061)	0.1543 (±0.0051)	0.106	Pentose phosphate pathway	Transaldolase
K00615	EC:2.2.1.1	0.4277 (±0.0141)	0.3979 (±0.0073)	0.104	Pentose phosphate pathway	Transketolase
K00090	EC:1.1.1.215	0.0099 (±0.0005)	0.0095 (±0.0004)	0.063	Pentose phosphate pathway	Gluconate 2-dehydrogenase
K01783	EC:5.1.3.1	0.0731 (±0.0017)	0.0652 (±0.0015)	0.165	Pentose phosphate pathway	Ribulose-phosphate 3-epimerase
K05774	EC:2.7.4.23	0.0018 (±0.0004)	0.0015 (±0.0002)	0.21	Pentose phosphate pathway	Ribose 1,5-bisphosphokinase
K01807	EC:5.3.1.6	0.0302 (±0.0026)	0.0271 (±0.0017)	0.16	Pentose phosphate pathway	Ribose 5-phosphate isomerase A
K01053	EC:3.1.1.17	0.1457 (±0.0071)	0.1109 (±0.0037)	0.394	Pentose phosphate pathway	Gluconolactonase
K00034	EC:1.1.1.47	0.0481 (±0.0057)	0.0319 (±0.002)	0.591	Pentose phosphate pathway	Glucose 1-dehydrogenase
K00115	EC:1.1.99.10	0.0063 (±0.0007)	0.0041 (±0.0005)	0.623	Pentose phosphate pathway	Glucose dehydrogenase (acceptor)
K06152	EC:1.1.99.3	0.0119 (±0.0008)	0.0076 (±0.0007)	0.638	Pentose phosphate pathway	Gluconate 2-dehydrogenase gamma chain
K06151	EC:1.1.99.3	0.0848 (±0.0083)	0.0505 (±0.0045)	0.748	Pentose phosphate pathway	Gluconate 2-dehydrogenase alpha chain
K00117	EC:1.1.5.2	0.1525 (±0.0099)	0.0984 (±0.0054)	0.633	Pentose phosphate pathway	Quinoprotein glucose dehydrogenase
K00101	EC:1.1.2.3	0.0864 (±0.0056)	0.0718 (±0.0029)	0.267	Pyruvate metabolism	L-lactate dehydrogenase (cytochrome)
K00116	EC:1.1.5.4	0.0291 (±0.0049)	0.0194 (±0.001)	0.581	TCA cycle	Malate dehydrogenase (quinone)
K01067	EC:3.1.2.1	0.0032 (±0.0002)	0.0101 (±0.0016)	-1.641	Pyruvate metabolism	Acetyl-coa hydrolase
K01225	EC:3.2.1.91	0.001 (±0.0003)	0.0024 (±0.0005)	-1.184	Starch and sucrose metabolism	Cellulose 1,4-beta-cellobiosidase
K01199	EC:3.2.1.39	0.0002 (±0.0001)	0.0004 (±0.0001)	-1.264	Starch and sucrose metabolism	Glucan endo-1,3-beta-D-glucosidase
K00706	EC:2.4.1.34	0.0001 (±0.0001)	0.0009 (±0.0003)	-2.649	Starch and sucrose metabolism	1,3-beta-glucan synthase
K05343	EC:5.4.99.16/ 3.2.1.1	0.2363 (±0.0133)	0.223 (±0.0051)	0.084	Starch and sucrose metabolism	Maltose alpha-D-glucosyltransferase/ alpha-amylase
K00697	EC:2.4.1.15	0.1141 (±0.0047)	0.0979 (±0.0036)	0.221	Starch and sucrose metabolism	Trehalose 6-phosphate synthase
K01178	EC:3.2.1.3	0.0404 (±0.0026)	0.0331 (±0.0032)	0.288	Starch and sucrose metabolism	Glucoamylase
K01194	EC:3.2.1.28	0.0159 (±0.0011)	0.0128 (±0.0015)	0.307	Starch and sucrose metabolism	Alpha,alpha-trehalase
K01196	EC:2.4.1.25/ 3.2.1.33	0.0002 (±0.0001)	0.0007 (±0.0003)	-2.221	Starch and sucrose metabolism	Glycogen debranching enzyme
K00690	EC:2.4.1.7	0.0013 (±0.0002)	0.0037 (±0.0003)	-1.478	Starch and sucrose metabolism	Sucrose phosphorylase
K00695	EC:2.4.1.13	0.0001 (±0)	0.0007 (±0.0002)	-2.846	Starch and sucrose metabolism	Sucrose synthase
K10118	unknown	0.0021 (±0.0005)	0.0046 (±0.0004)	-1.127	ABC transporters	Multiple sugar transport system permease protein
K10232	unknown	0.0028 (±0.0008)	0.0067 (±0.001)	-1.239	ABC transporters	Alpha-glucoside transport system substrate-binding protein
K10553	unknown	0.0035 (±0.0008)	0.0074 (±0.0002)	-1.094	ABC transporters	Fructose transport system permease protein
K10240	unknown	0.0019 (±0.0005)	0.0053 (±0.001)	-1.512	ABC transporters	Cellobiose transport system substrate-binding protein
K10117	unknown	0.0008 (±0.0002)	0.0018 (±0.0002)	-1.186	ABC transporters	Multiple sugar transport system substrate-binding protein
K10241	unknown	0.0024 (±0.0007)	0.0061 (±0.0011)	-1.344	ABC transporters	Cellobiose transport system permease protein
K10242	unknown	0.0023 (±0.0007)	0.0064 (±0.0012)	-1.503	ABC transporters	Cellobiose transport system permease protein
K10984	EC:2.7.1.69	0 (±0)	0 (±0)	-2.388	Phosphotransferase system (PTS)	PTS system, galactosamine-specific IIB component
K10986	unknown	0 (±0)	0 (±0)	-2.574	Phosphotransferase system (PTS)	PTS system, galactosamine-specific IID component
K02790	EC:2.7.1.69	0 (±0)	0.0002 (±0.0001)	-1.893	Phosphotransferase system (PTS)	PTS system, maltose and glucose-specific IIB component
K02791	unknown	0 (±0)	0.0002 (±0.0001)	-2.145	Phosphotransferase system (PTS)	PTS system, maltose and glucose-specific IIC component
K02761	unknown	0.0003 (±0.0001)	0.001 (±0.0001)	-1.525	Phosphotransferase system (PTS)	PTS system, cellobiose-specific IIC component
K02794	EC:2.7.1.69	0.0003 (±0.0001)	0.0009 (±0.0001)	-1.535	Phosphotransferase system (PTS)	PTS system, mannose-specific IIB component
K02795	unknown	0.0004 (±0.0001)	0.0011 (±0.0002)	-1.56	Phosphotransferase system (PTS)	PTS system, mannose-specific IIC component
K02793	EC:2.7.1.69	0.0122 (±0.0008)	0.0088 (±0.0005)	0.472	Phosphotransferase system (PTS)	PTS system, mannose-specific IIA component
K02760	EC:2.7.1.69	0 (±0)	0.0002 (±0)	-2.418	Phosphotransferase system (PTS)	PTS system, cellobiose-specific IIB component
K02757	unknown	0.0002 (±0)	0.0009 (±0.0003)	-2.124	Phosphotransferase system (PTS)	PTS system, beta-glucosides-specific IIC component
K02756	EC:2.7.1.69	0.0002 (±0.0001)	0.0009 (±0.0003)	-2.114	Phosphotransferase system (PTS)	PTS system, beta-glucosides-specific IIB component
K02755	EC:2.7.1.69	0.0002 (±0)	0.0009 (±0.0003)	-2.157	Phosphotransferase system	PTS system, beta-glucosides-specific IIA component

					(PTS)	
K02768	EC:2.7.1.69	0.0067 (±0.0019)	0.0191 (±0.0013)	-1.518	Phosphotransferase system (PTS)	PTS system, fructose-specific IIA component
K11198	EC:2.7.1.69	0 (±0)	0.0001 (±0)	-2.807	Phosphotransferase system (PTS)	PTS system, 2-O-A-mannosyl-D-glycerate-specific IIA component
K11199	EC:2.7.1.69	0 (±0)	0.0001 (±0)	-2.756	Phosphotransferase system (PTS)	PTS system, 2-O-A-mannosyl-D-glycerate-specific IIB component
K11200	unknown	0 (±0)	0.0001 (±0)	-2.721	Phosphotransferase system (PTS)	PTS system, 2-O-A-mannosyl-D-glycerate-specific IIC component
K02796	unknown	0.0003 (±0.0001)	0.0013 (±0.0003)	-2.009	Phosphotransferase system (PTS)	PTS system, mannose-specific IID component
K02777	EC:2.7.1.69	0.0002 (±0.0001)	0.0009 (±0.0001)	-1.856	Phosphotransferase system (PTS)	PTS system, glucose-specific IIA component
K02798	EC:2.7.1.69	0.0003 (±0)	0.0013 (±0.0002)	-2.107	Phosphotransferase system (PTS)	PTS system, mannitol-specific IIA component
K02800	unknown	0.0006 (±0.0001)	0.0026 (±0.0006)	-2.244	Phosphotransferase system (PTS)	PTS system, mannitol-specific IIC component
K02799	EC:2.7.1.69	0.0004 (±0.0001)	0.0023 (±0.0005)	-2.428	Phosphotransferase system (PTS)	PTS system, mannitol-specific IIB component
K02745	EC:2.7.1.69	0 (±0)	0.0001 (±0)	-4.79	Phosphotransferase system (PTS)	PTS system, N-acetylgalactosamine-specific IIB component
K02769	EC:2.7.1.69	0.0011 (±0.0003)	0.0047 (±0.0006)	-2.035	Phosphotransferase system (PTS)	PTS system, fructose-specific IIB component
K02770	unknown	0.0011 (±0.0002)	0.0046 (±0.0005)	-2.048	Phosphotransferase system (PTS)	PTS system, fructose-specific IIC component
K02764	EC:2.7.1.69	0.0003 (±0.0001)	0.0018 (±0.0003)	-2.629	Phosphotransferase system (PTS)	PTS system, D-glucosamine-specific IIB component
K02765	unknown	0.0003 (±0.0001)	0.0018 (±0.0003)	-2.65	Phosphotransferase system (PTS)	PTS system, D-glucosamine-specific IIC component
K02763	EC:2.7.1.69	0.0003 (±0.0001)	0.0018 (±0.0003)	-2.647	Phosphotransferase system (PTS)	PTS system, D-glucosamine-specific IIA component
K02809	EC:2.7.1.69	0.0001 (±0)	0.0008 (±0.0002)	-2.899	Phosphotransferase system (PTS)	PTS system, sucrose-specific IIB component
K02810	unknown	0.0001 (±0)	0.0008 (±0.0002)	-2.936	Phosphotransferase system (PTS)	PTS system, sucrose-specific IIC component
K03475	unknown	0 (±0)	0.0004 (±0.0001)	-3.331	Phosphotransferase system (PTS)	PTS system, ascorbate-specific IIC component
K02818	EC:2.7.1.69	0 (±0)	0.0006 (±0.0002)	-3.795	Phosphotransferase system (PTS)	PTS system, trehalose-specific IIB component
K02819	unknown	0 (±0)	0.0006 (±0.0002)	-3.687	Phosphotransferase system (PTS)	PTS system, trehalose-specific IIC component
K02778	EC:2.7.1.69	0.0007 (±0.0002)	0.0043 (±0.0006)	-2.729	Phosphotransferase system (PTS)	PTS system, glucose-specific IIB component
K02779	unknown	0.0006 (±0.0002)	0.0044 (±0.0006)	-2.743	Phosphotransferase system (PTS)	PTS system, glucose-specific IIC component
K01192	EC:3.2.1.25	0.0062 (±0.0019)	0.0133 (±0.0009)	-1.112	Lysosome	Beta-mannosidase
K00031	EC:1.1.1.42	0.1644 (±0.0144)	0.1575 (±0.011)	0.062	TCA cycle	Isocitrate dehydrogenase
K04565	EC:1.15.1.1	0.0156 (±0.0014)	0.0128 (±0.0011)	0.287	Peroxisome	Superoxide dismutase, Cu-Zn family
K04564	EC:1.15.1.1	0.136 (±0.0038)	0.1103 (±0.0029)	0.302	Peroxisome	Superoxide dismutase, Fe-Mn family
K00019	EC:1.1.1.30	0.058 (±0.0024)	0.0571 (±0.0019)	0.022	Butanoate metabolism	3-hydroxybutyrate dehydrogenase
K01040	EC:2.8.3.12	0.0399 (±0.0043)	0.0343 (±0.0025)	0.22	Butanoate metabolism	Glutaconate coa-transferase, subunit B
K01615	EC:4.1.1.70	0.0001 (±0)	0.0002 (±0)	-1.881	Butanoate metabolism	Glutaconyl-coa decarboxylase
K05973	EC:3.1.1.75	0.0357 (±0.0024)	0.0319 (±0.003)	0.164	Butanoate metabolism	Poly(3-hydroxybutyrate) depolymerase
K01039	EC:2.8.3.12	0.058 (±0.0058)	0.0471 (±0.0039)	0.301	Butanoate metabolism	Glutaconate coa-transferase, subunit A
K01799	EC:5.2.1.1	0.0086 (±0.002)	0.0072 (±0.0004)	0.265	Butanoate metabolism	Maleate isomerase
K01575	EC:4.1.1.5	0.0028 (±0.0003)	0.0021 (±0.0003)	0.404	Butanoate metabolism	Acetolactate decarboxylase
K00634	EC:2.3.1.19	0.0035 (±0.0008)	0.0092 (±0.0006)	-1.381	Butanoate metabolism	Phosphate butyryltransferase
K00929	EC:2.7.2.7	0.0002 (±0.0001)	0.0025 (±0.0006)	-3.999	Butanoate metabolism	Butyrate kinase
K03416	EC:2.1.3.1	0.0004 (±0.0001)	0.0012 (±0.0004)	-1.476	Propanoate metabolism	Methylmalonyl-coa carboxyltransferase
K01908	EC:6.2.1.17	0.1042 (±0.0079)	0.0908 (±0.0038)	0.199	Propanoate metabolism	Propionyl-coa synthetase
K01720	EC:4.2.1.79	0.0463 (±0.0018)	0.04 (±0.002)	0.211	Propanoate metabolism	2-methylcitrate dehydratase
K11264	EC:4.1.1.41	0.0006 (±0.0001)	0.0023 (±0.0002)	-1.903	Propanoate metabolism	Methylmalonyl-coa decarboxylase
K01574	EC:4.1.1.4	0.0374 (±0.0032)	0.0235 (±0.0019)	0.674	Propanoate metabolism	Acetoacetate decarboxylase
K01605	EC:4.1.1.41	0 (±0)	0.0002 (±0.0001)	-2.719	Propanoate metabolism	Methylmalonyl-coa decarboxylase alpha chain
K00101	EC:1.1.2.3	0.0864 (±0.0056)	0.0718 (±0.0029)	0.267	Pyruvate metabolism	L-lactate dehydrogenase (cytochrome)
K01067	EC:3.1.2.1	0.0032 (±0.0002)	0.0101 (±0.0016)	-1.641	Pyruvate metabolism	Acetyl-coa hydrolase

Supplementary Table 4-2. Frequencies of genes involved in energy metabolism between forest and pasture. The positive values of log_2foldchange indicate the extent of increased genes frequencies in forest and negative values of log_2foldchange indicate the extent of increased genes frequencies in pasture.

KO ID	EC	Forest	Pasture	log_2foldchange	KEGG Level 3	EC Annotation
K00401	EC:2.8.4.1	0 (±0)	0 (±0)	-3.828	Methane metabolism	Methyl-coenzyme M reductase beta subunit
K01070	EC:3.1.2.12	0.0115 (±0.0009)	0.0109 (±0.0004)	0.085	Methane metabolism	S-formylglutathione hydrolase
K00436	EC:1.12.1.2	0.0036 (±0.0002)	0.0099 (±0.0018)	-1.455	Methane metabolism	Hydrogen dehydrogenase
K03396	EC:4.4.1.22	0.024 (±0.0034)	0.0147 (±0.0015)	0.706	Methane metabolism	S-(hydroxymethyl)glutathione synthase
K04480	EC:2.1.1.90	0.0001 (±0)	0.0004 (±0.0001)	-2.264	Methane metabolism	Methanol---5-hydroxybenzimidazolylcobamide Co-methyltransferase
K03388	EC:1.8.98.1	0.0013 (±0.0003)	0.0057 (±0.002)	-2.133	Methane metabolism	Heterodisulfide reductase subunit A
K14126	EC:1.12.99.-	0.0001 (±0)	0.0011 (±0.0004)	-3.625	Methane metabolism	F420-non-reducing hydrogenase subunit A
K00198	EC:1.2.99.2	0.0001 (±0)	0.0033 (±0.0006)	-5.054	Methane metabolism	Carbon-monoxide dehydrogenase catalytic subunit
K00372	EC:1.7.99.4	0.06 (±0.0015)	0.0555 (±0.001)	0.114	Nitrogen metabolism	Nitrate reductase catalytic subunit
K04561	EC:1.7.2.5	0.0607 (±0.0046)	0.0499 (±0.0055)	0.285	Nitrogen metabolism	Nitric oxide reductase subunit B
K00366	EC:1.7.7.1	0.0737 (±0.0059)	0.0602 (±0.0032)	0.293	Nitrogen metabolism	Ferredoxin-nitrite reductase
K00376	EC:1.7.2.4	0.0195 (±0.0031)	0.0146 (±0.0023)	0.421	Nitrogen metabolism	Nitrous-oxide reductase
K00360	EC:1.7.1.1	0.0249 (±0.0026)	0.0157 (±0.0014)	0.671	Nitrogen metabolism	Nitrate reductase (NADH)
K00368	EC:1.7.2.1	0.0475 (±0.004)	0.0229 (±0.0025)	1.051	Nitrogen metabolism	Nitrite reductase (NO-forming)
K02586	EC:1.18.6.1	0.0001 (±0)	0.0125 (±0.0018)	-6.541	Nitrogen metabolism	Nitrogenase molybdenum-iron protein alpha chain
K02588	EC:1.18.6.1	0.0001 (±0)	0.0062 (±0.0011)	-6.084	Nitrogen metabolism	Nitrogenase iron protein nifh
K02591	EC:1.18.6.1	0.0001 (±0)	0.0148 (±0.0023)	-6.635	Nitrogen metabolism	Nitrogenase molybdenum-iron protein beta chain
K00343	EC:1.6.5.3	0.1607 (±0.0048)	0.1518 (±0.0058)	0.082	Oxidative phosphorylation	NADH-quinone oxidoreductase subunit N
K00333	EC:1.6.5.3	0.2334 (±0.0067)	0.2191 (±0.0013)	0.091	Oxidative phosphorylation	NADH-quinone oxidoreductase subunit D
K03885	EC:1.6.99.3	0.1813 (±0.0188)	0.158 (±0.0126)	0.199	Oxidative phosphorylation	NADH dehydrogenase
K02276	EC:1.9.3.1	0.0863 (±0.0022)	0.0771 (±0.0032)	0.163	Oxidative phosphorylation	Cytochrome c oxidase subunit III
K02828	EC:1.10.3.12	0.0003 (±0.0001)	0.0009 (±0.0003)	-1.615	Oxidative phosphorylation	Cytochrome aa3-600 menaquinol oxidase subunit III
K02297	EC:1.10.3.-	0.0143 (±0.0014)	0.0116 (±0.0013)	0.298	Oxidative phosphorylation	Cytochrome o ubiquinol oxidase subunit II
K05587		0.0016 (±0.0003)	0.004 (±0.0004)	-1.316	Oxidative phosphorylation	Bidirectional
K00356	EC:1.6.99.3	0.0351 (±0.002)	0.0291 (±0.0014)	0.269	Oxidative phosphorylation	NADH dehydrogenase
K02300	unknown	0.0077 (±0.0006)	0.0059 (±0.0007)	0.391	Oxidative phosphorylation	Cytochrome o ubiquinol oxidase operon protein cyod
K05586	unknown	0.001 (±0.0001)	0.0029 (±0.0003)	-1.475	Oxidative phosphorylation	Bidirectional
K05588	unknown	0.0024 (±0.0002)	0.007 (±0.0006)	-1.563	Oxidative phosphorylation	Bidirectional
K02258	unknown	0.0149 (±0.0011)	0.0105 (±0.0006)	0.506	Oxidative phosphorylation	Cytochrome c oxidase assembly protein subunit 11
K02124	EC:3.6.3.14	0.0012 (±0.0003)	0.0009 (±0.0001)	0.44	Oxidative phosphorylation	V-type H+-transporting ATPase subunit K
K02299	EC:1.10.3.-	0.0212 (±0.0013)	0.0145 (±0.0014)	0.542	Oxidative phosphorylation	Cytochrome o ubiquinol oxidase subunit III
K01535	EC:3.6.3.6	0.0229 (±0.0031)	0.0143 (±0.002)	0.686	Oxidative phosphorylation	H+-transporting ATPase
K02298	EC:1.10.3.-	0.0833 (±0.0099)	0.0569 (±0.0057)	0.551	Oxidative phosphorylation	Cytochrome o ubiquinol oxidase subunit I
K13378	EC:1.6.5.3	0.0148 (±0.0005)	0.0097 (±0.0003)	0.602	Oxidative phosphorylation	NADH-quinone oxidoreductase subunit C/D
K02118	EC:3.6.3.14	0.0079 (±0.0006)	0.0048 (±0.0004)	0.704	Oxidative phosphorylation	V-type H+-transporting ATPase subunit B
K02117	EC:3.6.3.14	0.0089 (±0.0009)	0.0053 (±0.0006)	0.745	Oxidative phosphorylation	V-type H+-transporting ATPase subunit A
K02123	EC:3.6.3.14	0.0025 (±0.0007)	0.0015 (±0.0002)	0.768	Oxidative phosphorylation	V-type H+-transporting ATPase subunit I
K02120	EC:3.6.3.14	0.0011 (±0.0003)	0.0005 (±0.0001)	1.085	Oxidative phosphorylation	V-type H+-transporting ATPase subunit D
K06019	EC:3.6.1.1	0.0002 (±0.0001)	0.0011 (±0.0003)	-2.458	Oxidative phosphorylation	Pyrophosphatase ppax
K02706	unknown	0 (±0)	0.0002 (±0)	-1.654	Photosynthesis	Photosystem II P680 reaction center D2 protein
K02690	unknown	0.0001 (±0)	0.0004 (±0.0001)	-1.493	Photosynthesis	Photosystem I P700 chlorophyll a apoprotein A2
K02689	unknown	0 (±0)	0.0002 (±0.0001)	-2.583	Photosynthesis	Photosystem I P700 chlorophyll a apoprotein A1
K08906	unknown	0 (±0)	0.0001 (±0)	-4.307	Photosynthesis	Cytochrome c6
K02705	unknown	0 (±0)	0.0001 (±0)	-3.946	Photosynthesis	Photosystem II CP43 chlorophyll apoprotein
K02634	unknown	0 (±0)	0.0001 (±0)	-3.797	Photosynthesis	Apocytochrome f
K02704	unknown	0 (±0)	0.0001 (±0)	-5.497	Photosynthesis	Photosystem II CP47 chlorophyll apoprotein
K08912	unknown	0 (±0)	0 (±0)	-3.095	Photosynthesis - antenna proteins	Light-harvesting complex II chlorophyll a/b binding protein 1
K00958	EC:2.7.7.4	0.0609 (±0.005)	0.0539 (±0.0041)	0.178	Sulfur metabolism	Sulfate adenylyltransferase

K00381	EC:1.8.1.2	0.0841 (±0.0042)	0.0709 (±0.0026)	0.247	Sulfur metabolism	Sulfite reductase (NADPH) hemoprotein beta-component
K00380	EC:1.8.1.2	0.0469 (±0.003)	0.0333 (±0.0018)	0.496	Sulfur metabolism	Sulfite reductase (NADPH) flavoprotein alpha-component
K00394	EC:1.8.99.2	0.0114 (±0.0005)	0.0078 (±0.0007)	0.552	Sulfur metabolism	Adenylylsulfate reductase, subunit A
K00395	EC:1.8.99.2	0.0052 (±0.0005)	0.0028 (±0.0004)	0.89	Sulfur metabolism	Adenylylsulfate reductase, subunit B
K01919	EC:6.3.2.2	0.05 (±0.0043)	0.0454 (±0.002)	0.139	Glutathione metabolism	Glutamate--cysteine ligase
K00383	EC:1.8.1.7	0.0531 (±0.0026)	0.0344 (±0.0012)	0.627	Glutathione metabolism	Glutathione reductase (NADPH)
K04565	EC:1.15.1.1	0.0156 (±0.0014)	0.0128 (±0.0011)	0.287	Peroxisome	Superoxide dismutase, Cu-Zn family
K04564	EC:1.15.1.1	0.136 (±0.0038)	0.1103 (±0.0029)	0.302	Peroxisome	Superoxide dismutase, Fe-Mn family

Supplementary Table 4-3. Frequencies of genes involved in amino acid metabolism between forest and pasture. The positive values of log_2foldchange indicate the extent of increased genes frequencies in forest and negative values of log_2foldchange indicate the extent of increased genes frequencies in pasture.

KO ID	EC	Forest	Pasture	log_2foldchange	KEGG Level 3	EC Annotation
K11358	EC:2.6.1.1	0.006 (±0.0013)	0.0053 (±0.0008)	0.165	Alanine, aspartate and glutamate metabolism	Aspartate aminotransferase
K01755	EC:4.3.2.1	0.1227 (±0.0026)	0.1237 (±0.0023)	-0.011	Alanine, aspartate and glutamate metabolism	Argininosuccinate lyase
K01425	EC:3.5.1.2	0.0405 (±0.0032)	0.0393 (±0.0012)	0.044	Alanine, aspartate and glutamate metabolism	Glutaminase
K01940	EC:6.3.4.5	0.1568 (±0.0088)	0.1525 (±0.0079)	0.04	Alanine, aspartate and glutamate metabolism	Argininosuccinate synthase
K01956	EC:6.3.5.5	0.1385 (±0.0046)	0.1297 (±0.0026)	0.094	Alanine, aspartate and glutamate metabolism	Carbamoyl-phosphate synthase small subunit
K01953	EC:6.3.5.4	0.3205 (±0.0235)	0.2759 (±0.0203)	0.216	Alanine, aspartate and glutamate metabolism	Asparagine synthase (glutamine-hydrolysing)
K09758	EC:4.1.1.12	0.0122 (±0.0016)	0.0074 (±0.0009)	0.709	Alanine, aspartate and glutamate metabolism	Aspartate 4-decarboxylase
K01581	EC:4.1.1.17	0.0617 (±0.0063)	0.0495 (±0.0035)	0.318	Arginine and proline metabolism	Ornithine decarboxylase
K02626	EC:4.1.1.19	0.0199 (±0.001)	0.011 (±0.0008)	0.854	Arginine and proline metabolism	Arginine decarboxylase
K00548	EC:2.1.1.13	0.4064 (±0.0141)	0.3977 (±0.009)	0.031	Cysteine and methionine metabolism	5-methyltetrahydrofolate--homocysteine methyltransferase
K01760	EC:4.4.1.8	0.0512 (±0.002)	0.0463 (±0.0015)	0.144	Cysteine and methionine metabolism	Cystathionine beta-lyase
K00641	EC:2.3.1.31	0.1426 (±0.0086)	0.1165 (±0.0029)	0.293	Cysteine and methionine metabolism	Homoserine O-acetyltransferase
K00549	EC:2.1.1.14	0.3282 (±0.0256)	0.2153 (±0.0173)	0.608	Cysteine and methionine metabolism	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
K00133	EC:1.2.1.11	0.1173 (±0.003)	0.1148 (±0.001)	0.031	Glycine, serine and threonine metabolism	Aspartate-semialdehyde dehydrogenase
K01733	EC:4.2.3.1	0.2224 (±0.0067)	0.1993 (±0.005)	0.158	Glycine, serine and threonine metabolism	Threonine synthase
K12524	EC:2.7.2.4 1.1.1.3	0.0137 (±0.0018)	0.0119 (±0.0007)	0.204	Glycine, serine and threonine metabolism	Bifunctional aspartokinase / homoserine dehydrogenase 1
K02204	EC:2.7.1.39	0.0294 (±0.0023)	0.0236 (±0.0009)	0.321	Glycine, serine and threonine metabolism	Homoserine kinase type II
K00808	EC:2.5.1.44	0.0509 (±0.004)	0.0317 (±0.0017)	0.681	Tropane, piperidine and pyridine alkaloid biosynthesis	Homospermidine synthase
K10004	EC:3.6.3.-	0.0156 (±0.0014)	0.0152 (±0.0008)	0.039	ABC transporters	Glutamate/aspartate transport system ATP-binding protein
K10001	unknown	0.0331 (±0.0032)	0.0287 (±0.0016)	0.207	ABC transporters	Glutamate/aspartate transport system substrate-binding protein
K01995	unknown	0.5088 (±0.046)	0.4425 (±0.0249)	0.201	ABC transporters	Branched-chain amino acid transport system ATP-binding protein
K01998	unknown	0.5033 (±0.0363)	0.4271 (±0.0209)	0.237	ABC transporters	Branched-chain amino acid transport system permease protein
K10013	unknown	0.0003 (±0.0001)	0.001 (±0.0003)	-1.555	ABC transporters	Lysine/arginine/ornithine transport system substrate-binding protein
K01996	unknown	0.5373 (±0.0485)	0.4497 (±0.0246)	0.257	ABC transporters	Branched-chain amino acid transport system ATP-binding protein
K01997	unknown	0.5097 (±0.0398)	0.4263 (±0.0213)	0.258	ABC transporters	Branched-chain amino acid transport system permease protein
K10003	unknown	0.0163 (±0.0017)	0.0133 (±0.0004)	0.288	ABC transporters	Glutamate/aspartate transport system permease protein
K10002	unknown	0.018 (±0.0021)	0.0146 (±0.0003)	0.299	ABC transporters	Glutamate/aspartate transport system permease protein
K01999	unknown	0.7108 (±0.0619)	0.5694 (±0.0202)	0.32	ABC transporters	Branched-chain amino acid transport system substrate-binding protein
K09972	EC:3.6.3.-	0.0244 (±0.0016)	0.0198 (±0.0017)	0.3	ABC transporters	General L-amino acid transport system ATP-binding protein
K11076	unknown	0.0267 (±0.0016)	0.0208 (±0.0008)	0.358	ABC transporters	Putrescine transport system ATP-binding protein
K09969	unknown	0.0422 (±0.0036)	0.0322 (±0.0027)	0.393	ABC transporters	General L-amino acid transport system substrate-binding protein
K09970	unknown	0.0265 (±0.0016)	0.0198 (±0.0009)	0.418	ABC transporters	General L-amino acid transport system permease protein
K09971	unknown	0.0275 (±0.0021)	0.0202 (±0.0008)	0.443	ABC transporters	General L-amino acid transport system permease protein
K02071	unknown	0.004 (±0.0009)	0.0106 (±0.0014)	-1.429	ABC transporters	D-methionine transport system ATP-binding protein
K11075	unknown	0.0236 (±0.0016)	0.0165 (±0.0008)	0.515	ABC transporters	Putrescine transport system permease protein
K11074	unknown	0.019 (±0.0015)	0.0126 (±0.0008)	0.593	ABC transporters	Putrescine transport system permease protein
K11073	unknown	0.0275 (±0.0027)	0.018 (±0.0017)	0.612	ABC transporters	Putrescine transport system substrate-binding protein

K02072	unknown	0.0008 (±0.0002)	0.004 (±0.0008)	-2.298	ABC transporters	D-methionine transport system permease protein
K02073	unknown	0.0011 (±0.0004)	0.0057 (±0.0011)	-2.382	ABC transporters	D-methionine transport system substrate-binding protein
K02234	unknown	0.0232 (±0.0023)	0.0158 (±0.0012)	0.554	Porphyrin and chlorophyll metabolism	Cobalamin biosynthesis protein cobw
K09882	EC:6.6.1.2	0.0316 (±0.0044)	0.0215 (±0.0022)	0.555	Porphyrin and chlorophyll metabolism	Cobaltochelatase cobs
K09883	EC:6.6.1.2	0.0395 (±0.0035)	0.0266 (±0.0011)	0.567	Porphyrin and chlorophyll metabolism	Cobaltochelatase cobt

Supplementary Table 4-4. Frequencies of genes involved in xenobiotic metabolism between forest and pasture. The positive values of log_2foldchange indicate the extent of increased genes frequencies in forest and negative values of log_2foldchange indicate the extent of increased genes frequencies in pasture.

KO ID	EC	Forest	Pasture	log_2foldchange	KEGG Level 3	EC Annotation
K01781	EC:5.1.2.2	0.0085 (±0.0012)	0.0079 (±0.0007)	0.108	Aminobenzoate degradation	Mandelate racemase
K09461	EC:1.14.13.40	0.0823 (±0.0057)	0.0693 (±0.002)	0.249	Aminobenzoate degradation	Anthraniloyl-coa monooxygenase
K01576	EC:4.1.1.7	0.0457 (±0.0027)	0.0375 (±0.0015)	0.285	Aminobenzoate degradation	Benzoylformate decarboxylase
K08295	EC:6.2.1.32	0.0523 (±0.0041)	0.0409 (±0.0017)	0.354	Aminobenzoate degradation	2-aminobenzoate-CoA ligase
K08710	EC:3.5.99.4	0.0009 (±0.0001)	0.0008 (±0.0001)	0.175	Atrazine degradation	N-isopropylammelide isopropylaminohydrolase
K01821	EC:5.3.2.-	0.0085 (±0.0005)	0.0083 (±0.0005)	0.031	Benzoate degradation	4-oxalocrotonate tautomerase
K00481	EC:1.14.13.2	0.0591 (±0.0026)	0.0538 (±0.0022)	0.135	Benzoate degradation	p-hydroxybenzoate 3-monooxygenase
K01857	EC:5.5.1.2	0.0386 (±0.0028)	0.0337 (±0.0024)	0.196	Benzoate degradation	3-carboxy-cis,cis-muconate cycloisomerase
K01607	EC:4.1.1.44	0.0492 (±0.0025)	0.0392 (±0.0023)	0.327	Benzoate degradation	4-carboxymuconolactone decarboxylase
K04101	EC:1.13.11.8	0.0276 (±0.0023)	0.0213 (±0.0012)	0.374	Benzoate degradation	Protocatechuate 4,5-dioxygenase, beta chain
K07823	EC:2.3.1.174	0.0232 (±0.002)	0.0164 (±0.0008)	0.497	Benzoate degradation	3-oxoadipyl-CoA thiolase
K01032	EC:2.8.3.6	0.0126 (±0.0008)	0.0092 (±0.0005)	0.453	Benzoate degradation	3-oxoadipate CoA-transferase, beta subunit
K05550	EC:1.14.12.10 1.14.12.-	0.0007 (±0.0001)	0.0006 (±0)	0.391	Benzoate degradation	Benzoate/toluate 1,2-dioxygenase subunit beta
K10220	EC:4.2.1.83	0.0503 (±0.0028)	0.0352 (±0.0018)	0.516	Benzoate degradation	4-oxalmesaconate hydratase
K10221	EC:3.1.1.57	0.0204 (±0.0016)	0.0135 (±0.0007)	0.597	Benzoate degradation	2-pyrone-4,6-dicarboxylate lactonase
K04108	EC:1.3.7.9	0.0302 (±0.0045)	0.015 (±0.0022)	1.012	Benzoate degradation	4-hydroxybenzoyl-CoA reductase subunit alpha
K01856	EC:5.5.1.1	0.0061 (±0.0014)	0.0059 (±0.0005)	0.065	Chlorocyclohexane and chlorobenzene degradation	Muconate cycloisomerase
K03384	EC:1.14.12.-	0.0129 (±0.0012)	0.0109 (±0.0006)	0.238	Chlorocyclohexane and chlorobenzene degradation	Unknown
K03380	EC:1.14.13.7	0.0167 (±0.0018)	0.0132 (±0.0013)	0.338	Chlorocyclohexane and chlorobenzene degradation	Phenol 2-monooxygenase
K01561	EC:3.8.1.3	0.0451 (±0.0044)	0.0341 (±0.0022)	0.407	Chlorocyclohexane and chlorobenzene degradation	Haloacetate dehalogenase
K03381	EC:1.13.11.1	0.0268 (±0.0021)	0.0206 (±0.001)	0.38	Chlorocyclohexane and chlorobenzene degradation	Catechol 1,2-dioxygenase
K01061	EC:3.1.1.45	0.1451 (±0.0054)	0.1081 (±0.0054)	0.425	Chlorocyclohexane and chlorobenzene degradation	Carboxymethylenebutenolidase
K01560	EC:3.8.1.2	0.057 (±0.005)	0.038 (±0.0034)	0.584	Chlorocyclohexane and chlorobenzene degradation	2-haloacid dehalogenase
K00480	EC:1.14.13.1	0.0654 (±0.0071)	0.0476 (±0.0045)	0.458	Dioxin degradation	Salicylate hydroxylase
K07519	EC:1.14.12.7	0.0026 (±0.0004)	0.0022 (±0.0002)	0.248	Polycyclic aromatic hydrocarbon degradation	Phthalate 4,5-dioxygenase
K11948	EC:1.13.11.38	0.0007 (±0.0002)	0.0004 (±0.0001)	0.802	Polycyclic aromatic hydrocarbon degradation	1-hydroxy-2-naphthoate dioxygenase
K04102	EC:4.1.1.55	0.0651 (±0.0091)	0.0344 (±0.0047)	0.921	Polycyclic aromatic hydrocarbon degradation	4,5-dihydroxyphthalate decarboxylase
K07546	EC:4.2.1.-	0.0002 (±0.0001)	0.0002 (±0)	0.448	Toluene degradation	E-phenylitaconyl-CoA hydratase
K10619	EC:1.14.12.-	0.0027 (±0.0002)	0.0023 (±0.0001)	0.236	Xylene degradation	p-cumate 2,3-dioxygenase subunit alpha

Supplementary Table 4-5. Frequencies of genes involved in sporulation between forest and pasture. The positive values of log_2foldchange indicate the extent of increased genes frequencies in forest and negative values of log_2foldchange indicate the extent of increased genes frequencies in pasture.

KO ID	EC	Forest	Pasture	log_2foldchange	KEGG Level 3	EC Annotation
K13533	EC:2.7.13.3	0.0002 (±0.0001)	0.0008 (±0.0002)	-1.96	Two-component system	Sporulation sensor kinase E
K07699	unknown	0.0002 (±0)	0.0013 (±0.0003)	-2.54	Two-component system	Response regulator, stage 0 sporulation protein A
K07697	EC:2.7.13.3	0.0002 (±0)	0.0013 (±0.0004)	-2.99	Two-component system	Sporulation sensor kinase B

Supplementary Table 4-6. Frequencies of genes involved in cell motility between forest and pasture. The positive values of log_2foldchange indicate the extent of increased genes frequencies in forest and negative values of log_2foldchange indicate the extent of increased genes frequencies in pasture.

KO ID	EC	Forest	Pasture	log_2foldchange	KEGG Level 3	EC Annotation
K02402	unknown	0.0004 (±0.0003)	0.0012 (±0.0002)	-1.44	Flagellar assembly	Flagellar transcriptional activator flhc
K02396	unknown	0.0055 (±0.0021)	0.0117 (±0.0006)	-1.09	Flagellar assembly	Flagellar hook-associated protein 1 flgk
K03408	unknown	0.0084 (±0.0026)	0.0176 (±0.0014)	-1.07	Bacterial chemotaxis	Purine-binding chemotaxis protein chew
K02406	unknown	0.0105 (±0.0051)	0.0254 (±0.0022)	-1.28	Flagellar assembly	Flagellin
K05874	unknown	0.0018 (±0.0009)	0.0046 (±0.0007)	-1.33	Bacterial chemotaxis	Methyl-accepting chemotaxis protein I, serine sensor receptor
K03406	unknown	0.0452 (±0.015)	0.1134 (±0.0159)	-1.33	Bacterial chemotaxis	Methyl-accepting chemotaxis protein
K05875	unknown	0.0004 (±0.0002)	0.0011 (±0.0002)	-1.53	Bacterial chemotaxis	Methyl-accepting chemotaxis protein II, aspartate sensor receptor
K03411	EC:3.5.1.44	0.0017 (±0.0008)	0.0051 (±0.0003)	-1.58	Bacterial chemotaxis	Chemotaxis protein ched
K13820	unknown	0.0001 (±0)	0.0003 (±0)	-1.93	Flagellar assembly	Flagellar biosynthetic protein flir/flhb
K02422	unknown	0.0004 (±0.0002)	0.0014 (±0.0002)	-1.82	Flagellar assembly	Flagellar protein flis

Supplementary Table 4-7. Frequencies of genes involved in the biosynthesis of other secondary metabolites between forest and pasture. The positive values of log_2foldchange indicate the extent of increased genes frequencies in forest and negative values of log_2foldchange indicate the extent of increased genes frequencies in pasture.

KO ID	EC	#Forest!	Pasture	log_2foldchange		EC Annotation
K11632	unknown	0 (±0)	0.0006 (±0.0004)	-3.571	ABC transporters	Bacitracin transport system permease protein
K11631	unknown	0.0001 (±0)	0.0004 (±0.0002)	-2.294	ABC transporters	Bacitracin transport system ATP-binding protein
K11629	EC:2.7.13.3	0 (±0)	0.0003 (±0.0002)	-3.218	Two-component system	OmpR family, bacitracin resistance sensor histidine kinase BceS
K11630	unknown	0 (±0)	0.0003 (±0.0002)	-3.288	Two-component system	OmpR family, bacitracin resistance response regulator BceR
K01434	EC:3.5.1.11	0.0461 (±0.0028)	0.0442 (±0.0028)	0.06	Penicillin and cephalosporin biosynthesis	Penicillin amidase
K01790	EC:5.1.3.13	0.0512 (±0.0029)	0.0514 (±0.0009)	0	Streptomycin biosynthesis	dTDP-4-dehydrorhamnose 3,5-epimerase

References

Allen, H.K., Donato, J., Wang, H.H., Cloud-Hansen, K.A., Davies, J. and Handelsman, J., 2010. Call of the wild: antibiotic resistance genes in natural environments. Nature reviews microbiology, 8(4), pp.251-259.

Ames, P. and Bergman, K., 1981. Competitive advantage provided by bacterial motility in the formation of nodules by *Rhizobium meliloti*. Journal of bacteriology, 148(2), pp.728-729.

Andrade, E.R., Jardim, J.G., Santos, B.A., Melo, F.P.L., Talora, D.C., Faria, D. and Cazetta, E., 2015. Effects of habitat loss on taxonomic and phylogenetic diversity of understory *Rubiaceae* in Atlantic forest landscapes. Forest ecology and management, 349, pp.73-84.

Bais, H.P., Weir, T.L., Perry, L.G., Gilroy, S. and Vivanco, J.M., 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. Annual review of plant biology, 57, pp.233-266.

Barabasi, A.L. and Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. Nature reviews genetics, 5(2), pp.101-113.

Barabâsi, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A. and Vicsek, T., 2002. Evolution of the social network of scientific collaborations. Physica A: Statistical mechanics and its applications, 311(3), pp.590-614.

Barberán, A., Bates, S.T., Casamayor, E.O. and Fierer, N., 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. The ISME journal, 6(2), pp.343-351.

Bascompte, J., and Stouffer, D.B. 2009. The assembly and disassembly of ecological networks. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364: 1781–1787.

Bastian, M., Heymann, S. and Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks. ICWSM, 8, pp.361-362.

Bell, T.H., Yergeau, E., Maynard, C., Juck, D., Whyte, L.G. and Greer, C.W., 2013. Predictable bacterial composition and hydrocarbon degradation in Arctic soils following diesel and nutrient disturbance. The ISME journal, *7*(6), pp.1200-1210.

Benjamini, Y., Krieger, A.M. and Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika, 93(3), pp.491-507.

Bergmann, G.T., Bates, S.T., Eilers, K.G., Lauber, C.L., Caporaso, J.G., Walters, W.A., Knight, R. and Fierer, N., 2011. The under-recognized dominance of *Verrucomicrobia* in soil bacterial communities. Soil biology and biochemistry, 43(7), pp.1450-1455.

Betts, R., Sanderson, M. and Woodward, S., 2008. Effects of large-scale Amazon forest degradation on climate and air quality through fluxes of carbon dioxide, water, energy, mineral dust and isoprene. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 363(1498), pp.1873-1880.

Bierregaard, R.O., Gascon, C., Lovejoy, T.E., and Mesquita, R. 2001. Lessons from Amazonia: The Ecology and Conservation of a Fragmented Forest. Yale University Press (New Haven, CT).

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. Applied and environmental microbiology, 63(7), pp.2647-2653.

Bowman, J., 2006. The methanotrophs—the families *Methylococcaceae* and *Methylocystaceae*. In The prokaryotes (pp. 266-289). Springer New York.

Brown, P.J., de Pedro, M.A., Kysela, D.T., Van der Henst, C., Kim, J., De Bolle, X., Fuqua, C. and Brun, Y.V., 2012. Polar growth in the Alphaproteobacterial order *Rhizobiales*. Proceedings of the national academy of sciences, 109(5), pp.1697-1701.

Büchi, L. and Vuilleumier, S., 2014. Coexistence of specialist and generalist species is shaped by dispersal and environmental factors. The American naturalist, 183(5), pp.612-624.

Caetano-Anollés, G., Wall, L.G., De Micheli, A.T., Macchi, E.M., Bauer, W.D. and Favelukes, G., 1988. Role of motility and chemotaxis in efficiency of nodulation by *Rhizobium meliloti*. Plant physiology, 86(4), pp.1228-1235.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Gonzalez, P.A., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nature methods, 7(5), pp.335-336.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M. and Gormley, N., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal, 6(8), pp.1621-1624.

Cardinale, B.J., Palmer, M.A. and Collins, S.L., 2002. Species diversity enhances ecosystem functioning through interspecific facilitation. Nature, 415(6870), pp.426-429.

Cenciani, K., Lambais, M.R., Cerri, C.C., Azevedo, L.C.B.D. and Feigl, B.J., 2009. Bacteria diversity and microbial biomass in forest, pasture and fallow soils in the southwestern Amazon basin. Revista brasileira de ciência do solo, 33(4), pp.907-916.

Chapin III, F.S., Zavaleta, E.S., Eviner, V.T., Naylor, R.L., Vitousek, P.M., Reynolds, H.L., Hooper, D.U., Lavorel, S., Sala, O.E., Hobbie, S.E. and Mack, M.C., 2000. Consequences of changing biodiversity. Nature, 405(6783), pp.234-242.

Chin, K.J., Liesack, W. and Janssen, P.H., 2001. *Opitutus terrae* gen. nov., sp. nov., to accommodate novel strains of the division *'Verrucomicrobia'* isolated from rice paddy soil. International journal of systematic and evolutionary microbiology, 51(6), pp.1965-1968.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B. and Hanspers, K., 2007. Integration of biological networks and gene expression data using Cytoscape. Nature protocols, 2(10), pp.2366-2382.

Cordovil, C.M.D.S., Coutinho, J., Goss, M. and Cabral, F., 2005. Potentially mineralizable nitrogen from organic materials applied to a sandy soil: fitting the one-pool exponential model. Soil use manage, 21(1), pp.65-72.

Costello, E.K., Lauber, C.L., Hamady, M., Fierer, N., Gordon, J.I. and Knight, R., 2009. Bacterial community variation in human body habitats across space and time. Science, 326(5960), pp.1694-1697.

Curtis, T.P., Sloan, W.T. and Scannell, J.W., 2002. Estimating prokaryotic diversity and its limits. Proceedings of the national academy of sciences, 99(16), pp.10494-10499.

Daims, H., Lebedeva, E.V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N.,

Palatinszky, M., Vierheilig, J., Bulaev, A. and Kirkegaard, R.H., 2015. Complete nitrification by Nitrospira bacteria. Nature, 528(7583), pp.504-509.

Dantas, G., Sommer, M.O., Oluwasegun, R.D. and Church, G.M., 2008. Bacteria subsisting on antibiotics. Science, 320(5872), pp.100-103.

Davidson, E.A., de Araújo, A.C., Artaxo, P., Balch, J.K., Brown, I.F., Bustamante, M.M., Coe, M.T., DeFries, R.S., Keller, M., Longo, M. and Munger, J.W., 2012. The Amazon basin in transition. Nature, 481(7381), pp.321-328.

Davidson, E.A., de Carvalho, C.J.R., Figueira, A.M., Ishida, F.Y., Ometto, J.P.H., Nardoto, G.B., Sabá, R.T., Hayashi, S.N., Leal, E.C., Vieira, I.C.G. and Martinelli, L.A., 2007. Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment. Nature, 447(7147), pp.995-998.

Deinum, B., Sulastri, R.D., Zeinab, M.H.J. and Maassen, A., 1996. Effects of light intensity on growth, anatomy and forage quality of two tropical grasses (*Brachiaria brizantha* and *Panicum maximum* var. *trichoglume*). NJAS wageningen journal of life sciences, 44(2), pp.111-124.

de Lima Procópio, R.E., da Silva, I.R., Martins, M.K., de Azevedo, J.L. and de Araújo, J.M., 2012. Antibiotics produced by Streptomyces. The Brazilian journal of infectious diseases, 16(5), pp.466-471.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P. and Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and environmental microbiology, 72(7), pp.5069-5072.

Ding, J., Zhang, Y., Deng, Y., Cong, J., Lu, H., Sun, X., Yang, C., Yuan, T., Van

Nostrand, J.D., Li, D. and Zhou, J., 2015. Integrated metagenomics and network analysis of soil microbial community of the forest timberline. Scientific reports, 5, p.7994.

Dirzo, R. and Raven, P.H., 2003. Global state of biodiversity and loss. Annual review of environment and resources, 28(1), pp.137-167.

Dixon, P., 2003. VEGAN, a package of R functions for community ecology. Journal of vegetation science, 14(6), pp.927-930.

Dridi, B., Fardeau, M.L., Ollivier, B., Raoult, D. and Drancourt, M., 2012. *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. International journal of systematic and evolutionary microbiology, 62(8), pp.1902-1907.

Duran-Pinedo, A.E., Paster, B., Teles, R. and Frias-Lopez, J., 2011. Correlation network analysis applied to complex biofilm communities. PLoS one, 6(12), p.e28438.

Eberspächer, J. and Lingens, F., 2006. The genus *Phenylobacterium*. In The Prokaryotes (pp. 250-256). Springer New York.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19), pp.2460-2461.

Eilers, K.G., Lauber, C.L., Knight, R. and Fierer, N., 2010. Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. Soil biology and biochemistry, 42(6), pp.896-903.

Elias, R.B. and Dias, E., 2009. Gap dynamics and regeneration strategies in Juniperus-Laurus forests of the Azores Islands. Plant ecology, 200(2), pp.179-189.

Erbilgin, O., McDonald, K.L. and Kerfeld, C.A., 2014. Characterization of a planctomycetal organelle: a novel bacterial microcompartment for the aerobic

degradation of plant saccharides. Applied and environmental microbiology, 80(7), pp.2193-2205.

Eva, H.D., Belward, A.S., De Miranda, E.E., Di Bella, C.M., Gond, V., Huber, O., Jones, S., Sgrenzaroli, M. and Fritz, S., 2004. A land cover map of South America. Global change biology, 10(5), pp.731-744.

Faust, K. and Raes, J., 2012. Microbial interactions: from networks to models. Nature reviews microbiology, 10(8), pp.538-550.

Faust, K., Lahti, L., Gonze, D., de Vos, W.M. and Raes, J., 2015. Metagenomics meets time series analysis: unraveling microbial community dynamics. Current opinion in microbiology, 25, pp.56-66.

Faust, K., Sathirapongsasuti, J.F., Izard, J., Segata, N., Gevers, D., Raes, J. and Huttenhower, C., 2012. Microbial co-occurrence relationships in the human microbiome. PLoS computational biology, 8(7), p.e1002606.

Fearnside, P.M., 1999. Biodiversity as an environmental service in Brazil's Amazonian forests: risks, value and conservation. Environmental conservation, 26(04), pp.305-321.

Feeley, K.J. and Silman, M.R., 2009. Extinction risks of Amazonian plant species. Proceedings of the national academy of sciences, 106(30), pp.12382-12387.

Feigl, B., Cerri, C., Piccolo, M., Noronha, N., Augusti, K., Melillo, J., Eschenbrenner, V. and Melo, L., 2006. Biological survey of a low-productivity pasture in Rondônia state, Brazil. Outlook on agriculture, 35(3), pp.199-208.

Fernandes, S.A.P., Bernoux, M., Cerri, C.C., Feigl, B.J. and Piccolo, M.C., 2002. Seasonal variation of soil chemical properties and CO_2 and CH_4 fluxes in unfertilized and P-fertilized pastures in an Ultisol of the Brazilian Amazon. Geoderma, 107(3), pp.227-

241.

Ferrenberg, S., O'Neill, S.P., Knelman, J.E., Todd, B., Duggan, S., Bradley, D., Robinson, T., Schmidt, S.K., Townsend, A.R., Williams, M.W. and Cleveland, C.C., 2013. Changes in assembly processes in soil bacterial communities following a wildfire disturbance. The ISME journal, 7(6), pp.1102-1111.

Fierer, N. and Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. Proceedings of the national academy of sciences, 103(3), pp.626-631.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H. and Caporaso, J.G., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proceedings of the national academy of sciences, 109(52), pp.21390-21395.

Foley, J.A., Asner, G.P., Costa, M.H., Coe, M.T., DeFries, R., Gibbs, H.K., Howard, E.A., Olson, S., Patz, J., Ramankutty, N. and Snyder, P., 2007. Amazonia revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin. Frontiers in ecology and the environment,5(1), pp.25-32.

Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R. and Ruppin, E., 2010. The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic acids research, 38(12), pp.3857-3868.

Fuhrman, J.A. and Steele, J.A., 2008. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. Aquatic Microbial ecology, 53(1), pp.69-81.

Fuhrman, J.A., 2009. Microbial community structure and its functional implications. Nature, 459(7244), pp.193-199.

Fuhrman, J.A., Hewson, I., Schwalbach, M.S., Steele, J.A., Brown, M.V. and Naeem, S., 2006. Annually reoccurring bacterial communities are predictable from ocean conditions. Proceedings of the national academy of sciences, 103(35), pp.13104-13109.

Garbeva, P., Van Veen, J.A. and Van Elsas, J.D., 2004. Microbial diversity in soil: selection of microbial populations by plant and soil type and implications for disease suppressiveness. Annual review of phytopathology, 42, pp.243-270.

Gianoulis, T.A., Raes, J., Patel, P.V., Bjornson, R., Korbel, J.O., Letunic, I., Yamada, T., Paccanaro, A., Jensen, L.J., Snyder, M. and Bork, P., 2009. Quantifying environmental adaptation of metabolic pathways in metagenomics. Proceedings of the national academy of sciences, 106(5), pp.1374-1379.

Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A.C., Knight, R., Joint, I. and Somerfield, P., 2012. Defining seasonal marine microbial community dynamics. The ISME journal, 6(2), pp.298-308.

Girvan, M.S., Bullimore, J., Pretty, J.N., Osborn, A.M. and Ball, A.S., 2003. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. Applied and environmental microbiology, 69(3), pp.1800-1809.

González, A.M.M., Dalsgaard, B. and Olesen, J.M., 2010. Centrality measures and the importance of generalist species in pollination networks. Ecological Complexity, 7(1), pp.36-43.

Grandy, A.S. and Neff, J.C., 2008. Molecular C dynamics downstream: the biochemical decomposition sequence and its impact on soil organic matter structure and function. Science of the total environment, 404(2), pp.297-307.

Greenblum, S., Chiu, H.C., Levy, R., Carr, R. and Borenstein, E., 2013. Towards a

predictive systems-level model of the human microbiome: progress, challenges, and opportunities. Current opinion in biotechnology, 24(4), pp.810-820.

Greenblum, S., Turnbaugh, P.J. and Borenstein, E., 2012. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. Proceedings of the national academy of sciences, 109(2), pp.594-599.

Griswold, A.R., Jameson-Lee, M. and Burne, R.A., 2006. Regulation and physiologic significance of the agmatine deiminase system of *Streptococcus mutans* UA159. Journal of bacteriology, 188(3), pp.834-841.

Hamaoui, G.S., Rodrigues, J.L., Bohannan, B.J., Tiedje, J.M. and Nüsslein, K., 2016. Land-use change drives abundance and community structure alterations of thaumarchaeal ammonia oxidizers in tropical rainforest soils in Rondônia, Brazil. Applied soil ecology, 107, pp.48-56.

Hartwell, L.H., Hopfield, J.J., Leibler, S. and Murray, A.W., 1999. From molecular to modular cell biology. Nature, 402, pp.C47-C52.

Herpin, U., Cerri, C.C., Carvalho, M.C.S., Markert, B., Enzweiler, J., Friese, K. and Breulmann, G., 2002. Biogeochemical dynamics following land use change from forest to pasture in a humid tropical area (Rondonia, Brazil): a multi-element approach by means of XRF-spectroscopy. Science of the total environment, 286(1), pp.97-109.

Hiraishi, A., and Imhoff, J.F. 1994. Genus *Rhodoplanes*, p. 545-549. In Brenner, D.J., Krieg, N.R., and Staley, J.T (Eds.). Bergey's Manual of Systematic Bacteriology: Volume Two: The Proteobacteria, Part 3. Springer Science & Business Media (New York, NY).

Horner-Devine, M.C., Carney, K.M. and Bohannan, B.J., 2004. An ecological perspective on bacterial biodiversity. Proceedings of the royal society of london B: biological sciences, 271(1535), pp.113-122.

Horner-Devine, M.C., Silver, J.M., Leibold, M.A., Bohannan, B.J., Colwell, R.K., Fuhrman, J.A., Green, J.L., Kuske, C.R., Martíny, J.B., Muyzer, G. and Øvreås, L., 2007. A comparison of taxon co-occurrence patterns for macro-and microorganisms. Ecology, 88(6), pp.1345-1353.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L. and Welch, D.M., 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome biology, 8(7), p.1.

INPE. 2011. Program for the Estimation of Amazon Deforestation (Projeto PRODES Digital). Brazilian national institute for space research.

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. The ISME journal, 3(9), pp.1004-1011.

Kanehisa, M. and Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), pp.27-30.

Kauffman, J.B., Cummings, D.L., Ward, D.E. and Babbitt, R., 1995. Fire in the Brazilian Amazon: 1. Biomass, nutrient pools, and losses in slashed primary forests. Oecologia, 104(4), pp.397-408.

Knights, D., Costello, E.K. and Knight, R., 2011. Supervised classification of human microbiota. FEMS microbiology reviews, 35(2), pp.343-359.

Koch, I.H., Gich, F., Dunfield, P.F. and Overmann, J., 2008. *Edaphobacter modestus* gen. nov., sp. nov., and *Edaphobacter aggregans* sp. nov., acidobacteria isolated from

alpine and forest soils. International journal of systematic and evolutionary microbiology, 58(5), pp.1114-1122.

Konings, W.N., Lolkema, J.S., Bolhuis, H., Van Veen, H.W., Poolman, B. and Driessen, A.J.M., 1997. The role of transport processes in survival of lactic acid bacteria, Energy transduction and multidrug resistance. Antonie van leeuwenhoek, 71(1-2), pp.117-128.

Krave, A.S., Lin, B., Braster, M., Laverman, A.M., Straalen, N.M.V., Röling, W.F. and Verseveld, H.W.V., 2002. Stratification and seasonal stability of diverse bacterial communities in a *Pinus merkusii* (pine) forest soil in central Java, Indonesia. Environmental microbiology, 4(6), pp.361-373.

Kuczynski, J., Stombaugh, J., Walters, W.A., González, A., Caporaso, J.G. and Knight, R., 2012. Using QIIME to analyze 16S rRNA gene sequences from microbial communities. Current protocols in microbiology, pp.1E-5.

Laland, S. and Zimmer, T.L., 1972. The protein thiotemplate mechanism of synthesis for the peptide antibiotics produced by *Bacillus brevis*. Essays in biochemistry, 9, pp.31-57.

Lange, M., Eisenhauer, N., Sierra, C.A., Bessler, H., Engels, C., Griffiths, R.I., Mellado-Vázquez, P.G., Malik, A.A., Roy, J., Scheu, S., and Steinbeiss, S. 2015. Plant diversity increases soil microbial activity and soil carbon storage. Nature communications, 6.

Langenheder, S. and Székely, A.J., 2011. Species sorting and neutral processes are both important during the initial assembly of bacterial communities. The ISME journal, 5(7), pp.1086-1094.

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Thurber, R.L.V., Knight, R. and Beiko, R.G., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene

sequences. Nature biotechnology, 31(9), pp.814-821.

Lauber, C.L., Hamady, M., Knight, R. and Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Applied and environmental microbiology, 75(15), pp.5111-5120.

Leibold, M.A. and McPeek, M.A., 2006. Coexistence of the niche and neutral perspectives in community ecology. Ecology, 87(6), pp.1399-1410.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C. and Schleper, C., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature, 442(7104), pp.806-809.

Lidicker, W.Z., 1979. A clarification of interactions in ecological systems. BioScience, 29(8), pp.475-477.

Liu, Y. and Whitman, W.B., 2008. Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. Annals of the new york academy of sciences, 1125(1), pp.171-189.

Lors, C., Ryngaert, A., Périé, F., Diels, L. and Damidot, D., 2010. Evolution of bacterial community during bioremediation of PAHs in a coal tar contaminated soil. Chemosphere, 81(10), pp.1263-1271.

Love, M.I., Huber, W. and Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology, 15(12), p.1.

Ma, B., Wang, H., Dsouza, M., Lou, J., He, Y., Dai, Z., Brookes, P.C., Xu, J. and Gilbert, J.A., 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. The ISME journal, 10, pp.1891–1901

MacLean, R.C., Bell, G. and Rainey, P.B., 2004. The evolution of a pleiotropic fitness

tradeoff in *Pseudomonas fluorescens*. Proceedings of the national academy of sciences, 101(21), pp.8072-8077.

Mäder, P., Fliessbach, A., Dubois, D., Gunst, L., Fried, P. and Niggli, U., 2002. Soil fertility and biodiversity in organic farming. Science, 296(5573), pp.1694-1697.

Madsen, E.L., 2011. Microorganisms and their roles in fundamental biogeochemical cycles. Current opinion in biotechnology, 22(3), pp.456-464.

Manojlovic, M., Cabilovski, R. and Bavec, M., 2010. Organic materials: sources of nitrogen in the organic production of lettuce. Turkish journal of agriculture and forestry, 34(2), pp.163-172.

Martens-Habbena, W., Berube, P.M., Urakawa, H., José, R. and Stahl, D.A., 2009. Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. Nature, 461(7266), pp.976-979.

Martiny, J.B.H., Bohannan, B.J., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R. and Morin, P.J., 2006. Microbial biogeography: putting microorganisms on the map. Nature reviews microbiology, 4(2), pp.102-112.

Mary, I., Tarran, G.A., Warwick, P.E., Terry, M.J., Scanlan, D.J., Burkill, P.H. and Zubkov, M.V., 2008. Light enhanced amino acid uptake by dominant bacterioplankton groups in surface waters of the Atlantic Ocean. FEMS microbiology ecology, 63(1), pp.36-45.

Melillo, J.M., Steudler, P.A., Feigl, B.J., Neill, C., Garcia, D., Piccolo, M.C., Cerri, C.C. and Tian, H., 2001. Nitrous oxide emissions from forests and pastures of various ages in the Brazilian Amazon. Journal of geophysical research: atmospheres, 106(D24),

pp.34179-34188.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A. and Wilkening, J., 2008. The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC bioinformatics, 9(1), p.386.

Mirza, B.S. and Rodrigues, J.L., 2012. Development of a direct isolation procedure for free-living diazotrophs under controlled hypoxic conditions. Applied and environmental microbiology, 78(16), pp.5542-5549.

Mirza, B.S., Potisap, C., Nüsslein, K., Bohannan, B.J. and Rodrigues, J.L., 2014. Response of free-living nitrogen-fixing microorganisms to land use change in the Amazon rainforest. Applied and environmental microbiology, 80(1), pp.281-288.

Montoya, J.M., Pimm, S.L. and Solé, R.V., 2006. Ecological networks and their fragility. Nature, 442(7100), pp.259-264.

Morris, B.E., Henneberger, R., Huber, H. and Moissl-Eichinger, C., 2013. Microbial syntrophy: interaction for the common good. FEMS microbiology reviews, 37(3), pp.384-406.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R. and Gordon, J.I., 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science, 332(6032), pp.970-974.

Mueller, R.C., Rodrigues, J.L., Nüsslein, K. and Bohannan, B.J., 2016. Land use change in the Amazon rainforest favors generalist fungi. Functional ecology.

Naclerio, G., Ricca, E., Sacco, M. and De Felice, M., 1993. Antimicrobial activity of a

newly identified bacteriocin of *Bacillus cereus*. Applied and environmental microbiology, 59(12), pp.4313-4316.

Navarrete, A.A., Taketani, R.G., Mendes, L.W., Cannavan, F.D.S., Moreira, F.M.D.S. and Tsai, S.M., 2011. Land-use systems affect archaeal community structure and functional diversity in western Amazon soils. Revista brasileira de ciência do solo, 35(5), pp.1527-1540.

Navarrete, A.A., Venturini, A.M., Meyer, K.M., Klein, A.M., Tiedje, J.M., Bohannan, B.J., Nüsslein, K., Tsai, S.M. and Rodrigues, J.L., 2015. Differential Response of Acidobacteria Subgroups to Forest-to-Pasture Conversion and Their Biogeographic Patterns in the Western Brazilian Amazon. Frontiers in microbiology, 6.

Navas-Molina, J.A., Peralta-Sánchez, J.M., González, A., McMurdie, P.J., Vázquez-Baeza, Y., Xu, Z., Ursell, L.K., Lauber, C., Zhou, H., Song, S.J. and Huntley, J., 2013. Advancing our understanding of the human microbiome using QIIME. Methods in enzymology, 531, p.371-444.

Neill, C., Melillo, J.M., Steudler, P.A., Cerri, C.C., de Moraes, J.F., Piccolo, M.C. and Brito, M., 1997a. Soil carbon and nitrogen stocks following forest clearing for pasture in the southwestern Brazilian Amazon. Ecological applications, 7(4), pp.1216-1225.

Neill, C., Piccolo, M.C., Cerri, C.C., Steudler, P.A., Melillo, J.M. and Brito, M., 1997b. Net nitrogen mineralization and net nitrification rates in soils following deforestation for pasture across the southwestern Brazilian Amazon Basin landscape. Oecologia, 110(2), pp.243-252.

Neill, C., Piccolo, M.C., Melillo, J.M., Steudler, P.A. and Cerri, C.C., 1999. Nitrogen dynamics in Amazon forest and pasture soils measured by 15 N pool dilution. Soil

biology and biochemistry, 31(4), pp.567-572.

Neill, C., Piccolo, M.C., Steudler, P.A., Melillo, J.M., Feigl, B.J. and Cerri, C.C., 1995. Nitrogen dynamics in soils of forests and active pastures in the western Brazilian Amazon Basin. Soil biology and biochemistry, 27(9), pp.1167-1175.

Neill, C., Steudler, P.A., Garcia-Montiel, D.C., Melillo, J.M., Feigl, B.J., Piccolo, M.C. and Cerri, C.C., 2005. Rates and controls of nitrous oxide and nitric oxide emissions following conversion of forest to pasture in Rondônia. Nutrient cycling in agroecosystems, 71(1), pp.1-15.

Newman, M.E., 2006. Modularity and community structure in networks. Proceedings of the national academy of sciences, 103(23), pp.8577-8582.

Ober, D. and Hartmann, T., 1999. Homospermidine synthase, the first pathway-specific enzyme of pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase. Proceedings of the national academy of sciences, 96(26), pp.14777-14782.

Ofiţeru, I.D., Lunn, M., Curtis, T.P., Wells, G.F., Criddle, C.S., Francis, C.A. and Sloan, W.T., 2010. Combined niche and neutral effects in a microbial wastewater treatment community. Proceedings of the national academy of sciences, 107(35), pp.15345-15350.

Olesen, J.M., Bascompte, J., Dupont, Y.L. and Jordano, P., 2007. The modularity of pollination networks. Proceedings of the national academy of sciences, 104(50), pp.19891-19896.

Pandit, S.N., Kolasa, J. and Cottenie, K., 2009. Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. Ecology, 90(8), pp.2253-2262.

Papke, R.T. and Ward, D.M., 2004. The importance of physical isolation to microbial

diversification. FEMS microbiology ecology, 48(3), pp.293-303.

Paul, K., Nonoh, J.O., Mikulski, L. and Brune, A., 2012. "*Methanoplasmatales*," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. Applied and environmental microbiology, 78(23), pp.8245-8253.

Paula, F.S., Rodrigues, J.L., Zhou, J., Wu, L., Mueller, R.C., Mirza, B.S., Bohannan, B.J., Nüsslein, K., Deng, Y., Tiedje, J.M. and Pellizari, V.H., 2014. Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities. Molecular ecology, 23(12), pp.2988-2999.

Pichard, B., Larue, J.P. and Thouvenot, D., 1995. Gavaserin and saltavalin, new peptide antibiotics produced by Bacillus polymyxa. FEMS microbiology letters, 133(3), pp.215-218.

Pollard, K.S., Dudoit, S. and van der Laan, M.J., 2005. Multiple testing procedures: the multtest package and applications to genomics. In Bioinformatics and computational biology solutions using R and bioconductor (pp. 249-271). Springer New York.

Prober, S.M., Leff, J.W., Bates, S.T., Borer, E.T., Firn, J., Harpole, W.S., Lind, E.M., Seabloom, E.W., Adler, P.B., Bakker, J.D. and Cleland, E.E., 2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. Ecology letters, 18(1), pp.85-95.

Raes, J. and Bork, P., 2008. Molecular eco-systems biology: towards an understanding of community function. Nature reviews microbiology, 6(9), pp.693-699.

Ranjan, K., Paula, F.S., Mueller, R.C., Jesus, E.D.C., Cenciani, K., Bohannan, B.J., Nüsslein, K. and Rodrigues, J.L., 2015. Forest-to-pasture conversion increases the

diversity of the phylum Verrucomicrobia in Amazon rainforest soils. Frontiers in microbiology, 6.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodrigues, J.L., Pellizari, V.H., Mueller, R., Baek, K., Jesus, E.D.C., Paula, F.S., Mirza, B., Hamaoui, G.S., Tsai, S.M., Feigl, B. and Tiedje, J.M., 2013. Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. Proceedings of the national academy of sciences, 110(3), pp.988-993.

Saavedra, S., Stouffer, D.B., Uzzi, B. and Bascompte, J., 2011. Strong contributors to network persistence are the most vulnerable to extinction. Nature, 478(7368), pp.233-235.

Sala, O.E., Chapin, F.S., Armesto, J.J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L.F., Jackson, R.B., Kinzig, A. and Leemans, R., 2000. Global biodiversity scenarios for the year 2100. Science, 287(5459), pp.1770-1774.

Sanford, R.A., Wagner, D.D., Wu, Q., Chee-Sanford, J.C., Thomas, S.H., Cruz-García, C., Rodríguez, G., Massol-Deyá, A., Krishnani, K.K., Ritalahti, K.M. and Nissen, S., 2012. Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. Proceedings of the national academy of sciences, 109(48), pp.19709-19714.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J. and Sahl, J.W., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology, 75(23), pp.7537-7541.

Schmitz, O., Boison, G., Salzmann, H., Bothe, H., Schütz, K., Wang, S.H. and Happe, T., 2002. HoxE—a subunit specific for the pentameric bidirectional hydrogenase complex (HoxEFUYH) of cyanobacteria. Biochimica et biophysica acta (BBA)-bioenergetics, 1554(1), pp.66-74.

Seneviratne, G., 2000. Litter quality and nitrogen release in tropical agriculture: a synthesis. Biology and fertility of soils, 31(1), pp.60-64.

Shade, A., Peter, H., Allison, S.D., Baho, D.L., Berga, M., Bürgmann, H., Huber, D.H., Langenheder, S., Lennon, J.T., Martiny, J.B. and Matulich, K.L., 2012. Fundamentals of microbial community resistance and resilience. Frontiers in microbiology, 3.

Shannon, C.E. 1948. A mathematical theory of communication. The Bell system technical journal. 27: 379–423 and 623–656.

Shannon, C.E., 2001. A mathematical theory of communication. ACM sigmobile mobile computing and communications review, 5(1), pp.3-55.

Silo-Suh, L.A., Lethbridge, B.J., Raffel, S.J., He, H., Clardy, J. and Handelsman, J., 1994. Biological activities of two fungistatic antibiotics produced by Bacillus cereus UW85. Applied and environmental microbiology, 60(6), pp.2023-2030.

Silver, L.L., 2011. Challenges of antibacterial discovery. Clinical microbiology reviews, 24(1), pp.71-109.

Sly, L.I., Hodgkinson, M.C. and Arunpairojana, V., 1988. Effect of water velocity on the early development of manganese-depositing biofilm in a drinking-water distribution system. FEMS microbiology ecology, 4(3-4), pp.175-186.

Smillie, C.S., Smith, M.B., Friedman, J., Cordero, O.X., David, L.A. and Alm, E.J., 2011. Ecology drives a global network of gene exchange connecting the human

microbiome. Nature, 480(7376), pp.241-244.

Soares-Filho, B.S., Nepstad, D.C., Curran, L.M., Cerqueira, G.C., Garcia, R.A., Ramos, C.A., Voll, E., McDonald, A., Lefebvre, P. and Schlesinger, P., 2006. Modelling conservation in the Amazon basin. Nature, 440(7083), pp.520-523.

Sole, R.V. and Montoya, M., 2001. Complexity and fragility in ecological networks. Proceedings of the royal society of london B: biological sciences, 268(1480), pp.2039-2045.

Spang, A., Poehlein, A., Offre, P., Zumbrägel, S., Haider, S., Rychlik, N., Nowka, B., Schmeisser, C., Lebedeva, E.V., Rattei, T. and Böhm, C., 2012. The genome of the ammonia-oxidizing *Candidatus Nitrososphaera gargensis*: insights into metabolic versatility and environmental adaptations. Environmental microbiology, 14(12), pp.3122-3145.

Stahl, D.A. and de la Torre, J.R., 2012. Physiology and diversity of ammonia-oxidizing archaea. Annual review of microbiology, 66, pp.83-101.

Steele, J.A., Countway, P.D., Xia, L., Vigil, P.D., Beman, J.M., Kim, D.Y., Chow, C.E.T., Sachdeva, R., Jones, A.C., Schwalbach, M.S. and Rose, J.M., 2011. Marine bacterial, archaeal and protistan association networks reveal ecological linkages. The ISME journal, 5(9), pp.1414-1425.

Stegen, J.C., Lin, X., Konopka, A.E. and Fredrickson, J.K., 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. The ISME journal, 6(9), pp.1653-1664.

Steinfeld, H., Gerber, P., Wassenaar, T., Castel, V., Rosales, M. and De Haan, C., 2006. Livestock's long shadow (p. 392). Rome: FAO.

Steudler, P.A., Melillo, J.M., Feigl, B.J., Neill, C., Piccolo, M.C. and Cerri, C.C., 1996. Consequence of forest-to-pasture conversion on CH4 fluxes in the Brazilian Amazon Basin. Journal of geophysical research: atmospheres, 101(D13), pp.18547-18554.

Stone, K.J. and Strominger, J.L., 1971. Mechanism of action of bacitracin: complexation with metal ion and C55-isoprenyl pyrophosphate. Proceedings of the national academy of sciences, 68(12), pp.3223-3227.

Sul, W.J., Cole, J.R., Jesus, E.D.C., Wang, Q., Farris, R.J., Fish, J.A. and Tiedje, J.M., 2011. Bacterial community comparisons by taxonomy-supervised analysis independent of sequence alignment and clustering. Proceedings of the national academy of sciences, 108(35), pp.14637-14642.

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. Microbial ecology, 59(4), pp.734-743.

Tettelin, H., Riley, D., Cattuto, C. and Medini, D., 2008. Comparative genomics: the bacterial pan-genome. Current opinion in microbiology, 11(5), pp.472-477.

Tiedje, J.M., 1988. Ecology of denitrification and dissimilatory nitrate reduction to ammonium. Biology of anaerobic microorganisms, 717, pp.179-244.

Torsvik, V., Goksøyr, J. and Daae, F.L., 1990. High diversity in DNA of soil bacteria. Applied and environmental microbiology, 56(3), pp.782-787.

Ude, S., Arnold, D.L., Moon, C.D., Timms-Wilson, T. and Spiers, A.J., 2006. Biofilm formation and cellulose expression among diverse environmental *Pseudomonas* isolates. Environmental microbiology, 8(11), pp.1997-2011.

Valiente-Banuet, A. and Verdú, M., 2007. Facilitation can increase the phylogenetic

diversity of plant communities. Ecology letters, 10(11), pp.1029-1036.

van Kessel, M.A., Speth, D.R., Albertsen, M., Nielsen, P.H., den Camp, H.J.O., Kartal, B., Jetten, M.S. and Lücker, S., 2015. Complete nitrification by a single microorganism. Nature, 528(7583), pp.555-559.

Větrovský, T., Steffen, K.T. and Baldrian, P., 2014. Potential of cometabolic transformation of polysaccharides and lignin in lignocellulose by soil *Actinobacteria*. PloS one, 9(2), p.e89108.

Vick-Majors, T.J., Priscu, J.C. and Amaral-Zettler, L.A., 2014. Modular community structure suggests metabolic plasticity during the transition to polar night in ice-covered Antarctic lakes. The ISME journal, 8(4), pp.778-789.

Wallenstein, M.D., McMahon, S. and Schimel, J., 2007. Bacterial and fungal community structure in Arctic tundra tussock and shrub soils. FEMS microbiology ecology, 59(2), pp.428-435.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology, 73(16), pp.5261-5267.

Ward, N.L., Challacombe, J.F., Janssen, P.H., Henrissat, B., Coutinho, P.M., Wu, M., Xie, G., Haft, D.H., Sait, M., Badger, J. and Barabote, R.D., 2009. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. Applied and environmental microbiology, 75(7), pp.2046-2056.

Watts, D.J. and Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature, 393(6684), pp.440-442.

Widder, S., Besemer, K., Singer, G.A., Ceola, S., Bertuzzo, E., Quince, C., Sloan, W.T.,

Rinaldo, A. and Battin, T.J., 2014. Fluvial network organization imprints on microbial co-occurrence networks. Proceedings of the national academy of sciences, 111(35), pp.12799-12804.

Williams, R.J., Howe, A. and Hofmockel, K.S., 2014. Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. Frontiers in microbiology, 5, p.358.

Yoon, S., Cruz-García, C., Sanford, R., Ritalahti, K.M. and Löffler, F.E., 2015. Denitrification versus respiratory ammonification: environmental controls of two competing dissimilatory NO3−/NO2− reduction pathways in *Shewanella loihica* strain PV-4. The ISME journal, 9(5), pp.1093-1104.

Zheng, Y.M., Cao, P., Fu, B., Hughes, J.M. and He, J.Z., 2013. Ecological drivers of biogeographic patterns of soil archaeal community. PloS one, 8(5), p.e63375.

Zhou, J., Deng, Y., Luo, F., He, Z. and Yang, Y., 2011. Phylogenetic molecular ecological network of soil microbial communities in response to elevated CO2. MBio, 2(4), pp.e00122-11.

Zhou, J., Xia, B., Treves, D.S., Wu, L.Y., Marsh, T.L., O'Neill, R.V., Palumbo, A.V. and Tiedje, J.M., 2002. Spatial and resource factors influencing high microbial diversity in soil. Applied and environmental microbiology, 68(1), pp.326-334

Biography

Md Abdul Wadud Khan was born in Sirajganj district, Bangladesh on November 20, 1984 to Md Joynul Abedin Khan and Rekhanur Begum. During his stay in Sirajganj situated by the river Jamuna, he completed his Secondary School Certificate (SSC; equivalent to 10^{th} grade) exam from a local high school with first division. Later, he moved to the capital city of Bangladesh, Dhaka, which is about 100 miles southeast of his birthplace for higher education. He got admitted in Science at Notre Dame College through a highly competitive admission test, where he completed Higher Secondary Certificate (HSC; equivalent to 12^{th} grade) exam with first division. He then got enrolled in the Department of Microbiology, University of Dhaka in 2003. Following graduation in 2008, he did master's from the same department in 2009, and then moved to Eastern Illinois University, USA for a second master's degree that he completed in 2011. While doing Doctor of Philosophy (PhD; Fall, 2012 - Fall, 2016) in Quantitative Biology (Microbiology track), he was working as a Graduate Teaching Assistant (GTA). Wadud's research focuses on the taxonomic and functional diversities of soil microbial communities in the Amazon rainforest, and response of these diversities to deforestation. Three chapters of dissertation are expected to be published in peer-reviewed journals. His long-term interests are to pursue postdoctoral research on Human Gut Microbiome, and then look for a faculty position in the field of biomedical sciences.