

A dense tropical forest scene with a person in the background. The person is wearing a light-colored shirt and dark pants, and is looking down at the ground. The forest is filled with various green plants and trees, with sunlight filtering through the canopy.

# **Lessons from the Amazon biome: impact of management practices on aboveground-belowground interactions**

**Márcio Fernandes Alves Leite**



NEDERLANDS INSTITUUT VOOR ECOLOGIE (NIOO-KNAW)  
NETHERLANDS INSTITUTE OF ECOLOGY (NIOO-KNAW)



**Lessons from the Amazon biome:  
impact of management practices on  
aboveground-belowground interactions**

Márcio Fernandes Alves Leite

## **Thesis committee**

Prof. dr. P.W. Crous

Prof. dr. H. Ter Steege

Prof. dr. J. Falcao-Salles

Prof. dr. ir. Max Rietkerk

Dr. Y. Hautier

Cover Design: Márcio Fernandes Alves Leite

Printing: ProefschriftMaken

ISBN: 978-94-6423-671-2

This thesis should be cited as: Leite, M.F.A. (2022) Lessons from the Amazon biome: impact of management practices on aboveground-belowground interactions. PhD thesis. Utrecht University, Utrecht, the Netherlands. DOI: <https://doi.org/10.33540/820>

# Lessons from the Amazon biome: impact of management practices on aboveground- belowground interactions

Lessen uit het Amazone: effecten van landbouwbeheer op boven- en  
ondergronds interacties

**(met een samenvatting in het Nederlands)**

## Proefschrift

ter verkrijging van de graad van doctor aan de  
Universiteit Utrecht  
op gezag van de  
rector magnificus, prof.dr. H.R.B.M. Kummeling,  
ingevolge het besluit van het college voor promoties  
in het openbaar te verdedigen op

woensdag 9 maart 2022 des middags te 2.15 uur

door

Márcio Fernandes Alves Leite

geboren op 29 augustus 1989  
te São Luís, Maranhão, Brazilië

## Promotoren:

Prof. dr. G.A. Kowalchuk  
Prof. dr. E.E. Kuramae

This thesis was (partly) accomplished with financial support from Coordination for the Improvement of Higher Education Personnel (CAPES), and the Research Fund of Maranhão State (FAPEMA).

# TABLE OF CONTENTS

<b>Chapter 1</b>	General Introduction	7
<b>Chapter 2</b>	Spatial organization of vegetation changes along secondary succession in central Amazonian forests	24
<b>Chapter 3</b>	The effect of land-use on aboveground biomass and soil quality indicators in spontaneous forests and agroforests of eastern Amazonia	50
<b>Chapter 4</b>	You must choose, but choose wisely: a review of model-based approaches for microbial community analysis	78
<b>Chapter 5</b>	Towards the Amazon mature forest microbiome: secondary succession convergence to fungi and agroforests divergence to bacteria	108
<b>Chapter 6</b>	General Discussion	132
<b>References</b>		152
<b>Summary</b>		175
<b>Samenvatting</b>		177
<b>Acknowledgments</b>		180
<b>About the author</b>		183
<b>Publication List</b>		184



CHAPTER 1  
General Introduction

The Amazon landscape is one of the most astonishing, biodiverse, and endangered biomes on Earth. Part of this complexity is due to the interplay between plants and their associated soil communities. In the light of rapid deforestation, but also restoration efforts, it is important to understand the role the soil microbiome in succession of plant community composition and the development of aboveground-belowground interactions. However, we currently have a general lack of understanding the influence of land-use changes in soil microbiomes and the aboveground-belowground interactions in which they are involved.

In the following sections of this introduction, I discuss the roles of aboveground-belowground interactions in maintaining ecosystem stability within the Amazon Forest. I also examine how the traditional agricultural practices circumvent the nutrient-poor soils in a short-term but failed to achieve a long-term sustainability.

## **Amazon ecosystem: high plant diversity, high plant biomass, poor soil, and challenges for sustainable crop production**

Tropical rainforests are among the world's most complex and diverse ecosystems (Condit *et al.*, 2002; Laurance, 2015; Novotny *et al.*, 2010). The Amazon forests, the largest tropical rainforest, contain a vast biodiversity (Condit *et al.*, 2002), ranging from the largest trees (Slik *et al.*, 2015; Steege *et al.*, 2013, 2020) to the tiniest microorganisms (Lima *et al.*, 2015; Navarrete *et al.*, 2015). Many factors explain this ecosystem diversity and complexity: the homogenous climate with high precipitation rates, narrow thermal amplitude, high solar radiation and reduced constraint on seasonal timing of plant reproduction (Usinowicz *et al.*, 2017). In those conditions, plants can grow and reach higher levels of biomass ranging from 150 to 300 Mg ha<sup>-1</sup> (Saatchi *et al.*, 2007) even in deeply weathered and nutrient-poor soils (Lucas, 2001). In the tropical rainforest, trees store nutrients in the form of biomass, but the maintenance of nutrient cycling depends strongly on a complex network of

aboveground-belowground interactions.

In general, the soils across the Amazon are highly weathered and nutrient poor (Davidson *et al.*, 2004). These characteristics play a crucial role in determining species distributions, which according to Figueiredo *et al.* (2017) overcomes the climatic conditions. Forest ecosystems growing under the constraint of nutrient-poor substrates have developed several mechanisms to conserve the nutrients (Jordan & Herrera, 1981), such as a dense root mat in the topsoil and intensification of nutrient cycling through root-fungi symbiosis. Nutrients are thus accumulated in plant biomass by a complex net of plant-soil interactions that maintain nutrient cycling. It has only recently become possible to explore the nature of these interactions and how they may change over time (Flores-Ferrer *et al.*, 2021).

The storage of nutrients in plant biomass has led the tropical farmers to develop a land use system called ‘slash-and-burn’ agriculture with shifting cultivation. Slash-and-burn consists of cutting the vegetation, leaving it to dry for several days and then burning it (Pollini, 2014). Not only does this provide the necessary open space, but the fire also quickly mineralizes plant nutrients, which are thereby made available for crops (Giardina *et al.*, 2000; Juo & Manu, 1996). However, the low capability of soil nutrient retention (Laurance *et al.*, 1999; Schaefer *et al.*, 2008), combined with the climatic conditions of the tropics, favors soil nutrient loss. Consequently, a few years after slash-and-burn, most of the deposited nutrients in soil are depleted through runoff and leaching (Kauffman *et al.*, 1995; Thomaz, 2013). This rapid loss of soil nutrients leads to the necessity for farmers to abandon the land and seek for another piece of forest to start a new cultivation cycle. First, slash-and-burn strategies typically result in a peak of soil fertility in the first 1-2 years, thereby supporting mostly the growth of short cycle crops (e.g., maize, cassava, beans) (Gehring, 2006), and crop productivity quickly drops after this initial nutrient burst, rendering the land unsuitable for future high-yield cropping (Pollini, 2014). In sum, tropical smallholder farmers, representing the majority of the rural community in the Amazon, face an on-going challenge of how

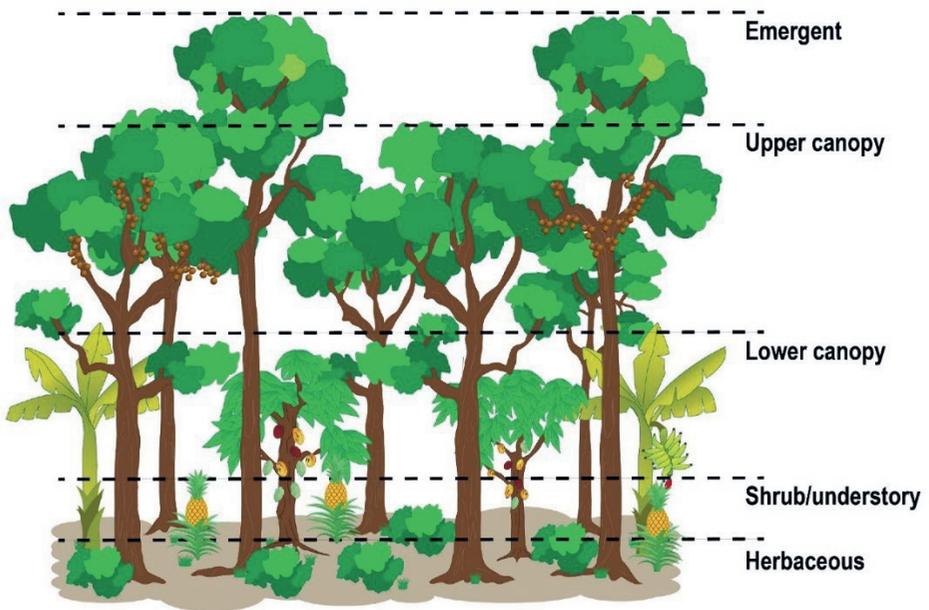
to manage Amazonian soils to allow for sustainable crop productivity without increasing deforestation (Jordan & Herrera, 1981).

Increased populations sizes and food demand bring the need to expand agricultural practices, thus resulting in land conversion via slash and burn (Brady, 1996). However, given the short-term nature of nutrient release, slash-and-burn agriculture is clearly not sustainable and come at a large environmental cost in terms of habitat loss, declines in biodiversity and reduction of carbon storage. Still, several steps can be taken to help ensure more sustainable use of Amazonian soil. According to Kang (1997), the maintenance of tropical soil productivity is possible by (i) retaining organic matter and adequate coverage of soil surface; (ii) minimizing disruption of the soil surface; (iii) appropriate use of fertilizers; and (iv) fallow and crop rotation. The first two topics depend strongly on the aboveground-belowground interactions. First, plants represent the input of organic matter and provide cover to the soil via littermass or plant canopy. Second, fallow allows plant diversity and plant biomass to recover, which contributes to strengthen the plant contribution to soil organic matter. Plant-soil interactions play a significant role in the ecosystem functioning and balance of Amazon tropical rainforest and are determinant of achieving sustainable crop production. To obtain sustainable crop production in tropical lands, the farmers should try to mimic the natural ecosystem functioning by (i) avoiding nutrient losses, (ii) promoting nutrient accumulation in plant biomass, and (iii) cycling nutrients through plant-soil feedbacks. One of the most promising agricultural practices to foster and mimic the plant-soil feedbacks found in tropical rainforests is the use of agroforestry systems (Nair, 2013).

## **Agroforestry systems: alternative land use systems for the Amazon**

Agroforestry systems are land use systems based on the combination of woody perennials trees with crops and/or animals to optimize and diversify crop

production per unit of area with an enhanced interaction between the components (Atangana, Khasa, Chang, Degrande, *et al.*, 2014; Nair, 2014) (Figure 1). This land-use system relies on the capacity of trees to capture nutrients from deep layers of soil and recycling them via root and litter deposition, a system called nutrient pumping (Isaac & Borden, 2019; Seneviratne *et al.*, 2006). Trees also provide protection against nutrient losses by runoff, erosion and leaching by reducing the impact of rain (Nair, 2013). The primary advantage of agroforestry systems is to control and manage ecosystem resources i.e., nutrient use efficiency and nutrient cycling in an integrated farm while delivering long-term food and crop production (Nair, 2014).

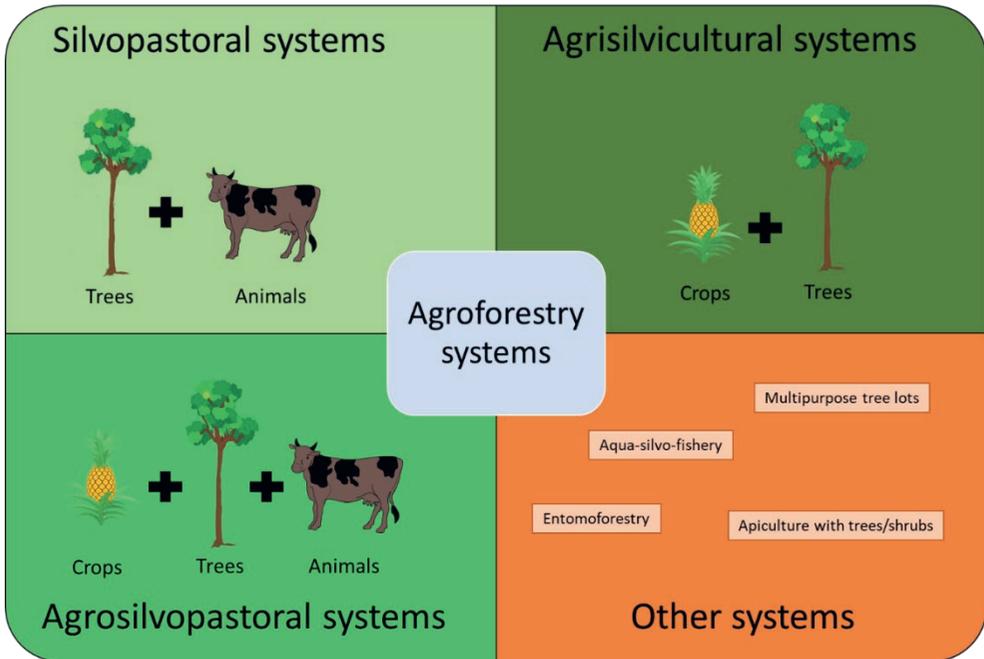


**Figure 1** – Representation of the multiple layers in an agroforestry system. Adapted from: Elevitch *et al.* (2018)

Agroforestry is a viable alternative to the slash-and-burn practices in the tropics (Atangana, Khasa, Chang, & Degrande, 2014a; Mbow *et al.*, 2014) because it can improve soil quality (Pinho *et al.*, 2012), provide ecosystem services such as sequestration of carbon, regulation of soil erosion, pest regulation, and pollination (Mohri *et al.*, 2013), and reduce economic risks by providing better market flexibility

via multiple products (H. Souza *et al.*, 2012). Furthermore, trees reduce the impact of rainfall that causes soil erosion and consequently nutrient loss (Labrière *et al.*, 2015), strengthening the so-called safety-net mechanisms (Seneviratne *et al.*, 2006). Therefore, trees in agroforestry systems might contribute significantly to the sustainability of crop production by playing two central roles: nutrient acquisition and soil conservation.

Agroforestry integrates biophysical, environmental, socio-economic and cultural factors. They can be functionally as well as economically complex (Atangana, Khasa, Chang, & Degrande, 2014b) with a range of designs and management strategies. This variety has been captured in several classification criteria that describe differences between agroforestry systems, thereby also highlighting potential advantages and limitations of various systems. Although no classification is all-encompassing, Atangana, Khasa, Chang, & Degrande (2014b) suggested grouping agroforestry systems according to (i) rank the systems in three main categories (agrosilvopastoral, silvopastoral and agrosilvicultural), and a specific category for systems such as multipurpose tree lots and aqua-silvo-fishery (combined into a single group of other systems); and, (ii) proceed with structural, functional, socio-economic or ecological classification (Figure 2).



**Figure 2** – Representation of the multiple types of an agroforestry system depending on the combination of trees, crops, and animals. Adapted from: Atangana, Khasa, Chang, & Degrande (2014b)

In this thesis, I focused on a subset of the rank categories, namely agrosilviculture with multipurpose trees (fruit crops, medicine, etc.). I further classified systems according to functional, socio-economic and ecological role. I focused on three main classes of agroforestry systems:

**Commercial agroforestry:** designed and optimized crop production for market/commercial purposes, operated in medium- or large-scale (>20 ha) with paid or contracted labor. Plants (trees and crops) are grown in a spatially regular arrangement, receive nutrient input, and are selected according to their capacity to yield high profits (Yamada & Gholz, 2002).

**Enriched agroforestry:** Natural secondary forests (5-15 years) are selected to for the addition of plants to produce high-value products such as fruit, medicines, or high-grade timber that provide economic benefits to households during the fallow

period (Brookfield & Padoch, 1994). This system takes advantage of the natural canopy recovered during secondary succession to grow crops and trees in the understory. Usually, these systems are designed and established with the help of governmental agencies or Non-Governmental Organization (NGO). Enriched agroforestry generates income in otherwise ‘unproductive’ forest. It thereby allows for crop cultivation in areas designated as legal reserves by governmental law or that would otherwise lie as fallow forest after land abandonment in slash-and-burn agriculture (Cardozo *et al.*, 2015).

**Homegarden agroforestry:** This is one of the oldest land-use systems in the world (Kumar & Nair, 2004). This type of agroforestry aims primarily to simply meet the basic needs of local households. Functionally, it is considered as a low-intense production system with a complex/random design and with an intensive presence of farmer management. Typically, a homegarden begins as an integral part of the small farmer’s house. Trees, shrubs, and herbs are selected for edible products and cash income, as well as for their aesthetic and ecological benefits. These characteristics allow homegarden agroforests to contain a wide range of species, which approximates the richness encountered in natural forests (Gajaseni & Gajaseni, 1999). The crop combinations in the homegardens are strongly influenced by the specific needs and preferences of the household and by the nutritional complementarity of the crops. In contrast to agriculture practices, homegardens typically have an irregular spatial arrangement that can vary over time, factors that contribute to the unique and multistrata characteristics of such systems.

In summary, agroforestry systems are designed, intentionally or not, to optimize resource use spatially, physically and temporally by maximizing positive and minimizing negative interactions between plants. The primary focus of these interactions relies on the tree-crop relationship or plant-animal interactions (Mohri *et al.*, 2013; Pinho *et al.*, 2011; Somarriba *et al.*, 2014). Only recently, have researchers begun to focus on plant-soil interactions at an ecosystem level to determine specifically how aboveground systems interact with the belowground

factors (Flores-Ferrer *et al.*, 2021). Such above-/belowground interactions are central to the sustainable functioning of agroforestry systems, with major goals of such systems including the improvement of soil physical properties, maintenance of soil organic matter, and promotion of nutrient cycling (Isaac & Borden, 2019; Schroth & Mota, 2014).

Given the importance of aboveground-belowground interactions on the sustainability of agroforestry systems, it is important to examine the role of agricultural management in reducing or intensifying these interactions. According to Hautier *et al.* (2015), the anthropogenic changes in the environment affects both plant species and soil organisms decreasing ecosystem stability, however for the tropical conditions, this remains largely unknown. Therefore, it is crucial to determine the degree of impact in the aboveground-belowground interactions by the different land-use systems to understand and evaluate the sustainability of agroforestry systems.

## **Plant-soil-microorganism interactions and ecosystem functioning**

The study of aboveground-belowground interactions is a relatively new area of ecology (Bardgett *et al.*, 2005). There have been relatively more advances with respect to the study of belowground biotic and abiotic components (Kuramae *et al.*, 2012) and plant-plant interactions (Brooker, 2006). In the field of plant-plant interactions, we now have an increasing understanding of the spatial scales of plant coexistence (Hart *et al.*, 2017) by using point pattern analysis to infer the nature of interactions between plants (Wiegand & Moloney, 2004). Previous studies have highlighted how such spatial patterns capture different aspects of forest dynamics and biodiversity structure such as mortality rate, species richness, species abundance distribution, beta-diversity, and the species-area relationship (May *et al.*, 2015). According to Haase (2001), spatial pattern analysis of plants uses coordinates to

evaluate the point-to-point (plant-to-plant) distribution patterns and determines if there is some interaction between them, the so-called point pattern analysis. Another motivation for studying point patterns, according to Velázquez *et al.* (2016), relies on the possibility of investigating the process determining the disposition of the plants together with their characteristics as an imprint. Therefore, spatial pattern analysis functions as an ‘ecological archive’ that contain valuable information on the underlying process, structure, and function of the system.

On the other hand, the studies of belowground soil communities in tropical soils has provided insights into populations distributions over space and time (Phillips *et al.*, 2019; Rodrigues *et al.*, 2013), responses to global change (Nottingham *et al.*, 2019) and consequences of biodiversity change on plant community dynamics (Too *et al.*, 2018), aboveground and trophic interactions (Barnes *et al.*, 2014), and biogeochemical cycles (Sang *et al.*, 2012).

Advances in molecular ecology and metagenomic approaches are providing new tools to elucidate the detailed structure and metabolic potential of soil-borne microbial communities, thereby allowing them to be linked with aboveground diversity and functional patterns. Such studies are able to yield information related to the ecosystem processes and provide clues into mechanisms driving community assembly in such tropical ecosystems (Goss-Souza *et al.*, 2020). Only recently have researchers acknowledged the crucial role of these interactions in the development and functioning of tropical ecosystems (Berkelmann *et al.*, 2020).

Plants represent the primary source of organic matter inputs into soil via litterfall and rhizodeposition (Ayres *et al.*, 2009). Plants affect the belowground subsystem via leaf litter quality (Wright *et al.*, 2004) and specific exudation patterns can be part of strategies to maximize resource gain or conserve nutrients. Such strategies are particularly important in Amazonian soils due to their low capability for nutrient retention (Laurance *et al.*, 1999; Schaefer *et al.*, 2008). Given the complex network of aboveground and belowground interactions that drive the functioning of Amazon Forest systems, sustainable use of such systems should seek

to mimic and recreate these vital linkages between above and belowground components.

Better understanding of aboveground-belowground relationships within agroforestry systems is crucial for developing better agricultural practices capable of enhancing both productivity and sustainability. To unify the aboveground and belowground sub-systems, especially the soil microbiome, research has attempted to evaluate how macroecological models can explain and predict patterns of soil microbial abundance and diversity-abundance scaling (Shoemaker *et al.*, 2017). However, one should be careful when applying concepts used in macroecology to the microbial world. The complexity of soil microbial community, the high level spatial and temporal heterogeneity, the plasticity of microbial genomes and the high capacity for gene transfer create a scenario where the available theories of ecology might not be fully applicable to describing microbial population patterns.

Thus, while the use of macroecological models may be a useful starting point in evaluating the ecological principles governing the microbial world, they must be modified and adjusted to take the unique features of microbial communities into account, including limitations of microbial community data. The main advantage of applying macroecological models consists in the possibility to assess both plant community patterns (via spatial analysis) and soil microbial community structure (via metagenomics) in a unified framework, thereby potentially gaining some insight into the complex nature of aboveground-belowground interactions. A promising unifying framework to understand the changes in plant-soil-microbe interactions consists in evaluating the context-dependence of interactions (Chamberlain *et al.*, 2014) to achieve a better understanding of the trade-offs between each component.

## **Trade-offs in aboveground-belowground interactions**

According to Chamberlain et al. (2014), interactions between different organisms may change in magnitude and/or sign as a function of the biotic or abiotic context, the so-called context-dependent interactions. A range of studies has been devoted to examine the relative importance of different environmental factors in steering soil-borne microbial community structure (Kuramae *et al.*, 2012; Rousk *et al.*, 2010; Yan *et al.*, 2019). Each community presents a different degree of co-dependence with the aboveground and belowground subsystems. Those different degrees of interactions suggest a potential trade-off between aboveground and belowground sub-systems. For example, in their evaluation of a long-term experiment, albeit not in a tropical system, Cassman et al. (2016) provided evidence that bacterial community shifts are mainly due to changes in soil factors (the belowground sub-system), whereas fungal communities were more linked to plant community diversity (the aboveground sub-system). For the tropics, Mueller et al. (2014) reported that the variability of plant community composition rather than soil properties or geographic distance affects fungal community composition, suggesting that land-use changes (e.g., deforestation) might interfere in the strong link between above- and belowground communities.

The environmental context has a strong influence on the soil microbiome. For example, Zhou et al. (2021) showed that co-occurrences between bacteria and fungi might shift from negative to positive following a gradient of altitude. Additionally, Peay et al. (2013) showed that Peruvian tropical forest possesses a fungal community diversity more dependent on the plant diversity than on soil characteristics. However, those authors did not account for the influence of plant biomass and the litter layer, thus missing the aboveground component. Similarly, Cline & Zak (2015) examined the effect of secondary succession on soil microbial communities and showed that different microbial communities may display disparate responses to edaphic properties and plant material. Altogether, the effect

on the microbial interactions and especially on aboveground-belowground interactions remains superficial. To understand the impacts of land-use changes it is crucial that we evaluate the aboveground-belowground interactions.

## Thesis outline

In this PhD thesis, I focus on identifying shifts in aboveground-belowground interactions resulting from either natural secondary succession or anthropic management. First, I investigated plant-plant interactions with focus on the spatial arrangement of leguminous plants (trees and lianas). After that, I focused on plant-soil interactions by evaluating the co-dependence between aboveground biomass and soil nutrients. Here, I also was interested in evaluating the differences between secondary forest succession and agroforestry systems to examine the impacts of agroforestry practices on plant-soil interactions. Because of the importance of microbial community structure for soil ecosystem functioning, I extended the previous approach to include plant-soil-microbe interactions. I included the analysis of soil microbial communities with novel statistical tools for community analysis (obtained after a literature review on statistical methods) that allowed me to determine the impact of management in the interdependency of these three components (plant, soil, and microbiota). Based on these findings, I further discuss the implications of different management practices in the aboveground-belowground interactions. I also reflect on possible new approaches to develop land-uses that strengthen aboveground-belowground interactions as a crucial step for the sustainable land use in the tropics. The studies described here are the first attempt to identify key land-use practices that impact aboveground-belowground interactions. It may help to develop and implement techniques to minimize the consequences of anthropic land use.

Plant communities affect soil subsystem and vice-versa. Anthropic management impacts both plants and soil via different land uses; thus, interfering in

the co-dependence between plant, soil, and microbial components. However, we still know little about how anthropic management practices interfere in the co-dependence structure and trade-offs between the aboveground and belowground sub-systems. In this thesis, I focus on *identifying potential land uses that induce or increase the co-dependence between taxa of plants and soil microorganisms* by evaluating (1) successional changes in plant community assemblage in both taxonomic and functional composition; and (2) the effects of forest management on above- and belowground interactions between plant biomass and soil communities in two types of land-use systems in the Amazon forest: natural fallow after slash-and-burn agriculture and agroforestry systems.

Firstly, I considered plant-plant interactions via spatial arrangements. Although many studies have evaluated plant-plant interactions via point pattern analysis, to date changes in plant spatial interactions under natural land recovery after slash-and-burn agriculture remain largely unknown. In **Chapter 2**, I evaluated how the spatial pattern of plants changes along natural land recovery gradient after slash-and-burn practices in the Brazilian Amazon Biome. I focused on how the spatial pattern of leguminous plants varies over a chronosequence of secondary forests in comparison with mature Amazonian Forest. The findings suggest more clustered arrangements and less stochastic distributions of leguminous plants in young secondary forests than in late successional stages or mature forests. Moreover, this chapter highlights the impact of repeated cycles of slash and burn agriculture and the potential of this practice to impact plant-plant interactions, suggesting that plant-plant interactions might increase under the stress induced by slash-and-burn degradation.

In **Chapter 3**, I studied at how land-use interferes in the plant-soil relationship by comparing secondary successional systems with agroforests. Here, I studied managed systems with characteristics that are similar to Amazon secondary forest, but with different intensities of management (i.e., intensely managed commercial plantation agroforests, and less intensely managed enriched

agroforests). I explored the effects of natural succession on plant and soil properties to understand the impact of land-use on aboveground-belowground interactions. Land use was found to have an impact on soil quality indicators and aboveground biomass. These findings suggest that management reduced the strength of aboveground-belowground interactions, pointing to a disruption in the plant-litter-soil co-dependence. The results of this study also suggest that traditional homegarden agroforestry may represent a strategy for sustainable rural development system with a significant potential for improved management. Management of soil and understory vegetation generates a situation in which plant performance (biomass) is less dependent on soil factors.

Next, by including microorganisms in the framework of this thesis, I first tackled the challenge of amplicon sequencing bias in **Chapter 4**. In this chapter, I investigate different approaches for microbial data analysis, highlighting the potential role of a hierarchical model based on imperfect detection to achieve more accurate estimates of microbial community structure. This chapter also highlights the importance of including more replicates to produce more reliable estimates of microbial community composition.

In **Chapter 5**, I applied the bioinformatic and statistical tools studied in the previous chapter to investigate the impact of management on plant-soil-microbe interactions in the Amazon. Again, I compared the interactions during the successional process with the management of agroforestry systems. Agroforestry systems reshape aboveground-belowground interactions and increased the abundance of the soil bacterial community. On the other hand, mature forests and different stages of the secondary succession presented a higher abundance of fungal communities. Here, I also found a reduced covariance between plant biomass and soil factors in agroforestry systems, suggesting weaker co-dependence between aboveground and belowground components, a disruption of plant-soil-microbial interactions. In summary, the disruption of plant-soil-microbial interactions by agroforestry system practices resulted in land-use systems that do not mimic the

aboveground-belowground interactions occurring in mature tropical forest or being developed along the path of secondary succession.

In **Chapter 6**, I integrate the findings of this thesis and discuss the role of land use in determining the intensity of plant-soil-microbial interactions. I highlight the impacts of both secondary succession and agroforestry practices in plant, soil, and microbial sub-systems, and, more generally, the potential for more sustainable agricultural practices to foster more intense plant-soil-microbe interactions.

## CHAPTER 2

# Plant spatial organization changes along secondary succession in central Amazonian forests

**Authors**

Márcio Fernandes Alves Leite, Christoph Gehring

Submitted to: *Forest Ecology and Management*

## **Abstract**

Slash-and-burn agriculture and its intensification affects tropical forest succession and may likewise impact plant spatial organization. However, little is known about the impact of secondary vegetation succession on the spatial organization of plants. This study explores spatial organization along succession based on mapping and taxonomic identification /functional classification of all legume trees and lianas, and - for comparison - of 3 abundant pioneer tree and 3 abundant liana species/genera in central Amazonia. We investigate successional dynamics in an 18-site chronosequence approach covering 2-25 yrs of regeneration after first-cycle shifting cultivation and compare these with six mature rainforest control sites. Additionally, we considered the impact of land use intensification with three second-cycle slash-and-burn forests ('degraded' sites). We explore univariate and bivariate spatial patterns via O-ring statistics for legume trees and lianas, and we investigate the role of focal abundant species / genera on legume richness and evenness via individual species-area relationship (ISAR) and spatially explicit Shannon index. Monovariate distributions of all functional groups and of an exemplary abundant legume liana were clustered at a scale of <5 m throughout succession, clustering was stronger for lianas than for trees. Bivariate point patterns reveal both co-occurrence of some functional groups and random relationships between others, and bivariate point patterns differed between young regrowth and mature rainforests. Diversity analysis revealed that focal species/genera had random or slightly negative effects on surrounding legume species richness and slightly positive impacts on legume evenness in first-cycle secondary succession. However, only a moderate land-use intensification (2<sup>nd</sup>-cycle regrowth) strongly affected all aspects of spatial organization of secondary vegetation, increasing clustering, co-occurrences of functionally distinct plants, and systematic increase in focal species' negative impacts on surrounding legume diversity. Thus, land-use intensification affect spatial organization of spontaneous vegetation far more than first-cycle secondary succession does.

---

## Introduction

Ongoing deforestation of mature rainforests together with slash-and-burn cultivation converts Amazon forests into expanding secondary forests (Hauser & Norgrove, 2013). Secondary forests are known to differ systematically from mature rainforests in their structure (Gehring, Denich, *et al.*, 2005), species diversity (Klanderud *et al.*, 2009), and functional vegetation composition (Gehring, Vlek, *et al.*, 2005). However, degradation caused by multi-cycle land use with frequent fires and shortening fallow phases can retard or alter such successional trajectories (Jakovac, Bongers, *et al.*, 2016). Previous studies identified several processes that affect spatial patterns in mature rainforests, such as the mode and range of seed dispersal and density-dependent survival of juveniles - the Janzen-Connell hypothesis (Peters, 2003), tree-fall gaps (Velázquez & Wiegand, 2020), and other disturbance patterns which benefit opportunistic undergrowth and favor clusters of heliophytes, soil fertility / edaphic niches, fire regime (Yu *et al.*, 2009), as well as facilitation and competition between plant species (Taubert *et al.*, 2018; Wiegand *et al.*, 2007). A wide range of theories strive to explain species' co-existence and the high species diversity of tropical rainforests (Brown *et al.*, 2013; Levine & HilleRisLambers, 2009; Rosindell *et al.*, 2011). However, we still lack an understanding on such spatial vegetation patterns along secondary forest succession.

Due to their sessile nature, the spatial distribution of plants represents the result of a time-integrated series of underlying biotic and abiotic processes and interactions which form vegetation composition and diversity (Brown *et al.*, 2013; Wiegand *et al.*, 2007). Wiegand *et al.* (2007) showed how some species are capable of shaping diversity in their surroundings by acting as diversity repellers or accumulators. Therefore, not only single species but also species assemblies and species diversity vary in space. Individual key plant taxonomic or functional groups will differentially impact the biotic and abiotic environment in their surroundings, with specific effects on surrounding light availability (Smith & Reynolds, 2015),

insect communities (Kos *et al.*, 2015), litter-layer (Freschet *et al.*, 2012), soil chemistry, and microbiota in their surroundings (Uriarte *et al.*, 2015). Consequently, different species are likewise expected to act differentially on plant species composition and diversity in their surroundings. Some plant species have been classified as ‘successional facilitators’ in secondary forests, whereas others as ‘successional inhibitors’ with contrasting associations with overall vegetation diversity (Mesquita *et al.*, 2001). Previous research (Davidson *et al.*, 2004, 2007) highlighted the role of legume plants for nitrogen-recovery in secondary regrowth. Other plant traits such as growth form can also explain how plant species influence their surroundings. For example, the role of lianas is specially ambivalent: lianas are known to suppress tree growth, and increase tree mortality due to their high competitiveness in capturing light (Schnitzer & Carson, 2010). On the other hand, lianas may also increase resource availability, as species capable of biological N<sub>2</sub>-fixation dominate within the liana community especially in the early stages of succession, suggesting a key role in the restoration of the heavy nitrogen-losses caused by preceding slash-and-burn shifting cultivation (Gehring *et al.*, 2008). In summary, it remains unclear how such successional changes in vegetation composition translate into plant spatial patterns along succession, and how degradation caused by repeated slash-and-burn cycles might interfere in the spatial organization of successional vegetation assembly. Therefore, a closer look at plant traits (e.g., growth form) and functionality (N<sub>2</sub>-fixation capacity) might bring new perspectives on the processes that govern plant spatial arrangements.

This study investigates the spatial organization of woody legume trees (total of 147 species) and lianas (total of 54 species), and for comparison of selected non-legume pioneer trees (*Goupia glabra*, *Laetia procera*, and *Cecropia* spp.) and lianas (*Davilla rugosa*, *Memora* sp., and *Rourea* sp.). In central Amazon, we evaluated chronosequences of secondary succession after shifting cultivation and of mature rainforests controls of. We explore univariate spatial patterns and bivariate interactions at different taxonomic levels and in a combination of two different

functional classifications (i) growth form: trees vs. lianas, and (ii) the capacity to fix nitrogen: potentially nodulating / N<sub>2</sub>-fixing legume plant vs. non- N<sub>2</sub>-fixing plants (both legumes and selected non-legumes). We also searched for diversity repelling or attracting species or species-groups, and for successional shifts in spatial patterns and interactions. Our results improve our understanding of the successional processes of plant functional assemblages following slash-and-burn shifting cultivation land-use.

## Methods

### *Study region and sites*

Research was conducted in 18 study sites in two recently-developed slash-and-burn colonization areas, 110 km to the north (Presidente Figueiredo County) and 70 km to the east (Rio Preto da Eva county) of Manaus, central Amazonia, Brazil. Fifteen sites (7.08 ha) represent 3–25 years of secondary forest succession (grouped into ‘young’ 2-3-yrs.-old secondary forest (YSF, 1.46 ha), ‘mid-aged’ 5 - 10 yrs.-old secondary forest (MSF, 3.43 ha), and ‘old’ 12-25 yrs.-old first-cycle secondary forest regrowth (OSF, 2.19 ha). These are compared with six mature rainforest sites (MF, 4.3 ha) and a further three 4-11 yrs.-old degraded secondary forest sites after intensified land-use (2<sup>nd</sup>-cycle slash-and-burn) (DSF, 1.48 ha). The study sites were dispersed over an area of approximately 300 km<sup>2</sup>, thereby reducing the potential of regional singularities. Sites were grouped into four chronosequences, each consisting of differently aged secondary regrowth and one or two mature rainforest sites (Table S1): (i) young secondary forest (YSF, 2-3 yrs.-old, 3 sites); (ii) mid-aged secondary forest (MSF, 5-10 yrs.-old, 5 sites); (iii) old secondary forest (OSF, 12-25 yrs.-old, 3 sites); (iv) and undisturbed mature rainforest (MF, >100 yrs.-old, 4 sites). We also included secondary forest following second-cycle slash-and-burn agriculture, classified as degraded secondary forests (DSF, 4-11 yrs.-old, 3 sites), and considered

them as a proxy for land use intensification in Amazon (Jakovac, Peña-Claros, *et al.*, 2016). Soils of all sites were acid, nutrient-poor and highly weathered kaolinitic Oxisols (haplic Acrorthox). These soils are derived from the tertiary Barreiras sediments and are remarkably homogenous throughout large parts of Amazonia, with little variations in their low-fertility status (Cerri *et al.*, 1991; Moraes *et al.*, 1995). Both first- and second-cycle land-use was standardized to manual slash-and-burn and 1.5 yrs cassava cultivation. All sites were level and rectangular, site size ranged between 0.2-1.1 ha and increased with forest age (Supplementary Table S1). Site selection excluded 20-50 m wide boarder-zones to neighboring vegetation (mature rainforests: > 100 m), more detailed site descriptions are given in previously published papers (Gehring *et al.*, 2008; Gehring, Vlek, *et al.*, 2005).

### *Data collection and mapping*

We mapped (i) all legume trees (total of 3453 plants) and lianas (total of 4910 plants), as well as (ii) all individuals of selected (abundant) non-legume trees (total of 5 species, 3746 plants) and lianas (total of 3 species, 1927 plants) >50 cm height occurring within the study sites to a precision of 1 m ( $\pm 0.50$  m). As boarder-effect correction algorithms require rectangular areas (Haase, 2001), we cut off irregular site-edges preceding analysis.

Within legume vegetation, we identified >98% of all individuals to the species-level, in a total of 138 legume tree and 48 legume liana species. We subsequently classified legume tree and liana species according to their capability to nodulate and biologically fix N<sub>2</sub> under favorable conditions as a taxonomically defined plant trait resulting from the evolution of legume-rhizobia symbiosis (Bryan *et al.*, 1997; Doyle, 1994; Sprent, 2008) using literature data when available (de Faria & de Lima, 1998; F. M. S. Moreira *et al.*, 1993; F. Moreira & Silva, 1993; L. Souza *et al.*, 1994), exploratory root excavations (establishing nodulation in 3 legume liana species), or else following plant systematic relationships derived from

legume:rhizobia coevolution within the respective legume genus or tribe (Werner *et al.*, 2014). We classify 75.4% of all legume trees and 77.1% of all legume lianas as belonging to potentially nodulating/N<sub>2</sub>-fixing species, these percentages were systematically higher in all secondary forest sites than in the mature rainforest sites.

We furthermore mapped all plants >50 cm height of selected abundant and characteristic non-legume tree species (*Goupia glabra* Aubl., *Laetia procera* (Poepp.) Eichl., *Cecropia sciadophylla* Mart, *C. distachya* Huber, *C. purpurescens* Berg), and liana species (*Davilla rugosa* Poir, *Memora* sp., *Rourea cuspidata* Benth. ex Baker). We subsequently combined the three similar *Cecropia*-species at the genus level for further analyses. We selected these species due to their omnipresence within Amazonian secondary forests, with some of these species also occurring in the mature rainforest sites. We report on species, and functionally group all plants into four groups: potentially nodulating legume trees and lianas and selected non-legume trees and lianas.

### *Data analysis*

A common approach to plant spatial analysis consists of performing spatial point pattern statistics where plants are considered points within a two-dimensional space (Wiegand & Moloney, 2004). Monovariate point pattern analysis evaluates if the arrangement of plants is clustered, regular or random, and bivariate point pattern analysis reveals if two groups of plants (species or functional groups) are arranged to one another closer than chance ('co-occurrence'), more distant than chance ('repellence') or at random.

One limitation of spatial point pattern statistics is the need of a minimum 70 points for each pattern analyzed (Wiegand & Moloney, 2013). Most studies therefore concentrate on abundant species as focal plants, frequently comparing them with all the other plants as a heterospecific group. Our study follows a different approach. We focus on woody legumes (Fabaceae) and grouped them according to their growth

form (tress vs. lianas) and their taxonomically defined capability to nodulate and biologically fix  $N_2$ . We expand the group of non-fixing species by including four abundant non-legume tree species/genera and three abundant non-legume liana species. For each of the 18 sites we explore spatial distribution of functional groups with univariate and bivariate O-ring point pattern statistics (Wiegand & Moloney, 2004). Due to insufficient number of plants, not all combinations of plant groups and successional stages were possible. We, therefore, limited our analysis to the cases where we had a minimum of 3 sites per successional stage to ensure sufficient accuracy of the analysis. We compare observed point pattern distributions with complete spatial randomness (CSR) generated by 399 Monte Carlo simulations for the O-ring analysis both of univariate and bivariate point patterns.

In a second step, we selected two focal potentially  $N_2$ -fixing legume genera, the tree genus *Inga* (32 species) and the liana genus *Machaerium* (13 species), as well as the three non-legume pioneer trees *Laetia procera*, *Goupia glabra*, and *Cecropia* spp, (all with  $\geq 70$  individuals per site). We explore the effects of these focal plants as potential repellors or accumulators of legume species diversity in their surroundings, using the ISAR function (Wiegand *et al.*, 2007) to measure the richness and the spatial Shannon index (Rajala & Illian, 2012) to represent the evenness of plant diversity.

We evaluate departures from the null hypothesis using the procedures presented by Getzin *et al.* (2014) and extend their application for a continuous distance rather than a single range. This approach allows us to estimate standardized effect sizes for both O-ring and diversity measures. Our measure of the standardized effect size (deviation from the null hypothesis / simulation envelope) allows us to obtain a comparable measurement of our successional stages at the same distances while also including different sites as replicates. We combine the replicated patterns of data from different sites with the same successional stage into one summary effect size coupled with 95% confidence intervals, thus providing inferential information for statistical comparison of successional spatial patterns. We interpret positive

deviations  $>1.96$  from CSR as significant monovariate clustering or bivariate co-occurrence, as opposed to deviations below  $-1.96$  as significant monovariate uniform distribution or bivariate over-dispersion.

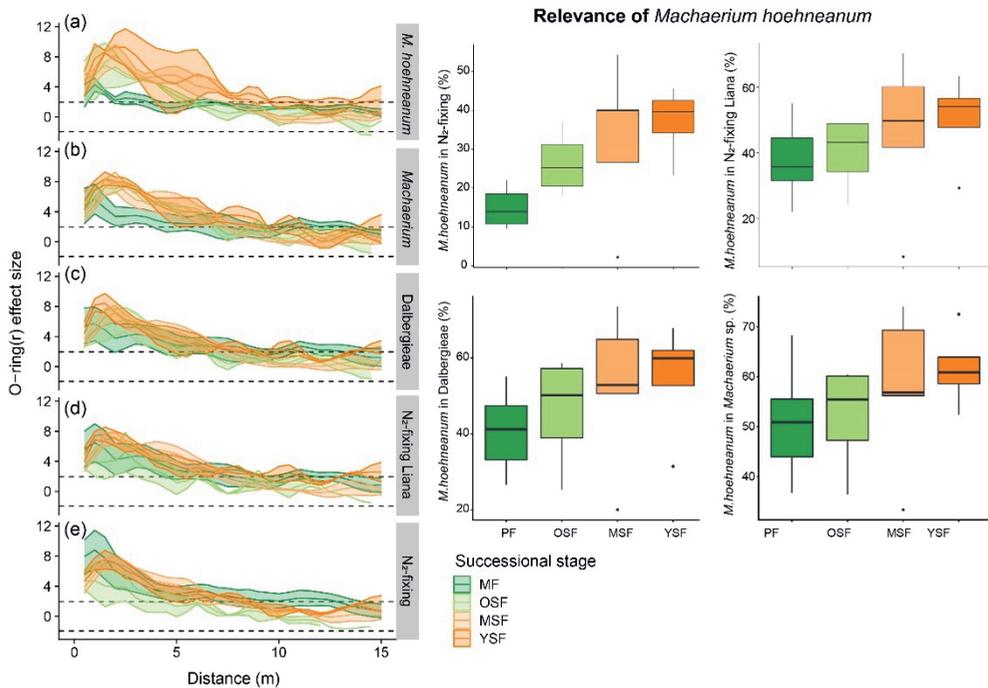
We also investigate possible effects of mixed species on the spatial patterns of the different groups. The diversity of plant species within each plant trait group ranged around 0.8-0.6 in the Pielou's J evenness index (Pielou, 1966). We furthermore explore the influence of successional stage on the diversity of each of our legume functional groups (nodulation potential + growth form). The J evenness index did not identify any statistically significant differences between successional stages, suggesting that no plant species dominated a single functional group (Figure S1). This makes us confident that the spatial patterns we observed refer more closely to the effect of the plant trait rather than a bias caused by a single outstanding species. All analyses were conducted in the R environment (R Development Core Team, 2007) using the spatstat (Baddeley & Turner, 2005) and spatalsegregation (Rajala & Illian, 2012) packages.

## Results

### *Taxonomic and functional levels of plant spatial organization*

We evaluated at which level of taxonomic hierarchy plants are spatially organized, based on the ubiquitous and abundant potentially  $N_2$ -fixing liana species *Machaerium hoehneanum* Ducke. Figure 1 shows (i) a spatial scale of significant aggregation at the first 5-10 m distance and random distribution beyond over all levels of taxonomic grouping over all stages of succession, but also (ii) a systematic reduction in effect sizes of clustering both with taxonomic aggregation and along succession and close to random distribution of *M. hoehneanum* in the mature rainforest controls. *M. hoehneanum* represented 37-74% of the abundance within the genus *Machaerium*, 17-73% of the tribe Dalbergieae, 8-70% of all legume lianas,

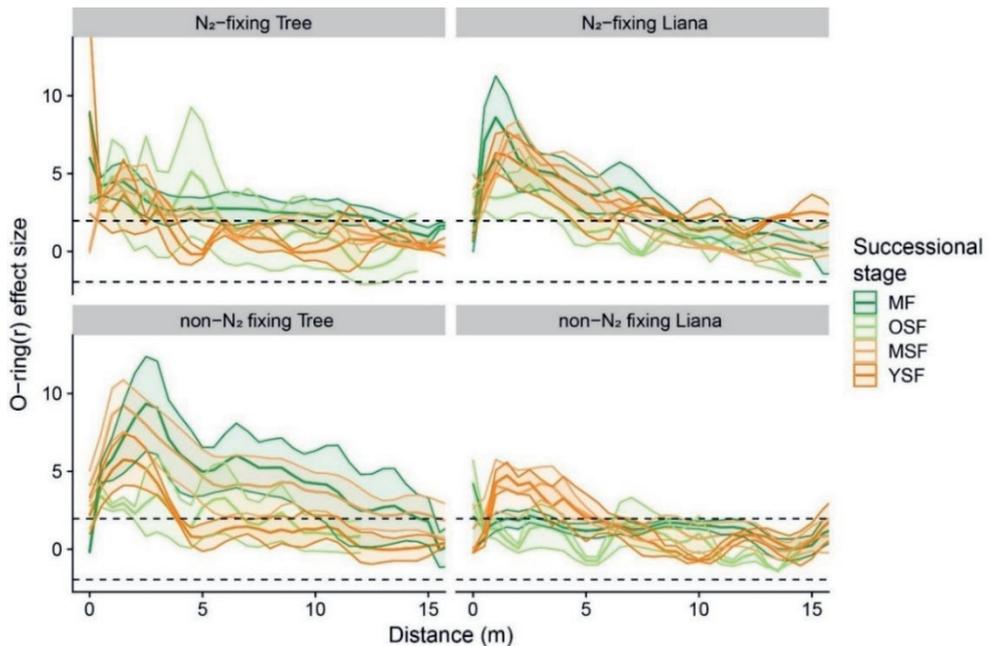
and 2-54% of all legume plants. Abundance-share of *M. hoehneanum* was highest in young secondary forest, decreased systematically along succession and was lowest in mature rainforest (Figure 1, boxplots).



**Figure 1** – Impact of taxonomic / functional level of species-grouping on spatial organization: Monovariate distribution of the abundant legume liana species *Machaerium hoehneanum* (a), of the genus *Machaerium* (b), the tribe Dalbergieae (c), and of the functional group of potentially N<sub>2</sub>-fixing legume lianas (d) and of all legume plants (e). Means  $\pm$  CI<sub>95%</sub> of sites within each group of 3-5 sites, only sites with  $\geq 70$  individuals included, positive deviation from dashed line +1.96 indicates significant clustering. YSF (young secondary forest, 2-3 yrs.-old, 3 sites), MSF (mid-aged secondary forest, 5-10 yrs.-old, 5 sites), OSF (old secondary forest, 12-25 yrs.-old, 3 sites), and MF (mature rainforest, >100 yrs.-old, 4 sites). Box-whisker-plots on the right depict *M. hoehneanum* abundance shares at the different levels of taxonomic aggregation along succession.

## *Impacts of secondary succession on plant spatial distributions and diversity*

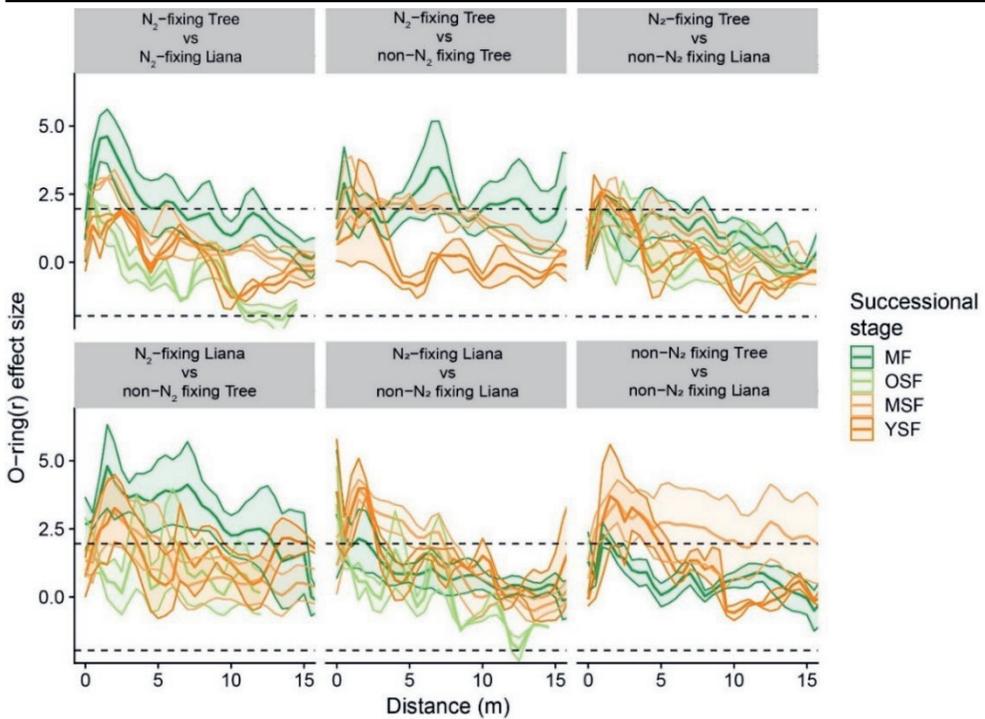
Spatial arrangement within functional groups (univariate point pattern analysis) revealed clustering around 5 m distance for most functional groups and stages of succession (Figure 2). However, this clustering tended to be stronger for potentially N<sub>2</sub>-fixing lianas. Maximum clustering of non N<sub>2</sub>-fixing trees in MF was likely caused by gap-opportunist concentration in former treefall-gaps. Effect-size of clustering was higher for potentially N<sub>2</sub>-fixing lianas than for non-N<sub>2</sub>-fixing lianas. Notably, the non-N<sub>2</sub>-fixing lianas showed stronger clustering during the early stages of succession (YSF and MSF) but shifts to random during the later stages (OSF and MF).



**Figure 2** – Departure from the null hypothesis (effect size) of univariate O-ring point pattern analysis for legume trees and lianas and for selected non-legume trees (3 genera) and lianas (3 genera) along a successional process in central Amazonia. Means  $\pm$  CI<sub>95%</sub> of sites within

each group of 3-5 sites, horizontal lines  $\pm 1.96$  indicate simulated CSR, only sites with  $\geq 70$  individuals included.  $N_2$ -fixing. = potential  $N_2$ -fixing; non- $N_2$ -fixing = non- $N_2$ -fixing plant, YSF (young secondary forest, 2-3 yrs.-old, 3 sites), MSF (mid-aged secondary forest, 5-10 yrs.-old, 5 sites), OSF (old secondary forest, 12-25 yrs.-old, 3 sites), and MF (mature rainforest, >100 yr.-old, 4 sites).

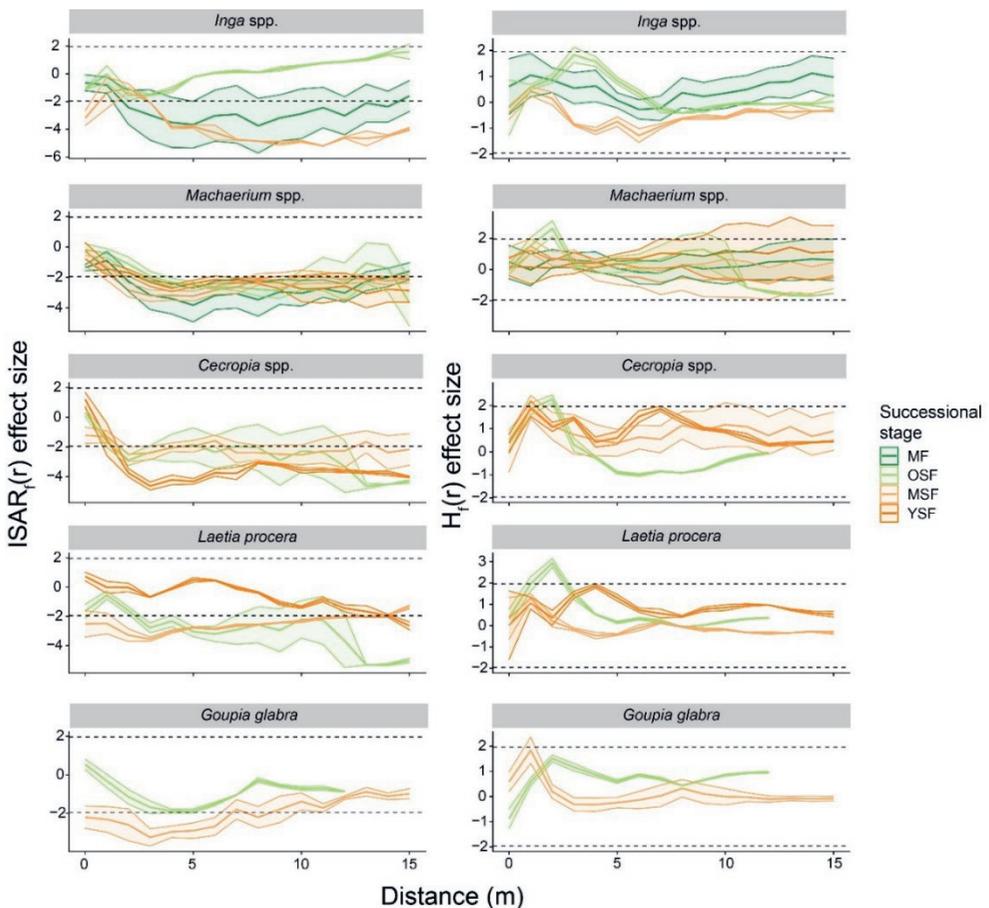
Plant spatial arrangement between functional groups (bivariate point pattern analysis) varied between random and co-occurrence, no evidence of repulsion for any bivariate point patterns (Figure 3). Positive departures from random indicate co-occurrence of  $N_2$ -fixing lianas and non- $N_2$ -fixing lianas and of non- $N_2$ -fixing trees and non- $N_2$ -fixing lianas in early stages of succession. For the MF, we observed a significant co-occurrence of  $N_2$ -fixing lianas and  $N_2$ -fixing trees, but  $N_2$ -fixing lianas were also associated with non- $N_2$ -fixing trees. Similarly, to univariate analysis, the non- $N_2$ -fixing lianas showed significant associations with non- $N_2$ -fixing trees. The results showed a strong clustering in YSF and MSF that declined to random associations in mature forest. On the other hand, the  $N_2$ -fixing lianas showed the same spatial organization regardless of the successional stage and the functional group that surrounds the plants. In summary, non- $N_2$ -fixing trees have a stronger co-occurrence with the non- $N_2$ -fixing lianas than the  $N_2$ -fixing lianas at earlier secondary stages.



**Figure 3** – Bivariate spatial organization of contrasting functional groups and growth forms. Departure from the null hypothesis (effect size) of bivariate O-ring point pattern analysis for potentially  $N_2$ -fixing and non- $N_2$ -fixing trees and lianas and aggregated functional groups along succession in central Amazonia, horizontal lines  $\pm 1.96$  indicate simulated CSR. Means  $\pm CI_{95\%}$  of sites within each successional group of 3-5 sites, only sites with  $\geq 70$  individuals included. YSF (young secondary forest, 2-3 yrs.-old, 3 sites), MSF (mid-aged secondary forest, 5-10 yrs.-old, 5 sites), OSF (old secondary forest, 12-25 yrs.-old, 3 sites), and MF (mature rainforest,  $>100$  yr.-old, 4 sites).

For both richness and evenness, we observed either random relations or significant reduction of legume richness (Figure 4). None of the five focal plants genera / species increased legume richness in their surroundings along secondary succession. In fact, the focal plants reduced the legume richness, but this effect depends on the successional stages. For example, *Cecropia* spp. and *Goupia glabra* are stronger repellers of legume species at earlier successional stages (YSF for *Cecropia* spp. and MSF for *Goupia glabra*) than at later successional stages. Notably, *Machaerium* spp. consistently repelled other legume species (ISAR index)

regardless of successional stage. However, this legume liana showed no significant impact on the evenness of other legume species. Focal genus/species-effects on legume species evenness were predominantly random, though we observed short-scale (<3 m) positive deviations pointing to an increase in evenness in some cases. Strong ISAR-reduction associated with focal legume genera *Inga* and *Machaerium* are partially compensated by the proper diversity of these two species-rich genera. In summary, richness reductions caused by focal non-legume pioneer trees were limited to young or mid-aged secondary regrowth.

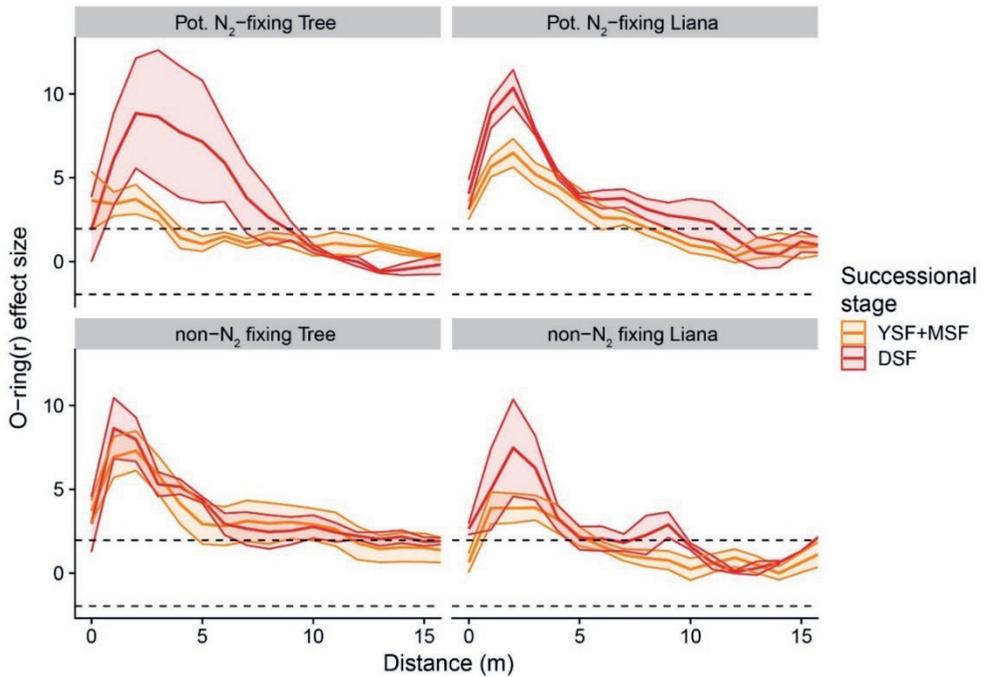


**Figure 4** – Departure from the null hypothesis (effect size) of focal abundant plant groups on spatial ISAR function of legume taxonomic diversity (left column) and on legume

taxonomic evenness (spatially explicit Shannon index -  $H_i(r)$ ) (right column) in their surroundings of: legume tree genus *Inga* (32 species) and legume liana genus *Machaerium* (13 species), as well as selected single abundant non-legume pioneer trees *Cecropia* spp. (three morphologically similar species), *Laetia procera* and *Goupia glabra* along succession in central Amazonia. Values below -1.96 indicate significant decrease in richness (left columns) and evenness (right columns) of legume diversity, and values above +1.96 indicate significant increase of legume richness / evenness. Means  $\pm$  CI<sub>95%</sub> of sites within each group of 3-5 sites, only sites with  $\geq 70$  individuals included. YSF (young secondary forest, 2-3 yrs.-old, 3 sites), MSF (mid-aged secondary forest, 5-10 yrs.-old, 5 sites), OSF (old secondary forest, 12-25 yrs.-old, 3 sites), and MF (mature rainforest, >100 yrs.-old, 4 sites).

### *Impact of land-use intensification on plant spatial organization and diversity*

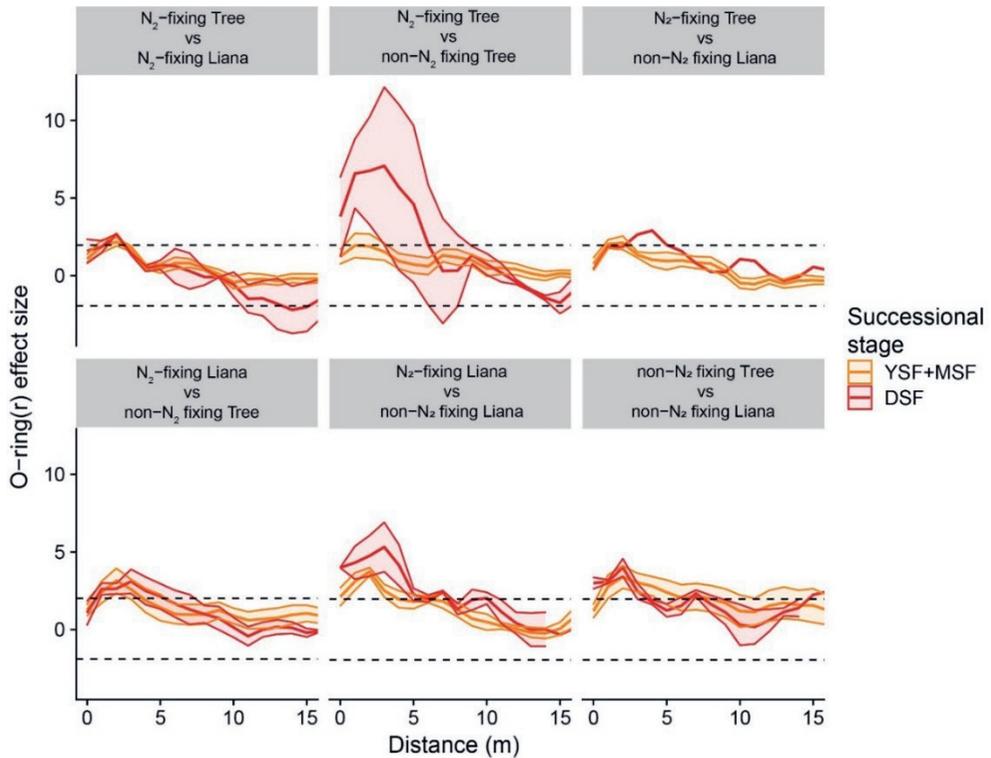
A second-cycle shifting cultivation led to profound changes in the spatial arrangement of  $N_2$ -fixing plants. Figure 5 compares monivariate point pattern distribution of the four functional plant groups in three 3-11 yrs.-old 'degraded' (second-cycle shifting cultivation) sites with young and mid-aged (2-10 yrs.-old) first-cycle secondary succession (eight sites). Although aggregation below 5 m distance remained significant for both vegetation types, effect-sizes were significantly stronger in the degraded sites for potentially  $N_2$ -fixing trees and lianas and non- $N_2$ -fixing lianas.



**Figure 5** – Departure from the null hypothesis (effect size) of univariate O-ring point pattern analysis for legume trees and lianas and for selected non-legume trees (3 genera) and lianas (3 genera) after a first-cycle (YSF+MSF) and second-cycle slash-and-burn agriculture in central Amazonia. Means  $\pm$  CI<sub>95%</sub> of sites within each group of 3-5 sites, horizontal lines  $\pm$  1.96 indicate simulated CSR, only sites with  $\geq 70$  individuals included. pot. N<sub>2</sub>-fixing. = potential N<sub>2</sub>-fixing; non-N<sub>2</sub>-fixing = non-N<sub>2</sub>-fixing plant, YSF (young secondary forest, 2-3 yrs.-old, 3 sites), MSF (mid-aged secondary forest, 5-10 yrs.-old, 5 sites), DSF (degraded secondary forest, 4-11 yrs.-old, 3 sites).

Figure 6 compares bivariate point patterns between the four functional plant groups of the three degraded (second-cycle) regrowth sites with those in young and mid-aged first-cycle secondary succession. Bivariate patterns showed no evidence of repulsion and ranged between random and co-occurrence. Co-occurrence scaled geographically up to 5 m distance. For different functional groups clustering was higher in the degraded sites than in first-cycle succession, notably between N<sub>2</sub>-fixing trees and non- N<sub>2</sub>-fixing trees, and N<sub>2</sub>-fixing lianas and non- N<sub>2</sub>-fixing lianas.

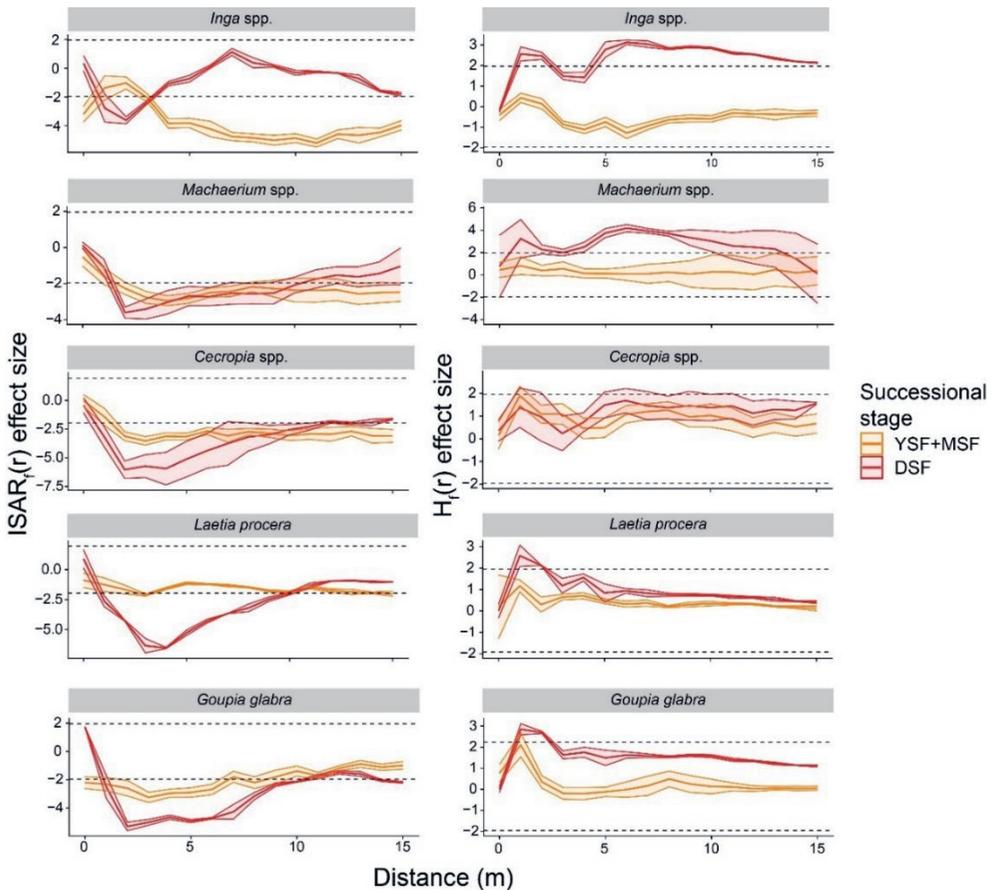
Therefore, co-occurrences tended to be stronger following land use intensification in Amazon Forest.



**Figure 6** – Departure from the null hypothesis (effect size) of bivariate O-ring point pattern analysis for legume trees and lianas and for selected non-legume trees (3 genera) and lianas (3 genera) after a first-cycle (YSF+MSF) and second-cycle (DSF) slash-and-burn agriculture in central Amazonia. Means  $\pm$  CI<sub>95%</sub> of sites within each group of 3-5 sites, horizontal lines  $\pm$  1.96 indicate simulated CSR, only sites with  $\geq 70$  individuals included. pot. N<sub>2</sub>-fixing. = potential N<sub>2</sub>-fixing; non-N<sub>2</sub>-fixing = non-N<sub>2</sub>-fixing plant, YSF (young secondary forest, 2-3 yrs.-old, 3 sites), MSF (mid-aged secondary forest, 5-10 yrs.-old, 5 sites), DSF (degraded secondary forest, 4-11 yrs.-old, 3 sites).

Land use intensification also reshaped the spatial arrangement of legume plant diversity. We found that, in general, focal species became stronger repellers of legume plants (ISAR index, Figure 7, left column) after a second-cycle of slash-and-

burn cultivation (DSF sites). However, *Inga* spp showed the opposite pattern and, at scales  $> 4$  m, played no significant influence in the richness of legume species. *Inga* spp and *Machaerium* spp. showed positive effect on legume evenness between 5-10m distance (Figure 7, right column). For the other focal three species/genera (*Cecropia* spp., *Laetia procera*, and *Goupia glabra*), the surrounding legume richness was systematically and significantly reduced following land use intensification (second cycle of slash-and-burn agriculture).



**Figure 7** – Departure from the null hypothesis (effect size) of focal abundant plant groups on spatial ISAR function of legume taxonomic diversity (left column) and on legume taxonomic evenness (spatially explicit Shannon index -  $H_1(r)$ ) (right column) in their surroundings of: legume tree genus *Inga* (32 species) and legume liana genus *Machaerium*

(13 species), as well as selected single abundant non-legume pioneer trees *Cecropia* spp. (three morphologically similar species), *Laetia procera* and *Goupia glabra* after a first-cycle (YSF+MSF) and second-cycle (DSF) slash-and-burn agriculture in central Amazonia. Values below -1.96 indicate significant decrease in richness (left columns) and evenness (right columns) of legume diversity, and values above +1.96 indicate significant increase of legume richness / evenness. Means  $\pm$  CI<sub>95%</sub> of sites within each group of 3-5 sites, only sites with  $\geq 70$  individuals included. YSF (young secondary forest, 2-3 yrs.-old, 3 sites), MSF (mid-aged secondary forest, 5-10 yrs.-old, 5 sites), DSF (degraded secondary forest, 4-11 yrs.-old, 3 sites).

## Discussion

We take spatial vegetation patterns to evaluate the impact of slash-and-burn cultivation by comparing secondary forests with mature rainforest controls. We interpreted the degree and velocity of re-approximation of secondary forests to control levels as an indicator of sustainable fallow time-spans and impact of land-use intensification.

After a first-cycle of slash-and-burn agriculture, most of the spatial interactions between plants ( $N_2$ -fixing vs non-  $N_2$ -fixing) are random or showed small-scale clustering. However, a second cycle of shifting cultivation is enough to induce a stronger clustering at both univariate and bivariate interactions. Univariate clustering is likely a result of vegetative resprouting (Bellingham & Sparrow, 2000). Additionally, our study provides evidence that functionally distinct plants ( $N_2$ -fixing vs non-  $N_2$ -fixing) tend to co-occur as a result of land use intensification. This result might indicate that after successive cycles of slash-and-burn cultivation, non-legume pioneer plants tend to associate more positively with legume plants, as they might become a source of nitrogen – a scarce nutrient after slash-and-burn agriculture (Davidson *et al.*, 2007). These findings contributed to support the stress gradient hypothesis (SGH) which suggests that under stress environmental conditions communities may shift from net negative to net positive interactions (He & Bertness,

2014; Passy, 2017). Future studies can advance our understanding on the stress gradient hypothesis in plant spatial interactions by verifying the impact of repeated cycles of slash-and-burn agriculture, which is left as an avenue for future research.

In our study, we found that focal species/genera reduced the legume richness without significant effect on their evenness. According to Ma (2005), at small scales (50 m<sup>2</sup>), species richness and evenness are not directly related and may result from different ecological process. Diversity is a legacy of past regional processes, but small-scale disturbances also play an essential role in shaping communities together with biotic and abiotic interactions (Comita *et al.*, 2009; Cornell & Lawton, 1992). Therefore, different ecological process might be in action to explain the spatial patterns found in our study. For example, the genus *Cecropia* is a pioneer tree that frequently dominates the seedbank (Uhl *et al.*, 1988) and the earlier stages of secondary succession (Villa *et al.*, 2019). Its less dense wood and hollow stems are decisive in its rapid growth, providing an advantage in the competition for sunlight (van Breugel *et al.*, 2012). Our results show the weakest effects of *Cecropia* spp. on legume diversity and evenness, likely the result of its uniform distribution in the overstory of all secondary forests. Therefore, *Cecropia* spp. contributed little to the recovery of legume plant diversity along succession. Additionally, we found that, at earlier successional stages, non-leguminous trees genera *Laetia procera* and *Goupia glabra* also become stronger repellers of legume plant diversity because of land use intensification. These findings suggest a stronger competitive exclusion in those degraded areas where the available nitrogen is a scarce resource. By reducing the richness of legume plants in their surroundings, these pioneer plants might slow forest recovery. On the other hand, the leguminous tree *Inga* spp. showed the opposite pattern and showed no significant impact on legume diversity after a second-cycle of slash-and-burn agriculture. *Inga* spp. are particularly interesting for the recovery of tropical forest due to their vast root system and potential for nodulation and the symbiotic fixation of atmospheric nitrogen (Sccoti *et al.*, 2019). This supports the view that leguminous plants play a crucial role in the initial stages

---

of secondary succession.

Our study considered the role of plant growth form in spatial organization by mapping trees and lianas. Liana spatial patterns challenge ecological studies due to their capability of spreading far away from their root base. Analysis of liana spatial patterns likely reveals their success in seed germination and initial plant survival rather than aboveground interactions of established plants, as shoot competition may occur far away from rooting points. Our results showed liana plant bases to be spatially more aggregated than of trees, though their bivariate effects of large lianas on their surroundings are likely to act aboveground far beyond their root bases.

Lianas have been frequently reported to suppress tree growth or fecundity, and to increase tree mortality (Schnitzer & Carson, 2010). Our sampling sites did not have many different species of lianas that sufficed the minimum abundance required for spatial analysis. However, our results for *Machaerium* spp. indicated a consistent reduction of legume plant richness regardless of successional stage. Lianas reduced richness presumably by a combination of aboveground competition, e.g. via competition for light and mechanical stress (Schnitzer & Carson, 2010) and belowground competition. Despite that, *Machaerium* spp. is also a nitrogen fixing legume plant, and, thus, can also help to recover the soil nitrogen content. This dual role could also explain why, in degraded forests, they promoted a more even diversity of leguminous plants. For the other non-N<sub>2</sub>-fixing plants, competition with other plants could limit tree establishment and slow forest recovery, and increased liana dominance in degraded sites could ultimately worsen the slash-and-burn degradation cycle.

The study of plant spatial pattern contributes to unravel how plants interact with each other and shape their surrounding diversity. Nowadays few studies adopted plant functional traits in plant spatial pattern investigation. Previous works explored the influence of fire on dead vs alive plants (Wiegand & Moloney, 2013; Yu *et al.*, 2009). Our results show how analysis of plant traits combined with plant diversity can provide insightful information not only about spatial impacts of species

assemblies along secondary forest succession, but they also depict the profound impacts of land use intensification on the spatial organization of functional groups and diversity in secondary regrowth in Amazonia.

## **Conclusion**

Although the effects of 2-25 years of succession on spatial organization of first-cycle secondary regrowth were small, an increase of land-use intensity (2<sup>nd</sup>-cycle slash-and-burn shifting cultivation) caused strong shifts of plant spatial organization within and between functional groups. Three findings are especially noteworthy: (i) strong increase in clustering, probably outcome of increased vegetative resprouting; (ii) partial increase in co-occurrences, suggesting competition at later stages of succession, but facilitation following increase of land use intensity; and (iii) strong reduction in richness of surrounding legume diversity. The latter effect is especially worrisome as it indicates species-reduction due to competitive exclusion in resource-limited degraded sites, ultimately, this will slow and/or compromise the process of natural forest recovery.

## **Acknowledgements**

This research was financed by the Research Fund of Maranhão State (FAPEMA). We gratefully acknowledge a fellowship by FAPEMA to the first author.

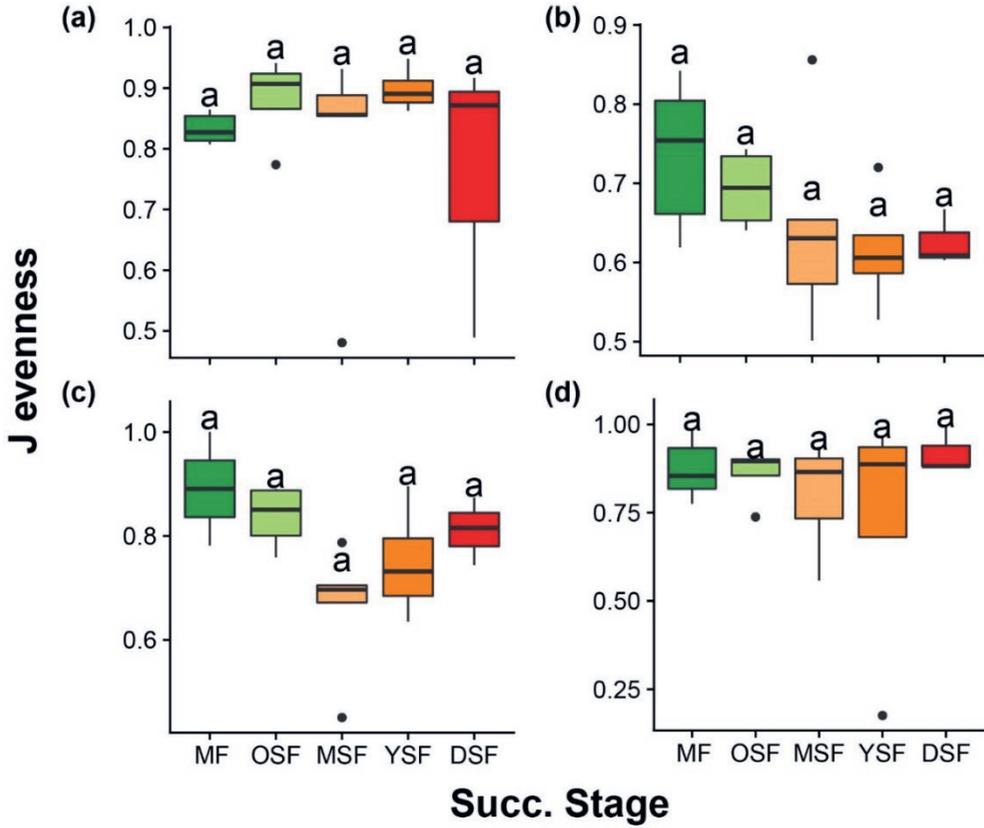
## Supplementary Results

Supplementary Table S1. Characterization of the 18 study sites

site name	age	CS <sup>(1)</sup>	size (ha)	mean vegetation total		median % contributions to TAGB						
				abund. (n ha <sup>-1</sup> )	TAGB (t ha <sup>-1</sup> )	lianas	palms	pioneer trees <sup>(2)</sup>	single-shoot plants	largest 5% plants	pot. N <sub>2</sub> -fixers	
young first-cycle secondary vegetation												
Joaquim	2	2	0.23	5867	61.7	10.2	2.0	11.7	35.9	62.8	9.6	
Gustavo (young)	3	3	0.60	8588	116.8	3.5	0.4	25.6	84.9	15.5	1.3	
raldo (young)	3	3	0.47	10600	141.5	4.8	0.3	12.6	81.5	52.4	14.1	
mid-aged first-cycle secondary vegetation												
Adventista (young)	5	4	0.33	10044	122.4	2.7	0	13.2	56.5	46.2	19.9	
Geraldo (old)	7	3	1.10	5505	147.7	2.8	0.4	8.3	73.5	35.4	9.1	
Ilson	8	4	0.89	5773	175.1	4.3	0	26.6	56.8	44.2	1.9	
Edson (young)	10	1	0.48	4011	164.6	4.2	0.1	11.7	45.0	38.9	1.4	
Guillermo	10	2	0.63	8450	178.7	1.4	0.4	58.6	94.3	17.3	1.9	
old first-cycle secondary vegetation												
Anna (old)	12	2	0.28	5580	220.1	1.0	0.3	37.4	66.2	42.0	5.7	
Jasiel	20	1	0.98	4624	202.4	1.2	0.1	15.1	48.6	44.0	0.5	
Edson (old)	25	1	0.63	5013	221.2	2.1	2.3	0	58.5	68.4	0.5	

Supplementary Table S1. Characterization of the 18 study sites (cont.)

site name	age	CS <sup>(1)</sup>	size (ha)	mean vegetation total		median % contributions to TAGB					pot. N <sub>2</sub> -fixers	
				abund. (t ha <sup>-1</sup> )	TAGB (t ha <sup>-1</sup> )	lianas	palms	pioneer trees <sup>(2)</sup>	single-shoot plants	largest 5% plants		
mature rainforest												
Anna	MF	2	0.70	5422	353.1	1.3	0.4	0	99.1	84.2	2.0	
Francisco	MF	2	0.63	5556	443.0	0.7	0	0	97.5	90.1	8.2	
Edson	MF	1	0.66	4689	418.1	3.5	0.3	0	99.0	83.7	0.2	
Geraldo	MF	3	1.16	5978	540.6	1.0	4.6	0	99.2	90.0	1.2	
'degraded' secondary vegetation												
Anna (2 <sup>nd</sup> cycle)	4	2	0.33	12433	81.0	10.2	0.2	14.7	29.5	46.2	5.2	
Guustavo (2 <sup>nd</sup> cycle)	5	3	0.45	6514	116.9	1.3	0	37.2	46.7	90.0	0.8	
Bispo (2 <sup>nd</sup> cycle)	11	2	0.70	7567	172.9	1.7	0.4	55.1	19.3	38.9	1.9	



**Figure S1.** J evenness index in different tropical forest successional stages in different functional groups (nodule potential and growth habit): (a) N<sub>2</sub>-fixing legume trees, (b) N<sub>2</sub>-fixing legume lianas, (c) non- N<sub>2</sub>-fixing trees, (d) non- N<sub>2</sub>-fixing lianas, 25/75-percentiles, non-outlier range and extremes.



## CHAPTER 3

# The effect of land-use on aboveground biomass and soil quality indicators in spontaneous forests and agroforests of eastern Amazonia

### **Authors**

Márcio Fernandes Alves Leite, Ronildson Lima Luz, Karol Henry Mavisoy  
Muchavisoy, Marcelo Luís Corrêa Zelarayán, Ernesto Gomez Cardoso,  
Flávio Henrique Reis Moraes, Guillaume Xavier Rousseau, Christoph  
Gehring

Published in: *Agroforestry Systems*

## Abstract

This study unites physicochemical indicators of aboveground vegetation, litter-layer and topsoil (0-20 cm) in contrasting vegetation types commonly found in the eastern Amazonia. We compare three agroforestry systems (enriched fallows, homegardens, and commercial plantations) with three spontaneous forest types (young and old secondary forests and mature rainforests) via one-way ANOVA, linear and non-linear regressions and multivariate analyses. Agroforests had significantly lower understory biomass when compared with young secondary forest. Commercial plantation agroforests had higher topsoil pH and Ca-contents and homegardens had higher K-contents and P-availability hotspots, as revealed by the higher variance in this treatment. Agroforests and spontaneous forests were similar in their litter biomass (both leaves and twigs) and C:N ratio, and in soil organic matter and P contents. The overstory negatively impacted the understory ( $r^2=0.20$ ,  $p<0.05$ ) while only the understory related significantly with the litter-layer ( $r^2=0.11$ ,  $p<0.07$ ), pointing to a major discontinuity between vegetation and topsoil. Principal component analysis depicted a successional sequence of systems, with homegardens closest to mature rainforests. According to co-inertia analysis, plant biomass was strongly related with topsoil in spontaneous forests than in agroforests. Altogether, agroforests were similar with mature rainforests in a wide range of variables of the vegetation–litter–topsoil continuum, and co-inertia analysis indicates that agroforestry management can alter this continuum. Our study has revealed that homegardens in the study region had higher aboveground biomass and nutrient-availability that may be an outcome of the traditional sweep-and-burn practice of Amazon communities.

## Introduction

Anthropogenic pressure on the land-base continues to be high throughout the humid tropics and notably in eastern and southern Amazonia, an area also known as the ‘arc of deforestation’. After the first deforestation cycle of pristine rainforests (de las Heras *et al.*, 2011) spontaneous regrowth of secondary forest develops on abandoned pastures and shifting cultivation fields. Shifting cultivation is the predominant land-use system adopted by the majority of low-income rural populations in the region.

However, this land-use system contributes to environmental degradation, slower secondary forest regeneration, lower biodiversity, increased dominance of aggressive ruderal species, impoverishment of soil fertility, altogether leading to rural poverty (Satyam Verma, 2012), bringing a socioecological crisis in the study region. The soils of the humid tropics are particularly vulnerable to degradation because of fast organic matter decomposition and mineralization as well as nutrient losses caused by leaching (Markewitz *et al.*, 2004). Degradation caused by repeated slash-and-burn practice and reduced fallow-periods lowers agricultural productivity and ultimately leads to rural poverty (Varma, 2003). Maintenance or restoration of the soil ‘productive potential’ is a central pillar of sustainable land use (Boddey *et al.*, 2003). Agroforestry has been proposed as viable alternative for shifting cultivation in the tropics, as it can improve soil quality (Pinho *et al.*, 2012), increase system stability (Mohri *et al.*, 2013) and reduce economic risks because of better market flexibility via multiple products (Souza *et al.*, 2012). Trees are a key factor for maintaining soil fertility in agroforestry due to nutrient pumping and safety-net mechanisms (Seneviratne *et al.*, 2006).

Depletion of soil nutrients as the outcome of non-sustainable land-use intensification can cause reductions in above- and belowground biodiversity, ecosystem functioning and stability (Thiele-Bruhn *et al.*, 2012). The diversity of agroforestry systems is astonishing (Atangana *et al.*, 2014c); therefore, the ecological and social

sustainability of these systems is likewise expected to vary widely. Here, we assess the impact of different land-use forms, which are three agroforestry systems: commercial plantations, enriched fallows and homegardens, and mature rainforests, and secondary forest regrowth on the physicochemistry of aboveground vegetation, litter-layer and topsoil in eastern Amazonia (north-central Maranhão and eastern Pará states) and we explore system-related differences in along the vegetation-litter-topsoil continuum.

## **Materials and methods**

### *Study region and site clusters*

Research was conducted on 32 study sites in the eastern periphery of Amazonia across five counties aggregated into two regional clusters (Table 1). Pastures and secondary forests have almost entirely replaced the original rainforests in central Maranhão and eastern Pará states. Maximum distance between sites within each county was less than 30 km while the maximum distance between counties in each regional cluster was less than 150 km. Seventeen of the 32 sites were located in central-northern Maranhão state and the others were located approximately 400 km further west in Tomé-açu County in eastern Pará state. Climate is classified according to Köppen as *Aw* and *Ami* and varies slightly between the two regional clusters, with 2100 mm and 2300 mm annual rainfall, and with 6- and 5-months hydric deficit in central Maranhão and eastern Pará states, respectively. Soils are nutrient-poor acid Oxisols and Ultisols (USDA, 2010). The topsoil texture was classified as ‘loamy/fine-sand’ and varied little between sites and clusters (sand average =  $72.88 \pm 17.11\%$ , silt average =  $12.97 \pm 18.63\%$  and clay average =  $14.15 \pm 7.21\%$ ). There were no differences in soil texture between systems, counties and regional clusters.

**Table 1.** Number (in brackets) and distribution of the 32 study sites within the two regional clusters and five counties. Vegetation characteristics and site use histories of the study sites in Eastern Amazonia

Forest system	System type	Region	County	Location*	
<b>Spontaneous forest (9)</b>	Young secondary forest (4)	Maranhão	São Luís (2)	2°59'12.0" S - 44°20'52.2" W	
				2°59'11.2" S - 44°20'52.6" W	
		Pará	Tomé-açu (2)	2°24'06.6" S - 48°11'48.4" W	
				2°24'29.0" S - 48°11'52.3" W	
	Old secondary forest (3)	Maranhão	São Luís (2)	2°58'44.5" S - 44°21'14.5" W	
				2°58'44.2" S - 44°20'52.6" W	
		Pará	Tomé-açu (1)	2°24'45.7" S - 48°11'51.2" W	
	Mature forest (2)	Pará	Tomé-açu (2)	2°23'39.1" S - 48°09'51.6" W	
	<b>Agroforest (23)</b>	Enriched fallow (6)	Maranhão	Anajatuba (2)	3°26'58.2" S - 44°23'13.3" W
Morros (3)				3°11'31.3" S - 44°35'29.3" W	
				2°53'10.0" S - 44°09'23.5" W	
Homegarden (11)		Maranhão	Pará	Tomé-açu (1)	2°59'48.9" S - 43°55'05.0" W
			Anajatuba (3)	2°32'38.4" S - 48°16'26.7" W	
				Arari (5)	3°14'59.0" S - 44°36'46.8" W
				3°15'00.6" S - 44°36'43.5" W	
				3°33'40.8" S - 44°49'57.3" W	

			3°33'08.6"S - 44°49'10.5"W
	Pará	Tomé-açu (3)	2°11'31.9"S - 48°08'41.9"W 2°11'55.9"S - 48°09'17.2"W
Commercial Plantation (6)	Pará	Tomé-açu (6)	2°12'32.2"S - 48°17'44.2"W 2°33'11.0 "S - 48°07'39.2"W

\*For counties with more than two sampling sites per system we give regional coordinates.

### *Spontaneous forest and agroforest system classification*

We compared three types of spontaneous forest with three types of agroforests, which were distributed into two regional clusters and five counties (Table 1). Site classification was based on previous work by E. Cardozo *et al.* (2015) as follows:

#### (i) Spontaneous forest types

**Secondary Forests:** Spontaneous secondary forest regrowth following slash-and-burn shifting cultivation or on abandoned pastures. We divided these into ‘young’ ( $\leq 12$ -year-old; **SFY**) and ‘old’ (20- to 26-year-old; **SFO**). The most common species were the babassu palm (*Attalea speciosa* Mart.) as well as *Cecropia* sp., *Inga* sp., *Astrocaryum vulgare* Mart., *Lecythis* sp., and *Carica microcarpa* Jacq.

**Mature rainforest (MF):** Original mature forest without visible human perturbation (one site) or with low-intensity selective logging  $>60$  yrs ago (one site). The most common species were *Holopyxidium latifolium* (Ducke) R. Knuth, *Manilkara* sp., *Copaifera langsdorffii* Desf., *Carapa guianensis* Aubl., and *Protium* sp.

#### (ii) Agroforest types

**Enriched Fallow Agroforest (EFA):** Established by enrichment planting of fruit and timber species in the understory of 15- to 25-year-old secondary forest. The most common crops were cupuassu (*T. grandiflorum* K. Schum), açaí (*Euterpe*

*oleracea* Mart.), and cacao (*Theobroma cacao* L.), spaced at distances of 3-4 m.

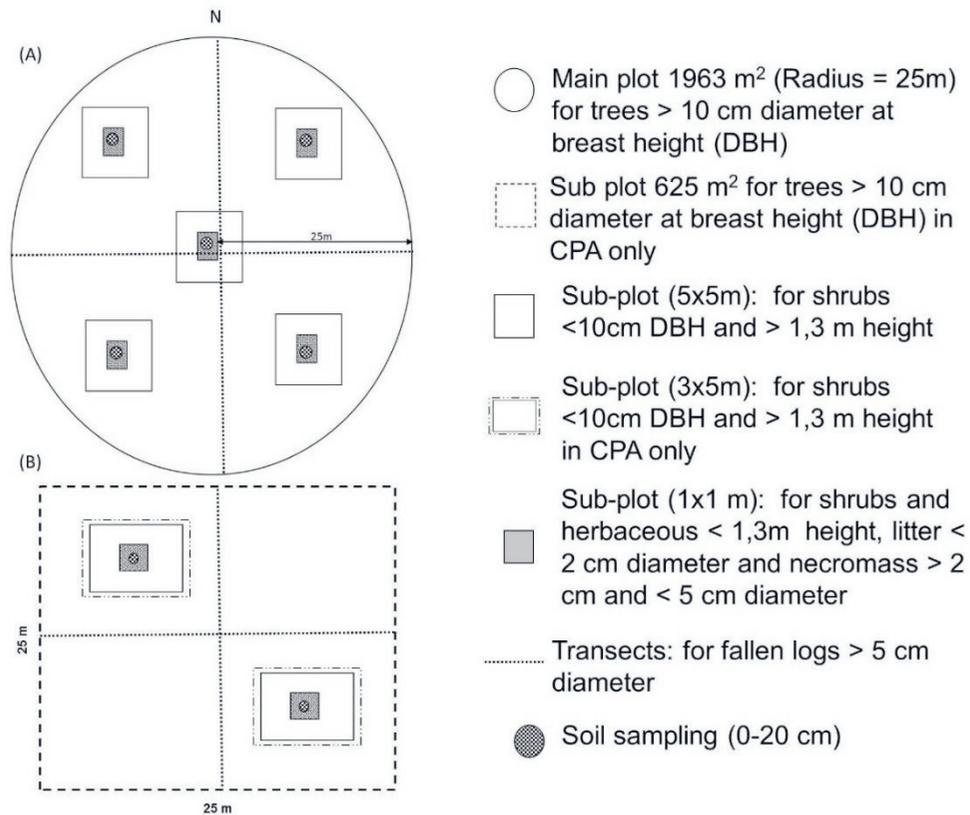
**Homegarden Agroforest (HA):** Tall multistrata agroforests surrounding houses, virtually omnipresent in our study region and elsewhere. This system has a particular fire regime called ‘sweep-and-burn’ (Winklerprins, 2009), the regular sweeping and subsequent burning of litter piles in order to keep the surroundings of the houses ‘clean’ and to avoid mosquitoes, spiders, snakes etc. Dominant overstory plants are mango (*Mangifera indica* L.), jackfruit (*Artocarpus integrifolia* L.) and açai, then banana (*Musa* spp.), cupuassu and cacao in the understory,

**Commercial Plantation Agroforest (CPA):** Regularly spaced plantations with inorganic fertilization and liming. This system was inspired and developed by Japanese immigrants. Main products are fruits of cacao, cupuassu, açai and black pepper (*Piper nigrum* L.).

Not all systems occurred in both regional clusters. Implications of the resulting unbalanced scheme are further discussed in sections 2.4 and 3.1.

### *Within-site sampling scheme*

We adopted a joint sampling scheme to guarantee compatibility between all of our datasets. Vegetation and litter sampling methods strive to capture ‘plant influence zones’, *sensu* Rhoades (1996). We estimated large tree aboveground biomass in a main circular plot and minor vegetation and litter in five subplots (Coomes *et al.*, 2002) (see Figure 1a). Topsoil (0-20 cm) was collected as composite samples from the centers of the five subplots.



**Figure 1** Sampling units with differently-sized and replicated sampling areas for vegetation, litter and topsoil for (a) the spontaneous forests, enriched fallow and homegardens and for (b) CPA only.

In the CPA sites, we adapted our sampling-scheme to the predominating regular spacing. Instead of a circle, we used six sub-plots (3 x 5 m), in three quadrangular main plots of 25x25 m (Fig 1b). This sampling scheme has been shown to generate reliable estimates of system-totals and to be superior to circular sampling geometry in regularly-spaced plantations (Kato *et al.*, 2009). We subsequently corrected the total area for the different sampling-sizes (4.5% smaller total sampling area) in this alternate scheme. During site selection we discarded large border-zones to neighboring vegetation.

### *Variables under investigation*

Large biomass components were estimated allometrically by Muchaviso (2013) by utilizing diameter-based equations as well as conversions between diameter at breast height (dbh) and diameter measured at 30 cm height for smaller vegetation components (Gehring, Park, *et al.*, 2008). These and further species-specific equations utilized in this study are detailed in Annex 1.

Small vegetation (<1.30 m height) was estimated destructively jointly with the litter-layer in the 1x1 m subplots, by complete harvesting and subsequent *in-situ* separation into biomass fractions wood and leaf of trees, lianas and palms, herbaceous, gramineous, as well as leaf and twig litter. We estimated dry-matter content in representative subsamples of each of these fractions by comparing the fresh weights (*in-situ*) and the dry weights, which were taken after 2 weeks at 65°C.

We distinguish biomass of the following components: large vegetation (trees with dbh  $\geq 10$  cm and palms  $> 2$  m height) was grouped as AboveGround Biomass of plants larger than 10 cm dbh (AGB $\geq 10$  cm dbh'), mid-sized vegetation (trees, shrubs and lianas with dbh  $< 10$ cm and palms  $< 2$  m height), and small vegetation (destructively sampled herbaceous and shrubs smaller than 1.30 m height). For statistical analyses we combined mid-sized and small vegetation as the AboveGround Biomass smaller than 10 cm dbh ('AGB $< 10$  cm dbh'). We estimated the necromass of fallen logs in four transects (Van Wagner, 1968) and standing dead logs in the circular main plots, following the methods of Arevalo *et al.* (2002). We destructively quantified and sampled the litter-layer during the rainy season, in five 1x1 m sub-quadrants within each sampling-unit, distinguishing between leaf and twig litter biomass. Total aboveground biomass (TAGB) was considered as the sum of all above-mentioned components.

To characterize topsoil physical quality, we measured (i) soil bulk density (volumetric rings), and (ii) soil texture (pipette method), utilizing procedures described in Klute *et al.* (1986). To characterize topsoil chemistry, we measured the

pH (0.01 M CaCl<sub>2</sub> suspension), soil organic matter content (Walkley-Black method), available P (extraction with synthetic anion exchange resin Amberlite IRA-400) and exchangeable K (Mehlich I), Ca, Mg (KCl extraction) and H+Al (SMP Method), following IAC (2001) routines. C- and N-concentrations in leaf-litter samples were determined with Walkley-Black (Walkley *et al.*, 1934) and Kjeldåhl digestion (Tedesco *et al.*, 1995), respectively.

### *Statistical analyses*

#### *Basic and univariate statistics*

Kolmogorov-Smirnov and Lilliefors's tests were used for checking the normality of the data. Most variables followed a normal distribution or could be normalized via  $\log_{10}$  or  $\log_{10}(x+1)$  transformation, with the exception of the leaf litter C:N-ratio, for which we utilized non-parametric procedures (Spearman correlations, Kruskal-Wallis ANOVA with rank-based Dunn's test). Homogeneity of variance was checked by using Brown-Forsythe test for unequal replication numbers. Outliers and extremes ( $>1.5*SE$ ) were excluded: one value for organic matter, two for twig biomass and one for total aboveground biomass (in a CPA with Brazil-nut trees). For between-group comparisons and regressions, significance level as 5% and tendencies with  $<10\%$  significance-level were established as default.

Types of single systems were directly compared by two or in groups of vegetation-types via t-tests, and jointly via one-way ANOVAs and subsequent *post-hoc* Spjøttvoll-Stoline test (HSD Tukey for unequal replication numbers). Relationships between pairs of variables via linear and non-linear regressions were investigated. These analyses were conducted with Statistica 8.0 (StatSoft, 2007), graphs were generated with SigmaPlot 11.0 (Systat, 2008) and in R.

### Multivariate statistics

Principal Component Analysis (PCA), Between-Class Analysis (BCA) (Chessel *et al.*, 2004), Co-Inertia Analysis (CIA) (Dray *et al.*, 2003) and permutation Monte Carlo significance-tests were used in the analyses. PCA was adopted to investigate variation of the entire dataset and BCA to isolate the system grouping contributions to data variability.

CIA investigates relationships between two or more datasets (Dray *et al.*, 2003), which in our case were the biomass and soil datasets. We utilized a Between-class CIA, which is a procedure that maximizes the covariance between groups, rather than between individual cases. We measured overall similarities using a multivariate extension of the Pearson correlation, the Rv coefficient. R environment (R Development Core Team, 2007) and “ade4” library were used for the multivariate analyses (Chessel *et al.*, 2004).

### Analysis of regional variance

Our experimental design was unbalanced, due to the non-occurrence of some systems in some clusters (chap. 2.2). We addressed the unbalanced design by searching for between-region (states) differences for the two systems which occurred in sufficient replications in both regions, via t-tests between the regions central Maranhão and western Pará (Tomé Açu approx. 400 km westwards) via between-class analysis (BCA).

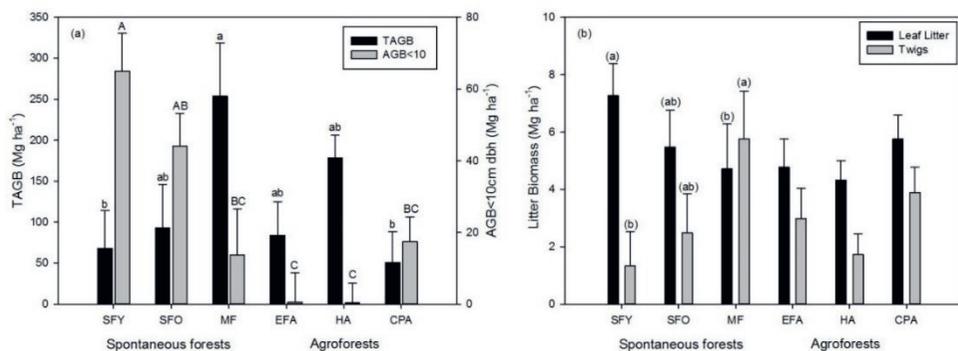
A multivariate between-class analysis (BCA) was used to investigate the potential effect of regional differences on secondary forests and on homegarden agroforests (which occurred in both regions, see Table 1). Furthermore, the regional effect on the sum of variation of the entire dataset was quantified by using Monte Carlo randomization procedure with 9,999 permutations for significance testing.

## Results

### *Vegetation, litter and topsoil variables across land use systems and between regions*

Total aboveground biomass (TAGB) differed significantly between systems (Figure 2a, left scale). Within the spontaneous forests, TAGB increased from 68.08 Mg ha<sup>-1</sup> in SFY to 92.98 Mg ha<sup>-1</sup> in SFO, and it was 2.73-fold higher in MF (253.55 Mg ha<sup>-1</sup>). TAGB likewise differed strongly among agroforests. TAGB of HA (178.52 Mg ha<sup>-1</sup>) was 3.5-fold higher than in CPA (50.95 Mg ha<sup>-1</sup>).

The agroforest and spontaneous forest systems differed significantly in their shrubs and herbaceous biomass (AGB<10 cm dbh, right scale in Figure 2a). SFY had an approximately 100-fold higher shrub biomass than HA (64.97 Mg ha<sup>-1</sup> and 0.42 Mg ha<sup>-1</sup>, respectively), and MF a nearly five-fold higher shrub biomass than CPA. There was no significant difference in necromass of standing and fallen logs greater than 5 cm, nor in litter mass (twigs or leaf litter, data not shown). Within the spontaneous forests (i.e., excluding agroforests), there was a tendency of decreasing leaf litter and increasing twig litter along the succession gradient ( $p < 0.10$ ) (Figure 2b). A t-test comparison between mature forest and homegardens revealed a significant ( $p < 0.03$ ) difference in total litter biomass (Table 2, as indicated by the capital letters).

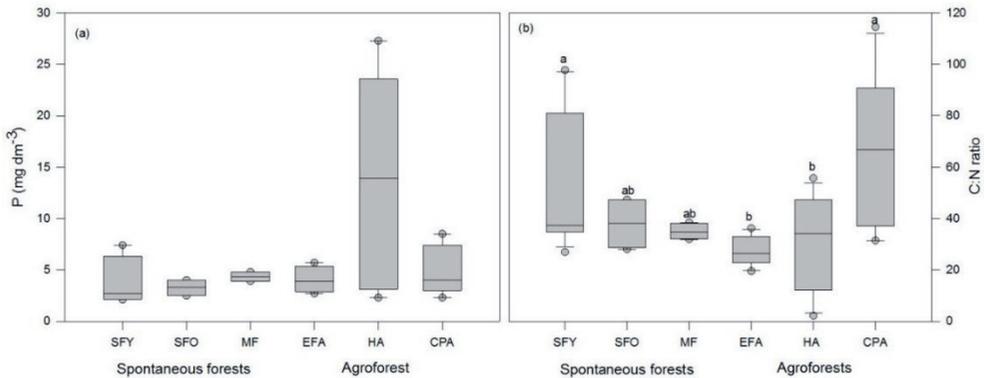


**Figure 2** – (a) Total aboveground biomass (TAGB) and biomass of small (<10 cm dbh) trees, lianas, shrubs and herbaceous plants and (b) twigs and leaf litter in the understory of spontaneous forests and agroforests. Note differences for the black (left scale) and grey (right scale) columns. SFY=Young Secondary Forest; SFO = Old Secondary Forest; MF= Mature Forest; EFA = Enriched Fallow Agroforest; HA = Homegarden Agroforest; CPA = Commercial Plantation Agroforest. Means+SE followed by the same letter do not differ between another, as indicated by Spjøtfol-Stoline test at the 5% probability-level. Letters in brackets correspond to Spjøtfol-Stoline test for spontaneous forests only (analysis excluding agroforests), at a 10% probability-level.

T-tests between the two regional clusters (central Maranhão state and Tomé-açu in eastern Pará state) did not detect any significant differences in any variable. Furthermore, the Between-class analysis (BCA) of clusters on the land use systems partitioned by region indicates that the difference between regions explains only 7.72% of the total variance for the data set of secondary forests fallow (SFY and SFO) and homegardens agroforests (HA), because both of which occurred in sufficient replications in both regions. According to Monte Carlo test of 9,999 random simulations between-region variation was neither relevant nor significant ( $P=0.114$ ).

We found no overall between-system differences in topsoil organic matter (OM) concentration. Excluding agroforests, i.e., concentrating on spontaneous succession, there was a significant ( $p<0.05$ ) increase of OM-concentration from

young to old secondary forests and highest values in MF (Table 2, letters in brackets). Bulk soil density was significantly higher in SFY than in SFO, MF and HA ( $p < 0.05$ ). Topsoil pH was highest in CPA and lowest in MF and HA ( $p < 0.05$ ) (Table 2). Available P-concentration was non-normally distributed and, due to high data variability, did not differ between systems, according to Dunn's non-parametric test (Figure 3a).



**Figure 3** – (a) Topsoil (0-20 cm) resin-extractable P-concentrations and (b) Carbon:Nitrogen ratios in leaf litter in spontaneous forests and agroforests. SFY=Young Secondary Forest; SFO = Old Secondary Forest; MF= Mature Forest; EFA = Enriched Fallow Agroforest; HA = Homegarden Agroforest; CPA = Commercial Plantation Agroforest. Lines and boxes represent medians and 25-75 percentiles of P concentrations, whiskers are the 90 and 10 percentiles. Systems with the same letter do not differ between another, as indicated by Dunn's nonparametric test at the 5% probability-level.

Topsoil K concentration was three times higher in HA than in the spontaneous forests (SFY, SFO and MF) ( $p < 0.05$ ), and topsoil Ca concentration was higher in CPA and SFO than in HA and MF ( $p < 0.05$ ). Soil potential acidity ( $\text{H}+\text{Al}$ ) was higher in MF and HA than in CPA ( $p < 0.05$ ) (Table 2).

**Table 2.** Topsoil (0-20 cm) organic matter (OM) concentration, bulk soil-density, pH-value, concentrations of K and Ca, and of littermass and leaf litter N-concentration in spontaneous forests and in agroforests.

	Spontaneous forests			Agroforestry systems		
	SFY	SFO	MF	EFA	HA	CPA
<b>Organic Matter (OM) (g dm<sup>-3</sup>)</b>	22.63 ±	31.03 ±	32.30 ±	30.64 ±	28.53 ±	26.00 ±
	2.95 (b)	3.41(ab)	4.17(a)	2.64	1.78	2.23
<b>Bulk density (BD) (g cm<sup>-3</sup>)</b>	1.37 ±	1.21 ±	1.17 ±	1.27 ±	1.22 ±	1.32 ±
	0.03a	0.04b	0.05b	0.03ab	0.02b	0.03ab
<b>pH</b>	4.28 ±	4.34 ±	3.77 ±	4.19 ±	4.03 ±	4.68 ±
	0.14ab	0.17ab	0.20b	0.13ab	0.09b	0.11a
<b>K (mg dm<sup>-3</sup>)</b>	9.75 ±	14.02 ±	12.82 ±	26.05 ±	37.95 ±	25.64 ±
	5.49 b	6.34 b	7.76 b	4.91 ab	3.31 a	4.15 ab
<b>Ca (mg dm<sup>-3</sup>)</b>	115.00 ±	246.67 ±	10.00 ±	212.00 ±	112.73 ±	251.43 ±
	43.47 ab	50.20 a	61.48 b	38.88 ab	26.22 b	32.86 a
<b>H+Al (mmolc dm<sup>-3</sup>)</b>	40.25 ±	50.33 ±	85.00 ±	41.80 ±	63.09 ±	34.86 ±
	8.82 ab	10.19 ab	12.48 a	7.89 ab	5.32 a	6.67 b
<b>Total Litter (Mg ha<sup>-1</sup>)</b>	8.61 ±	7.97 ±	10.47 ±	7.76 ±	6.07 ±	9.63 ±
	1.54	1.78	2.18 (A)	1.38	0.93(B)	1.17
<b>Leaf Litter N (g kg<sup>-1</sup>)</b>	9.75 ±	12.53 ±	13.32 ±	9.58 ±	10.64 ±	10.74 ±
	1.04 c	1.48 ab	1.48 a	1.04 c	0.64 bc	0.85 bc

SFY=Young Secondary Forest; SFO = Old Secondary Forest; MF= Mature Forest; EFA = Enriched Fallow Agroforest; HA = Homegarden Agroforest; CPA = Commercial Plantation Agroforest. Means+SE followed by the same letter do not differ between another, as indicated by Spjøtfoll-Stoline test at the 5% probability-level. Letters in brackets correspond to Spjøtfoll-Stoline test for spontaneous forests only (analysis excluding agroforests), at a 10% probability-level. Capital Letters in brackets correspond to the t-test applied for Mature

Forest and Homegardens only at a 3% level of probability.

Neither leaf nor twig litter-layer C-content differed significantly between systems (data not shown). By contrast, leaf-litter N-content was highest in MF (average 20% above CPA and HA, and 30% above SFY and EFA) (Table 2). Leaf-litter C:N was significantly higher in both SFY and CPA compared to EFA (Figure 3b)

### *Relationships between variables*

Here, we explore the bivariate relationships between aboveground vegetation–litter–topsoil over the 32 study sites.

#### *Within-vegetation and vegetation-litter relationships*

We found a negative logarithmic relationship in spontaneous forests between  $AGB \geq 10$  cm dbh and  $AGB < 10$  cm dbh over the 32 study sites ( $r^2=0.19$ ,  $p<0.04$ ). Our results also established a positive logarithmic relationship between  $AGB < 10$  cm dbh and total litter biomass ( $r^2=0.11$ ,  $p<0.07$ ). However, when distinguished between leaves and twigs, this relationship was significant only for leaf litter ( $r^2=0.14$ ,  $p<0.05$ ). There was no apparent relationship between  $AGB \geq 10$  cm dbh and leaf, twig or total litter mass, though there was a positive relationship (Spearman rank  $R=+0.61$ ,  $p<0.05$ ) between total litter mass and its C:N-ratio.

#### *Litter-topsoil relationships*

Neither total nor leaf-litter biomass were significantly related with topsoil organic matter (all  $p > 0.57$ ). In contrast, there was a significant positive relationship between twig biomass and topsoil organic matter content ( $r^2=0.13$ ,  $p<0.05$ ). Plant-available P- and K-concentration were negatively related with total litter biomass ( $r^2=0.22$ ,  $p<0.01$  and  $r^2=0.20$ ,  $p<0.02$ ).

### Within-topsoil relationships

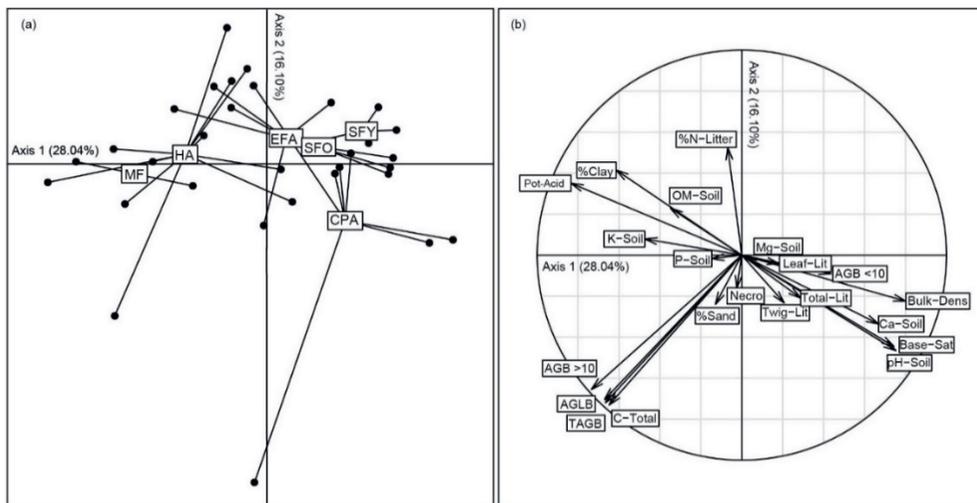
Soil organic matter (OM) content was positively related with clay content ( $r^2=0.20$ ,  $p<0.01$ ) and negatively related with soil bulk density ( $r^2=0.20$ ,  $p<0.01$ ). Clay-content related negatively with soil bulk density ( $r^2 = 0.18$ ,  $p<0.03$ ), whereas the other granulometric fractions did not significantly affect soil density.

We detected no significant relationships between OM and pH or nutrients (data not shown). Within the soil chemical indicators, we found a positive relationship between P and K ( $r^2 = 0.32$ ,  $p<0.01$ ), as well as between pH and Ca ( $r^2=0.67$ ,  $p<0.01$ ).

### Multivariate synthesis of variables and systems

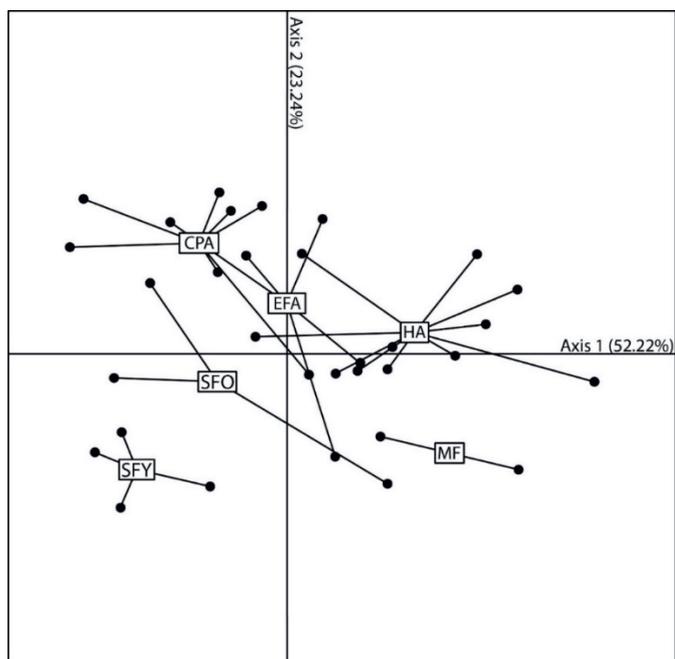
We explored the degree of covariation via PCA (Figure 4). The two main ‘factors’ (axis 1 and 2) together explained 44.14% of total variability of the combined vegetation, litter and topsoil physicochemical data. Figure 4a shows a successional trajectory of spontaneous forests from SFY via SFO and EFA to MF, aligned along axis 1 as a result of the influence of the aboveground biomass (i.e., TAGB and AGB <10 cm dbh) and of soil physicochemical properties (bulk soil density, pH, K- and Ca-content). Remarkably, HA were closest to MF. CPA formed a distinct group clearly separated from the other systems, with one extreme value presumably caused by the high biomass of Brazil-nut trees (*Bertholia excelsa* Humb. & Bonpl.) on this site only. Figure 4b delimits the PCA correlation circle over all variables of this study. Axis 1 captured most of the variability related to the soil physicochemistry (base saturation, soil potential acidity, pH, K concentration, clay, bulk soil density) and the litter layer and small vegetation components, whereas axis 2 represented the variability driven by OM, litter-layer N content and total and large vegetation biomass (these latter two also strongly contribute to axis 1). In summary, our results of PCA (i) partitioned two main axes (‘soil physicochemistry’ vs. ‘large

vegetation'), (ii) indicated successional pathways, and (iii) positioned HA close to MF and CPA as a distinct group distant from all others.



**Figure 4** – (a) Principal Component Analysis (PCA) over the 32 sites of spontaneous forests and agroforests in eastern Amazonia, and (b) PCA correlation circle of soil physical and chemical indicators and vegetation components. SFY=Young Secondary Forest; SFO = Old Secondary Forest; MF= Mature Forest; EFA = Enriched Fallow Agroforest; HA = Homegarden Agroforest; CPA = Commercial Plantation Agroforest.

We investigated differences between land-use systems using BCA (Figure 5). The between-group inertia corresponded to 30.3% of the total PCA variation ( $P=0.0001$ ). The first and second axes together represented 75.5% of the total data variation. The successional trajectory from SFY to MF was less apparent than in the PCA, which may be expected since the BCA method maximizes the inertia between-groups rather than between-sites. Further similarities were found with the PCA result - HA systems were most similar to MF; CPA again formed a distinct group. EFA were similar to SFO, HA and CPA; and SFY formed a separate group. Homegardens had the most widespread spread of points in the two axes, reflecting a high variability, whereas the market-oriented CPA exhibited the lowest data-spread.



**Figure 5** – Between-class analysis (BCA) of the 32 sites of spontaneous forests and agroforestry systems in eastern Amazonia. SFY=Young Secondary Forest; SFO = Old Secondary Forest; MF= Mature Forest; EFA = Enriched Fallow Agroforest; HA = Homegarden Agroforest; CPA = Commercial Plantation Agroforest. The inertia between classes was of 30.29%, the Monte Carlo permutation level of significance was  $P=0.001$ .

Between-class co-inertia analysis measured the overall similarity between aboveground biomass and litter biomass with soil physicochemistry datasets. Over all systems, the similarity between aboveground and litter biomass with soil physicochemistry was significant but weak ( $R_v=21.76\%$ ,  $P<0.05$ ). When considering only spontaneous forests (SFY, SFO and MF), the multiple correlation coefficient was three times higher ( $R_v=75.50\%$ ,  $P<0.01$ ), whereas the same correlation coefficient in agroforest systems (EFA, HA and CPA) was only 25.73% and also less significant ( $P<0.09$ ) (Table 3). The spontaneous forests retained more information of variability in the first two axes (95.14%) than the agroforests.

**Table 3.** Coefficient of matrix correlation ( $R_v$ ) between datasets of soil and biomass for spontaneous forests and agroforests.

Systems	Co-inertia ratio ( $R_v$ )	Cumulative project inertia (%) <sup>a</sup>	$P^b$
All systems	0.2176*	81.96	0.0429
Spontaneous forests	0.7550**	95.14	0.0026
Agroforests	0.2573	87.72	0.0817

<sup>a</sup>Total inertia percentage of the two principal axis of CIA; <sup>b</sup>Significance-level based on 9,999 Monte Carlo permutations; \* Co-inertia ratio was significant at  $P < 0.05$ ; \*\* Co-inertia was significant at the level of  $P < 0.01$ .

## Discussion

### *Regional variance and plot-size effects*

We investigated the possible problem of regional variability unduly affecting our results in secondary forests and in homegardens, both of which present in sufficient replications in central Maranhão and in Pará states. According to between-region BCA, the regional effects on data variability were small (only 7.7% of total variance), and t-tests between regions did not detect any significant differences. Thus, regional variability did not bring a relevant bias to our key findings.

We estimated  $AGB \geq 10$  cm dbh in a 0.20 ha sampling unit per site, similar in size to sampling units successfully employed in studies of tall spontaneous forests (Keller *et al.*, 2001). However, we recorded two very high ( $> 80$  Mg ha<sup>-1</sup>) estimates of shrubs, treelets and lianas ( $AGB$  1-10 cm dbh), one for SFY and another for CPA. A possible explanation for these outliers may be insufficient subplot-size for reliable estimation of mid-sized vegetation. This problem was likely of limited extent, as for a young secondary forest this may represent the main size range of the trees and for the commercial plantations this represents the size of the perennial tree crop (cacao, cupuassu, açaí, etc).

*Vegetation and land use-related differences*

Aboveground biomass of large ( $\geq 10$  cm dbh) trees and palms and of understory ( $< 10$  cm dbh) plants had a wide variation between systems, though frequently in opposite directions in consequence of the negative logarithmic relationship between large and small plants, presumably the outcome of light and nutrient competition (Schwinning *et al.*, 1998). Within the spontaneous forests, we observed successional trajectories of increasing overstory biomass and of decreasing understory biomass (Figure 2a). Within agroforests, we observed direct management effects, such as differing regimes of understory clearing, and of sweep-and-burn in the homegardens (Figure 2a). Overall, understory biomass was much lower in agroforests than in spontaneous forests, reflecting the management objective of reducing competition and maximizing crop resource-use. Neither topsoil OM nor bulk density differed significantly between spontaneous forests and agroforests (Table 2). Soil organic matter content was the lowest in SFY, intermediate in SFO and the highest in MF (Table 2), pointing to organic matter restoration along succession (Silver *et al.*, 2000). Management also influenced topsoil chemistry, with higher pH and Ca-concentrations in CPA (Table 2) presumably caused by liming and significant chemical fertilizer-input.

The increase of leaf litter N-content from young secondary forests to mature forests (Table 2) reflected a decreasing N-limitation and also a decreasing relevance of biological N<sub>2</sub>-fixation along secondary forest succession (Davidson *et al.*, 2007). N-concentrations in the litter-layer of agroforests (Table 2) were lower than reported elsewhere (Moço *et al.*, 2010), probably due to the low quantity of N<sub>2</sub>-fixing legumes in our agroforests.

Next to N, litter-layer C:N-ratios are key for litter-topsoil N-dynamics (Constantinides *et al.*, 1994). Plant tissues tend to mineralize with C:N-ratios of less than 25 while those greater than 25 tend to immobilize mineral N (Hobbie *et al.*, 2006). Thus, N-immobilization predominated in litter-layers (Fig 3b), with the

exception of one HA and two EFA sites. We observed a significant positive relationship between twig littermass (but not leaf littermass) and topsoil OM contents (chap. 3.2.2), which may be due to the elevated lignin/polyphenol-contents typical for twigs (Tu *et al.*, 2011). Twigs decompose slower than leaves within the litter layer, due to the higher concentrations of less degradable components, especially lignins and other polyphenols, which form precursors for the buildup of soil organic matter (Rahman *et al.*, 2013). Thus, a higher portion of twigs in littermass can be more important for organic matter buildup than the leaf litter layer. Agroforestry species with larger twig input could therefore improve soil carbon sequestration. Multivariate analysis permitted deeper insights into the interrelationships between variables, and the appreciation of land-use and management-effects on such relationships. Here, our results of PCA (i) partitioned two main axes ('soil physicochemistry' vs. 'large vegetation') (Fig 5b), (ii) indicated successional pathways, and (iii) positioned HA close to MF, whereas CPA formed a distinct proper group (Fig 5a).

According to BCA analysis (Fig 6), the homegardens had the most widespread point projection in both axes, pointing to high management variability. By contrast, the market-oriented commercial plantation agroforests exhibited the lowest within-system variability. This high variability within homegardens suggests a large potential of management improvement, by selecting sites with low indicator values and implementing agricultural practices to increase agronomic efficiency. Such beneficial effects could be maximized by combining plots with contrasting functions in a strategically designed landscape mosaic (S. T. Lovell *et al.*, 2009).

### *Plant-litter-topsoil interactions*

Plants and soils interact via a complex network of relationships between the vegetation components, the litter layer and the topsoil (van der Putten *et al.*, 2013). We observed (i) a negative relationship between aboveground biomass of large and

small plants, presumably the outcome of competition for light and soil resources (Schwinning *et al.*, 1998). However, this direct effect of canopy tree on understory is likely more complex since the spatial patterns of the light environment are not entirely dependent on stem height and the diameter at breast height (Montgomery *et al.*, 2001), the input variables measured to estimate aboveground plant biomass in this study; (ii) litter-layer biomass was affected by small vegetation but not by large trees, and was influenced mainly by the portion of leaves in the litter. This could be a consequence of the tradeoffs between light use efficiency and light interception efficiency (Onoda *et al.*, 2014) which results in fast changes of leaves in the understory rather than in the overstory; (iii) litter-layer twigs but not leaves contributed to soil organic matter build-up, probably due to their contrasting decomposition dynamics, with fast leaf litter decomposition contributing little to soil organic matter buildup, as opposed to the more recalcitrant twigs which are less easily degraded and form precursors of soil organic matter (Wang *et al.*, 2010). A further explanation for the weak relationships between the litter layer and topsoil organic matter is in the decisive role of (mainly fine) root biomass to soil organic matter formation (Zhang *et al.*, 2013), not covered in our study; (iv) leaf litter mass increased with a widening of its C:N-ratio and were negatively related with 'plant-available' topsoil P- and K-concentrations; similar results have been found in other studies throughout Amazonia (review by McGrath *et al.* (2001); and (v) expected relationships within topsoil, such as the positive relationship between clay-content and soil organic matter (Desjardins *et al.*, 2004), and the negative relationship between OM and bulk density (Feller *et al.*, 1997).

Management impacted strongly plant-soil interactions and may influenced the strength of the plant-litter-topsoil relationships. One conspicuous result of our study was the absence of significant relationships between topsoil OM-content and nutrient-concentrations or pH. In CPA (and partially in EFA), this can be attributed to liming (CPA only) and localized synthetic fertilizer-inputs. Low biomass-estimates of small plants and litter in the agroforest systems likely reflect the

management practice of regular weeding (for the CPA and the EFA systems) and in the homegardens of ‘sweep-and-burn’ in the house surroundings (Benjamin *et al.*, 2001). On the other hand, the lower values found in the mature forests are probably a result of the influence of the overstory (Schwinning *et al.*, 1998).

Relationships between components of plant biomass within vegetation, vegetation-litter and litter-topsoil physicochemistry were altogether weak. Several explanations are likely responsible for this: (i) Compared to regularly ploughed fields or species-poor grasslands, spatial variability in tropics may be high in both vegetation (Punchi-Manage *et al.* 2014) and soil (Fromin *et al.* 2013). (ii) We are confident that our sampling scheme well reflects the ‘grain’ of spatial variability of the different components covered, and previous study (Somarriba *et al.*, 2013) have shown similar sampling schemes to successfully represent the different vegetation, litter and soil components under investigation. However, an unavoidable problem of any sampling scheme is the assumption of ‘homogenous’ characteristics throughout the sampling units, whereas both above- and belowground plant-soil interactions vary in strength and expression from the plant center (stem position used for biomass estimation) towards the crown rim and neighboring vegetation (Rhoades 1996). (iii) Our study showed how management can affect soil nutrient availability via inputs such as liming, fertilization, and in homegardens the addition of crop residues from the outside (Rugalema *et al.*, 1994), and how it can affect its spatial variability via pruning, weeding, and sweeping and burning of litter piles. Thus, systems-specific management will alter plant-soil interactions and – over all systems – reduce the strength of these relationships.

Co-inertia analysis (Table 3) helped us to synthesize the multivariate relationships between vegetation-litter-soil. Overall vegetation impacts on topsoil were stronger in spontaneous forests than in agroforests and lacked (i.e., discontinuity) in two agroforestry-systems (CPA and HA). Our results indicate that even agroforestry systems, commonly accepted as a low-impact land use, may face sustainability challenges (P. K. R. Nair, 2014), as specific management practices can

affect the relationships within the plant-litter-soil interface.

### *The special case of homegardens*

Omnipresent homegardens are complex multistrata systems, typically with high plant diversity and are subsistence- rather than market-oriented (Mohri *et al.*, 2013).

Homegardens are also special in their specific ‘sweep-and-burn’ fire regime of the litter layer (Benjamin *et al.*, 2001; Winklerprins, 2009) practiced by the house residents in order to reduce the occurrence of insects and snake, and to keep the homegarden ‘clean’ around the farmer’s houses. Benjamin *et al.* (2001) suggested that this practice could offset carbon liberation and negatively affect nutrient-cycling. We found elevated K-concentrations (Table 2) and widespread P-concentrations or P-hotspots (Figure 3a) in the topsoil of homegardens, similar to results reported by Pinho *et al.* (2011) in homegardens of Roraima, Brazil. Potassium in the soil is easily lost via leaching (He *et al.*, 2013), whereas P is immobile but subject to P-fixation especially in highly weathered tropical soils (Gama-Rodrigues *et al.*, 2014; Melo *et al.*, 2015). The higher topsoil K-content in homegardens may result from leaching protection by the large trees, whereas the widespread values of topsoil P-availability point to high soil heterogeneity in homegardens. Both findings could be the outcome of spatially irregular burning of litter piles (the burning of litter piles which releases K and P into the soil, which is protected against leaching and concentrated in the location of the burn). Thus, sweep-and-burn may be a fire regime capable of creating a mosaic of edaphic conditions, and such heterogeneity could benefit ecosystem-stability (Tittonell *et al.*, 2013). By contrast, the lower pH values in both homegardens compared to CPA only (Table 2) may result from the low burning temperatures in the litter piles—insufficient to increase pH (Certini, 2005). Sweep-and-burn may combine quick nutrient mineralization and the creation of spatial heterogeneity with protection against leaching by the large trees with their

dense canopies and root systems. Further nutrient-relevant mechanisms in homegardens include the influence of nutrient pumping by larger plants, and nutrient inputs by the addition of crop and food residues from outside plantations and the house (crop residues, household refuse, animal manure) (Rugalema *et al.*, 1994; Seneviratne *et al.*, 2006).

In spite of their omnipresence (Mohri *et al.*, 2013) and socioeconomic success (E. Cardozo *et al.*, 2015), homegarden agroforests remain under-researched (P. K. R. Nair, 2014). Complex geometries and interactions have so far largely evaded science (Seneviratne *et al.*, 2006) and homegardens have never been part of systematic agronomic improvement efforts. The high between-site data variability found here (Fig 5 and 6) points to a substantial potential for management improvements (optimized composition, spacing, pruning, etc).

## Conclusions

We found strong impacts of land-use on soil-quality indicators and aboveground biomass, within agroforests partly related to contrasting effects of understory and soil management, whereas changes in spontaneous forests are related to successional trajectories. We also detected a disruption in the plant-litter-soil continuum for the agroforests, presumably the result of management. Altogether, biomass in agroforests was statistically similar with mature rainforests and litter and topsoil nutrients were higher than in young secondary forests, confirming their ecological sustainability and potential as alternative to slash-and-burn in eastern Amazonia.

Our study highlights the outstanding characteristics of traditional homegarden agroforestry in terms of biomass, homegardens ranked next to mature rainforests. The traditional ‘sweep-and-burn’ practice combines rapid nutrient mineralization with protection against leaching by large trees and a closed canopy. The large between-site variability pointed to a significant potential of improved

management, homegardens need to be part of strategies for sustainable rural development.

## **Acknowledgments**

Research was financed by the Research Fund of Maranhão State (FAPEMA) and by the Brazilian Council of Higher Education (CAPES). We thank INCRA (Federal Colonization and Land Reform Agency), Embrapa Eastern-Amazonia, MST (Movement of the Landless) and the Mixed Agricultural Cooperative of Tomé-açu (CAMTA) for their invaluable practical and infrastructure support. We also thank Waldeir Brito for his aid with multivariate vector graphics, and Noriko Cassman for English language editing of the manuscript.



## CHAPTER 4

# You must choose, but choose wisely: a review of model-based approaches for microbial community analysis

**Authors**

Márcio Fernandes Alves Leite, Eiko Eurya Kuramae

Published in: *Soil Biology and Biochemistry*

## Abstract

Soil microbial community data produced by next-generation sequencing platforms has introduced a new era in microbial ecology studies but poses a challenge for data analysis: huge tables with highly sparse data combined with methodological limitations leading to biased analyses. Methodological studies have attempted to improve data interpretation via data transformation and/or rarefaction but usually neglect the assumptions required for an appropriate analysis. Advances in both mathematics and computation are now making model-based approaches feasible, especially latent variable modeling (LVM). LVM is a cornerstone of modern unsupervised learning that permits the evaluation of evolutionary, temporal, and count structure in a unified approach that directly incorporates the data distribution. Despite these advantages, LVM is rarely applied in data analyses of the soil microbiome. Here, we review available methods to handle the characteristics of soil microbial data obtained from next-generation sequencing and advocate for model-based approaches. We focus on the importance of assumption checking for guiding the selection of the most appropriate method of data analysis. We also provide future directions by advocating for the consideration of the dataset produced by sequencing as a representation of microbial detections instead of abundances and for the adoption of hierarchical models to convert these detections into estimated abundances prior to evaluating the microbial community. In summary, we show that model assessment is important for qualifying interpretations and can further guide refinements in subsequent analyses. We have only begun to understand the factors regulating soil microbial communities and the impacts of this microbiota on the environment/ecosystem. Understanding the assumptions of new methods is essential to fully harness their power to test hypotheses using high-throughput sequencing data.

## Introduction

Understanding soil microbial community data (bacteria, archaea, fungi, and protists) has become the holy grail of microbial ecology. Advances in sequencing technologies and molecular biology methods have improved how microbiome studies are conducted, from sampling design to biological interpretation (Knight *et al.*, 2018), and enabled observation periods ranging from short to long (Silverman, Shenhav, *et al.*, 2018). However, the use of inappropriate methods of analysis in soil microbiome studies might generate insufficient or, even worse, misleading biological inferences (Sohn *et al.*, 2017) due to the high-dimensionality of sequence data (high number of variables representing all amplicons of phylogenetic markers or partial genes in metagenome sequencing), overrepresentation of zeros, variable effort of data collection, imperfect detection, and the need for ancillary data (environmental variables, phylogenetic information, etc.) (Li, 2015). In addition, bias can arise from the cumulative effect of both systematic and random errors throughout the whole process, including DNA extraction, sampling, amplicon, sequencing, and bioinformatics (Hugerth *et al.*, 2017; McLaren *et al.*, 2019; Pollock *et al.*, 2018). Therefore, conclusions cannot be drawn from comparisons of raw signal values for the same microbial taxa across multiple samples because the observed differences in sequence abundance are not due solely to differences in the true amount of DNA in a sample.

Another crucial characteristic of the soil microbiome is the complex interactions between microbes (Armada *et al.*, 2018; Ho *et al.*, 2016). These interactions cannot be represented solely by correlations because the compositional nature of microbial community data biases correlations by inducing false significant negative relationships: whenever a taxon become more abundant, it reduces the proportions of other groups, which current methods falsely interpret as a statistically significant negative correlation (Gloor *et al.*, 2017). Friedman *et al.* (2012) observed that when the population contains dominant taxa, the problem of spurious associations is very

severe.

Given these challenges and available methods, here we review current approaches for the analysis of microbial communities with the aim of synthesizing developments in model-based approaches. Our main goal is to highlight the importance of correctly understanding the assumptions underlying each method. In particular, we focus on (i) statistical implementations and applications for community-level soil microbiome analysis, (ii) scenarios in which various current approaches and recent methods may have tangible benefits, and (iii) the assumption checking required for appropriate analyses. We also advocate the potential of the largely unexplored approach of imperfect detection for handling sequencing bias, thus suggesting directions for future applications. We believe that soil microbiome analysis should provide not only inferential information but also information on the sources of bias and how bias was considered in the whole analysis. To better illustrate the concepts discussed here, we use the dataset on the endophytic microbiome from host tree seedlings available in the ‘*gjam*’ R package developed by Clark *et al.* (2017) and simulated data from Sohn and Li (Sohn *et al.*, 2017) in the ‘*GOMMS*’ R package.

## **What do soil microbiologists hope to learn from amplicon sequence datasets?**

According to Prosser (2020), microbial ecologists aim to gain an understanding of the relationships and interactions between microorganisms and their environments. Similarly, the major goals of soil microbiologists applying molecular approaches can be classified into five main topics: (i) assessing soil diversity (i.e. phylogenetic diversity); (ii) identifying dominant and low-occurrence microbes; (iii) comparing soil communities subjected to an experimental condition or geographical distance; (iv) identifying microbial taxa (OTUs, ASVs, or species) that are more sensitive to edaphic factors; and (v) establishing relationships within

and between microbial taxa to understand community structure. Developments in high-throughput sequencing technology have propelled this scientific field into a golden age, and soil microbiology has greatly profited from the information gained by these methods. In turn, the enormous amount of information generated has created a demand for data analysis approaches capable of synthesizing the results for thousands of different microorganisms into more tractable information.

Initial data analysis methods originated from classic approaches in community ecology. (i) To assess soil microbial diversity, microbiologists first used diversity indexes such as alpha (e.g., Shannon, Pielou) and beta (Bray-Curtis dissimilarity) and later developed indexes to include phylogenetic diversity (e.g., UniFrac) (Lozupone *et al.*, 2005; Lozupone *et al.*, 2010). (ii) To identify both dominant and low-occurring microbes, relative abundance and other measures of proportions (log-ratios) have been explored. To (iii) compare soil communities and (iv) identify sensitive microbial taxa, multivariate analysis tools such as cluster analysis, principal coordinate analysis, and constrained analysis have been applied. The main method of analysis is distance-based ordination (constrained or unconstrained), which summarizes the community variation in a concise, low-dimensional form, e.g., as a scatterplot (Hui *et al.*, 2015). These ordination methods can be followed by permutation-based analysis (PERMANOVA, ANOSIM, Mantel test, etc.). Ordination also permits the identification of correlations, but more direct analyses of correlation such as Spearman or Pearson were initially adopted to evaluate microbial taxa relationships and perform network analyses (Lupatini *et al.*, 2014).

Distance-based methods are now expanding to include ecologically meaningful metrics aimed at tackling multiple goals. For example, Schmidt *et al.* (2016) proposed the indices TINA and PINA, which combine phylogenetic information to measure and evaluate community similarity while accounting for the interactions between taxa. Given the increasing number of available metrics, (Paliy *et al.*, 2016) emphasized the need for testing different indices to quantify the different

aspects of the microbial community. However, only general guidance on index selection is available (Buttigieg *et al.*, 2014), with no formal test to ensure an appropriate choice (Warton, Foster, *et al.*, 2015; Warton *et al.*, 2012). A critical look at many studies of microbial ecology reveals that the choice of a metric typically focuses on the potential information offered by the approach. For example, the Bray-Curtis dissimilarity reflects the percentage of organisms belonging to different taxa in two samples (e.g., a value of 0.27 indicates that 27% of all organisms in one sample are different from those in the other). This is analogous to justifying the choice of a parametric approach based on a normal distribution because this would permit the evaluation of differences between means and standard errors. Based on the potential information offered, one may decide to use the Bray-Curtis dissimilarity without knowing that this metric is not suitable for compositional data (Jost *et al.*, 2011). To better guide the selection of distance-based methods, we first need to understand the nature of the sequencing data and how its characteristics limit or even bias currently available methods for data analysis.

## **The nature of sequencing data limits analysis using traditional approaches**

Advances in microbiology studies have revealed that the nature of sequencing data is distinct from that of data in traditional ecological studies of macro-organisms. First, sequencing the genes in a soil sample is not the same as counting (Gloor *et al.*, 2017). Thus, if a distance-based method takes into account the proportion, it is prone to errors (Fernandes *et al.*, 2014; Gloor *et al.*, 2017). Moreover, each step of a metagenomics workflow (DNA extraction, PCR, and sequencing) adds an error that seems to be taxa-specific (McLaren *et al.*, 2019).

In an attempt to control the influence of sequencing bias, microbiome data are often converted into proportions by dividing the number of observed sequencing reads (counts) by its the total (Gloor *et al.*, 2017) or into some dissimilarity measure

(e.g., Bray-Curtis, weighted UniFrac). Another widespread practice is to rarefy the count data to force the samples to have the same number of total sequencing reads (Hughes *et al.*, 2005), at the expense of discarding vast amounts of information (McMurdie *et al.*, 2014). The aim of these approaches is to fulfil the criteria for classic multivariate analysis and ANOVA-like tests, which use random permutations for inferential analysis (Anderson *et al.*, 2013). Therefore, these approaches focus on bending the data to fit a particular type of analysis instead of using the analytical approaches to better understand the data variability.

## **Data essence precedes data existence: the key properties of microbial community data**

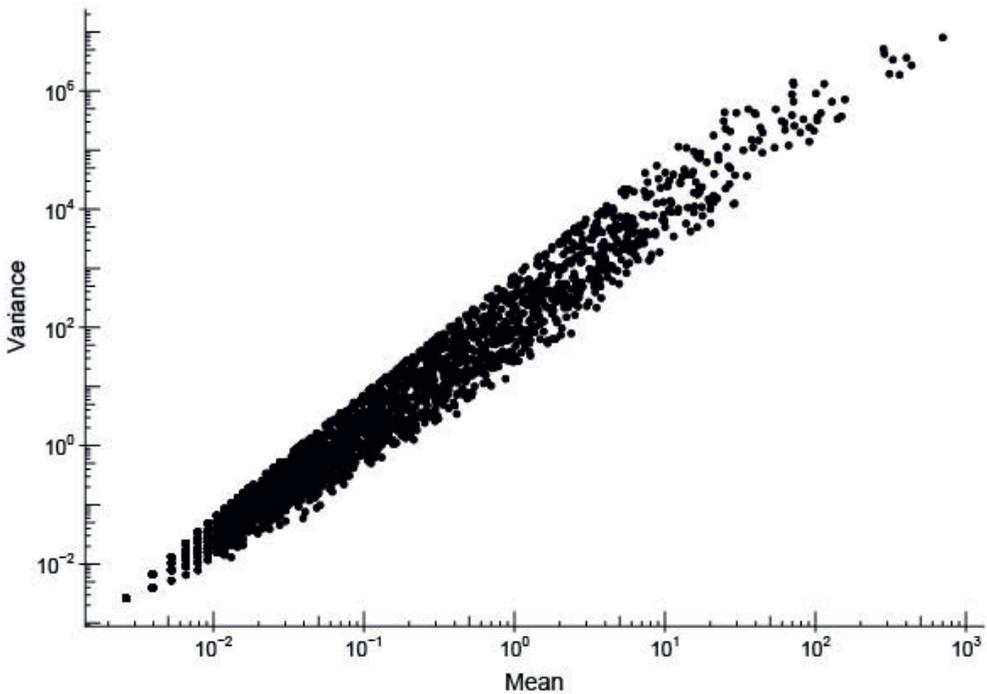
Soil microbial community data is the output of molecular approaches. Several sources of potential error arising before this microbial data is obtained could interfere with the final observed microbiome. Dimitrov *et al.* (2017) reported differences in the efficiency of soil DNA extraction and showed that successive extraction can increase the DNA yield from 1-374%, depending on the soil type. Subsequent amplicon sequencing remains strongly limited by the choice of primers. Moreover, shotgun metagenome approaches identify 1.5 to 10 times as many phyla and genera as 16S rRNA-based amplicon strategies (Poretzky *et al.*, 2014). 16S rRNA amplicon approaches can also introduce exaggerated between-sample diversity (Edgar, 2017). Furthermore, the standard sequence clustering approach is not free from criticism (Nguyen *et al.*, 2016). Given the limitations of sequencing platforms (Gloor *et al.*, 2017) and molecular approaches (Dimitrov *et al.*, 2017; McLaren *et al.*, 2019), we must assume that OTU (operational taxonomic unit) or ASV (amplicon sequence variant) counts do not reflect microbial community abundance but microbial detection in the soil environment. In essence, the sequencing data are compositional, sparse, and overdispersed, and each sample presents a different sequencing depth.

The compositional nature of soil microbial community data originates from the lack of independence imposed by high-throughput sequencing (Fernandes *et al.*, 2014). Every sequencing platform has a memory size limit, which is never reached because of differences in efficiency. Therefore, regardless of the amount of information available in the DNA sample, the output of a sequencing analysis will be determined by the limit of the sequencing platform. Consequently, to fulfill the goals of studies in soil microbiology, data analysis should focus on differences in proportions rather than counts. More advanced statistical approaches are required to accommodate the compositional nature of microbial community data (Lovell *et al.*, 2015; Rivera-Pinto *et al.*, 2018), as compositional effects can result in significant bias if a few taxa dominate the soil community (Delgado-Baquerizo *et al.*, 2018).

Microbiome datasets are also characterized by sparsity or a high number of zeros. In sequencing data, a zero count can be a result of either true absence or undetected presence. Silverman, Roche, *et al.* (2018) suggested three classes of zeros: (i) sampling zeros originating from undetected organisms in a given sampling method; (ii) technical zeros caused by technical bias (e.g., limitations on methods of DNA extraction, PCR or the sequencing platform); and (iii) biological zeros due to true absence in the evaluated sample (Gloor *et al.*, 2017; Hugerth *et al.*, 2017; Paliy *et al.*, 2016). The lack of clarity in detection has led to arbitrary approaches for avoiding this error, such as discarding taxa (OTUs or ASVs) with fewer than an arbitrarily chosen number or proportion of reads or presence in fewer than an arbitrary number of samples. These cutoffs might ultimately bias the analysis, especially for alpha diversity, as reviewed by Hugerth *et al.* (2017). In summary, the microbial abundance detected by high-throughput sequencing is far from a true picture of the microbiome.

The sparse nature of microbial community data becomes evident when evaluating mean-variance relationships (Björk *et al.*, 2018; Nelder *et al.*, 2004). The mean-variance relationship measures the degree of heterogeneity between different taxa and shows that differences between means are followed by differences in

variances (Warton *et al.*, 2016). The mean-variance relationship increases greatly with the sparsity of data, particularly for count-based data, generating so-called overdispersion (Warton, Blanchet, *et al.*, 2015; Warton *et al.*, 2012). This overdispersion affects the location of each sample in the ordination plot, leading to a biased interpretation of the results (Warton *et al.*, 2012). Therefore, the mean-variance relationship plays a central role in any multivariate analysis (Warton *et al.*, 2017): if it is neglected, the microbial community analysis will claim changes that do not exist or fail to detect differences that do. Different variables (taxon abundance, metabolite concentration, etc.) usually display a sparsity of data (many species tend to be present in large numbers or not at all), leading to variance greater than the mean (an overdispersed mean-variance relationship) (Warton, 2017). Figure 1 illustrates a strong positive mean-variance relationship of the endophytic microbiome dataset.



**Figure 1** – Mean-variance relationship of the endophytic microbiome from host tree seedlings available in the ‘gjam’ package published by Clark *et al.* (2017). The most abundant Operational Taxonomical Units (OTUs) (high values on the x-axis) also display a very large variance (high values on the y-axis).

*Assumptions first*

Each distance-based metric makes an implicit assumption about data variability that is rarely considered, which increases the likelihood of choosing an inappropriate dissimilarity metric (Warton *et al.*, 2012). For example, the Bray-Curtis dissimilarity distance only weights different taxa appropriately if they have similar variability (Warton *et al.*, 2012). The Bray-Curtis dissimilarity was originally developed for surveys with similar areas or volumes and therefore requires even sampling. If the sampling is uneven, as is the case with varying sequencing depth (Gloor *et al.*, 2017), the Bray-Curtis dissimilarity will fail to identify differences in the microbial community unless a vast amount of information is wasted in rarefaction (McMurdie *et al.*, 2014). An inappropriate dissimilarity metric will fail to account for the true mean-variance relationship present in the data, a key statistical property of multi-taxa data (Warton *et al.*, 2012).

An inappropriate choice of distance-based methods compromises the capacity of ordination analysis to distinguish differences between samples (Warton *et al.*, 2012). For example, a simulation performed by Sohn *et al.* (Sohn *et al.*, 2017) showed that if groups of samples (e.g., two or more different treatments) differ in their variability, such as a much higher variance of one group compared with the others, non-metric multidimensional scaling (NMDS) fails to display differences between them (clustering groups that are in fact different), while principal coordinate analysis (PCoA) tends to present false differences. Therefore, it is crucial to confirm that the method of choice is suitable for the data under study.

Another important issue is the flexibility of some methods in terms of certain assumptions, such as non-parametric multivariate approaches (Anderson, 2001). This flexibility might lead many researchers to neglect an important prerequisite for appropriate analysis: checking the method's assumptions. For instance, PERMANOVA is used to compare differences between two or more different groups or treatments of samples via dissimilarity. This method takes advantage of permutation and randomization procedures to deliver a robust analysis that does not

depend on data normality. However, as pointed out by Anderson (2006), many dissimilarity metrics are sensitive to differences in dispersion between groups of samples (e.g., microbial communities from different treatments). Therefore, to avoid a biased PERMANOVA, the dispersion between groups should first be checked via a homogeneity test. Accordingly, Anderson (2004) developed the PERMDISP test to evaluate the dispersion and test the homogeneity within each group of samples or treatments being evaluated. This crucial step is often neglected in published PERMANOVAs.

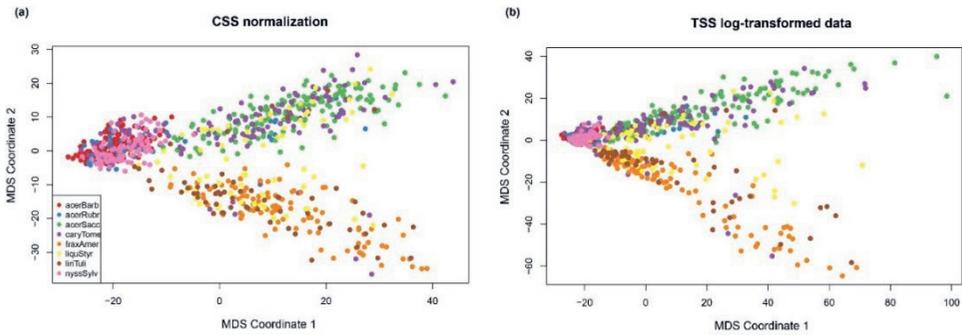
The need to check assumptions also applies to correlation analysis. Pearson's correlation between two compositional components cannot be calculated without considering the compositional nature of the sequencing data (Gloor *et al.*, 2017). Because of their compositional nature, taxa abundances are not independent and thus fail to fulfil the assumption of independence required by correlation analysis. Another major issue of correlation analysis is the isolation of the true signal of biotic interactions from other possible influences (shared environmental response), which is discussed in section 4.4.

### *Data transformations are often inappropriate or counterproductive*

As a first attempt, one might want to perform data transformation to reduce the influence of unequal mean-variance relationships. The sparse nature of microbial community data initially suggested data transformation to the log scale. However, concerns about the efficiency of this approach (Ives, 2015; O'Hara *et al.*, 2010; Warton *et al.*, 2016) and the required inclusion of a pseudo-count were raised, as the appropriate value of the pseudo-count is not clear (Costea *et al.*, 2014). Warton (2017) demonstrated that data transformation cannot be reasonably expected to stabilize variances for small counts and thus satisfy linear modeling assumptions. Consequently, a strong bias is imposed on the analysis when data transformation or normalization becomes a source of bias instead of allowing a fair comparison.

McMurdie *et al.* (2014) showed that the most common normalization technique, relative abundance (also known as total-sum scaling, TSS), fails to control overdispersion and also disrupts data independence, thereby also increasing the rate of false-positive correlations. Gloor *et al.* (2017) subsequently argued that the compositional nature of high-throughput sequencing data itself breaks data independence. The cumulative-sum scaling (CSS) approach (Paulson *et al.*, 2013) admits that some samples may be preferentially sequenced and focuses on removing the biases by dividing the raw counts by the cumulative sum of counts up to a specific percentile. This data-driven approach reduces the influence of preferentially sequenced samples to avoid creating an unbalanced sampling effort. Paulson *et al.* (2013) also used a mixture model that implements a zero-inflated Gaussian (ZIG) distribution to explicitly account for data sparsity. This model distinguishes true absences from technical absences, a common situation when some samples have a much smaller number of reads.

However, Costea *et al.* (2014) noted that, in terms of ordination analysis, the main advantage of CSS is the log-transformation. A simple log-transformation of TSS data in which the scale of the pseudo-count is adjusted produces similar results, as illustrated in Figure 2. Moreover, the adoption of a zero-inflated Gaussian as the distribution assumption is controversial. Warton (2005) argues that a zero-inflated Gaussian distribution is not always necessary, and that model diagnosis should determine its adequacy. In summary, CSS partially controls the influence of unbalanced sequencing data, but its adequacy should be checked prior to considering its results.



**Figure 2** – First two dimensions of multidimensional-scaling (MDS) analysis of endophytic microbiome data normalized by cumulative-sum scaling (CSS) (a) and total-sum scaling (TSS) log-transformed (b) normalization of the endophytic fungal community present in different plant hosts. For the TSS, we performed the log-transformation adding 0.000001 as a pseudo-count.

Data transformation also contributes to converting the data to proportions and tackling the compositionality constraint. The tool for this transformation is the center log-ratio (clr). First, the ratio converts the sequence counts to proportions, and then the log makes all the different samples symmetric and linearly related (Pawlowsky-Glahn *et al.*, 2015). However, the clr transformation remains subject to the sparsity problem. Some promising alternatives exist to handle the zeros and allow the use of this transformation, such as combining the clr transformation with the matrix completion method as proposed by Martino *et al.* (2019).

Approaches other than data transformation and distance-based methods also exist and offer an opportunity to perform analysis when the data presents a high and/or heterogeneous dispersion. Generalized linear modeling (GLM) is a statistical approach designed to handle different forms of data dispersion and data formats, such as counts (Cameron *et al.*, 2013), continuous variables (Foster *et al.*, 2013), and compositional data (Aitchison, 1982). Basically, GLM selects a probability distribution that best describes the data instead of trying to adjust the data to a specific distribution like the dissimilarity measures mentioned earlier. However, GLM remains dependent on the proposed assumptions of the model, which are not always clear to a broad audience. Consequently, the use of models that are more

complicated than those commonly employed by microbial ecologists may increase the risk of modeling errors. This risk is magnified by the traditional omission of advanced mathematics and statistics from the core curricula for aspiring microbial ecologists.

## **Tools and approaches for unbiased data analysis of the soil microbiome**

To circumvent the challenges imposed by the data's nature, the analysis method should be capable of handling the different sources of bias. We advocate that the method should (i) investigate the microbial community differences under a compositional constraint (the limitation imposed by the sequencing) and, consequently, a lack of independence of correlation analysis; (ii) determine the influence of sequencing bias on data variability, considering that some samples are preferentially sequenced due to interference in the detection of some specific microbes; (iii) adapt to the data mean-variance relationship present in the data; (iv) add additional information (traits, environmental covariates, phylogenetic distances) in a unified framework; (v) account for the influence of rare taxa; and (vi) disentangle biotic interactions from environmental responses.

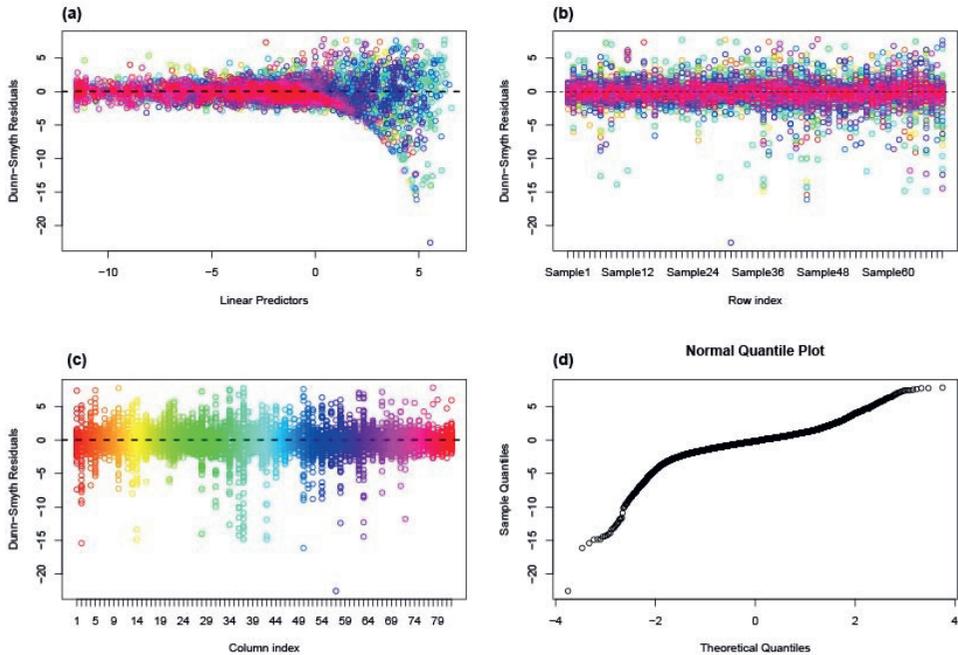
The solution for accurate microbial community analysis should not rely on a large arsenal of measures for community (dis)similarities. It should focus on powerful and versatile alternatives that appropriately handle high-throughput data and provide tools for checking model adequacy. The most appropriate analysis of microbial community data will account for the influence of the mean-variance relationship. Currently, the most suitable approach is GLM, which separates out the true effect from random noise by using a linear predictor (systematic component) and a distribution assumption (stochastic component). We propose the use of model-based approaches for both constrained and unconstrained analysis given their capacity to explicitly model the mean-variance relationship in the data.

## *Model-based approaches*

Model-based approaches involve the construction of a generalized linear model that explicitly dictates how the species response is generated. GLM distinguishes the data into observed variables (e.g., OTU or ASV read number) and explanatory variables or predictors, that is, any variable the researcher chooses to explain the observed variables (e.g., treatments, temperature, pH). According to Nelder *et al.* (2004), GLM requires three basic conditions: (i) there must be a probability distribution that fits the observed variable in order to generate a model in which the predictors explain the observed values; (ii) the effects of two or more explanatory variables are a combination of their effects (e.g., heating may increase some microbial groups if there are enough nutrients to support that growth); (iii) the fitted model provides the expected value as a result of the explanatory variables. In summary, GLM describes the linear relationship between the observed variables and explanatory variables as a linear function that follows a specific probability distribution tailored to the type of data under analysis. This specific probability distribution is used to link the observed data with the explanatory variables (e.g., treatments, environmental covariates) and is called the link function (Jiang, 2007). Hence, GLM permits the selection of the probability distribution that best fits the data, thus enabling analysis under a specific mean-variance relationship (e.g., if the variance equals the mean, the counts follow a Poisson distribution). After that, it is possible to check whether the GLM appropriately fits the data.

Warton *et al.* (2012) concluded that the best approach to select an appropriate mean-variance relationship for the analysis is to use GLM together with diagnostic tools to guide model specification. In summary, a specific link function is selected for the model, and the model fit is then checked to determine if this link function captured the data variability. For example, to handle overdispersed counts, a negative binomial (NB) distribution can be used as the link function in GLM. Other distributions might also be considered, including quasi-Poisson, zero-inflated

Poisson, and zero-inflated NB distributions (Zeileis *et al.*, 2008). Numerous distributions are available, but the residual analysis guides the choice of the one with the best fit (Dunn *et al.*, 1996), the so-called user-defined link function. To illustrate this process, we fit a model for data with overdispersion by selecting a Poisson distribution as the link function and evaluate its residual. This effect will be easily detected in a residual vs. fit plot or Dunn-Smyth residuals analysis, a visual approach that shows how the original data varies from its estimated value. If the data is overdispersed, the residual from a Poisson model will show a variance that continues to increase in microbial taxa with large counts (Figure 3a). Other auxiliary analyses allow us to check the homogeneity of residuals according to samples (Figure 3b) and taxa (Figure 3c). Additionally, we can evaluate the quantile plot (Figure 3d). If it does not follow a straight line, then the data diverges from the assumed distribution, thus demonstrating that the previously selected link function might not be appropriate for the given dataset. Therefore, assumption checking should be mandatory in model-based approaches (Warton *et al.*, 2016). The analysis should first capture the key data properties (using a priori knowledge and diagnostic tools) and then consider how to use these properties to make valid inferences.



**Figure 3** – Residual plot of the microbial community under the assumption that the data follow a Poisson distribution, as indicated by the Dunn-Smyth residual analysis of (a) linear predictors, (b) samples, and (c) taxa, combined with the (d) normal quantile plot to check how the data fits to the Poisson probability distribution.

GLM can be extended to evaluate dissimilarity measures. Generalized dissimilarity modeling (GDM) combines matrix regression and GLM to relate dissimilarities between pairs of sites to environmental conditions (Elith *et al.*, 2007). Hence, the GDM approach represents a potential solution to handle location-dispersion bias in dissimilarity metrics, such as in Bray-Curtis and weighted UniFrac distances. Therefore, GLM extends distance-based methods beyond the special cases for which it was developed.

*Model-based approaches to ordination*

Ordination analysis (e.g., PCA, PCoA, and NMDS) reduces data with many variables (high dimensionality) to a set of two or three responsive variables (Ter Braak *et al.*, 1988). Model-based approaches also offer many tools to reduce data dimensionality (Hawinkel *et al.*, 2018), including GLM with latent factors or latent variable modeling (LVM). Model-based approaches have the advantage of overcoming the limitations imposed by the nature of sequencing data. Analogous to classic ordination, a latent variable model creates a set of axes (here named the latent factor) to model the data variability, but the axes use a specified probability distribution that best captures the data variability (Hui *et al.*, 2015). Here again, the best choice is defined via residual analysis.

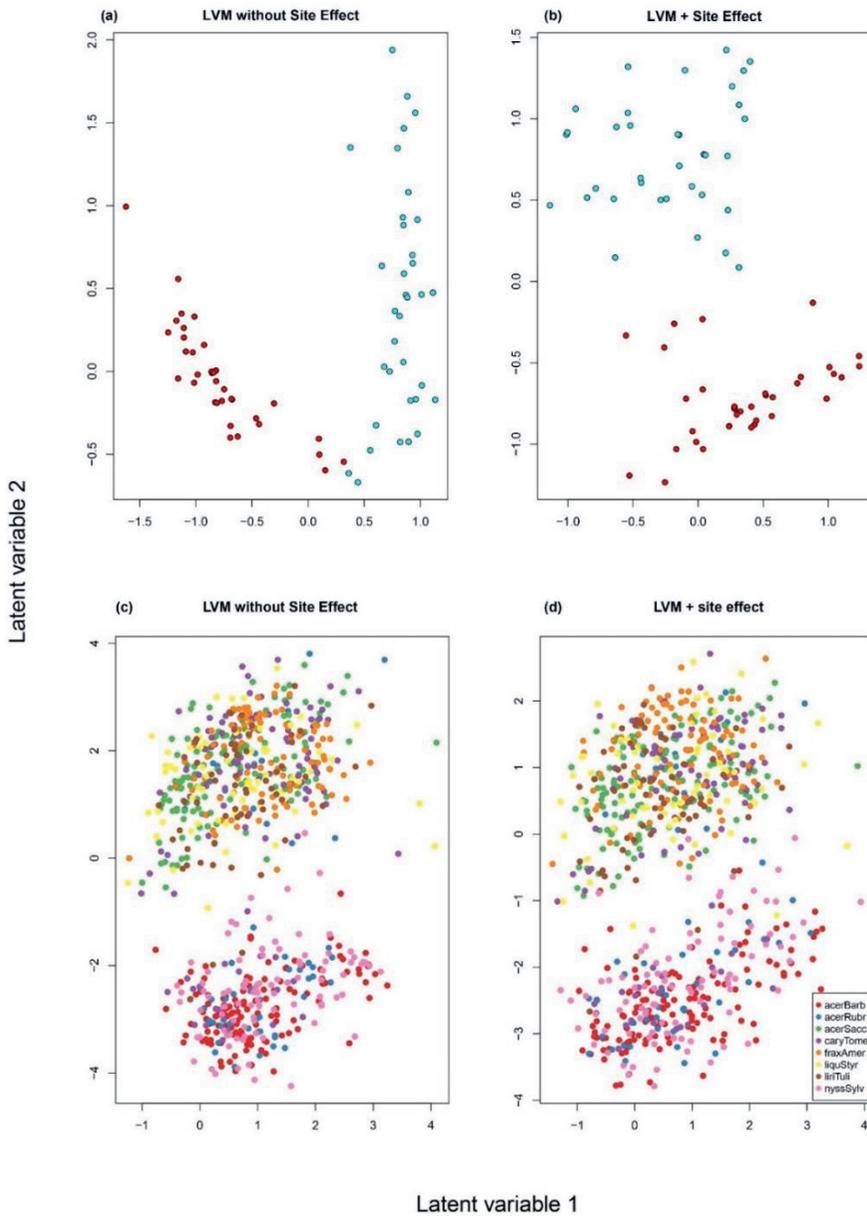
Sohn *et al.* (2017) proposed a GLM-based ordination method for microbiome samples (GOMMS) that uses as a link function a zero-inflated quasi-Poisson (ZIQP) distribution. However, the GOMMS package (<https://cran.r-project.org/web/packages/gomms/index.html>) does not yet offer the possibility of including environmental covariates, which limits the evaluation of the microbial community to unconstrained ordinations. If covariates are necessary (e.g., edaphic factors), the best option is general LVM (Verbeke *et al.*, 2017; Warton, Blanchet, *et al.*, 2015), which allows both constrained (predictor or ‘environmental’ variable) and unconstrained (response variable alone) ordinations.

LVM can also account for compositionality by including site effects. Here, a site effect is any factor that influences data at the sample level, such as differences in sample intensity (e.g., different numbers of sequencing reads in each sample). By including site effects, LVM evaluates the data in terms of relative taxa abundance (Hui *et al.*, 2015) even though the input is the number of sequences, and thus data transformation is not required. To illustrate, we performed LVM of the simulated data from Sohn *et al.* (2017) and found an arch effect (Figure 4a). This arch effect suggests that the data variability was not adequately captured by the latent factors.

However, after accounting for the site effects (Figure 4b), the bias in ordination disappeared, and the differences between sample groups became more evident. This result suggests that accounting for the compositional nature of data reduces bias in the analysis.

LVM can also handle complex relations by including polynomial terms (e.g.,  $x^2$ ,  $x^3$ ) in the linear predictor (covariates). For example, some soil factors such as pH might influence the microbial community with an optimum value, similar to a bell-shaped curve (where both low and high values are toxic). LVM can easily incorporate such complex relations by adding a quadratic term ( $x^2$ ) to pH in the model equation. Importantly, the need to account for any quadratic response could be diagnosed as an ‘arch effect’ in LVM ordinations (Hui *et al.*, 2015). Therefore, one can initially assume no quadratic relationship and, after visualizing the data, identify the presence of the arch effect. In classic multivariate methods, the presence of an arch effect usually invalidates an analysis and requires a change in approach. In LVM, an arch indicates the presence of nonlinearity, which might itself be a potential result and indicate a direction for continuing the investigation and model selection.

Like any statistical approach, LVM is founded upon assumptions (Hui *et al.*, 2015). Therefore, to ensure these assumptions are met, diagnostic tools should be used to select the best model; if the choice of probability distribution as the link function does not capture the data variability, this would become apparent during analysis. Alternative link functions can then be checked to ensure the best fit. In some cases, the neglected characteristic of the data may not be easy to identify. For example, for the data on endophytic fungi published by Clark *et al.* (2017), we found that the differences between considering or omitting the site effect may not be strong (Figure 4c-d), but because sequencing data is compositional (Gloor *et al.*, 2017), we should include the site effect as a default option for microbiome analysis.



**Figure 4** – Latent variable modeling (LVM) with no (a and c) and fixed (b and d) effects of sample sites. Addition of the fixed sample effect removed the ‘arch effect’, indicating that nonlinear effects influence the microbial community structure.

More importantly, model-based approaches to ordination provide a unified

framework for many different data types. LVM offers a parsimonious way to incorporate taxa interactions or other sources of correlation (such as missing covariates), thus enabling the development of joint species distribution models. For example, Björk *et al.* (2018) used a joint species distribution model to analyze the host-associated microbiome, which allowed adequate modeling of compositional aspects while answering ecological questions (e.g., proportion of microbial variance explained by elevation, sex, etc.).

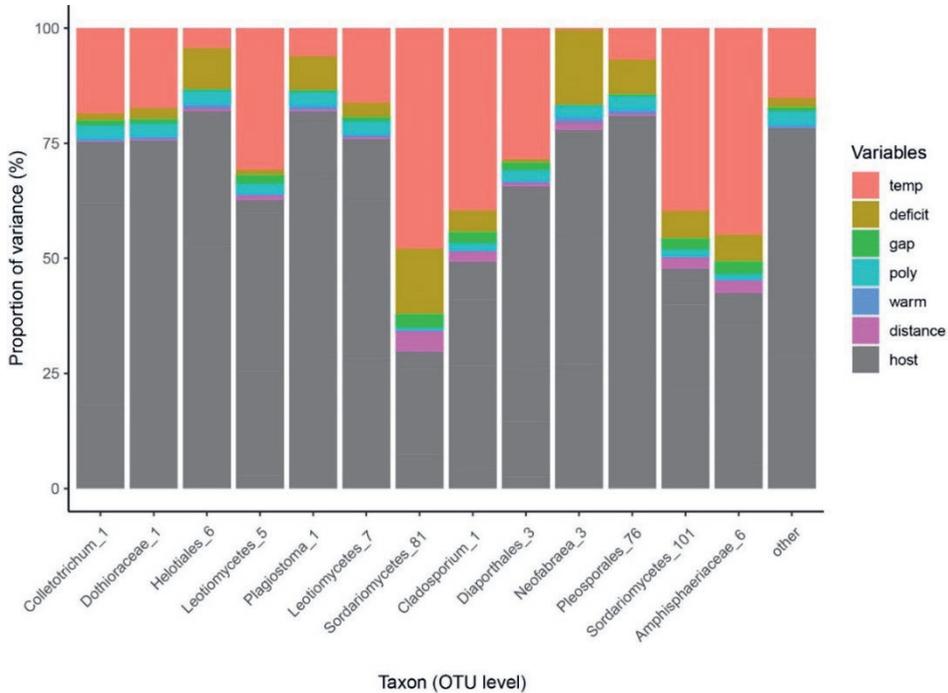
### *Joint species distribution modeling*

Joint species distribution modeling (JSDM) determines the influence of the covariates in determining the abundance of each taxon while also taking into account the presence of interactions between organisms (Ovaskainen *et al.*, 2017). By accounting for taxa interactions, JSDM provides more accurate estimates of the changes in communities (Clark *et al.*, 2014). However, JSDM remains comparatively unknown and underutilized in the field of microbial ecology, with few studies exploring its merits and limitations (Clark *et al.*, 2017; Ovaskainen *et al.*, 2011; Sankaran *et al.*, 2017). Recently, hierarchical Bayesian modeling (HBM) promoted an advance in JSDM by accounting for potential correlations between species using a residual term and introducing structured errors (Warton, Blanchet, *et al.*, 2015). The rationale for considering structured errors is the assumption that the residual term of a model analysis (the variability not explained by the treatments or covariates) is not just random noise but can also represent unobserved variables. First, the model identifies how the covariates explain the differences in organisms' abundance. Then the residual variability is considered as a result of potential interactions or non-included variables (e.g., phylogenetic distance, metabolism). HBM of communities encompasses several algorithms, e.g., the boral package (Hui, 2016) and mistnet package (Harris *et al.*, 2017), that use a hierarchical structure to infer signals of biotic interactions or improve community structure analysis.

The application of HBM culminated in the general framework proposed by Ovaskainen *et al.* (2017), the Hierarchical Modeling of Species Communities (HMSC). HMSC provides a hierarchical joint species distribution approach that offers a unified framework to combine different components (environmental factors, taxon traits, phylogeny, temporal, and spatial differences) to understand their influence on community assembly. HMSC assumes that (i) abiotic filtering is properly captured by the environmental covariates included in the model; (ii) biotic filtering is properly captured by the residual species-to-species correlation matrix or – in the case of time-series data – by matrices modelling the influence of each species in the previous year on the occurrence of every other species in the following year; and (iii) any residual variation (i.e., variation not predicted by environmental or biotic filtering) can be attributed to random processes such as dispersal limitations, environmental stochasticity or ecological drift.

HMSC offers the possibility of partitioning the explained variance into all components included in the model. Thus, it permits the measurement of the percentage of variability in the microbial community that can be explained by differences in soil factors, spatial distance, time, etc. As an example, Figure 5 shows the use of HMSC to evaluate the variance explained by the hosts and the other covariates (temperature, water deficit, shade, polyculture, warm, and distance) in the endophytic fungal dataset. This partitioning shows that the host plays the major role in selecting specific OTUs as endophytes (average of 66.2%), but changes in temperature also explain a large proportion of the endophytes found in the plants under study (22.6% in average). Notably, for a Sordariomycetes OTU, changes in temperature explain 47.9% of its variability (Figure 5). Therefore, this procedure informs how different processes shape the microbial community by assigning the proportion of variance of each microbial taxon occurrence resulting from environmental filtering, biotic filtering or random processes. This unified framework also decomposes the observed co-occurrences, leading to better hypotheses on true biotic interactions (Pollock *et al.*, 2014). In our example, the residual variance ranged

between 2.8% and 23.9%, suggesting that other non-measured variables also play an important role in determining the abundance of some endophytic fungi.



**Figure 5** – Variance partitioning obtained from hierarchical modelling of species communities (HMSC) using the example dataset of the endophytic microbiome from host tree seedlings available in the ‘gjam’ package published by Clark *et al.* (2017). Temp: temperature, deficit: water deficit, gap: presence of shade, poly: poly or monoculture, warm: room temperature or warming, far: distance (near or far) from the adult plant host. Host: plant species of the host.

Similar to HMSC, one can perform JSDM by modeling the observed taxa count data using Dirichlet multinomial (DM) regression, which has a different set of assumptions. This approach automatically accounts for measurement errors and the compositional constraint associated with counts data (Xia *et al.*, 2013). Therefore, at first glance, this approach seems more appropriate than HMSC for soil microbial data. However, the correlations between dependent variables in a Dirichlet model are always negative and thus may not represent the nature of the soil microbial

community.

Another interesting form of JSJM is generalized joint attribute modeling (GJAM) (Clark *et al.*, 2017). GJAM allows relationships between species to be inferred and interpreted on the observation scale while avoiding distorted correlations. Moreover, GJAM works well for datasets with more than 100 species based on the dimensional reduction approach developed by Taylor-Rodríguez *et al.* (2016). GJAM also accommodates sampling effort by treating observed counts as censored versions of true abundance, and parameters can be interpreted on the observation scale or the effort scale. This approach measures the quality of data and then determines the strength of the results and inferences. However, GJAM still does not offer a unified framework that includes the phylogenetic distance between different microbes. In summary, GJAM provides new flexibility for inference and prediction from ecological data by aligning the scales for multiple observation types and fitting the model on the observation scales. Recently, Norberg *et al.* (2019) evaluated the performance of 15 species distribution modeling strategies, including GJAM and other JSJM approaches, according to their accuracy, discrimination power, calibration, and precision for different levels of biological information (species occurrence, species richness, and community composition). Their results showed that no model performs well at all prediction tasks, further emphasizing the importance of model selection for achieving a more accurate analysis of community data. Therefore, the choice of a given model should be based on its capacity to best describe the data obtained from the soil community.

### *Understanding microbial co-occurrences*

Microorganisms in soil live as integrated microbial communities that process energy and materials and consequently affect their surroundings (Gilbert *et al.*, 2012; Lima-Mendez *et al.*, 2015; Schmidt *et al.*, 2016). Therefore, projections of taxon distribution and community structure across large spatial gradients must

include the effects of multispecies interactions. However, the compositional nature of microbial community data requires more advanced statistical approaches (Lovell *et al.*, 2015; Rivera-Pinto *et al.*, 2018), potentially in combination with a variable selection approach (Fang *et al.*, 2015). Compositional effects can result in significant bias if a few taxa dominate the soil community (Delgado-Baquerizo *et al.*, 2018). Moreover, the revealed associations among variables and patterns of object distribution do not inherently imply causality (unless the explanatory variables are the experimental treatments). Correlation is a powerful tool for helping researchers generate hypotheses (Dormann *et al.*, 2018), such as which interactions are biologically relevant in the systems under study and warrant further research (for example, through co-culturing or whole-genome sequencing). The models that we discuss here do not claim to specify or quantify interactions between species mechanistically; they simply allow for the detection of co-occurrences of different taxa that are potentially explained by true biotic interactions between microbes or common - but independent - responses to environmental variables.

Weiss *et al.* (2016) benchmarked the performance of eight correlation techniques (Bray-Curtis, Correlation Networks - CoNet, Local Similarity Analysis - LSA, Maximal Information Coefficient - MIC, Pearson, Random Matrix Theory - RMT, SparCC, and Spearman). Simulations with sparse data showed that the majority of correlation approaches are overly sensitive for OTUs with more than 50% absence. Based on these results, Weiss *et al.* recommended that an approach similar to SparCC be adopted to remove from the analysis any taxa that may actually compromise the investigation of community interactions. Again, such an approach neglects assumption checking and diagnosis. Taxa expected to be very rare and seldom should not be included in estimates of community network interactions; other studies have shown that increased rarefaction of data (leading to an increase in the proportion of zero counts) greatly increases the rate of false positives for all methods tested (Weiss *et al.*, 2016). Recently, Cougoul *et al.* (2019) demonstrated that classical associations and the abovementioned common network analyses do not

provide a reliable test for rare microbes (e.g., low-prevalence OTUs), and they recommend limiting these tests to only non-problematic pairs. Thus, the authors proposed implementing a filtering strategy to identify and select only testable associations. Therefore, once again, checking assumptions can provide better guidance on whether a specific method is reliable or not for a particular dataset.

Regardless of the procedure adopted, the tools for correlation analysis confound the effects originating from a shared environmental response (Pollock *et al.*, 2014), which occurs when two microbes co-occur due to some common response to a resource or environmental factor instead of a true biotic interaction. To address this point, JSDM determines if a co-occurrence results from a shared response (originating from an environmental, ecological or evolutionary process) or from a missing variable based on the residual correlation. The idea is that once the community is fit to a determined set of environmental factors, any departure from that fit (residue) might result from the potential interaction of any non-measured variable or an undetected or discarded microbe or from a biotic interaction (Dormann *et al.*, 2018; Pollock *et al.*, 2014). Therefore, residual correlation offers a statistical framework for reducing the influence of any measured environmental covariate (and hence the treatments) as a potential inducer of microbial co-occurrence.

Lastly, to apply model-based approaches to microbiome data analysis, microbial ecologists should focus on two questions: (1) Does the model fit the data well, and what techniques are available for performing this evaluation? (2) If the model fits well, does it provide informative summaries of the original data? We advocate that future research on microbial ecology provide answers to these questions prior to digging into scientific questions and biological explanations. Such an approach is essential for informing the reader about the quality of the findings of microbial community analyses.

## Future directions: the challenge of imperfect detection

To improve our understanding of the soil microbial community, it is important to recognize not only that microbiome datasets are compositional but also that detection is imperfect. Imperfect detection implies uncertainty about the abundance recorded (Dorazio *et al.*, 2015). Consequently, the observed abundance is a result of the efficiency in detecting all organisms in the sample and thus should be disentangled (Yamaura *et al.*, 2016). Soil sampling usually presents imperfect detection, resulting in the creation of the three classes of zeros discussed earlier (section 3.1) (Guillera-Arroita, 2017). Common species with low detection probability and rare species with high detection probability should not receive equal weight.

Model-based approaches deal with imperfect detection in taxa distribution by disentangling inferences about the taxa distribution from noise in observed abundances (Guillera-Arroita, 2017). Separating the detection from observed abundance requires technical replicates. For the soil microbial community, it is possible to increase the number of replicates by performing multiple DNA extractions from the same soil sample. Moreover, PCR can be performed multiple times from the same extracted DNA and contribute information on the variation induced within this process. PCR replication has also been suggested to provide a better snapshot of the microbial community (Ficetola *et al.*, 2015; Knight *et al.*, 2012).

By recording separately, the detections and non-detections in a sample in each DNA extraction and/or PCR, the model gains information about how detectable a species is and hence can account for changes due to missing the species at some time points (Iknayan *et al.*, 2014). However, conducting separate visits in time is not the only approach for collecting information about detectability, and we refer to Mackenzie *et al.* (2005) for a more in-depth discussion of currently available methods. In this review, we restrict our discussion to the case of multiple visits, since

it can be easily extended to molecular methods, and consider the N-mixture model, which requires data on replicated counts of individuals (Royle, 2004), adapted to the estimation of counts from sequencing reads.

Royle *et al.* (2003) developed the N-mixture model as a mathematical approach for estimating the abundance of a sample based on detection. This model allows sampling data to be evaluated as the number of individuals detected and associates a probability value with each taxon (each taxa occurrence consists of the outcome of a detection probability). The N-mixture model permits the development of a hierarchical detection-based model that accounts for imperfect detection by including many sources of bias, such as the distance between samples, differences in repeated measurements, and differences in time (Yamaura *et al.*, 2016). The model analyzes the experimental design only after determining the detection probability of each taxon in the table. Therefore, the hierarchical approach provides a framework for investigating the detection bias occurring in DNA extraction, PCR, and sequencing, if we include technical replicates in these processes.

If technical replicates are not possible or too expensive, another interesting alternative is generalized joint attribute modeling (GJAM), which classifies each sample according to its sampling effort (Clark *et al.*, 2017). In compositional data such as microbiome data, effort corresponds to the total number of reads per sample, and thus it is possible to put each sample into an effort scale. As a result, samples that were better sequenced will be considered more informative in the model. This weighted information enables a more accurate analysis of the microbial community.

In summary, accounting for imperfect detection of microbial communities implies greater sequencing effort; otherwise, detection would become an uninterpretable parameter. The effects of imperfect detection are a fruitful new area for microbial ecology to explore, and our focus is to introduce these simple methods for adjusting statistical models. Future research may use simulations such as the one reviewed by Escalona *et al.* (2016) to evaluate the impact of imperfect detection models on improving the estimation of the microbial community from next-

generation sequencing technologies. Comparisons with other approaches are left as an avenue for future work.

## Conclusion

Studies of the soil microbiome are a source of richly structured, high-dimensional data but bring novel scientific problems and setups. The increasing number of techniques and methods to overcome these challenges generates several possibilities for microbiome data analysis, and this large arsenal of choices increases the importance of assumption checking as the first step in appropriate model selection.

Model-based approaches can explicitly account for key statistical properties of data (compositional nature, imperfect detection, mean-variance relationship, lack of independence between taxa, rare taxa, and disentangling biotic from abiotic driven co-occurrences). By explicitly modeling the mean-variance relationship, LVM ordinations cover the true underlying site patterns. However, we must highlight that model-based approaches usually demand much more computational power and time than distance-based approaches, especially software that uses Markov chains to obtain regression coefficients. When specifying a statistical model, the primary goal should be choosing the model for the data at hand. As a substantial improvement that is currently available to provide more accurate analyses of microbial communities, we advocate models capable of estimating taxa abundances while accounting for imperfect detection, the effects of environmental factors, and associations between species. We have only begun to realize the potential of probabilistic methods to guide careful scientific reasoning – emphasizing both insights and their degree of uncertainty – in the analysis of complex microbiome datasets.

## **Acknowledgments**

The authors thank Johannes A. van Veen and Christoph Gehring for critical reading of the manuscript. We also would like to thank the anonymous reviewers for suggestions that significantly improved the manuscript. This research was supported by The Netherlands Organization for Scientific Research (NWO) grant number 870.015.022 and TKI Topsector Tuinbouw & Uitgangsmaterialen grant number TU17008. Publication 7057 of the Netherlands Institute of Ecology (NIOO-KNAW).

## CHAPTER 5

# Towards the Amazon mature forest microbiome: secondary succession convergence to fungi and agroforests divergence to bacteria

### **Authors**

Márcio F. A. Leite, Mattias de Hollander, Binbin Liu, Ernesto G. Cardozo,  
Hulda R. Silva, Ronildson L. Luz, Karol H. M. Muchavisoy, Flávio H. R.  
Moraes, Guillaume X. Rousseau, George A. Kowalchuk, Christoph  
Gehring, Eiko E. Kuramae

Submitted to: Science Advances

## **Abstract**

An alarming and increasing deforestation threaten Amazon tropical ecosystems. Notably, this occurs because shifting cultivation. Shifting cultivation is trapped in a vicious cycle of unsustainable intensification and degradation, due to the shortened period of forest recovery following natural succession (secondary forest). Agroforestry systems may offer a sustainable alternative, reportedly mimicking the plant-soil interactions of the natural mature forest. However, the role of microbial community, the key player in the above- and belowground linkages in tropical agroforestry systems (AFS) remains not known. Here, the land use systems (mature and secondary forest; AFS) affected both plant biomass and soil factors, and the soil microbial community. The current study provided evidence on how land use systems interfere with the aboveground and belowground interactions at ecosystem level, with a particular focus on these effects to the soil microbial communities (bacteria, and fungi). Collectively, this study demonstrates that secondary forests can converge to the same soil microbiome as found in the mature forest, by favoring the same groups of fungi. However, the different agroforestry systems foster distinct groups of bacterial community, likely a result of the differences in management practices. Additionally, our analysis suggests that AFS require more profound changes in the aboveground biomass and soil factors to reach the same microbiome as found in the mature forests. In summary AFS practices do not result in an ecosystem mimicking natural forest plant-soil interactions but reshape the ecosystem to a completely different relation between the aboveground biomass, soil properties, and soil microbiome.

## Introduction

Tropical rainforests are the planet's most species-rich biomes (Armstrong, 2016), a feature that is especially impressive given that these forests inhabit heavily weathered, nutrient-poor, and acid soils (Armstrong, 2020). Collectively, the combination of high diversity, high biomass, and nutrient-poor soil is made possible by an efficient aboveground-belowground interaction (Camenzind *et al.*, 2017; Milton *et al.*, 2007). Nutrients slowly accumulate in the plant biomass and efficiently cycle in the soil through a complex network of soil organisms.

Over the past decade Amazon tropical forest has been converted to commodity production (pasture and soybean) at a rate of 6.54M hectares per year (Kim *et al.*, 2015). To circumvent the limitations presented by nutrient-poor soils, many farmers adopt slash-and-burn practices, which use fire to quickly mineralize nutrients stored in the plant biomass, making them available for subsequent crops. However, the soils of the humid tropics are particularly vulnerable to degradation as the warm and humid environment promotes rapid organic matter decomposition and mineralization, nutrient loss caused by leaching and runoff (Markewitz *et al.*, 2004), and gaseous nitrogen losses (Brookshire *et al.*, 2017). Production thus declines rapidly after burning, causing farmers to abandon such land and move to a different plot of the forest, leading to further deforestation. Ultimately, repeated slash-and-burn cycles and shortened fallow periods (Lawrence *et al.*, 2010) lead to reduced agronomic productivity (Runyan *et al.*, 2012; Styger *et al.*, 2007), thereby exacerbating rural poverty (Jakovac, Peña-Claros, *et al.*, 2016; Satyam Verma, 2012).

Alternative to slash and burn agriculture in the tropics is agroforestry. The core principle of agroforestry systems lies in combining trees with crops, and/or animals in the same plot of land (a multistrata system) to mimic plant succession in the natural forest, while including crop production. As a potentially sustainable agronomic solution, agroforestry systems in tropical forests have been shown to have

the capacity to improve soil quality (Pinho *et al.*, 2012), increase system stability (Mohri *et al.*, 2013), and reduce economic risk by enabling production of multiple products and enhanced market flexibility (Souza *et al.*, 2012). However, the agroforestry systems can be very varied in their goals, crops and level of external impacts (Atangana *et al.*, 2014c). Consequently, the ecological and social sustainability of these systems varies widely.

Intentionally or unintentionally, agroforestry systems are designed to spatially, physically, and temporally optimize resource use by maximizing the positive interactions, and minimizing the negative interactions, between plants and soil subsystems (aboveground-belowground interactions). According to the central agroforestry hypothesis (Cannell *et al.*, 1996), the trees acquire and/or provide resources otherwise unavailable to the crops, especially nutrients from the soil deep layers (nutrient pumping). This also can reduce leaching and runoff (the safety net mechanisms) (Seneviratne *et al.*, 2006), thereby playing a crucial role in connecting the aboveground and belowground subsystems. However, we recently showed that, compared with the natural forest, agroforestry practices reduce the intensity of the aboveground-belowground interactions and the interdependence between plant and soil subsystems, which is likely an outcome of management practices (plant selection and arrangement, nutrient input, pruning, weeding, etc.) (Leite *et al.*, 2016). Thus, while the non-sustainable land use intensification in slash-and-burn practices clearly has negative impacts on soil nutrient recycling, above- and belowground biodiversity and ecosystem functioning and stability (Thiele-Bruhn *et al.*, 2012), intensely managed agroforestry systems may likewise interfere in the aboveground-belowground linkages that impact ecosystem function, especially nutrient cycling. The challenge in investigating the aboveground-belowground interactions in an agroforestry system begins with the multiple components or subsystems that play a major role in determining system functioning. Research on Amazon Forest to date has generally focused on tree-crop interactions or plant-animal interactions, while investigations of plant-soil interactions at the ecosystem level have only recently

been initiated (Leite *et al.*, 2016) and no studies on impact of agroforestry practices on soil microbiota - key component of soil ecosystem functioning (Bender *et al.*, 2016)- have carried out.

Highly efficient nutrient cycle is essential to tropical forest maintenance, what services are provided by soil microbiome. Therefore, here we examined soil microbial communities of three different agroforestry systems across the north-central Maranhão and eastern Pará states of eastern Amazonia and compared them those of immature secondary forest and mature rainforest sites. Our major goal was to investigate the capacity of the agroforestry systems to mimic the aboveground-belowground interactions found in mature forests. We linked microbiome features to measures of aboveground vegetation biomass, litter mass and the topsoil physico-chemical properties. By including the soil microbiome, we contribute to the design of more sustainable systems that better mimic the aboveground-belowground interactions of mature forests.

## Materials and Methods

### *Field survey, biomass estimation, and soil sample collection*

The study was conducted in the eastern periphery of Amazonia, at 56 study sites in six counties (Anajatuba-Itinga, Arari, Morros-Rosário, São Luís, Gurupi, and Tomé-açu), and aggregated into two regional clusters (Maranhão and Pará) in Brazil. The studied area covered 67,500 km<sup>2</sup> in north-central Maranhão and eastern Pará states of eastern Amazonia (Figure S1). We compared three different agroforestry systems with natural secondary forest succession and mature rainforest controls and explored system-associated differences in the aboveground-belowground interactions, specifically focusing on the impact of agroforestry system on these interactions. The maximum distance between sites within each county was <30 km, and the maximum distance between counties within each regional cluster was <150 km. Forty of the 56 sites were located in the central-northern Maranhão state, while

the others were approximately 400 km further westward in the Tomé-açu County in the eastern Pará state. The original rainforests in central Maranhão and in the majority of eastern Pará are almost entirely replaced by pastures and secondary forests. Lack of occurrence of mature rainforests in every county rendered geographic distance as a potential factor affecting the results and was therefore included in modeling. According to the Köppen classification, the climate is Aw and Am, and varies slightly between the two regional clusters (2100 mm of annual rainfall in the central Maranhão state and 2300 mm in the eastern Pará state, with 6 and 5 months of hydric deficit, respectively). The soil is nutrient-poor acid Oxisol or Ultisol (USDA, 2010), and the topsoil texture is loamy/fine-sand. Although the soil texture and density did not vary much between sites and clusters, we did detect significant soil physical/textural differences between some land use classes. We therefore considered soil texture as a covariate in the analyses of soil nutrients to investigate system-related differences. Interactions between the aboveground vegetation and soil physico-chemistry of a subset of 32 of these sites were previously analyzed and published (Leite *et al.*, 2016).

Four types of natural forests and three types of agroforests were compared. Site selection and classification was based on the work of E. G. Cardozo *et al.* (2015) and Leite *et al.* (2016), as follows:

(i) *Natural forests*: Secondary forests following slash-and-burn shifting cultivation or on abandoned pastures. Young (5–12-yr-old; SFY, 5 sites), mid-age (15–20-yr-old; SFM, 6 sites), and old (>30 yr-old; SFO, 7 sites) secondary forests were distinguished. Mature forests (MF, 12 sites) were also distinguished, and indicated original mature forests without any visible human perturbation or with low-intensity selective logging >60-yrs ago.

(ii) *Agroforests*: The following different systems were distinguished: enriched fallow agroforests (EFA, 6 sites), established by enrichment planting of fruit and timber species in the understory of 15–25-yr-old secondary forests; home-garden agroforests (HA, 13 sites), tall multistrata agroforests surrounding houses, virtually

omnipresent in the study region and throughout the tropics (Kumar *et al.*, 2004); and commercial plantation agroforests (CPA, 7 sites), regularly-spaced plantations with inorganic fertilization and liming, established or inspired by Japanese immigrants. The main products of CPA are cocoa (*Theobroma cacao* L.) and cupuassu (*T. grandiflorum* K. Schum), açaí (*Euterpe oleracea* Mart.), and black pepper (*Piper nigrum* L.).

Most of tropical rainforests are comprised of three to four canopy layers that agroforests tend to mimic: the canopy emergent, the main canopy, the understory, and the forest floor (Armstrong, 2020). In such complex systems, multiple interactions exist between these layers. Besides, spatial complexity occurs not only vertically but also horizontally. To capture that complexity, we adopted a joint sampling scheme to guarantee the compatibility of datasets for all investigated components (Leite *et al.*, 2016). Vegetation and litter sampling strived to capture plant influence zones, as suggested by Rhoades *et al.* (1996). The large tree aboveground biomass [AGB $\geq$ 10 cm diameter at breast height (dbh)] within each site was determined in the circular main plot (1,963 m<sup>2</sup>), and the minor vegetation and litter were determined in five subplots (25 m<sup>2</sup> and 1 m<sup>2</sup> for minor vegetation and litter, respectively). The topsoil (0–20 cm) was obtained as composite samples from the centers of the five subplots.

In CPA, the sampling scheme was adapted to the predominant regular spacing. Instead of a circle, six sub-plots (3 × 5 m) were used, in three quadrangular main plots of 25 × 25 m, similarly to Kato *et al.* (2009). The sub-plots and transects were sampled as above.

Large biomass components were assessed allometrically via diameter-based equations for mature rainforest trees (Overman *et al.*, 1994), secondary forest trees (Nelson *et al.*, 1999), lianas (Gehring *et al.*, 2004) and, when present, babassu palms (Gehring *et al.*, 2011), and also via conversions between the dbh and the diameter measured at a 30-cm height for smaller vegetation components (Gehring, Park, *et al.*, 2008). The following were distinguished: large vegetation (trees with dbh  $\geq$ 10

cm and palms >2 m high) ( $AGB \geq 10$  cm dbh); mid-sized vegetation (trees, shrubs, and lianas with dbh <10 cm, and palms <2 m high); and small vegetation (herbaceous and shrubs <1.30 m high). Small vegetation was estimated destructively and jointly with the litter layer. For statistical analyses, mid-sized and small vegetation were combined ( $AGB < 10$  cm dbh). The biomass of fallen logs in transects (Van Wagner, 1968) and standing dead logs in the circular main plots were quantified following the methods of Arevalo et al. (2002). The litter layer was destructively quantified and sampled in five  $1 \times 1$  m sub-quadrants within each sampling unit. The litter layer mass estimations distinguished between leaves and twigs. All the sampling was performed during the rainy season.

Soil bulk density (volumetric rings) and soil texture (via a pipette method) were determined as indicators of topsoil physical quality, according to the procedures of Klute *et al.* (1986). The following indicators of topsoil chemistry were evaluated: pH, determined via soil suspension in 0.01 M  $CaCl_2$ ; soil organic matter, determined by the Walkley-Black digestion method; the available P, estimated via extraction with a synthetic anion exchange resin Amberlite IRA-400; exchangeable K, determined via Mehlich I extraction; Ca and Mg, determined via KCl extraction; and H+Al, determined by the Shoemaker-McLean-Pratt (SMP) method. All procedures followed the specifications of Agronomic Institute of Campinas-IAC (2001).

### *Amplicon-based 16S and 18S rRNA gene analyses*

The soil subsamples (20 g) were stored at  $-80^\circ C$  prior to molecular analyses. Total soil DNA was extracted from 0.25 g of the soil using the Power Soil kit (Mobio, Carlsbad, CA, USA), following the manufacturer's instructions. To assess the impact of treatments on the bacterial community, the 16S and 18S rRNA genes were sequenced from total DNA extracted from the soil samples. The V4 region of the 16S rRNA gene was amplified using the archaeal/bacterial primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-

GGACTACHVGGGTWTCTAAT-3'), and the 18S rRNA gene marker was selected to investigate the fungal community using the primers FR1 (5'-AICCATTCAATCGGTAIT-3') and FF390.1 (5'-CGWTAACGAACGAGACCT-3'). The sequences were PCR-amplified using barcoded primers (Caporaso *et al.*, 2012).

The 16S rRNA gene amplifications for library preparation were performed using the C1000 thermocycler (Biorad, Hercules, CA, USA) with the following thermal conditions: 95 °C for 5 min; 35 cycles of 95 °C for 30 s, 53 °C for 30 s, and 72 °C for 60 s; and 72 °C for 10 min. A 25- $\mu$ l reaction contained 2.5  $\mu$ l of 10 $\times$  PCR buffer, 2.5  $\mu$ l of dNTPs (200  $\mu$ M), 0.25  $\mu$ l of each primer (0.1 pmol/ $\mu$ l), 0.2  $\mu$ l of FastStart Exp polymerase (0.056 U), and 1  $\mu$ l of DNA (0.6 ng). The 18S rRNA gene amplification reactions were performed using 5 micromolar of each primer, 2 mM dNTPS (Invitrogen, Carlsbad, CA), 0.5 microliters of BSA, 10 PCR buffer, 0.56 units of Fast Start Exp-Polymerase and 1 microliter of sample DNA template in a total reaction volume of 25 microliters. The PCR was conducted with initial incubation of 5 min at 95 °C followed by 25 cycles of 30 sec. at 95 °C, 1 min. at the annealing temperature of 57 °C, 1 min. at the extension temperature of 72 °C, followed by a final extension for 10 min. at 72 °C. The reactions were performed in duplicate and a negative control was included. The amplicon sizes were checked by gel electrophoresis. PCR products were purified using the Agencourt AMPURE XP system (Beckman Coulter, Brea, CA, USA) to remove primer dimers, quantified using Fragment Analyzer (Perkin-Elmer Corp.), and mixed in equimolar amounts for sequencing using Illumina MiSeq (Illumina Inc., San Diego, CA, USA).

Sequences of the 16S and 18S rRNA partial gene amplicons were processed using dada2 workflow (Callahan *et al.*, 2016) on a 32-node server running Linux Ubuntu 14.4. The forward and reverse primer sequences were removed from the FASTQ file of each sample using Flexbar version 2.5. Reads were filtered based on sequence quality by running the Sickle tool (minimum quality score of 25 and minimum length of 150). Taxonomic information for each ASV was added to the BIOM file using the SILVA 132 database. The sequences were deposited in ENA

database. In total, the sequencing resulted in 3,308,164 reads for bacteria and 3,334,355 for fungi.

### *Data analysis and statistics*

Studying the microbial community composition at high taxonomic resolution based solely on short reads is challenging. In order to provide a clear but still detailed analysis and avoid overinterpretation at species/ASV level, all the analyses of bacterial and fungal community were done at the taxonomic level of genus. The rarefaction curves for both bacterial and fungal communities are presented in Supplementary Figures S1 and S2.

The composition of microbial communities in soil is tightly connected with soil characteristics (Cassman *et al.*, 2016), nutrient availability (Delgado-Baquerizo *et al.*, 2017; Pan *et al.*, 2014), and plant biomass (Aponte *et al.*, 2013). These parameters are in turn connected with land use and management practices (Barnes *et al.*, 2014). These relationships make use of environmental variables as predictors of the microbiome prone to bias towards collinearity and overfitting (Dormann *et al.*, 2013). To circumvent this problem, we adopted the Generalized Joint Attribute Modeling (GJAM) (Clark *et al.*, 2017). This model allows one to include variables of different types and to analyze them jointly, thus revealing the regression coefficients of the effects of different land-uses in the relative abundance of taxa within the soil microbiome, as constrained by the compositionality (Gloor *et al.*, 2017), the aboveground biomass and soil factors. Subsequently, we performed a hierarchical clustering analysis (ward distance) of the different regression coefficient to identify similarities in the responses to the land-use. GJAM also allows the possibility to include random effects, which accounts for the within site replicates. Another advantage of the GJAM approach is the possibility to simulate scenarios for some specific set of dependent variables. This tool was adopted to simulate a scenario where all the land-use systems (agroforests and secondary regrowth) have

the same microbiome found in the mature forests. The intention here was to compare how much the soil factors and plant biomass should differ from the original value recorded during the measurements in each sampling site to achieve that specific microbiome. The level of change for each variable was summarized as a ratio of change (the ratio between the simulated value and the original value found in each site).

## Results

### *Interdependency between soil microbiome, plant biomass and soil factors across land use systems*

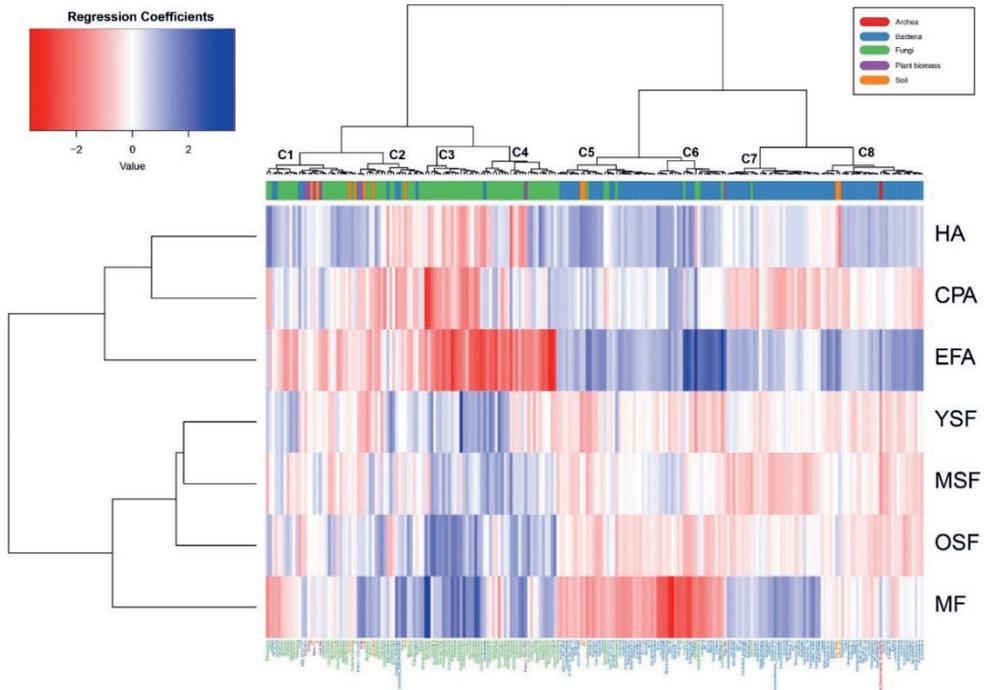
Mature forest (MF), without any visible human perturbation, was used as a tropical rainforest standard to which we compared three different agroforestry practices, as well as changes associated with a gradient of natural secondary succession following slash-and-burn agriculture. GJAM analysis was used to determine the effect of the different land-use systems on the microbial community, plant biomass and soil factors, and cluster analysis of the resulting regression coefficients allowed us to identify groups of land-use with similar characteristics and groups of response variables with similar shifts according to the land-use (Figure 1). First, we observed that the system with the most similar characteristics across sites was the secondary succession (YSF, MSF, and OSF). These natural regrowth systems are characterized by an increasing number of Ascomycota and Basidiomycota fungi (Figure 2) that matches with the mature forest (MF). Another cluster grouped the home-gardens (HA) with the commercial plantation agroforests (CPA), and the enriched agroforests (EFA), likely due to reduced proportion of some specific genera from fungi (*Ascomycota*, *Basidiomycota*) and bacteria (*Proteobacteria*), especially as compared to MF.

Interestingly, the MF sites showed more similarities with secondary forest regrowth than the agroforestry systems. In general, the natural forest exhibited a high proportion of different genera of *Ascomycota* (YSF = 8, MSF = 8, and OSF = 16),

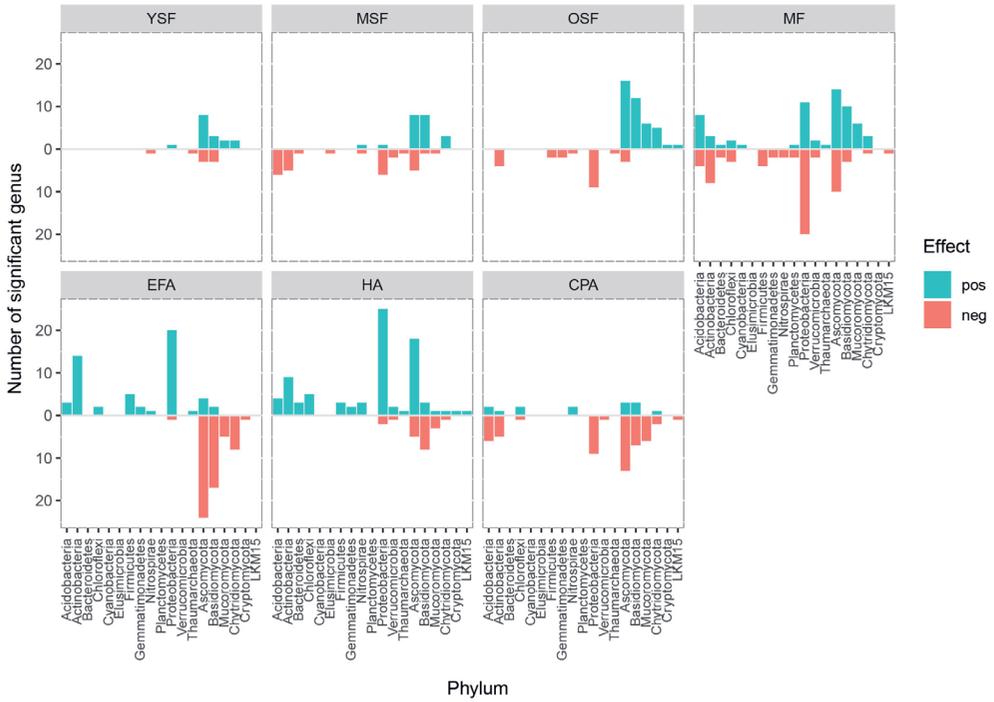
*Basidiomycota* (YSF = 3, MSF = 6, and OSF = 9) (Figure S1) and the fungal phylum *Mucoromycota* (YSF = 2 and OSF = 6). Within the *Mucoromycota* phylum, we found two unclassified genera of mycorrhizal fungi (one classified as *Glomeromycetes* and the other as *Glomerales*). These two genera were significantly more abundant in MF and the secondary succession forests (OSF and YSF) as compared to the agroforestry systems. In contrast, the agroforestry systems had a higher relative abundance of numerous bacterial genera (EFA = 44, CPA = 7, and HA = 53), yet only few fungal taxa (EFA = 6, CPA = 7, and HA = 25). The bacteria belonged to the phyla of *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Chloroflexi*, *Chytridiomycota*, *Planctomycetes*, *Proteobacteria*, *Thaumarchaeota*, and *Verrucomicrobia*.

The collective changes in plant biomass and soil factors also contributed to the distinction between the land-use systems (Figure 1 and S1). Overall, the MF systems had the highest values for total aboveground biomass (TAGB), followed by the OSF and HA. We also found that the coefficient of aboveground biomass shifts from negative to positive from the YSF to the OSF, suggesting a gain of plant biomass along the natural regrowth. Our model also revealed which microbial genera exhibited a similar response to the shifts found in the variables of plant biomass (*Trichoderma*, *Apiotrichum*, *Mortierella*, *Geastrum*, *Myrothecium*, and *Glomeromycetes*) that clustered together in C3 with the variables of total aboveground biomass (TAGB), living aboveground biomass (LAGB), and the biomass of plants >10cm dbh (the group of plants that represent the canopy) (Figure 1). Overall, we found that many genera of fungi were associated with the variables that described the plant aboveground biomass (leaf litter, dead logs, and living aboveground biomass) while bacterial community responses generally tracked trends found in the soil factors (Ca, Mg, soil organic matter). However, we also observed bacterial taxa that were associated with the changes in twigs biomass (cluster C8). Other associations with soil factors included changes in pH and soil P-content generally being associated with shifts in two genera of uncultured

*Bacteroidetes* (one from the family AKYH767 and the other from the family *Cytophagaceae*).



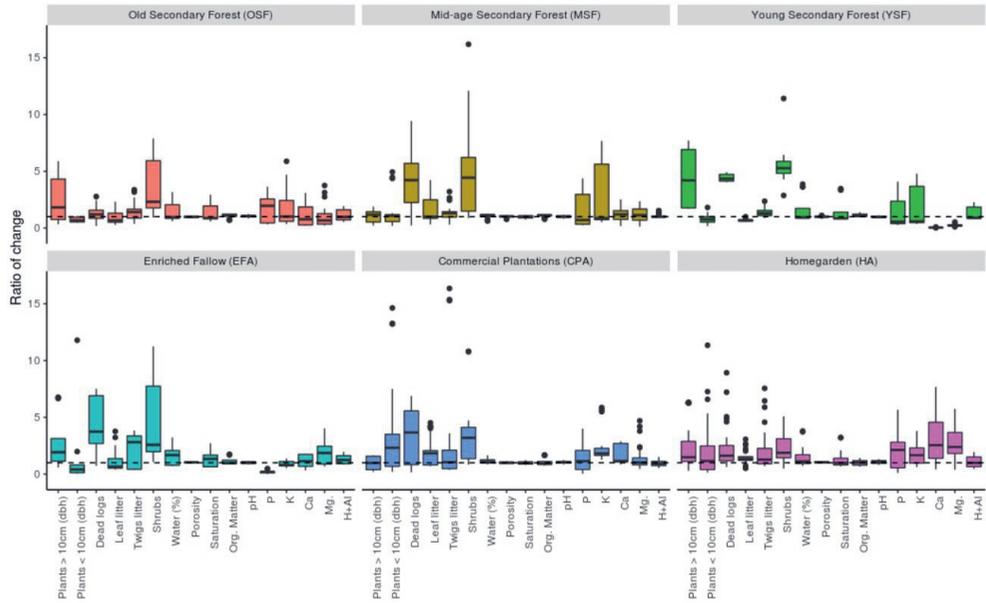
**Figure 1** – Heatmap of the regression coefficient showing the shifts in the soil microbiome, plant biomass, and soil factors. MF, mature forest (>100-yr-old, 12 sites); SFO, old (>30-yr-old, 7 sites) secondary forest; SFM, mid-age (15–20-yr-old, 6 sites) secondary forest; SFY, young (5–12-yr-old, 5 sites) secondary forest; EFA, enriched fallow agroforestry (6 sites); HA home-garden agroforestry (13 sites); CPA, commercial plantation agroforestry (7 sites). Blue and red refer to positive and negative values for the regression coefficients, respectively.



**Figure 2** – Number of genera that exhibited positive (blue) or negative (red) significant regression coefficients within the phyla for bacteria, archaea, and fungi according to the different land use systems. MF, mature forest (>100-yr-old, 12 sites); SFO, old (>30-yr-old, 7 sites) secondary forest; SFM, mid-age (15–20-yr-old, 6 sites) secondary forest; SFY, young (5–12-yr-old, 5 sites) secondary forest; EFA, enriched fallow agroforestry (6 sites); HA home-garden agroforestry (13 sites); CPA, commercial plantation agroforestry (7 sites).

The joint analysis of the microbiome, the aboveground biomass, and the soil factors allowed us to generate simulated scenarios to evaluate how those variables would shift under a specific condition. Since our goal was to investigate the capacity of the agroforestry systems to mimic the aboveground-belowground interactions found in mature forests, we simulated a scenario where all the different land use systems have the same microbiome: the estimated microbiome of the mature forests (MF). With that, we let all the variables of plant biomass and soil factors free to vary and obtained the expected values of aboveground biomass and soil factors where each one of the agroforestry systems and the successional forests possessed the same

microbiome estimated for the mature forest (MF). This analysis allowed us to identify the site-specific variables that should change in order to achieve the same microbiome found in MF. The results from these simulations allowed us to determine the ratio of changes from each system (Figure 3).



**Figure 3** – Ratio of changes from the simulated scenario where all the land-use systems mimic the same estimated microbiome found in the mature forest. The closer the value is to one (horizontal dashed line in the graph) the closer the system already resembles the microbiome estimated for the mature forest (MF). SFO, old (>30-yr-old, 7 sites) secondary forest; SFM, mid-age (15–20-yr-old, 6 sites) secondary forest; SFY, young (5–12-yr-old, 5 sites) secondary forest; EFA, enriched fallow agroforestry (6 sites); HA home-garden agroforestry (13 sites); CPA, commercial plantation agroforestry (7 sites).

The secondary forests (YSF, MSF, and OSF) showed the lowest ratios of change. For the OSF, the biggest changes should occur in the biomass of shrubs with a median increase of 2.32 times the original value, and with an interquartile range (IQR) of 1.78-5.94, followed by the biomass of plants > 10cm (dbh) (median = 1.82, IQR = [0.75-4.32]). For the MSF, the microbiome would require bigger changes in

the shrubs (median = 4.44, IQR = [1.48-6.23]) and dead logs (median = 4.22, IQR = [2.25-5.7]). Finally, for the YSF, the microbiome would require mainly an increase in shrubs biomass (media = 5.28, IQR = [4.82-5.88]), deadlogs (media = 4.35, IQR = [4.14-4.74]), and plants > 10cm (dbh) (median = 4.2, IQR = [1.78-6.92]).

Interestingly, all the agroforestry systems had very large ratios of change, especially for the variables of plant biomass. Our simulation suggests that the EFA systems required a median of 3.74 (IQR = [2.7-6.9]) more dead logs than found within the sites, as well as 2.81 (IQR = [0.43-3.36]) times more litter mass, and 2.58 (IQR = [1.97-7.75]) times more shrub biomass. Similarly, the CPA systems required similar increases in the dead logs (median = 3.67, IQR = [0.84-5.56]), shrubs (median = 3.2, IQR = [1.36-4.12]), and olants < 10cm (dbh) (median = 2.32, IQR = [0.69-3.5]). Finally, the HA showed the lowest medians in the ratios of changes, but many more variables with ratios of change bigger than 1. From all the three agroforestry systems under investigation, the home-garden agroforestry showed the lowest ratios of change, but still would require many adjustments in the plant and soil factors in order to achieve the microbiome of a mature forest.

## Discussion

An increasing land use pressure does not allow for strategies relying purely on natural secondary forest succession. Agroforestry represents an alternative land use strategy that can potentially provide crop, fruit and wood production while allowing for ecosystem development. However, our study shows that this capacity to mimic the complex interactions found in the mature forest is not realized by the agroforestry practices in terms of soil microbiome development. The soil microbiome in the mature forest and secondary successional forests harbored more fungal-driven communities, whereas agroforestry systems consisted of mainly bacteria-driven microbiomes.

Our results provide evidence that the soil microbiome did not fully recover after 30-60 years of secondary regrowth. Studies suggest that forest structure and species composition on shifting cultivation fallows recover faster than on other abandoned agricultural lands, such as pastures and monoculture plantations (Ferguson *et al.*, 2003; Poorter *et al.*, 2016). However, it should be emphasized that fallows do not completely revert to the original forests, since not all parameters recover at the same speed. This is especially true for plant diversity, which seems to take longer than plant biomass recovery.

We reported clustered responses of the *Glomeromycetes* and the variables of aboveground biomass, but not much with the available P-content in soil (different cluster). Mycorrhizal fungi strongly affect plant population and community biology (Tedersoo *et al.*, 2020), and our results suggest that there is a stronger dependence between AMF and aboveground biomass than towards the available P-content in soil. A possible explanation for that relates to the capacity of mycorrhizal fungi access less available sources of P in soil (Guo *et al.*, 2016). We also found *Glomerales* that was more prevalent in the mature forest as compared to the agroforestry systems, with the most pronounced difference found for CPA systems. One important contribution of the mycorrhizal fungi is to form the common mycelia network that plays a crucial role in plant population dynamics, diversity, and coexistence (Tedersoo *et al.*, 2020). By favoring more groups of mycorrhizal fungi, the mature forest and secondary succession may have developed more interactions between the aboveground and belowground communities than AFS. On the other hand, the changes in abiotic and biotic environments as a result of the agroforestry practices altered the costs and benefits of mycorrhizal associations and shifted the competitive balance offered to the plants.

The increase in the regression coefficient of TAGB from YSF to OSF and its similarity with the MF reflected the potential resilience of tropical forests to slash-and-burn disturbance. Likewise, it has been previously reported that aboveground biomass (AGB) recovers more quickly than plant biodiversity in tropical forests

(Martin *et al.*, 2013). Such recovery along natural forest secondary succession points to the large carbon sequestration potential of secondary forest regrowth, with an estimated gain of 8.48 petagrams of carbon (Pg C) for low-cost natural regeneration (Chazdon *et al.*, 2016). Further, considering only the AGB and soil factors, SFO (>30 yr) clustered together with MF. Our results show that the soil microbiome along secondary succession also seems to recover more in the direction of the mature forest microbiome. Remarkably, the agroforestry systems (EFA, CPA, and HA) clustered together and differed markedly from the secondary forest successional trajectory. This was likely an outcome of the increased importance of bacterial communities in the agroforestry systems followed by the changes in the understory biomass [plants <10 cm diameter at breast height (dbh)] and in the soil nutrients, mainly high K levels and a tendency toward low Mg levels.

By showing the importance of soil microbial diversity for ecosystem multifunctionality, Delgado-Baquerizo *et al.* (2020) highlighted the value of including soil biodiversity in the political and management agenda to protect the functioning of terrestrial ecosystems worldwide. Our results contribute to this discussion. Even for so-called low impact land-use practices, it is a major challenge to develop a microbiome structure similar to that observed for mature forest targets sites, and this goal may be unattainable for commonly used agroforestry practices.

By modeling the soil microbial community jointly with the plant biomass and soil factors, we moved beyond the mere identification of the impacts of each land-use system. Our results allowed us to explore potential scenarios of land-use systems that might mimic the soil microbiome of a mature forest. Our study therefore provides strong evidence of multiple associations between the soil microbiome in aboveground-belowground interactions. Overall, our findings revealed the relative importance of the aboveground biomass and soil factors in contributing to shaping the soil microbiome for each of the different land-use systems. Our model results suggest that several fold changes in key soil factors and plant biomass would be

necessary to support a target microbiome in each of the agroforestry systems examined in our study.

Despite the fact the secondary forest succession presented a more similar microbiome than the agroforestry systems, the microbial communities were not recovered in our study after more than 30 years of secondary regrowth (OSF systems). According to Bonner *et al.* (2020) two phenomena limit the recovery of the soil microbial community: (1) indirectly through deficiencies in the ecological conditions provided by plants or (2) directly via barriers to dispersal that hinder assembly of a microbial community characteristic of reference forest. By looking at the ratio of changes from the set of plant and soil variables measured in our study, we highlight the importance of plant biomass and soil factors and confirmed that ecological conditions provided by the plants are crucial for the rainforest microbiome.

Apparently, agroforestry practices created new habitat conditions that fostered a microbial community (mostly bacteria) that was divergent from those in the mature forest soils. Our findings suggest that these communities may be difficult to displace, making it difficult to restore aboveground-belowground interactions (plant biomass and soil factors) and related functionality. The main drivers of these changes were the loss of plant biomass (shrubs and littermass), increasing soil nutrient content (P, K, Ca, and Mg), and other variables indicative of increased land-use intensity (e.g., pruning, weeding, clearing the area for increasing the light in the understory, nutrient input in CPA systems). Notably, other factors tested here such as pH, and potential soil acidity had little effect on soil microbial community composition.

Manzoni *et al.* (2012) showed that the presence of plant residues that are chemically too homogenous to promote functionally diverse microbial community. Therefore, by selecting plants only based on their cash value, agroforestry systems may have detrimental effects on the new microbiome, for instance by reducing the role of fungi in linking above- and belowground elements of the ecosystem.

Our findings provide observational evidence that the soil microbiome is critically important for maintaining ecosystem function across secondary regrowth forest and in the mature forest, particularly the increasing importance of mycorrhizal fungi in association with aboveground plant biomass. We also show that in order to mimic a mature forest soil microbiome, agroforestry practices must be adjusted to better host this complex and multifunctional microbiome. Our multi-faced approach suggests that multiple interactions in the ecosystem are supported by the soil biodiversity that should not be overlooked by the low-impact agroforestry practices, as they probably have key roles for system sustainability. Increasing knowledge of the soil microbiome could provide an emerging cornerstone for sustainable agricultural practices and become a key component of decision making for sustainable tropical land-use management.

Although agroforestry systems are thought to mimic actual natural or secondary forests, it appears to be important that such strategies also take soil microbial communities into account, especially for more intense agroforestry management practices (such as commercial plantations and home-gardens). These types of ecosystems still fail to maintain or achieve the strong aboveground-belowground interdependence found in mature rainforests. Consequently, agroforests systems may claim to be a low impact land use for the aboveground systems, but such management practices interfere with aboveground- belowground linkages that may compromise system sustainability. Increased knowledge of the impacts of land use systems on soil microbial community structure is needed to allow for more efficient use of existing agricultural land and to balance ecological and economic goals.

### *Land use effects on plant-soil-microbe interactions*

Our findings suggest that human impact via agroforestry systems reduced the interdependence of the soil microbiome from the vegetation and soil. This stems from a reduction of plant-soil interactions caused by land management (nutrient

input, pruning, weeding, etc.), as previously reported by us (Leite *et al.*, 2016), which reflects the efforts in regulating ecosystem productivity towards autoconsumption or market-related products. With our joint model, we presented evidence that the complexity of the soil bacterial community is primarily governed by soil nutrients, and the fungal community relates is more associated with variables related to plant aboveground biomass (AGB). This extends the findings of previous reports on Scotland (Delgado-Baquerizo *et al.*, 2017) to a tropical region, by disentangling the role of soil nutrients from pH as the driver of population variability. Consequently, the management strategies in the agroforestry systems should focus on enhancing plant-soil interactions to better mimic the interdependence observed in mature rainforests.

Collectively, this multisite study suggested that changes in land use, whether agriculturally manipulated or as a natural succession in response to shifting cultivation, cause consistent alterations in the tripartite plant-soil-microbe interactions. The design of complex and sustainable agricultural systems in the tropics remains challenging, especially in light of the ecosystem balance of plant-soil-microbe interactions. Soil and climatic conditions in the humid tropics challenge the development of continuous crops and require the design of agroecosystems capable of maintaining aboveground-belowground interactions.

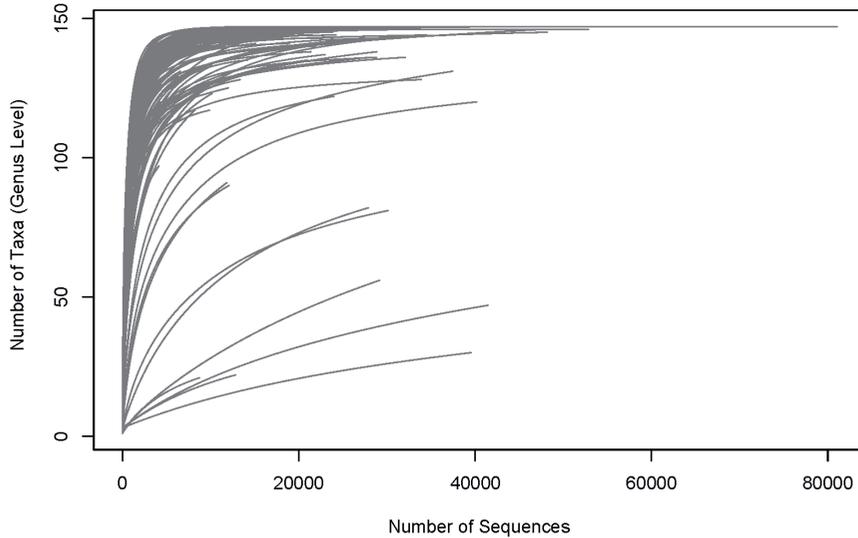
## Conclusions

We focused on the microbiological aspects of land-use changes and combined those with the plant biomass and soil factors to build a more complete framework on the impact of anthropic management in the aboveground-belowground co-dependence within the Amazon biome.

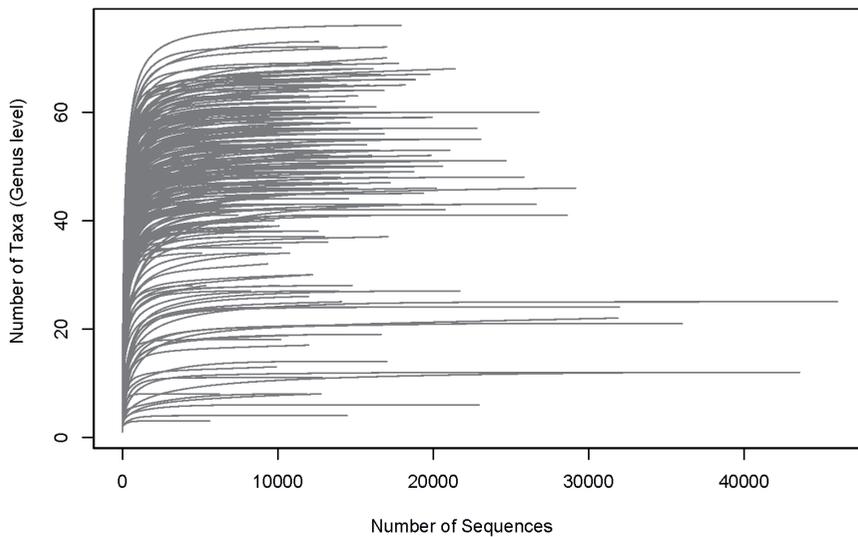
We conclude that agroforestry practices that help to improve aboveground structure are not sufficient to promote belowground communities towards those found in mature rainforests. Mature forest and secondary succession stages (YSF, MSF, and OSF) favored more fungal groups, and the high abundance of mycorrhizal

fungi in the mature forest suggests a key role played by these fungi in the cycling of phosphorus in such tropical ecosystems. In contrast, the agroforests were characterized by higher relative abundance of certain bacterial taxa. These systems thus apparently do not serve to mimic true tropical forest features with respect to promoting above-belowground linkages. Even though natural secondary succession appeared to show a transition toward the mature forest microbiome, which was not that case for agroforestry, we still observed marked differences in the secondary forest after more than 30 years of laying fallow. Full recovery of aboveground and belowground components of tropical rainforest systems after slash and burn agriculture thus requires either additional measures or time.

## Supplementary Results



**Supplementary Figure S1** – Rarefaction curve of the bacterial community for the subsamples from the 56 study sites in the Eastern Amazon.



**Supplementary Figure S2** – Rarefaction curve of the fungal community for the subsamples from the 56 study sites in the Eastern Amazon.



## CHAPTER 6

### General Discussion

This thesis assessed the impact of land-use changes on aboveground and belowground interactions in the Brazilian Amazon region. To this end, I integrated multiple approaches across different soil land use systems. I combined allometric estimation of plant biomass and soil physico-chemical analysis with next generation sequencing technology of the soil microbiome to assess the shifts in aboveground-belowground interactions as a result of land-use changes.

In this chapter, I discuss:

- a) The methodology used to explore the total microbial community, including sampling strategy, high-throughput sequencing and the generalized joint attribute model that allowed me to assess microbial community shifts and identify keystone microbial groups;
- b) The importance of soil type and plant biomass in shaping the structure of microbial communities and how they determine their interactions;
- c) How land-use changes, through the recovery of natural vegetation (from young secondary forest to mature rainforest) and introduction of anthropogenic uses (agroforestry systems), may affect the structure of the microbiome across the Amazon region;
- d) Ideas and directions for future studies on land-use changes in the Amazon ecosystem.

## **Methodological approaches**

The field studies in Chapters 2, 3, and 5 can be classified as observational in which randomization is restricted solely to selecting samples from the study sites of interest. In Chapter 4, I presented a summary of some classical approaches of microbial data analysis, the distanced-based methods. I then presented some practical examples of the application of these methods to both real and simulated datasets. Here, I highlight that every statistical approach possesses its assumptions

and neglecting these may lead to a biased analysis that is not so easy to recognize in the final output of the data. In this chapter, I also suggested the advantages of adopting model-based approaches to circumvent several challenges imposed by the nature of the high-throughput sequencing (e.g., compositionality constraint, high number of zeros, and overdispersion).

My field study provided a unique opportunity to investigate the effect of land use changes on aboveground-belowground interactions in areas that are otherwise nearly identical in terms of physiography and microclimate characteristics. Thus, the sampling strategy used in Chapters 2, 3, and 5 provided an opportunity to investigate the effects of land-use changes on plant community, soil factors, and soil microbial community in the Amazon biome. Interactions between plants, soil factors, and soil microbes can be positive, neutral, or negative. The full set of relations represents the aboveground-belowground interactions.

## **Factors that drive the spatial arrangement of plants during secondary succession**

6

In Chapter 2, we took spatial vegetation patterns as a way to detect whether or not plant-plant interactions changed along secondary succession and compared it with the mature forest. To achieve that, we select the group of leguminous plants as a model to identify these shifts. Then, by looking at the similarities of the spatial arrangements between leguminous plants, we could evaluate how close from the mature forest each of the successional stages are placed. This allowed us to identify the fallow timespans and the impact of land use intensification.

The study of plant spatial patterns unraveled how plants interact with each other and shape the surrounding diversity. To date, few studies investigated plant spatial patterns according to their functional traits. Our results contributed to fill this gap by analyzing plant traits combined with plant diversity. Leguminous trees and lianas were all clustered at small scales, confirming spatial studies from throughout

the world, including eastern African savanna trees (Dohn et al. 2016), temperate forests of China and the USA (Wang et al., 2011), and tropical forests of Sri Lanka (Wiegand et al. 2007b). Our results also confirm previous findings on a stronger clustering of lianas relative to trees (Clark and Schnitzer, 2018). This approach provided insightful information not only about species strategies, but also fundamental ecological processes and the impact of anthropic land cover changes in Amazon forests.

Spatial patterns of legumes were random beyond the initial  $\pm 5$  m and changed along succession, pointing to a shift in the ecological importance of this plant group. Our power to detect and correctly reflect bivariate species interactions is expected to be high because of the high number of leguminous plants and the different sampling sites involved (Rajala et al., 2019). We take leguminous trees and lianas as taxonomic (Martins et al., 2018) and functionally (Clark and Schnitzer, 2018) similar groups resulting in similar spatial distribution patterns between them, though we have to expect some blurring of spatial patterns within these groups on smaller taxonomic scales not covered in this study.

A second cycle of slash-and-burn agriculture showed to be enough for inducing strongest associations between  $N_2$ -fixing and non- $N_2$ -fixing plants, with a weaker effect of the surrounding diversity. This result might indicate that after slash-and-burn plants tend to associate more positively or neutrally, which might be especially prominent for leguminous plants given their ability to procure fixed nitrogen – a scarce nutrient after slash-and-burn agriculture. We provide evidence that functionally distinct plants ( $N_2$ -fixing vs non-  $N_2$ -fixing) tend to co-occur as a result of land use intensification. These results provide evidence to support the stress gradient hypothesis (SGH), which suggests that under stress environmental conditions communities may shift from net negative to net positive interactions.

Our target liana (*Machaerium* spp.) consistently reduced tree diversity for all the successional stages. This result occurred presumably by a combination of above and belowground competition, e.g., through mechanical stress (Schnitzer and

Carson 2010). Competition with lianas may limit trees establishment and slow forest recovery. Ultimately, this may worsen the slash-and-burn degradation cycle.

The spatial distribution of sessile plants by themselves or in interaction with others is the outcome of a wide range of ecological processes over longer periods of time. Seed dispersal limitation results in clustered univariate distribution (Seidler and Plotkin, 2006; Obiang et al., 2019), and clonality likewise causes strong univariate clustering (Shouten and Houseman, 2019). Density-dependent mortality should affect plant occurrence in creating more uniform distribution patterns further along succession (Larson et al., 2015) or with increasing plant size (Nguyen et al., 2014). Plant survival (traditional O-ring statistics) also will vary according to the variation of edaphic factors such as chemical soil fertility (Vleminckx et al., 2017), topography (Punchi-Manage et al., 2014) or water availability (Browning et al., 2014). Species-specific differences in such edaphic associations indicate niche segregation and thus, presumably reduced interspecific competition.

The *Cecropia* genus include pioneer trees that usually dominate the seedbank and the earlier stages of secondary succession (Uhl et al., 1988; Villa et al., 2019). Its less dense wood allows for rapid growth, which provide an advantage in the competition for sunlight (van Breugel et al., 2012). Our results support the view that at earlier stages, *Cecropia* dominates its surroundings and starts to mix with other different species later, likely as an outcome of new species being established in the understory. We also found that, at earlier successional stages, non-leguminous trees genera *Cecropia* and *Goupia* are stronger repellers of plant diversity than leguminous plant as *Inga*. *Inga* spp. play a crucial role forest recovery due to their vast root system and potential for nodulation and the symbiotic fixation of atmospheric nitrogen (Scoti et al., 2019). This supports the view that leguminous plants play a crucial role in the initial stages of secondary succession. In summary, our analysis of spatial diversity revealed that plants change their influence on the surrounding diversity, but this capacity depends on plant traits and growth strategy.

At younger stages of secondary succession, leguminous trees and lianas co-occur formed spatial clusters of plant individuals. Such co-occurrence is most likely caused by the occupation of similar ecological niches, which would result in direct competition (Wilson, 2011; Soliveres and Mestre, 2014). Facilitation between plants provides a second alternative. Facilitation consists of a niche expansion as a result of positive interactions with one or more different species. Many species facilitate neighboring species by modifying the local environment i.e., trees cast shade on the forest floor, and leguminous plants increase soil nitrogen levels, thus allowing other plant species to grow in the soil otherwise unsuitable. We explored the possibility of facilitation by legume trees and lianas (via biological nitrogen fixation (BNF) in N-limited landscapes) via legume-fix bivariate analysis and the diversity shifts at different spatial scales.

Under land use intensification legume trees and lianas are more clustered among themselves and are strongly segregated, suggesting niche separation for the avoidance or/and reduction of competition. However, this segregation quickly declines at older successional stages. Legume species aggregation is expected to decrease along forest regrowth, as species clustering has been shown to promote species coexistence (Pedersen and Guichard, 2016). The dominance of the leguminous plants is bound to further increase throughout degraded Amazonia with each new burn cycle and with each new shifting-cultivation cycle, the spatial patterns of the remaining vegetation are predicted to reorganize to become more clustered and segregated.

Chapter 2 advanced our knowledge of how plant spatial interactions reflect the sum of all potential interactions with the environment. However, our analysis of spatial interactions did not include the interactions with the soil factors and the plant biomass. Plant biomass is the major reservoir of the aboveground-belowground interactions within the tropical ecosystem such as the Amazon Forest. In the following Chapter 3 of my thesis, we begin to understand the other potential factors that impact plant-soil systems. I investigated the role of plant biomass and its

dependence on soil factors to better understand how soil factors can explain differences in plant biomass. Moreover, I expanded the land use systems by including agroforestry practices.

## **The impact of agroforestry practices above- and belowground**

Trees has been widely recognized for their impact on soil nutrient levels due to their low-cost-input in shifting cultivation (Døckersmith *et al.*, 1999) or by feedback mechanisms in agroforestry systems (Aponete *et al.*, 2013; Brandt *et al.*, 2013). Hence, trees are the main drivers of increasing soil fertility (Rhoades, 1996) in such systems. Plants and soils interact via a continuum of multiple individual relationships between components of vegetation, the litter-layer and (top) soil (van der Putten *et al.*, 2013).

In Chapter 3, we observed (i) a negative logarithmic relationship between the aboveground biomass of large and small trees, presumably caused by competition for light and soil resources between these two components (Battaglia *et al.*, 2002; Beer *et al.*, 1998; Schwinning *et al.*, 1998); (ii) litter-layer biomass affected by small vegetation components, but not by the large trees >10 cm dbh ; (iii) topsoil (in our case 0-20 cm) OM content affected by twig but not by leaf litter mass; (iv) litter mass related to a widening litter C:N-ratio, and to reduced 'plant-available' P and K contents in the topsoil (McGrath *et al.*, 2000); and (v) some relationships within soil, what somehow was expected, such as the positive relationship between clay content and soil organic matter (McGrath *et al.*, 2001; Six *et al.*, 2002), and the negative relationship between soil OM and bulk density (Soane, 1990).

Management strongly influences the plant-soil interactions, which could partially explain the lack of clear and detectable relationships between these two components. One conspicuous result of our study is the absence of any significant relationship between topsoil OM content and nutrient concentrations or soil pH. In the Commercial Plantation Agroforests (and to a lesser degree in Fallow Enrichment

Agroforests), this can be attributed to liming practices (CPA only) and localized synthetic fertilizer inputs. The low values of AGB<10 cm dbh and litter biomass in the homegardens likely reflect the management practice of keeping the undergrowth and soil surface clean around the houses. Other significant pathways of plant-soil nutrient interactions such as root exudates, stemflow and throughfall are likewise to be expected, as reviewed by Rhoades (1996).

At this point, the co-inertia analysis helped us to understand the multivariate relation between aboveground biomass and soil indicators in a way that was not possible for the univariate analysis, separating the forests systems and helping to understand that soil is more strongly related with biomass in naturally regenerating forests than in agroforests systems. Furthermore, we were able to identify the systems that yielded plant-soil discontinuity, the commercial plantations and homegardens. As homegardens have a low disturbance fire regime, they may create a different route of nutrient flux and deserves further investigations.

The results of PCA supported the analysis of physico-chemical indicators, which showed strong statistical differences between the commercial agroforestry systems and the natural fallows and mature forests. Based on PCA, the irregularly spaced agroforestry systems are closer to mature forest with respect to soil resources and aboveground biomass.

As we found a grouping of the irregularly systems, we could suggest that there is some transition in the agroforestry systems studied, and this changing is mainly influenced by the physico-chemical characteristics of soil and the litter-layer biomass.

In this Chapter, we confirmed the role of aboveground biomass in both natural and managed ecosystems. The mature forests exhibited the strongest dependence, especially for the taller trees (DBH > 10). For those plants, where it is the most important characteristic of this environment and in the commercial plantation agroforestry, since it's the growing of the crop plants which drives all the

---

management practices.

High variability within each system, as revealed for our PCA and BCA analyses, indicates the potential to make improvements in fields with the lowest indicator values, while differences between systems suggest the potential to mitigate negative impacts of land use changes by combining plots with contrasted functions in a strategically designed landscape mosaic (Lavelle *et al.*, 2014). The separation of commercial plantations agroforests systems from the other systems and the natural succession might be explained by the high degree of intensification of management practices. The commercial plantations agroforest systems studied here represent strategies with relatively low- costs and strong profits, according to Cardozo (2013).

Beyond local system alternatives, limits to large-scale land conversion should be set, while taking into consideration the agricultural territory at the landscape scale (Lavelle *et al.*, 2014; O'Farrell *et al.*, 2010). Lovell and Johnston (2009) suggested that to improve ecological performance and service delivery it is necessary to increase spatial heterogeneity. Based in our findings, we can provide suggestions for the landscape design for eastern Amazonia with respect to agroforestry systems: (i) land use should apply different levels of the system management observed to construct local land use gradients, from the more conservative use to the most market-oriented management and (ii) use this landscape planning to mitigate anthropogenic impacts under and near primary forests through national policies.

Trees in agroecosystems can be present in an infinite number of arrangements and species combinations as I detailed in Chapter 1. But for managed systems such as agroforests that arrangement depends mostly on the farmers' objectives as well as the environmental characteristics of the region (Nair, 2013; Pinho *et al.*, 2012). This creates an enormous possibility to establish spatial heterogeneities and construct not just efficient and productive systems, but also to design sustainable landscapes.

Vegetation structure contributes to enhanced soil nutrients, mainly via litter deposition (Jensen, 1993). The concentration of K under homegarden management was higher than the natural fallow and conserved forest. A high potassium level in homegarden soils was previously reported (Jensen, 1993; Pinho *et al.*, 2010). However, this relation seems to be more complex, since, according to our results, the litter showed a negative relation with main soil nutrients. The contribution of the litter layer to soil fertility improvement depends on numerous factors, including the quantity, quality, timing, and placement of litter (Van Noordwijk *et al.*, 1995). For the special case of homegardens, another important factor to consider is the traditional practice of sweep-and-burn of the littermass, as reported for Pinho *et al.* (2010). The frequent burning of the litter layer in small piles in different parts of the homegarden is different from the intensive fires used in slash-and-burn systems, which may affect the extent of leaching. These little piles of burning organic materials are common practice, in what Winkleprinks (2009) refers to as a sweep-and-burn approach. This practice has been suggested as one of possible explanation for the fertility of Amazonian anthropogenic dark earths (Glaser *et al.*, 2012).

Our study identifies homegardens agroforestry as an efficient and promising land use system, both in terms of soil use resources and vegetational biomass stocks. Such homegardens have been developed independently by many different cultures and farmers throughout the world (Nair, 2001), especially in tropics. Even though they still are definitely under-researched (Kumar *et al.*, 2004) and have been main separated from systematic agronomic improvement efforts (R. P. Nair, 2014), these systems persist to today and, thus, constitute a remarkable success story both in ecological and economic terms.

Soil improvement under agroforestry systems is to a great degree related to anthropogenic different practices of soil and aboveground management, mainly in the form of surface litter, whether keeping it for land cover or burning for accelerate nutrients cycling.

The sustainability in the tropical biomes relies on the development of

---

agricultural systems with optimized nutrient cycling and continuous production with reducing dependence of external inputs. To achieve this, it is crucial to understand the environmental contribution of each management strategy in order to adopt better strategies of landscape design of heterogeneities and establish policies to protect the conserved environmental and improve the sustainability of land use management systems.

Besides considering the role of dependence between plant biomass and soil factors interactions, I saw the need to include the role of soil microbiome. Therefore, in the next chapters of my thesis, I expanded the sampling design in Chapter 3 and also include metagenomic approaches to characterize both bacterial and fungal communities from secondary successional forest and agroforestry systems.

## **Land-use impact on plant-soil-microbiome systems and interactions**

For the microbial community, I choose the explore the bulk soil because this represents the microbial pool. However, prior to going straight to the output of metagenomics analysis, I started with a reflection on the limitations of current methods used in microbial community analysis in Chapter 4. Based on that, I argue that the principles of hierarchical model and latent variables promote a unified approach to a more accurate understanding of the shifts in the aboveground-belowground systems, because it allows one to circumvent several limitations from the sequencing technologies (Gloor *et al.*, 2017). At the same time, it offers a well-developed approach to combine community data with soil factors and plant variables. These conclusions, prompted my use of a generalized joint attribute modeling approach as the most suitable one for handling the complex relations between plant biomass, soil factors, and the microbiome under different land use systems.

In Chapter 5, I assumed that mature rainforest is the best representation of

long-term soil sustainability in the tropics. In mature forest there is an intense aboveground-belowground interactions that helps to maintain nutrient cycling in nutrient-poor tropical soils. Then, I considered the soil microbiome as an important component of agroforestry practices that aim to mimic the ecology of mature forests. Observations presented in the current study also indicate that different agroforestry systems alter the aboveground-belowground interactions and, despite the efforts to mimic the ecological structure of tropical forests, agroforestry management still reduced plant-soil-bacterium interactions. This reduction may compromise the sustainability of agroforestry, especially as related to nutrient cycling.

With this in mind, I used modelling approaches to estimate what plant and soil variables could be managed to convert the soil microbiome to the same community found in mature rainforest. To my knowledge, this is the first attempt to design management practices devoted to achieving a specific soil microbiome.

The need of the natural fallow after a slash-and-burn cultivation is one of the main characteristics of slash-and-burn systems. As this fire regime severely changes the agroecosystem, it is required to wait for the natural recovery of the plant biomass and soil resources until the land is able to reach the same levels of crop production levels. To achieve this the ecosystem restoration takes time that frequently exceeds 20 years (Comte *et al.*, 2012). However, the findings in Chapter 5 revealed that natural fallows can more easily achieve the microbiome of a mature forest, the number of both plant and soil factors that must be adjusted are smaller than in the managed agroforestry systems. Thus, the slash-and-burn practice may be sustainable if the shifting cultivation involves long periods for land recovery, such as 20 years of natural fallow (Hamer *et al.*, 2013). Unfortunately, the continuous use of the soil after slash-and-burn is nearly impossible and the time needed to the land restoration is too long. So, there is a sustainability crisis in the practice of slash-and-burn agriculture: it results in a short period of cropping and needs long periods for land recovery. It is also important to note that the modern technology of agricultural intensification (fertilizer, irrigation, mechanization and capital to unit area of

production) has a limited success in tropics. So, to tackle this sustainability crisis, alternative practices to slash-and-burn have been proposed to achieve a more productive and less disturbing agricultural system.

If an increasing land use pressure does not allow for strategies relying purely on natural secondary forest succession, agroforestry may represent an alternative land use. Agroforestry systems provide crop, fruit and wood production while the agroecosystem develops increased complexity. However, our analysis of the soil microbiome revealed that this capacity of mimic the complex interactions found in the mature forest is highly compromised by the agroforestry practices. In summary, while the soil microbiome in the mature forest and secondary succession fostered a more fungal-driven community, the agroforestry systems consisted mainly of bacteria-driven communities.

## **Future opportunities of sustainable land-use in Amazon**

### *The challenge of replacing slash-and-burn practices in tropical agroecosystems*

It is well accepted that slash-and-burn in shifting cultivation causes great disturbances to ecosystem equilibrium (Giardina *et al.*, 2000). The changes in biomass and the soil nutrient inputs can cause a multitude of complex responses in the ecosystem: higher content of carbon emissions, high production of pyrogenic compounds, high nutrient input, death of small animals, soil microbiome primer effect. These initial effects also change in time: nutrient losses by runoff, leaching, and harvesting; soil organic matter depletion; selection of the soil seed bank to more fire-adapted plant. All these changes lead to the agroecosystem impoverishment affecting the land production and, in consequence, to the land abandonment to natural recovery. Therefore, it is necessary to find alternatives of land-use to replace the slash-and-burn practice.

The challenge of replacing the slash-and-burn cultivation system begins when most of the land use alternatives to slash-and-burn aim to eliminate the use of fire, pointing this as the main factor responsible for the land degradation after slash-and-burn. But the fire itself does not produce all these consequences, and its real role needs to be specified to select land use alternatives that do not result in the same problem. For example, the slash-and-mulch approach is a promising alternative to slash-and-burn (Reichert *et al.*, 2014) but causes higher soil carbon emission and immobilizes soil nutrients during the early stages (Davidson *et al.*, 2008). According to Davidson *et al.* (2008), slash-and-mulch carbon emissions were five times lower as compared with slash-and-burn. This work had a monitoring window of two years, and we should consider that while the slash-and-burn has high initial greenhouse gasses emissions which reduces considerably after burning, the slash-and-mulch will have lower but more continuous emissions over time for as long as the mulch remains. So, despite the lack of long-term measurements on CO<sub>2</sub> emission for slash-and-mulch systems, these agricultural practices essentially change a short and high intensity greenhouse gas-emitting system into a slower, more long-lasting one.

When searching for alternatives to slash-and-burn that strengthen the aboveground-belowground interactions, it may not be that use of fire is the main problem. The crucial fire problem in tropical rainforest is not the fire usage, but rather the changes in fire regime with an increase in burn frequencies (Cochrane (2003). According to archeological evidence (Denevan, 1992), only after the introduction of iron axes and later steel axes and machetes, did the slash-and-burn practice become widespread as it is known today. So, the native Americans practiced a different regime of land use and fire management. Fertility was achieved mainly by mulching, composting, intercropping and agroforestry, and these practices are continued today, as suggested by Hecht (Hecht, 2003) when describing a singular fire regime named “in-field burn” practiced by the Kayapó tribe.

Thus, it is necessary to consider the context in which the fire acts to, not only include slash-and-burn, but also other fire practices in the tropics, and consider

---

environmental interactions in the light of the current knowledge of fire regimes before mitigation recommendations can be made.

### *The design of sustainable agroforestry systems: mimicking the mature forest*

In the current thesis I showed that agroforestry systems failed to mimic the aboveground-belowground interactions found in both secondary succession and mature forest (Chapter 3 and 5). The aboveground biomass in agroforest presented similar values as found in the mature forest (especially for the homegardens). However, agroforestry systems favored a distinct soil microbiome with more bacteria whereas the secondary forested converged to the same fungal-rich community as observed in the mature forest. These findings reflect the impact of agroforestry practices (e.g., species selection, spatial arrangement, nutrient input, pruning, clearing of the understory) in shaping the soil microbiome. Each one of the agroforestry systems present management practices with distinct intensity this like contributed to explain the divergence found in the soil microbiome of each system (homegardens x commercial agroforests x enriched agroforests).

If the agroforestry practices reduced the co-dependence between aboveground and belowground subsystems, they can also be adjusted to foster stronger interactions. The current thesis also showed that, depending on the type of agroforestry system, different factors of plant biomass and soil variables must change to allow the agroforestry system to achieve the same soil microbiome as found in the mature forest. For example, for the enriched agroforests (EFA) and commercial plantation agroforestry (CPA) most of the changes are related with plant variables. EFA systems need, in median, 3.74 times more dead logs and 2.6 times more littermass. CPA are also required to increase the dead logs biomass (in median 3.2x), but also need to increase the plants from the lower canopy (plants <10cm dbh) by 2.32 times. This lower canopy represents the plants that deliver most of the

commercial products (e.g., cacao, açai, banana) and where the management efforts are concentrated. So, increasing the biomass of this group of plants might be challenging. CPA and EFA make use of the knowledge already available from monoculture of these plants to define the spatial arrangement that provides better production while allow other plants to grow between plants and between lines. Therefore, the increase in plant biomass here by adding more plants might compromise the system production. If the spatial arrangement does not consider the spatial interaction between those plants (diversity attractors and repellers), the design of agroforestry systems might not achieve the same level of complexity as found in the mature forest. In Chapter 2, I found that the spatial arrangements of trees change along secondary succession and become more complex in the mature forest. Based on that, the design of agroforestry systems must also include changes in the spatial arrangement of plants. This type of practice will help to increase the dead logs biomass and help to build spatial complexity as found in the mature forest.

For the homegardens, the analysis presented in this thesis suggested that most focus on the agroforestry practices should be done in increasing the availability of soil nutrients. One interesting management practice is the sweep-and-burn. The practices of sweep-and-burn consist of the regular sweeping of plant litter deposition, accumulating all this biomass in one or more litter piles and then burning the litter. According to Benjamin *et al.* (2001) it is practiced mainly in homegardens by the house residents in order to keep the surroundings clean and to avoid infestation by insects (mainly mosquitos), spiders, snakes and scorpions. Due to a general lack of research, there are still little date regarding the benefits of sweep-and-burn practices for soil nutrient status. To our knowledge, the work of Benjamin *et al.* (2001) as Pinho *et al.* (2010), as well as the data presented in Chapter 3 of this thesis (Leite *et al.*, 2016) show a higher nutrient availability (mainly phosphorus and potassium) that could be attributed to sweep-and-burn. Moreover, archeologists and anthropologists suggested that this semi-intensive agriculture with frequent burn cultivation may be responsible for the formation of dark brown *terra mulata* soils,

probably a precursor of Amazon Dark Earth (Winklerprins, 2009). To dig deep into that topic requires more controlled experimental designs that goes beyond the topic of this thesis, but as an alternative use of fire, the sweep-and-burn might increase the toolkit of anthropic management in a movement towards more sustainable and continuous crop production in the tropics. The sweep-and-burn strategy may become a more popular strategy for adjusting the nutrient availability as suggested by the model, as farmers are able to adjust the nutrient availability according to the physiological necessities of their woody perennials and crop plants. However, the impact of sweep-and-burn in contributing to the microbial community of agroforestry systems remains to be determined and is left as an avenue for future research.

In summary, by modeling the soil microbial community jointly with the plant biomass and soil factors, we explored potential scenarios of land-use systems mimicking the soil microbiome of a mature forest. The design of complex and sustainable agricultural systems in the tropics remains challenging, especially considering the plant-soil-microbe interactions. Soil and climatic conditions in the humid tropics challenge the development of continuous crops and require the design of agroecosystems capable of strengthening the aboveground-belowground interactions via more fungal rich soil microbiome. Consequently, management strategies in the agroforestry systems should focus on enhancing plant-soil interactions to better mimic the interdependence observed in mature rainforests.

## **Concluding remarks**

In this thesis, I took advantage of many different approaches (plant spatial arrangement, estimation of aboveground biomass, analysis of soil physico-chemical properties, and soil metagenomics) to investigate the influence of land use and anthropic management on aboveground and belowground communities and their

interactions. With respect to plant-plant interactions, I found that a second cycle of slash-and-burn agriculture is enough to induce a stronger clustering at both univariate and bivariate interactions. When looking into plant-soil relationships, I found an increasing co-dependence between soil factors and plant biomass in successional systems. Finally, when I included the soil microbiome into the picture of aboveground-belowground interactions, the results revealed a fungal-driven relationship along succession. On the other hand, agroforestry systems exhibited a weaker co-dependence between plant biomass and soil factors, and their soil microbiomes appear to promote more specific bacterial populations as opposed to fungal taxa.

In addition, by jointly modeling aboveground and belowground components, I was able to better compare different land use systems and model which variables could be adjusted to steer microbiome composition to become more similar as found in target native tropical rainforest soils. This analysis offers a new perspective on which agricultural practices might better mimic natural systems. Because there are many factors that influence soil-borne microbial communities, it remains a challenge to disentangle the mechanistic details driving the microbial community composition and aboveground-belowground interactions. The first challenges lie in the choice and robustness of the methodological approaches applied. However, the combination of advanced statistical methods with informed experimental design and improved representation of soil will contribute to close this knowledge gap.

In this thesis, I focused on secondary successional forest and agroforestry systems and considered the mature forest as the reference level. Secondary succession is used to understand the long-term effect of the slash-and-burn agriculture, because the slash-and-burn practice was the cause of the successional stages in my sampling. On the other hand, the agroforestry practices are pointed as a promising alternative to replace slash-and-burn agriculture. Successful and sustainable land use management of tropical soils depends on a strengthening of aboveground-belowground interactions. Hence, it is important to note that

agroforestry systems may not be a sustainable land use, given the observed break in the co-dependence between plant-soil-microbes.

We still need to improve our knowledge on the impact of agroforestry practices on the ecosystems characteristics to inform management to avoid the loss of connection between aboveground and belowground subsystems. At this point, integrating high-resolution genomic data with complex environmental data in observational studies is an important step towards identifying ecological differentiation and shifts in aboveground-belowground interactions. Only a combined approach that accounts for the plants' biomass, the soil factors, and the soil microbial community allows one to disentangle the relative importance of each component of the system - a central challenge in ecology of macro and microorganisms.



## References

- Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2), 139-177. <http://www.jstor.org/stable/2345821>
- Anderson, M. (2004). PERMDISP: a FORTRAN computer program for permutational analysis of multivariate dispersions (for any two-factor ANOVA design) using permutation tests. *Department of Statistics, University of Auckland, New Zealand*, 24.
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32-46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Anderson, M. J. (2006). Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics*, 62(1), 245-253. <https://doi.org/10.1111/j.1541-0420.2005.00440.x>
- Anderson, M. J., & Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing? *Ecological Monographs*, 83(4), 557-574. <https://doi.org/10.1890/12-2010.1>
- Aponte, C., García, L. V., & Marañón, T. (2013). Tree species effects on nutrient cycling and soil biota: A feedback mechanism favouring species coexistence. *Forest Ecology and Management*, 309(0), 36-46. <https://doi.org/http://dx.doi.org/10.1016/j.foreco.2013.05.035>
- Arevalo, L. A., Alegre, J. C., & Vilcahuaman, L. J. M. (2002). *Metodologia para estimar o estoque de carbono em diferentes sistemas de uso da terra*. Embrapa Floresta.
- Armada, E., Leite, M. F. A., Medina, A., Azcón, R., & Kuramae, E. E. (2018). Native bacteria promote plant growth under drought stress condition without impacting the rhizomicrobiome. *FEMS Microbiology Ecology*, fiy092-fiy092. <https://doi.org/10.1093/femsec/fiy092>
- Armstrong, A. (2016). Tropical Rainforest Ecosystems. *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, 1-16.
- Armstrong, A. H. (2020). International Encyclopedia of Geography. 1-16. <https://doi.org/10.1002/9781118786352.wbieg0644.pub2>
- Atangana, A., Khasa, D., Chang, S., & Degrande, A. (2014a). Agroforestry Modeling. In A. Atangana, D. Khasa, S. Chang, & A. Degrande (Eds.), *Tropical Agroforestry* (pp. 367-373). Springer Netherlands. [https://doi.org/10.1007/978-94-007-7723-1\\_21](https://doi.org/10.1007/978-94-007-7723-1_21)
- Atangana, A., Khasa, D., Chang, S., & Degrande, A. (2014b). Definitions and Classification of Agroforestry Systems. In *Tropical Agroforestry* (pp. 35-47). Springer Netherlands. [https://doi.org/10.1007/978-94-007-7723-1\\_3](https://doi.org/10.1007/978-94-007-7723-1_3)
- Atangana, A., Khasa, D., Chang, S., & Degrande, A. (2014c). Major Agroforestry Systems of the Humid Tropics. In *Tropical Agroforestry* (pp. 49-93). Springer Netherlands. [https://doi.org/10.1007/978-94-007-7723-1\\_4](https://doi.org/10.1007/978-94-007-7723-1_4)
- Ayres, E., Steltzer, H., Berg, S., Wallenstein, M. D., Simmons, B. L., & Wall, D. H. (2009). Tree Species Traits Influence Soil Physical, Chemical, and Biological Properties in High Elevation Forests. *PLoS ONE*, 4(6), e5964. <https://doi.org/10.1371/journal.pone.0005964>
- Bardgett, R. D., Bowman, W. D., & Kaufmann, R. (2005). A temporal approach to linking aboveground and belowground ecology. *Trends in Ecology & ....* <https://doi.org/10.1016/j.tree.2005.08.005>
- Barnes, A. D., Allen, K., Kreft, H., Corre, M. D., Jochum, M., Veldkamp, E., Clough, Y., Daniel, R., Darras, K., Denmead, L. H., Farikhah Haneda, N., Hertel, D., Knohl, A., Kotowska, M. M., Kurniawan, S., Meijide, A., Rembold, K., Edho Prabowo, W.,

- Schneider, D., Tschardtke, T., & Brose, U. (2017). Direct and cascading impacts of tropical land-use change on multi-trophic biodiversity. *Nature Ecology & Evolution*, *1*(10), 1511-1519. <https://doi.org/10.1038/s41559-017-0275-7>
- Barnes, A. D., Jochum, M., Mumme, S., Haneda, N. F., Farajallah, A., Widarto, T. H., & Brose, U. (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning [Article]. *Nat Commun*, *5*. <https://doi.org/10.1038/ncomms6351>
- Battaglia, M. A., Mou, P., Palik, B., & Mitchell, R. J. (2002). The effect of spatially variable overstory on the understory light environment of an open-canopied longleaf pine forest. *Canadian Journal of Forest Research*, *32*(11), 1984-1991. <https://doi.org/10.1139/x02-087>
- Beer, J., Muschler, R., Kass, D., & Somarriba, E. (1998). Shade management in coffee and cacao plantations. In *Directions in Tropical Agroforestry Research* (pp. 139-164). Springer.
- Bender, S. F., Wagg, C., & van der Heijden, M. G. A. (2016). An Underground Revolution: Biodiversity and Soil Ecological Engineering for Agricultural Sustainability. *Trends in Ecology & Evolution*, *31*(6), 440-452. <https://doi.org/https://doi.org/10.1016/j.tree.2016.02.016>
- Benjamin, T. J., Montañez, P. I., Jaménez, J. J. M., & Gillespie, A. R. (2001). Carbon, water and nutrient flux in Maya homegardens in the Yucatán peninsula of México. *Agroforestry Systems*, *53*(2), 103-111. <https://doi.org/10.1023/a:1013312217471>
- Berkelmann, D., Schneider, D., Meryandini, A., & Daniel, R. (2020). Unravelling the effects of tropical land use conversion on the soil microbiome. *Environmental Microbiome*, *15*(1), 5. <https://doi.org/10.1186/s40793-020-0353-3>
- Björk, J. R., Hui, F. K. C., O'Hara, R. B., & Montoya, J. M. (2018). Uncovering the drivers of host-associated microbiota with joint species distribution modelling. *Molecular Ecology*, *27*(12), 2714-2724. <https://doi.org/doi:10.1111/mec.14718>
- Boddey, R. M., Xavier, D. F., Alves, B. J. R., & Urquiaga, S. (2003). Brazilian Agriculture: The Transition to Sustainability. *Journal of Crop Production*, *9*(1-2), 593-621. [https://doi.org/10.1300/J144v09n01\\_10](https://doi.org/10.1300/J144v09n01_10)
- Bonner, M. T. L., Allen, D. E., Brackin, R., Smith, T. E., Lewis, T., Shoo, L. P., & Schmidt, S. (2020). Tropical Rainforest Restoration Plantations Are Slow to Restore the Soil Biological and Organic Carbon Characteristics of Old Growth Rainforest. *Microbial Ecology*, *79*(2), 432-442. <https://doi.org/10.1007/s00248-019-01414-7> PMID - 31372686
- Bonner, M. T. L., Shoo, L. P., Brackin, R., & Schmidt, S. (2018). Relationship between microbial composition and substrate use efficiency in a tropical soil. *Geoderma*, *315*, 96-103. <https://doi.org/https://doi.org/10.1016/j.geoderma.2017.11.026>
- Brady, N. C. (1996). Alternatives to slash-and-burn: a global imperative. *Agriculture, Ecosystems & Environment*, *58*(1), 3-11. [https://doi.org/https://doi.org/10.1016/0167-8809\(96\)00650-0](https://doi.org/https://doi.org/10.1016/0167-8809(96)00650-0)
- Brandt, A. J., de Kroon, H., Reynolds, H. L., & Burns, J. H. (2013). Soil heterogeneity generated by plant-soil feedbacks has implications for species recruitment and coexistence. *Journal of Ecology*, *101*(2), 277-286. <https://doi.org/10.1111/1365-2745.12042>
- Brooker, R. W. (2006). Plant-plant interactions and environmental change. *New Phytologist*, *171*(2), 271-284. <https://doi.org/10.1111/j.1469-8137.2006.01752.x>
- Brookfield, H., & Padoch, C. (1994). Appreciating Agrodiversity: A Look at the Dynamism

- and Diversity of Indigenous Farming Practices. *Environment: Science and Policy for Sustainable Development*, 36(5), 6-45. <https://doi.org/10.1080/00139157.1994.9929164>
- Brookshire, E. N. J., Gerber, S., Greene, W., Jones, R. T., & Thomas, S. A. (2017). Global bounds on nitrogen gas emissions from humid tropical forests. *Geophysical Research Letters*, 44(5), 2502-2510. <https://doi.org/10.1002/2017GL072867>
- Brown, C., Burslem, D. F. R. P., Illian, J. B., Bao, L., Brockelman, W., Cao, M., Chang, L. W., Dattaraja, H. S., Davies, S., Gunatilleke, C. V. S., Gunatilleke, I. A. U. N., Huang, J., Kassim, A. R., LaFrankie, J. V., Lian, J., Lin, L., Ma, K., Mi, X., Nathalang, A., Noor, S., Ong, P., Sukumar, R., Su, S. H., Sun, I. F., Suresh, H. S., Tan, S., Thompson, J., Uriarte, M., Valencia, R., Yap, S. L., Ye, W., & Law, R. (2013). Multispecies coexistence of trees in tropical forests: spatial signals of topographic niche differentiation increase with environmental heterogeneity. *Proceedings of the Royal Society B: Biological Sciences*, 280(1764), 20130502. <https://doi.org/doi:10.1098/rspb.2013.0502>
- Buttigieg, P. L., & Ramette, A. (2014). A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. *FEMS microbiology ecology*, 90(3), 543-550. <https://doi.org/10.1111/1574-6941.12437>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods*, 13(7). <https://doi.org/10.1038/nmeth.3869>
- Camenzind, T., Hättenschwiler, S., Treseder, K., Lehmann, A., & Rillig, M. C. (2017). Nutrient limitation of soil microbial processes in tropical forests. *Ecological Monographs*. <https://doi.org/10.1002/ecm.1279>
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression Analysis of Count Data* (2 ed.). Cambridge University Press. <https://doi.org/DOI:10.1017/CBO9781139013567>
- Cannell, M. G. R., Van Noordwijk, M., & Ong, C. K. (1996). The central agroforestry hypothesis: the trees must acquire resources that the crop would not otherwise acquire [journal article]. *Agroforestry Systems*, 34(1), 27-31. <https://doi.org/10.1007/bf00129630>
- 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., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms [Short Communication]. *The ISME journal*, 6, 1621. <https://doi.org/10.1038/ismej.2012.8>  
<https://www.nature.com/articles/ismej20128#supplementary-information>
- Cardozo, E., Muchavisoy, H., Silva, H., Zelarayán, M., Leite, M., Rousseau, G., & Gehring, C. (2015). Species richness increases income in agroforestry systems of eastern Amazonia. *Agroforestry Systems*, 1-16. <https://doi.org/10.1007/s10457-015-9823-9>
- Cardozo, E. G., Muchavisoy, H. M., Silva, H. R., Zelarayán, M. L. C., Leite, M. F. A., Rousseau, G. X., & Gehring, C. (2015). Species richness increases income in agroforestry systems of eastern Amazonia [journal article]. *Agroforestry Systems*, 89(5), 901-916. <https://doi.org/10.1007/s10457-015-9823-9>
- Cassman, N. A., Leite, M. F. A., Pan, Y., de Hollander, M., van Veen, J. A., & Kuramae, E. E. (2016). Plant and soil fungal but not soil bacterial communities are linked in long-term fertilized grassland [Article]. *Scientific Reports*, 6, 23680. <https://doi.org/10.1038/srep23680>
- Certini, G. (2005). Effects of fire on properties of forest soils: a review. *Oecologia*, 143(1),

- 1-10. <https://doi.org/10.1007/s00442-004-1788-8>
- Chamberlain, S. A., Bronstein, J. L., & Rudgers, J. A. (2014). How context dependent are species interactions? *Ecology Letters*, *17*(7), 881-890. <https://doi.org/10.1111/ele.12279>
- Chazdon, R. L., Broadbent, E. N., Rozendaal, D. M. A., Bongers, F., Zambrano, A. M. A., Aide, T. M., Balvanera, P., Becknell, J. M., Boukili, V., Brancalion, P. H. S., Craven, D., Almeida-Cortez, J. S., Cabral, G. A. L., de Jong, B., Denslow, J. S., Dent, D. H., DeWalt, S. J., Dupuy, J. M., Durán, S. M., Espirito-Santo, M. M., Fandino, M. C., César, R. G., Hall, J. S., Hernández-Stefanoni, J. L., Jakovac, C. C., Junqueira, A. B., Kennard, D., Letcher, S. G., Lohbeck, M., Martínez-Ramos, M., Massoca, P., Meave, J. A., Mesquita, R., Mora, F., Muñoz, R., Muscarella, R., Nunes, Y. R. F., Ochoa-Gaona, S., Orihuela-Belmonte, E., Peña-Claros, M., Pérez-García, E. A., Piotto, D., Powers, J. S., Rodríguez-Velazquez, J., Romero-Pérez, I. E., Ruíz, J., Saldarriaga, J. G., Sanchez-Azofeifa, A., Schwartz, N. B., Steininger, M. K., Swenson, N. G., Uriarte, M., van Breugel, M., van der Wal, H., Veloso, M. D. M., Vester, H., Vieira, I. C. G., Bentos, T. V., Williamson, G. B., & Poorter, L. (2016). Carbon sequestration potential of second-growth forest regeneration in the Latin American tropics. *Science Advances*, *2*(5). <https://doi.org/10.1126/sciadv.1501639>
- Chessel, D., Dufour, A. B., & Thioulouse, J. (2004). The ade4 package-I- One-table methods. *R News*(4), 5-10.
- Clark, J. S., Gelfand, A. E., Woodall, C. W., & Zhu, K. (2014). More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications*, *24*(5), 990-999. <https://doi.org/10.1890/13-1015.1>
- Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs*, *87*(1), 34-56. <https://doi.org/10.1002/ecm.1241>
- Cline, L. C., & Zak, D. R. (2015). Soil Microbial Communities are Shaped by Plant-Driven Changes in Resource Availability During Secondary Succession. *Ecology*. <https://doi.org/10.1890/15-0184.1>
- Cochrane, M. A. (2003). Fire science for rainforests [10.1038/nature01437]. *Nature*, *421*(6926), 913-919. <http://dx.doi.org/10.1038/nature01437>
- Comte, I., Davidson, R., Lucotte, M., de Carvalho, C. J. R., de Assis Oliveira, F., da Silva, B. P., & Rousseau, G. X. (2012). Physicochemical properties of soils in the Brazilian Amazon following fire-free land preparation and slash-and-burn practices. *Agriculture, Ecosystems & Environment*, *156*(0), 108-115. <https://doi.org/10.1016/j.agee.2012.05.004>
- Condit, R., Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., Núñez, P., Aguilar, S., Valencia, R., Villa, G., Muller-Landau, H. C., Losos, E., & Hubbell, S. P. (2002). Beta-Diversity in Tropical Forest Trees. *Science*, *295*(5555), 666-669. <https://doi.org/10.1126/science.1066854>
- Constantinides, M., & Fownes, J. H. (1994). Nitrogen mineralization from leaves and litter of tropical plants: Relationship to nitrogen, lignin and soluble polyphenol concentrations. *Soil Biology and Biochemistry*, *26*(1), 49-55. [https://doi.org/http://dx.doi.org/10.1016/0038-0717\(94\)90194-5](https://doi.org/http://dx.doi.org/10.1016/0038-0717(94)90194-5)
- Coomes, D. A., Allen, R. B., Scott, N. A., Goulding, C., & Beets, P. (2002). Designing systems to monitor carbon stocks in forests and shrublands. *Forest Ecology and*

## References

---

- Management*, 164(1–3), 89-108. [https://doi.org/http://dx.doi.org/10.1016/S0378-1127\(01\)00592-8](https://doi.org/http://dx.doi.org/10.1016/S0378-1127(01)00592-8)
- Costea, P. I., Zeller, G., Sunagawa, S., & Bork, P. (2014). A fair comparison. *Nature methods*, 11(4), 359. <https://doi.org/10.1038/nmeth.2897>
- Cougoul, A., Bailly, X., Vourc'h, G., & Gasqui, P. (2019). Rarity of microbial species: In search of reliable associations. *PLoS ONE*, 14(3), e0200458. <https://doi.org/10.1371/journal.pone.0200458> PMID - 30875367
- Davidson, E., de Carvalho, C., Figueira, A., Ishida, F., Ometto, J., Nardoto, G., Sabá, R., Hayashi, S., Leal, E., Vieira, I., & Martinelli, L. (2007). Recuperation of nitrogen cycling in Amazonian forests following agricultural abandonment. *Nature*, 447(7147), 995-998. <https://doi.org/10.1038/nature05900>
- Davidson, E. A., Carvalho, J. R., Vieira, I. C. G., Figueiredo, P. M., Ishida, F. Y., Santos, M. T. P., Guerrero, J. B., Kalif, K., & Sabá, R. T. (2004). Nitrogen and phosphorus limitation of biomass growth in a tropical secondary forest. *Ecological Applications*, 14(4), S150-S163. <http://www.esajournals.org/doi/abs/10.1890/01-6006>
- Davidson, E. A., SÁ, D. A., DEANE, T., REIS CARVALHO, C. J., De OLIVEIRA FIGUEIREDO, R., KATO, M. d. S. A., KATO, O. R., & ISHIDA, F. Y. (2008). An integrated greenhouse gas assessment of an alternative to slash-and-burn agriculture in eastern Amazonia. *Global Change Biology*, 14(5), 998-1007.
- de las Heras, A., Lake, I. R., Lovett, A., & Peres, C. (2011). Future deforestation drivers in an Amazonian ranching frontier. *Journal of Land Use Science*, 7(4), 365-393. <https://doi.org/10.1080/1747423x.2011.590234>
- Delgado-Baquerizo, M., Oliverio, A. M., Brewer, T. E., Benavent-González, A., Eldridge, D. J., Bardgett, R. D., Maestre, F. T., Singh, B. K., & Fierer, N. (2018). A global atlas of the dominant bacteria found in soil. *Science*, 359(6373), 320-325. <https://doi.org/10.1126/science.aap9516>
- Delgado-Baquerizo, M., Reich, P. B., Khachane, A. N., Campbell, C. D., Thomas, N., Freitag, T. E., Abu Al-Soud, W., Sørensen, S., Bardgett, R. D., & Singh, B. K. (2017). It is elemental: soil nutrient stoichiometry drives bacterial diversity. *Environmental Microbiology*, 19(3), 1176-1188. <https://doi.org/10.1111/1462-2920.13642>
- Delgado-Baquerizo, M., Reich, P. B., Trivedi, C., Eldridge, D. J., Abades, S., Alfaro, F. D., Bastida, F., Berhe, A. A., Cutler, N. A., Gallardo, A., García-Velázquez, L., Hart, S. C., Hayes, P. E., He, J.-Z., Hseu, Z.-Y., Hu, H.-W., Kirchmair, M., Neuhauser, S., Pérez, C. A., Reed, S. C., Santos, F., Sullivan, B. W., Trivedi, P., Wang, J.-T., Weber-Grullon, L., Williams, M. A., & Singh, B. K. (2020). Multiple elements of soil biodiversity drive ecosystem functions across biomes. *Nature Ecology & Evolution*, 4(2), 210-220. <https://doi.org/10.1038/s41559-019-1084-y> PMID - 32015427
- Denevan, W. M. (1992). Stone vs. metal axes: The ambiguity of shifting cultivation in prehistoric Amazonia. *Journal of the Steward Anthropological Society*, 20(1-2), 153-165.
- Desjardins, T., Barros, E., Sarrazin, M., Girardin, C., & Mariotti, A. (2004). Effects of forest conversion to pasture on soil carbon content and dynamics in Brazilian Amazonia. *Agriculture, Ecosystems & Environment*, 103(2), 365-373. <https://doi.org/http://dx.doi.org/10.1016/j.agee.2003.12.008>
- Dimitrov, M. R., Veraart, A. J., de Hollander, M., Smidt, H., van Veen, J. A., & Kuramae, E. E. (2017). Successive DNA extractions improve characterization of soil

- microbial communities. *PeerJ*, 5, e2915. <https://doi.org/10.7717/peerj.2915>
- Døckersmith, I., Giardina, C., & Sanford, R., Jr. (1999). Persistence of tree related patterns in soil nutrients following slash-and-burn disturbance in the tropics. *Plant and Soil*, 209(1), 137-157. <https://doi.org/10.1023/a:1004503023973>
- Dorazio, R. M., Connor, E. F., & Askins, R. A. (2015). Estimating the Effects of Habitat and Biological Interactions in an Avian Community. *PLoS ONE*, 10(8), e0135987. <https://doi.org/10.1371/journal.pone.0135987>
- Dormann, C. F., Bobrowski, M., Dehling, D. M., Harris, D. J., Hartig, F., Lischke, H., Moretti, M. D., Pagel, J., Pinkert, S., & Schleuning, M. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, 27(9), 1004-1016.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Dray, S., Chessel, D., & Thioulouse, J. (2003). Co-inertia analysis and the linking of ecological data tables. *Ecology*, 84(11), 3078-3089. <http://www.esajournals.org/doi/abs/10.1890/03-0178>
- Dunn, P. K., & Smyth, G. K. (1996). Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236-244. <https://doi.org/10.1080/10618600.1996.10474708>
- Edgar, R. C. (2017). Accuracy of microbial community diversity estimated by closed- and open-reference OTUs. *PeerJ*, 5, e3889. <https://doi.org/10.7717/peerj.3889> PMID - 29018622
- Elevitch, C., Mazaroli, D., & Ragone, D. (2018). Agroforestry Standards for Regenerative Agriculture. *Sustainability*, 10(9), 3337. <https://doi.org/10.3390/su10093337>
- Elith, J., & Leathwick, J. (2007). Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions*, 13(3), 265-275. <https://doi.org/10.1111/j.1472-4642.2007.00340.x>
- Escalona, M., Rocha, S., & Posada, D. (2016). A comparison of tools for the simulation of genomic next-generation sequencing data. *Nature Reviews Genetics*, 17(8), 459-469. <https://doi.org/10.1038/nrg.2016.57>
- Fang, H., Huang, C., Zhao, H., & Deng, M. (2015). CCLasso: correlation inference for compositional data through Lasso. *Bioinformatics*, 31(19), 3172-3180.
- Feller, C., & Beare, M. H. (1997). Physical control of soil organic matter dynamics in the tropics. *Geoderma*, 79(1-4), 69-116. [https://doi.org/http://dx.doi.org/10.1016/S0016-7061\(97\)00039-6](https://doi.org/http://dx.doi.org/10.1016/S0016-7061(97)00039-6)
- Ferguson, B. G., Vandermeer, J., Morales, H., & Griffith, D. M. (2003). Post-Agricultural Succession in El Petén, Guatemala. *Conservation Biology*, 17(3), 818-828. <https://doi.org/10.1046/j.1523-1739.2003.01265.x>
- Fernandes, A. D., Reid, J. N., Macklaim, J. M., McMurrough, T. A., Edgell, D. R., & Gloor, G. B. (2014). Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*, 2, 15. <https://doi.org/10.1186/2049-2618-2-15> PMID - 24910773

## References

---

- Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguët-Covex, C., Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G., & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, *15*(3), 543-556. <https://doi.org/10.1111/1755-0998.12338>
- Figueiredo, F. O. G., Zuquim, G., Tuomisto, H., Moulatlet, G. M., Balslev, H., & Costa, F. R. C. (2017). Beyond climate control on species range: The importance of soil data to predict distribution of Amazonian plant species. *Journal of Biogeography*. <https://doi.org/10.1111/jbi.13104>
- Flores, B. M., Oliveira, R. S., Rowland, L., Quesada, C. A., & Lambers, H. (2020). Editorial special issue: plant-soil interactions in the Amazon rainforest. In: Springer.
- Foster, S. D., & Bravington, M. V. (2013). A Poisson-Gamma model for analysis of ecological non-negative continuous data [journal article]. *Environmental and Ecological Statistics*, *20*(4), 533-552. <https://doi.org/10.1007/s10651-012-0233-0>
- Freschet, G. T., Cornwell, W. K., Wardle, D. A., Elumeeva, T. G., Liu, W., Jackson, B. G., Onipchenko, V. G., Soudzilovskaia, N. A., Tao, J., & Cornelissen, J. H. C. (2013). Linking litter decomposition of above- and below-ground organs to plant-soil feedbacks worldwide. *Journal of Ecology*, *101*(4), 943-952. <https://doi.org/10.1111/1365-2745.12092>
- Friedman, J., & Alm, E. J. (2012). Inferring Correlation Networks from Genomic Survey Data. *PLOS Computational Biology*, *8*(9), e1002687. <https://doi.org/10.1371/journal.pcbi.1002687>
- Gajaseni, J., & Gajaseni, N. (1999). Ecological rationalities of the traditional homegarden system in the Chao Phraya Basin, Thailand [journal article]. *Agroforestry Systems*, *46*(1), 3-23. <https://doi.org/10.1023/a:1006188504677>
- Gama-Rodrigues, A. C., Sales, M. V. S., Silva, P. S. D., Comerford, N. B., Cropper, W. P., & Gama-Rodrigues, E. F. (2014). An exploratory analysis of phosphorus transformations in tropical soils using structural equation modeling. *Biogeochemistry*, *118*(1-3), 453-469. <https://doi.org/10.1007/s10533-013-9946-x>
- Gehring, C. (2006). O ambiente do trópico úmido e o manejo sustentável dos agrossistemas. In E. G. Moura & A. C. F. Aguiar (Eds.), *O desenvolvimento rural como forma de ampliação dos direitos no campo: Princípios e tecnologias* (Vol. 2, pp. 101-140). Estação Produções Ltda.
- Gehring, C., Denich, M., & Vlek, P. L. G. (2005). Resilience of secondary forest regrowth after slash-and-burn agriculture in central Amazonia. *Journal of Tropical Ecology*, *21*(5), 519-527. <https://doi.org/10.1017/S0266467405002543>
- Gehring, C., Muniz, F. H., & Gomes de Souza, L. A. (2008). Leguminosae along 2–25 years of secondary forest succession after slash-and-burn agriculture and in mature rain forest of Central Amazonia. *The Journal of the Torrey Botanical Society*, *135*(3), 388-400. <https://doi.org/10.3159/08-RA-032.1>
- Gehring, C., Park, S., & Denich, M. (2004). Liana allometric biomass equations for Amazonian primary and secondary forest. *Forest Ecology and Management*, *195*(1–2), 69-83. <https://doi.org/10.1016/j.foreco.2004.02.054>
- Gehring, C., Park, S., & Denich, M. (2008). Close relationship between diameters at 30cm height and at breast height (DBH). *Acta Amazonica*, *38*, 71-76. [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0044-59672008000100008&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0044-59672008000100008&nrm=iso)
- Gehring, C., Vlek, P. L. G., de Souza, L. A. G., & Denich, M. (2005). Biological nitrogen

- fixation in secondary regrowth and mature rainforest of central Amazonia. *Agriculture, Ecosystems & Environment*, 111(1–4), 237-252. <https://doi.org/http://dx.doi.org/10.1016/j.agee.2005.06.009>
- Gehring, C., Zelarayán, M. L. C., Almeida, R. B., & Moraes, F. H. R. (2011). Allometry of the babassu palm growing on a slash-and-burn agroecosystem of the eastern periphery of Amazonia. *Acta Amazonica*, 41, 127-134. [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0044-59672011000100015&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0044-59672011000100015&nrm=iso)
- Giardina, C. P., Sanford, R. L., Døckersmith, I. C., & Jaramillo, V. J. (2000). The effects of slash burning on ecosystem nutrients during the land preparation phase of shifting cultivation. *Plant and Soil*, 220(1-2), 247-260. <https://doi.org/10.1023/A:1004741125636>
- Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrück, L., Reeder, J., Temperton, B., Huse, S., McHardy, A. C., Knight, R., Joint, I., Somerfield, P., Fuhrman, J. A., & Field, D. (2012). Defining seasonal marine microbial community dynamics. *The ISME journal*, 6(2), 298-308. <https://doi.org/10.1038/ismej.2011.107>
- Glaser, B., & Birk, J. J. (2012). State of the scientific knowledge on properties and genesis of Anthropogenic Dark Earths in Central Amazonia (terra preta de Índio). *Geochimica et Cosmochimica Acta*, 82, 39-51. <https://doi.org/http://dx.doi.org/10.1016/j.gca.2010.11.029>
- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology*, 8, 2224. <https://doi.org/10.3389/fmicb.2017.02224> PMID - 29187837
- Goss-Souza, D., Mendes, L. W., Borges, C. D., Baretta, D., Tsai, S. M., & Rodrigues, J. L. M. (2017). Soil microbial community dynamics and assembly under long-term land use change. *FEMS microbiology ecology*, 93(10). <https://doi.org/10.1093/femsec/fix109>
- Guillera-Aroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, 40(2), 281-295. <https://doi.org/10.1111/ecog.02445>
- Guo, Y., Du, Q., Li, G., Ni, Y., Zhang, Z., Ren, W., & Hou, X. (2016). Soil phosphorus fractions and arbuscular mycorrhizal fungi diversity following long-term grazing exclusion on semi-arid steppes in Inner Mongolia. *Geoderma*, 269, 79-90. <https://doi.org/https://doi.org/10.1016/j.geoderma.2016.01.039>
- Haase, P. (2001). Can isotropy vs. anisotropy in the spatial association of plant species reveal physical vs. biotic facilitation? *Journal of Vegetation Science*. <https://doi.org/10.1111/j.1654-1103.2001.tb02623.x>
- Hamer, U., Potthast, K., Burneo, J., & Makeschin, F. (2013). Nutrient stocks and phosphorus fractions in mountain soils of Southern Ecuador after conversion of forest to pasture. *Biogeochemistry*, 112(1-3), 495-510. <https://doi.org/10.1007/s10533-012-9742-z>
- Harris, D. J., Taylor, S., & White, E. P. (2017). Forecasting biodiversity in breeding birds using best practices. *PeerJ*, 6, e4278. <https://doi.org/10.1101/191130>
- Hart, S. P., Usinowicz, J., & Levine, J. M. (2017). The spatial scales of species coexistence. *Nature Ecology & Evolution*, 1(8), 1066. <https://doi.org/10.1038/s41559-017-0230-7>
- Hauser, S., & Norgrove, L. (2013). Slash-and-burn agriculture, effects of. *Encyclopedia of biodiversity*, Volume 6(Ed. 2), 551-562.
- Hautier, Y., Tilman, D., Isbell, F., Seabloom, E. W., Borer, E. T., & Reich, P. B. (2015).

- Anthropogenic environmental changes affect ecosystem stability via biodiversity. *Science*, 348(6232), 336-340. <https://doi.org/10.1126/science.aaa1788>
- Hawinkel, S., Kerckhof, F.-M., Bijmens, L., & Thas, O. (2018). A unified framework for unconstrained and constrained ordination of microbiome read count data. *bioRxiv*, 429340.
- He, W., & Chen, F. (2013). Evaluating Status Change of Soil Potassium from Path Model. *PLoS ONE*, 8(10), e76712. <https://doi.org/10.1371/journal.pone.0076712>
- Hecht, S. B. (2003). Indigenous Soil Management and the Creation of Amazonian Dark Earths: Implications of Kayapó Practice. In J. Lehmann, D. Kern, B. Glaser, & W. Wodos (Eds.), *Amazonian Dark Earths* (pp. 355-372). Springer Netherlands. [https://doi.org/10.1007/1-4020-2597-1\\_18](https://doi.org/10.1007/1-4020-2597-1_18)
- Herrera, R., Jordan, C., Klinge, H., & Medina, E. (1978). Amazon ecosystems. Their structure and functioning with particular emphasis on nutrients. *Interciencia*, 3(4), 223-231.
- Ho, A., Angel, R., Veraart, A. J., Daebeler, A., Jia, Z., Kim, S., Kerckhof, F.-M., Boon, N., & Bodelier, P. L. E. (2016). Biotic Interactions in Microbial Communities as Modulators of Biogeochemical Processes: Methanotrophy as a Model System. *Frontiers in Microbiology*, 7, 1285. <https://doi.org/10.3389/fmicb.2016.01285>
- Hobbie, S. E., Reich, P. B., Oleksyn, J., Ogdahl, M., Zytkowski, R., Hale, C., & Karolewski, P. (2006). Tree species effects on decomposition and forest floor dynamics in a common garden. *Ecology*, 87(9), 2288-2297. [https://doi.org/10.1890/0012-9658\(2006\)87\[2288:tseoda\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[2288:tseoda]2.0.co;2)
- Hugerth, L. W., & Andersson, A. F. (2017). Analysing Microbial Community Composition through Amplicon Sequencing: From Sampling to Hypothesis Testing. *Frontiers in Microbiology*, 8, 1561. <https://doi.org/10.3389/fmicb.2017.01561>
- Hughes, J. B., & Hellmann, J. J. (2005). The Application of Rarefaction Techniques to Molecular Inventories of Microbial Diversity. In *Methods in Enzymology* (Vol. 397, pp. 292-308). Academic Press. [https://doi.org/https://doi.org/10.1016/S0076-6879\(05\)97017-1](https://doi.org/https://doi.org/10.1016/S0076-6879(05)97017-1)
- Hui, F., Taskinen, S., Pledger, S., Foster, S. D., & Warton, D. I. (2015). Model-based approaches to unconstrained ordination. *Methods in Ecology and Evolution*, 6(4), 399-411. <https://doi.org/10.1111/2041-210X.12236>
- Hui, F. K. C. (2016). boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. *Methods in Ecology and Evolution*, 7(6), 744-750. <https://doi.org/10.1111/2041-210X.12514>
- IAC. (2001). *Análise química para avaliação da fertilidade de solos tropicais*. Instituto Agrônomico.
- Iknayan, K. J., Tingley, M. W., Furnas, B. J., & Beissinger, S. R. (2014). Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology & Evolution*, 29(2), 97-106. <https://doi.org/10.1016/j.tree.2013.10.012>
- Isaac, M. E., & Borden, K. A. (2019). Nutrient acquisition strategies in agroforestry systems. *Plant and Soil*, 444(1-2), 1-19.
- Ives, A. R. (2015). For testing the significance of regression coefficients, go ahead and log-transform count data. *Methods in Ecology and Evolution*, 6(7), 828-835. <https://doi.org/10.1111/2041-210X.12386>
- Jakovac, C. C., Bongers, F., Kuyper, T. W., Mesquita, R. C. G., & Peña-Claros, M. (2016). Land use as a filter for species composition in Amazonian secondary forests. *Journal of Vegetation Science*, 27(6), 1104-1116. <https://doi.org/10.1111/jvs.12457>

- Jakovac, C. C., Peña-Claros, M., Mesquita, R., Bongers, F., & Kuyper, T. W. (2016). Swiddens under transition: Consequences of agricultural intensification in the Amazon. *Agriculture, Ecosystems & Environment*, 218, 116-125. <https://doi.org/10.1016/j.agee.2015.11.013>
- Jensen, M. (1993). Soil conditions, vegetation structure and biomass of a Javanese homegarden. *Agroforestry Systems*, 24(2), 171-186. <https://doi.org/10.1007/BF00706890>
- Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. Springer Science & Business Media.
- Jordan, C. F., & Herrera, R. (1981). Tropical Rain Forests: Are Nutrients Really Critical? *The American Naturalist*, 117(2), 167-180. <http://www.jstor.org/stable/2460498>
- Jost, L., Chao, A., & Chazdon, R. L. (2011). Compositional similarity and  $\beta$  (beta) diversity. *Biological diversity: frontiers in measurement and assessment*, 66-84.
- Juo, A. S. R., & Manu, A. (1996). Chemical dynamics in slash-and-burn agriculture. *Agriculture, Ecosystems & Environment*, 58(1), 49-60. [https://doi.org/http://dx.doi.org/10.1016/0167-8809\(95\)00656-7](https://doi.org/http://dx.doi.org/10.1016/0167-8809(95)00656-7)
- Kang, B. T. (1997). Alley cropping—soil productivity and nutrient recycling. *Forest Ecology and Management*, 91(1), 75-82. [https://doi.org/https://doi.org/10.1016/S0378-1127\(96\)03886-8](https://doi.org/https://doi.org/10.1016/S0378-1127(96)03886-8)
- Kato, O. R., Lunz, A. M., Bispo, C. J. C., Carvalho, C. J. R. d., Miranda, I. S., Takamatsu, J., Maués, M. M., Gerhard, P., Azevedo, R. d., Vasconcelos, S. S., Honhwal, S., & Lemos, W. d. P. (2009). Projeto dendê: sistemas agroflorestais na agricultura familiar (Oilpalm Project: Agroforestry systems in smallholder agriculture). *Embrapa Eastern Amazonia*. <http://www.alice.cnptia.embrapa.br/bitstream/doc/657476/1/05tema19.pdf>
- C:\Users\João Batista\Desktop\Mestrado\05tema19.pdf
- Kauffman, J. B., Cummings, D. L., Ward, D. E., & Babbitt, R. (1995). Fire in the Brazilian Amazon: 1. Biomass, nutrient pools, and losses in slashed primary forests. *Oecologia*, 104(4), 397-408. <https://doi.org/10.1007/BF00341336>
- Keller, M., Palace, M., & Hurr, G. (2001). Biomass estimation in the Tapajós National Forest, Brazil: Examination of sampling and allometric uncertainties. *Forest Ecology and Management*, 154(3), 371-382. [https://doi.org/http://dx.doi.org/10.1016/S0378-1127\(01\)00509-6](https://doi.org/http://dx.doi.org/10.1016/S0378-1127(01)00509-6)
- Kim, D. H., Sexton, J. O., & Townshend, J. R. (2015). Accelerated deforestation in the humid tropics from the 1990s to the 2000s. *Geophysical Research Letters*, 42(9), 3495-3501. <https://doi.org/10.1002/2014GL062777>
- Klanderud, K., Mbolatiana, H. Z. H., Vololomboahangy, M. N., Radimbison, M. A., Roger, E., Totland, Ø., & Rajeriarison, C. (2009). Recovery of plant species richness and composition after slash-and-burn agriculture in a tropical rainforest in Madagascar. *Biodiversity and Conservation*, 19(1), 187. <https://doi.org/10.1007/s10531-009-9714-3>
- Klute, A., Campbell, G. S., Nielsen, D. R., Jackson, R. D., & Mortland, M. N. (1986). *Methods of soil analysis – Physical and mineralogical methods*. Soil Science Society of America.
- Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J. A., Hugenholtz, P., van der Lelie, D., Meyer, F., Stevens, R., Bailey, M. J., Gordon, J. I., Kowalchuk, G. A., & Gilbert, J. A. (2012). Unlocking the potential of metagenomics through replicated experimental design. *Nature biotechnology*, 30(6), 513.

- <https://doi.org/10.1038/nbt.2235>
- Knight, R., Vrbanac, A., Taylor, B. C., Akseonov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolok, T., McCall, L.-I., & McDonald, D. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, *16*(7), 410-422.
- Kos, M., Tuijl, M. A. B., de Roo, J., Mulder, P. P. J., & Bezemer, T. M. (2015). Species-specific plant–soil feedback effects on above-ground plant–insect interactions. *Journal of Ecology*, *103*(4), 904-914. <https://doi.org/10.1111/1365-2745.12402>
- Kumar, B., & Nair, P. (2004). The enigma of tropical homegardens. *New Vistas in Agroforestry*. [https://doi.org/10.1007/978-94-017-2424-1\\_10](https://doi.org/10.1007/978-94-017-2424-1_10)
- Kuramae, E. E., Yergeau, E., Wong, L. C., Pijl, A. S., van Veen, J. A., & Kowalchuk, G. A. (2012). *Soil characteristics more strongly influence soil bacterial communities than land-use type* (Vol. 79) [Journal Article]. <https://doi.org/10.1111/j.1574-6941.2011.01192.x>
- Labrière, N., Locatelli, B., Laumonier, Y., Freycon, V., & Bernoux, M. (2015). Soil erosion in the humid tropics: A systematic quantitative review. *Agriculture, Ecosystems & Environment*, *203*, 127-139. <https://doi.org/10.1016/j.agee.2015.01.027>
- Laurance, W. F. (2015). Emerging Threats to Tropical Forests. *Annals of the Missouri Botanical Garden*, *100*(3), 159-169. <https://doi.org/10.3417/2011087>
- Laurance, W. F., Fearnside, P. M., Laurance, S. G., Delamonica, P., Lovejoy, T. E., Rankin-de Merona, J. M., Chambers, J. Q., & Gascon, C. (1999). Relationship between soils and Amazon forest biomass: a landscape-scale study. *Forest Ecology and Management*, *118*(1–3), 127-138. [https://doi.org/http://dx.doi.org/10.1016/S0378-1127\(98\)00494-0](https://doi.org/http://dx.doi.org/10.1016/S0378-1127(98)00494-0)
- Lavelle, P., Rodríguez, N., Arguello, O., Bernal, J., Botero, C., Chaparro, P., Gómez, Y., Gutiérrez, A., Hurtado, M. d. P., Loaiza, S., Pullido, S. X., Rodríguez, E., Sanabria, C., Velásquez, E., & Fonte, S. J. (2014). Soil ecosystem services and land use in the rapidly changing Orinoco River Basin of Colombia. *Agriculture, Ecosystems & Environment*, *185*(0), 106-117. <https://doi.org/http://dx.doi.org/10.1016/j.agee.2013.12.020>
- Lawrence, D., Radcliff, C., Tully, K., Schmoor, B., & Schneider, L. (2010). Untangling a Decline in Tropical Forest Resilience: Constraints on the Sustainability of Shifting Cultivation Across the Globe. *Biotropica*, *42*(1), 21-30. <https://doi.org/10.1111/j.1744-7429.2009.00599.x>
- Leite, M. F. A., Luz, R. L., Muchavisoy, K. H. M., Zelarayán, M. L. C., Cardoso, E. G., Moraes, F. H. R., Rousseau, G. X., & Gehring, C. (2016). The effect of land use on aboveground biomass and soil quality indicators in spontaneous forests and agroforests of eastern Amazonia [journal article]. *Agroforestry Systems*, *90*(6), 1009-1023. <https://doi.org/10.1007/s10457-015-9880-0>
- Levine, J. M., & Lammers, J. H. R. (2009). The importance of niches for the maintenance of species diversity. *Nature*, *461*(7261), 254-257. <https://doi.org/10.1038/nature08251>
- Li, H. (2015). Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis. *Annual Review of Statistics and Its Application*, *2*(1), 73-94. <https://doi.org/10.1146/annurev-statistics-010814-020351>
- Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L., Cornejo-Castillo, F. M., d'Ovidio, F., De Meester, L., Ferrera, I., Garet-Delmas, M.-J., Guidi, L., Lara, E., Pesant, S., Royo-Llonch, M., Salazar, G., Sánchez, P., Sebastian, M., Souffreau,

- C., Dimier, C., Picheral, M., Searson, S., Kandels-Lewis, S., Gorsky, G., Not, F., Ogata, H., Speich, S., Stemmann, L., Weissenbach, J., Wincker, P., Acinas, S. G., Sunagawa, S., Bork, P., Sullivan, M. B., Karsenti, E., Bowler, C., de Vargas, C., & Raes, J. (2015). Determinants of community structure in the global plankton interactome. *Science*, *348*(6237), 1262073. <https://doi.org/10.1126/science.1262073>
- Lima, A. B., Cannavan, F. S., Navarrete, A. A., Teixeira, W. G., Kuramae, E. E., & Tsai, S. M. (2015). Amazonian Dark Earth and Plant Species from the Amazon Region Contribute to Shape Rhizosphere Bacterial Communities. *Microbial Ecology*, *69*(4), 855-866. <https://doi.org/10.1007/s00248-014-0472-8>
- Lovell, D., Pawlowsky-Glahn, V., Egozcue, J. J., Marguerat, S., & Bähler, J. (2015). Proportionality: a valid alternative to correlation for relative data. *PLoS computational biology*, *11*(3), e1004075.
- Lovell, S. T., & Johnston, D. M. (2009). Designing Landscapes for Performance Based on Emerging Principles in Landscape Ecology. *Ecology and Society*, *14*(1), Article 44. <http://www.ecologyandsociety.org/vol14/iss1/art44/>
- Lovell, S. T., & Johnston, D. M. (2009). Designing Landscapes for Performance Based on Emerging Principles in Landscape Ecology. *Ecology and Society*, *14*(1), 44. <Go to ISI>://WOS:000267846300050
- Lozupone, C., & Knight, R. (2005). UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*, *71*(12), 8228-8235. <https://doi.org/10.1128/aem.71.12.8228-8235.2005>
- Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., & Knight, R. (2010). UniFrac: an effective distance metric for microbial community comparison. *The ISME journal*, *5*(2). <https://doi.org/10.1038/ismej.2010.133>
- Lucas, Y. (2001). THE ROLE OF PLANTS IN CONTROLLING RATES AND PRODUCTS OF WEATHERING: Importance of Biological Pumping. *Annual Review of Earth and Planetary Sciences*, *29*(1), 135-163. <https://doi.org/10.1146/annurev.earth.29.1.135>
- Lupatini, M., Suleiman, A. K. A., Jacques, R. J. S., Antonioli, Z. I., de Siqueira Ferreira, A., Kuramae, E. E., & Roesch, L. F. W. (2014). Network topology reveals high connectance levels and few key microbial genera within soils [Original Research]. *Frontiers in Environmental Science*, *2*(10). <https://doi.org/10.3389/fenvs.2014.00010>
- Mackenzie, D. I., & Royle, A. J. (2005). Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, *42*(6), 1105-1114. <https://doi.org/10.1111/j.1365-2664.2005.01098.x>
- Manzoni, S., Taylor, P., Richter, A., Porporato, A., & Ågren, G. I. (2012). Environmental and stoichiometric controls on microbial carbon-use efficiency in soils. *New Phytologist*, *196*(1), 79-91.
- Markewitz, D., Davidson, E., Moutinho, P., & Nepstad, D. (2004). Nutrient loss and redistribution after forest clearing on a highly weathered soil in amazonia. *Ecological Applications*, *14*(sp4), 177-199. <https://doi.org/10.1890/01-6016>
- Martin, P. A., Newton, A. C., & Bullock, J. M. (2013). Carbon pools recover more quickly than plant biodiversity in tropical secondary forests. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1773). <https://doi.org/10.1098/rspb.2013.2236>
- Martino, C., Morton, J. T., Marotz, C. A., Thompson, L. R., Tripathi, A., Knight, R., & Zengler, K. (2019). A Novel Sparse Compositional Technique Reveals Microbial

## References

---

- Perturbations. *mSystems*, 4(1), e00016-00019.  
<https://doi.org/10.1128/mSystems.00016-19>
- May, F., Huth, A., & Wiegand, T. (2015). Moving beyond abundance distributions: neutral theory and spatial patterns in a tropical forest. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1802), 20141657.  
<https://doi.org/10.1098/rspb.2014.1657>
- Mbow, C., Van Noordwijk, M., Luedeling, E., Neufeldt, H., Minang, P. A., & Kowero, G. (2014). Agroforestry solutions to address food security and climate change challenges in Africa. *Current Opinion in Environmental Sustainability*, 6(Supplement C), 61-67.  
<https://doi.org/https://doi.org/10.1016/j.cosust.2013.10.014>
- McGrath, D., Smith, C., Gholz, H., & Oliveira, F. (2001). Effects of Land-Use Change on Soil Nutrient Dynamics in Amazônia. *Ecosystems*. <https://doi.org/10.1007/s10021-001-0033-0>
- McGrath, D. A., Comerford, N. B., & Duryea, M. L. (2000). Litter dynamics and monthly fluctuations in soil phosphorus availability in an Amazonian agroforest. *Forest Ecology and Management*, 131(1-3), 167-181.  
[https://doi.org/http://dx.doi.org/10.1016/S0378-1127\(99\)00207-8](https://doi.org/http://dx.doi.org/10.1016/S0378-1127(99)00207-8)
- McLaren, M. R., Willis, A. D., & Callahan, B. J. (2019). Consistent and correctable bias in metagenomic sequencing experiments. *bioRxiv*, 559831.  
<https://doi.org/10.1101/559831>
- McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol*, 10(4), e1003531.  
<https://doi.org/10.1371/journal.pcbi.1003531>
- Melo, V. F., Uchôa, S. C. P., Senwo, Z. N., & Amorim, R. J. P. (2015). Phosphorus Adsorption of Some Brazilian Soils in Relations to Selected Soil Properties. *Open Journal of Soil Science*, 5, 101-109. <https://doi.org/10.4236/ojss.2015.55010>
- Mesquita, R. C., Ickes, K., Ganade, G., & Williamson, G. B. (2001). Alternative successional pathways in the Amazon Basin. *Journal of Ecology*, 89(4), 528-537.
- Milton, Y., & Kaspari, M. (2007). Bottom-up and top-down regulation of decomposition in a tropical forest. *Oecologia*, 153(1), 163-172. <https://doi.org/10.1007/s00442-007-0710-6> PMID - 17375326
- Moço, M. K. S., Gama-Rodrigues, E. F., Gama-Rodrigues, A. C., Machado, R. C. R., & Baligar, V. C. (2010). Relationships between invertebrate communities, litter quality and soil attributes under different cacao agroforestry systems in the south of Bahia, Brazil. *Applied Soil Ecology*, 46(3), 347-354.  
<https://doi.org/10.1016/j.apsoil.2010.10.006>
- Mohri, H., Lahoti, S., Saito, O., Mahalingam, A., Gunatilleke, N., Irham, Hoang, V. T., Hitinayake, G., Takeuchi, K., & Herath, S. (2013). Assessment of ecosystem services in homegarden systems in Indonesia, Sri Lanka, and Vietnam. *Ecosystem Services*, 5(0), 124-136.  
<https://doi.org/http://dx.doi.org/10.1016/j.ecoser.2013.07.006>
- Montgomery, R. A., & Chazdon, R. L. (2001). Forest structure, canopy architecture, and light transmittance in tropical wet forests. *Ecology*, 82(10), 2707-2718.  
[https://doi.org/10.1890/0012-9658\(2001\)082\[2707:FSCAAL\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2001)082[2707:FSCAAL]2.0.CO;2)
- Muchavisoy, H. M. (2013). *Estoque de carbono em florestas, capoeiras e sistemas agroflorestais da Amazônia Oriental* [Dissertation, Universidade Estadual do Maranhão]. São Luís.

- Mueller, R. C., Paula, F. S., Mirza, B. S., Rodrigues, J. L. M., Nusslein, K., & Bohannan, B. J. M. (2014). Links between plant and fungal communities across a deforestation chronosequence in the Amazon rainforest [Short Communication]. *ISME J*, 8(7), 1548-1550. <https://doi.org/10.1038/ismej.2013.253>
- Nair, P. K. R. (2001). Do tropical homegardens elude science, or is it the other way around? *Agroforestry Systems*, 53(2), 239-245. <https://doi.org/10.1023/a:1013388806993>
- Nair, P. K. R. (2005). Agroforestry. In *Encyclopedia of Soils in the Environment* (pp. 35-44). Elsevier. <https://doi.org/https://doi.org/10.1016/B0-12-348530-4/00244-7>
- Nair, P. K. R. (2013). Agroforestry: Trees in Support of Sustainable Agriculture. In *Reference Module in Earth Systems and Environmental Sciences*. Elsevier. <https://doi.org/http://dx.doi.org/10.1016/B978-0-12-409548-9.05088-0>
- Nair, P. K. R. (2014). Grand Challenges in Agroecology and Land Use Systems [Specialty Grand Challenge]. *Frontiers in Environmental Science*, 2. <https://doi.org/10.3389/fenvs.2014.00001>
- Nair, R. P. (2014). Grand Challenges in Agroecology and Land Use Systems [Specialty Grand Challenge]. *Frontiers in Environmental Science*, 2. <https://doi.org/10.3389/fenvs.2014.00001>
- Navarrete, A. A., Kuramae, E. E., de Hollander, M., Pijl, A. S., van Veen, J. A., & Tsai, S. M. (2013). *Acidobacterial community responses to agricultural management of soybean in Amazon forest soils* (Vol. 83) [Journal Article]. <https://doi.org/10.1111/1574-6941.12018>
- Nelder, J. A., & Baker, R. J. (2004). Generalized Linear Models. In *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc. <https://doi.org/10.1002/0471667196.ess0866.pub2>
- Nelson, B. W., Mesquita, R., Pereira, J. L. G., Garcia Aquino de Souza, S., Teixeira Batista, G., & Bovino Couto, L. (1999). Allometric regressions for improved estimate of secondary forest biomass in the central Amazon. *Forest Ecology and Management*, 117(1-3), 149-167. [https://doi.org/http://dx.doi.org/10.1016/S0378-1127\(98\)00475-7](https://doi.org/http://dx.doi.org/10.1016/S0378-1127(98)00475-7)
- Nguyen, N.-P., Warnow, T., Pop, M., & White, B. (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *npj Biofilms and Microbiomes*, 2(1), 16004. <https://doi.org/10.1038/npjbiofilms.2016.4> PMID - 28721243
- Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., Husby, M., Kålås, J. A., Lehtikoinen, A., Luoto, M., Mod, H. K., Newell, G., Renner, I., Roslin, T., Soininen, J., Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N. E., Gravel, D., & Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, 89(3), e01370. <https://doi.org/10.1002/ecm.1370>
- Nottingham, A. T., Bååth, E., Reischke, S., Salinas, N., & Meir, P. (2019). Adaptation of soil microbial growth to temperature: Using a tropical elevation gradient to predict future changes. *Global Change Biology*, 25(3), 827-838. <https://doi.org/10.1111/gcb.14502>
- Novotny, V., Miller, S. E., Baje, L., Balagawi, S., Basset, Y., Cizek, L., Craft, K. J., Dem, F., Drew, R. A. I., Hulcr, J., Leps, J., Lewis, O. T., Pokon, R., Stewart, A. J. A., Allan Samuelson, G., & Weiblen, G. D. (2010). Guild-specific patterns of species

- richness and host specialization in plant–herbivore food webs from a tropical forest. *Journal of Animal Ecology*, 79(6), 1193-1203. <https://doi.org/10.1111/j.1365-2656.2010.01728.x>
- O'Farrell, P. J., & Anderson, P. M. L. (2010). Sustainable multifunctional landscapes: a review to implementation. *Current Opinion in Environmental Sustainability*, 2(1–2), 59-65. <https://doi.org/http://dx.doi.org/10.1016/j.cosust.2010.02.005>
- O'Hara, R. B., & Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, 1(2), 118-122. <https://doi.org/10.1111/j.2041-210X.2010.00021.x>
- Onoda, Y., Saluñga, J. B., Akutsu, K., Aiba, S.-i., Yahara, T., & Anten, N. P. R. (2014). Trade-off between light interception efficiency and light use efficiency: implications for species coexistence in one-sided light competition. *Journal of Ecology*, 102(1), 167-175. <https://doi.org/10.1111/1365-2745.12184>
- Ovaskainen, O., & Soininen, J. (2011). Making more out of sparse data: hierarchical modeling of species communities. *Ecology*, 92(2), 289-295. <https://doi.org/10.1890/10-1251.1>
- Ovaskainen, O., Tikhonov, G., Norberg, A., Blanchet, G. F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 561-576. <https://doi.org/10.1111/ele.12757>
- Overman, J. P. M., Witte, H. J. L., & Saldarriaga, J. G. (1994). Evaluation of regression models for above-ground biomass determination in Amazon rainforest. *Journal of Tropical Ecology*, 10(2), 207-218. <http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=5248636>
- Paliy, O., & Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology*, 25(5), 1032-1057. <https://doi.org/10.1111/mec.13536>
- Pan, Y., Cassman, N., de Hollander, M., Mendes, L. W., Korevaar, H., Geerts, R. H., van Veen, J. A., & Kuramae, E. E. (2014). Impact of long-term N, P, K, and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil. *FEMS microbiology ecology*, 90(1), 195-205. <http://femsec.oxfordjournals.org/content/femsec/90/1/195.full.pdf>
- Paulson, J. N., Stine, O. C., Bravo, H. C. C., & Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature methods*, 10(12), 1200-1202. <https://doi.org/10.1038/nmeth.2658>
- Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data*. John Wiley & Sons.
- Peay, K. G., Baraloto, C., & Fine, P. V. A. (2013). Strong coupling of plant and fungal community structure across western Amazonian rainforests [Original Article]. *ISME J*, 7(9), 1852-1861. <https://doi.org/10.1038/ismej.2013.66>
- Peters, H. A. (2003). Neighbour-regulated mortality: the influence of positive and negative density dependence on tree populations in species-rich tropical forests. *Ecology Letters*, 6(8), 757-765. <https://doi.org/10.1046/j.1461-0248.2003.00492.x>
- Phillips, H. R. P., Guerra, C. A., Bartz, M. L. C., Briones, M. J. I., Brown, G., Crowther, T. W., Ferlian, O., Gongalsky, K. B., van den Hoogen, J., Krebs, J., Orgiazzi, A., Routh, D., Schwarz, B., Bach, E. M., Bennett, J. M., Brose, U., Decaëns, T., König-Ries, B., Loreau, M., Mathieu, J., Mulder, C., van der Putten, W. H., Ramirez, K. S., Rillig, M. C., Russell, D., Rutgers, M., Thakur, M. P., de Vries, F. T., Wall, D. H., Wardle, D. A., Arai, M., Ayuke, F. O., Baker, G. H., Beauséjour, R., Bedano, J.

- C., Birkhofer, K., Blanchart, E., Blossey, B., Bolger, T., Bradley, R. L., Callahan, M. A., Capowiez, Y., Caulfield, M. E., Choi, A., Crotty, F. V., Crumsey, J. M., Dávalos, A., Diaz Cosin, D. J., Dominguez, A., Duhour, A. E., van Eekeren, N., Emmerling, C., Falco, L. B., Fernández, R., Fonte, S. J., Fragoso, C., Franco, A. L. C., Fugère, M., Fusilero, A. T., Gholami, S., Gundale, M. J., López, M. G., Hackenberger, D. K., Hernández, L. M., Hishi, T., Holdsworth, A. R., Holmstrup, M., Hopfensperger, K. N., Lwanga, E. H., Huhta, V., Hurisso, T. T., Iannone, B. V., Iordache, M., Joschko, M., Kaneko, N., Kanianska, R., Keith, A. M., Kelly, C. A., Kernecker, M. L., Klaminder, J., Koné, A. W., Kooch, Y., Kukkonen, S. T., Lalthanzara, H., Lammel, D. R., Lebedev, I. M., Li, Y., Jesus Lidon, J. B., Lincoln, N. K., Loss, S. R., Marichal, R., Matula, R., Moos, J. H., Moreno, G., Morón-Ríos, A., Muys, B., Neiryneck, J., Norgrove, L., Novo, M., Nuutinen, V., Nuzzo, V., Pansu, J., Paudel, S., Pérès, G., Pérez-Camacho, L., Piñeiro, R., Ponge, J.-F., Rashid, M. I., Rebollo, S., Rodeiro-Iglesias, J., Rodríguez, M. Á., Roth, A. M., Rousseau, G. X., Rozen, A., Sayad, E., van Schaik, L., Scharenbroch, B. C., Schirrmann, M., Schmidt, O., Schröder, B., Seeber, J., Shashkov, M. P., Singh, J., Smith, S. M., Steinwandter, M., Talavera, J. A., Trigo, D., Tsukamoto, J., de Valença, A. W., Vanek, S. J., Virto, I., Wackett, A. A., Warren, M. W., Wehr, N. H., Whalen, J. K., Wironen, M. B., Wolters, V., Zenkova, I. V., Zhang, W., Cameron, E. K., & Eisenhauer, N. (2019). Global distribution of earthworm diversity. *Science*, 366(6464), 480-485. <https://doi.org/10.1126/science.aax4851>
- Pinho, R. C., Alfaia, S. S. A., Miller, R. P., Uguen, K., Magalhães, L. D., Ayres, M., Freitas, V., & Trancose, R. (2010). Islands of fertility: Soil improvement under indigenous homegardens in the savannas of Roraima, Brazil. *Agroforestry Systems*, 81. <https://doi.org/10.1007/s10457-010-9336-5>
- Pinho, R. C., Alfaia, S. S. A., Miller, R. P., Uguen, K., Magalhães, L. D., Ayres, M., Freitas, V., & Trancose, R. (2011). Islands of fertility: Soil improvement under indigenous homegardens in the savannas of Roraima, Brazil. *Agroforestry Systems*, 81. <https://doi.org/10.1007/s10457-010-9336-5>
- Pinho, R. C., Miller, R. P., & Alfaia, S. S. (2012). Agroforestry and the Improvement of Soil Fertility: A View from Amazonia. *Applied and Environmental Soil Science*, 2012. <https://doi.org/10.1155/2012/616383>
- Pollini, J. (2014). Slash-and-Burn Agriculture. In P. Thompson & D. Kaplan (Eds.), *Encyclopedia of Food and Agricultural Ethics* (pp. 1648-1657). Springer Netherlands. [https://doi.org/10.1007/978-94-007-0929-4\\_87](https://doi.org/10.1007/978-94-007-0929-4_87)
- Pollock, J., Glendinning, L., Wisedchanwet, T., & Watson, M. (2018). The madness of microbiome: Attempting to find consensus “best practice” for 16S microbiome studies. *Applied and Environmental Microbiology*, 84(7), e02627-02617. <https://doi.org/10.1128/aem.02627-17>
- Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Veski, P. A., & McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5(5), 397-406. <https://doi.org/10.1111/2041-210X.12180>
- Poorter, L., Bongers, F., Aide, T. M., Almeyda Zambrano, A. M., Balvanera, P., Becknell, J. M., Boukili, V., Brancalion, P. H. S., Broadbent, E. N., Chazdon, R. L., Craven, D., de Almeida-Cortez, J. S., Cabral, G. A. L., de Jong, B. H. J., Denslow, J. S., Dent, D. H., DeWalt, S. J., Dupuy, J. M., Durán, S. M., Espirito-Santo, M. M., Fandino, M. C., César, R. G., Hall, J. S., Hernandez-Stefanoni, J. L., Jakovac, C. C.,

- Junqueira, A. B., Kennard, D., Letcher, S. G., Licona, J.-C., Lohbeck, M., Marín-Spiotta, E., Martínez-Ramos, M., Massoca, P., Meave, J. A., Mesquita, R., Mora, F., Muñoz, R., Muscarella, R., Nunes, Y. R. F., Ochoa-Gaona, S., de Oliveira, A. A., Orihuela-Belmonte, E., Peña-Claros, M., Pérez-García, E. A., Piotto, D., Powers, J. S., Rodríguez-Velázquez, J., Romero-Pérez, I. E., Ruíz, J., Saldarriaga, J. G., Sanchez-Azofeifa, A., Schwartz, N. B., Steininger, M. K., Swenson, N. G., Toledo, M., Uriarte, M., van Breugel, M., van der Wal, H., Veloso, M. D. M., Vester, H. F. M., Vicentini, A., Vieira, I. C. G., Bentos, T. V., Williamson, G. B., & Rozendaal, D. M. A. (2016). Biomass resilience of Neotropical secondary forests. *Nature*, *530*, 211. <https://doi.org/10.1038/nature16512>
- Poretzky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D., & Konstantinidis, K. T. (2014). Strengths and Limitations of 16S rRNA Gene Amplicon Sequencing in Revealing Temporal Microbial Community Dynamics. *PLoS ONE*, *9*(4), e93827. <https://doi.org/10.1371/journal.pone.0093827> PMID - 24714158
- Prosser, J. I. (2020). Putting science back into microbial ecology: a question of approach. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1798), 20190240. <https://doi.org/doi:10.1098/rstb.2019.0240>
- R Development Core Team. (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.
- Rahman, M. M., Tsukamoto, J., Rahman, M. M., Yoneyama, A., & Mostafa, K. M. (2013). Lignin and its effects on litter decomposition in forest ecosystems. *Chemistry and Ecology*, *29*(6), 540-553. <https://doi.org/10.1080/02757540.2013.790380>
- Rajala, T., & Illian, J. (2012). A family of spatial biodiversity measures based on graphs [journal article]. *Environmental and Ecological Statistics*, *19*(4), 545-572. <https://doi.org/10.1007/s10651-012-0200-9>
- Reichert, J. M., Bervald, C. M. P., Rodrigues, M. F., Kato, O. R., & Reinert, D. J. (2014). Mechanized land preparation in eastern Amazon in fire-free forest-based fallow systems as alternatives to slash-and-burn practices: Hydraulic and mechanical soil properties. *Agriculture, Ecosystems & Environment*, *192*(0), 47-60. <https://doi.org/http://dx.doi.org/10.1016/j.agee.2014.03.046>
- Rhoades, C. (1996). Single-tree influences on soil properties in agroforestry: lessons from natural forest and savanna ecosystems. *Agroforestry Systems*, *35*(1), 71-94. <https://doi.org/10.1007/bf02345330>
- Rivera-Pinto, J., Egozcue, J., Pawlowsky-Glahn, V., Paredes, R., Noguera-Julian, M., & Calle, M. (2018). Balances: a new perspective for microbiome analysis. *mSystems*, *3*(4), e00053-00018.
- Rodrigues, J. L. M., Pellizari, V. H., Mueller, R., Baek, K., Jesus, E. d. C., Paula, F. S., Mirza, B., Hamaoui, G. S., Tsai, S. M., Feigl, B., Tiedje, J. M., Bohannan, B. J. M., & Nüsslein, K. (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), 988-993. <https://doi.org/10.1073/pnas.1220608110>
- Rosindell, J., Hubbell, S. P., & Etienne, R. S. (2011). The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. *Trends in Ecology & Evolution*, *26*(7), 340-348. <https://doi.org/https://doi.org/10.1016/j.tree.2011.03.024>
- Rousk, J., Baath, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., Knight, R., & Fierer, N. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J*, *4*(10), 1340-1351. <https://doi.org/10.1038/ismej.2010.58>
- Royle, A. J. (2004). N-Mixture Models for Estimating Population Size from Spatially

- Replicated Counts. *Biometrics*, 60(1), 108-115. <https://doi.org/10.1111/j.0006-341x.2004.00142.x>
- Royle, A. J., & Nichols, J. D. (2003). Estimating abundance from repeated presence-absence data or point counts. *Ecology*, 84(3), 777-790. [https://doi.org/10.1890/0012-9658\(2003\)084\[0777:EAFRPA\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2)
- Rugalema, G. H., Okting'ati, A., & Johnsen, F. H. (1994). The homegarden agroforestry system of Bukoba district, North-Western Tanzania. 1. Farming system analysis. *Agroforestry Systems*, 26(1), 53-64. <https://doi.org/10.1007/BF00705152>
- Runyan, C. W., D'Odorico, P., & Lawrence, D. (2012). Effect of repeated deforestation on vegetation dynamics for phosphorus-limited tropical forests. *Journal of Geophysical Research: Biogeosciences* (2005–2012), 117(G1). <https://doi.org/10.1029/2011JG001841>
- Saatchi, S. S., Houghton, R. A., Dos Santos Alvalá, R. C., Soares, J. V., & Yu, Y. (2007). Distribution of aboveground live biomass in the Amazon basin. *Global Change Biology*, 13(4), 816-837. <https://doi.org/10.1111/j.1365-2486.2007.01323.x>
- Sankaran, K., & Holmes, S. (2017). Latent Variable Modeling for the Microbiome. *Latent Variable Modeling for the Microbiome*, 20(4), 599-614. [http://scholar.google.com/scholar?q=Latent Variable Modeling for the Microbiome&btnG=&hl=en&num=20&as\\_sdt=0%2C22](http://scholar.google.com/scholar?q=Latent+Variable+Modeling+for+the+Microbiome&btnG=&hl=en&num=20&as_sdt=0%2C22)
- Satyam Verma, S. J. (2012). Impact of forest fire on physical, chemical and biological properties of soil: A review. *Proceedings of the International Academy of Ecology and Environmental Sciences*, 2(3), 168-176. [http://www.iaees.org/publications/journals/piaees/articles/2012-2\(3\)/impact-of-forest-fire.pdf](http://www.iaees.org/publications/journals/piaees/articles/2012-2(3)/impact-of-forest-fire.pdf)
- Schaefer, C. E. G. R., do Amaral, E. F., de Mendonça, B. A. F., Oliveira, H., Lani, J. L., Costa, L. M., & Fernandes Filho, E. I. (2008). Soil and vegetation carbon stocks in Brazilian Western Amazonia: relationships and ecological implications for natural landscapes. *Environmental Monitoring and Assessment*, 140(1-3), 279-289. <https://doi.org/10.1007/s10661-007-9866-0>
- Schmidt, T. S. B., Rodrigues, J. F. M., & Mering, C. v. (2016). A family of interaction-adjusted indices of community similarity. *The ISME journal*, 11(3), 791-807. <https://doi.org/10.1038/ismej.2016.139>
- Schnitzer, S. A., & Carson, W. P. (2010). Lianas suppress tree regeneration and diversity in treefall gaps. *Ecology Letters*, 13(7), 849-857. <https://doi.org/10.1111/j.1461-0248.2010.01480.x>
- Schroth, G., & Mota, M. d. S. S. d. (2014). Agroforestry: Complex Multistrata Agriculture. In N. Van Alfen (Ed.), *Encyclopedia of Agriculture and Food Systems* (Vol. 1, pp. 195-207): Elsevier Science & Technology.
- Schwinning, S., & Weiner, J. (1998). Mechanisms determining the degree of size asymmetry in competition among plants. *Oecologia*, 113(4), 447-455. <https://doi.org/10.1007/s004420050397>
- Seneviratne, G., Kuruppuarachchi, K. A. C. N., Somaratne, S., & Seneviratne, K. A. C. N. (2006). Nutrient Cycling and Safety-net Mechanism in the Tropical Homegardens. *International Journal of Agricultural Research*, 1(2), 169-182. <https://doi.org/10.3923/ijar.2006.169.182>
- Shoemaker, W. R., Locey, K. J., & Lennon, J. T. (2017). A macroecological theory of microbial biodiversity [Article]. *1*, 0107. <https://doi.org/10.1038/s41559-017-0107> <https://www.nature.com/articles/s41559-017-0107#supplementary-information>

## References

---

- Silver, W. L., Ostertag, R., & Lugo, A. E. (2000). The Potential for Carbon Sequestration Through Reforestation of Abandoned Tropical Agricultural and Pasture Lands. *Restoration Ecology*, 8(4), 394-407. <https://doi.org/10.1046/j.1526-100x.2000.80054.x>
- Silverman, J. D., Roche, K., Mukherjee, S., & David, L. A. (2018). Naught all zeros in sequence count data are the same. *bioRxiv*, 477794.
- Silverman, J. D., Shenhav, L., Halperin, E. A., Mukherjee, S. A., & David, L. A. (2018). Statistical considerations in the design and analysis of longitudinal microbiome studies. *bioRxiv*, 448332.
- Six, J., Conant, R. T., Paul, E. A., & Paustian, K. (2002). Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant and Soil*, 241(2), 155-176. <https://doi.org/10.1023/a:1016125726789>
- Slik, J. W. F., Arroyo-Rodríguez, V., Aiba, S.-I., Alvarez-Loayza, P., Alves, L. F., Ashton, P., Balvanera, P., Bastian, M. L., Bellingham, P. J., van den Berg, E., Bernacci, L., da Conceição Bispo, P., Blanc, L., Böhning-Gaese, K., Boeckx, P., Bongers, F., Boyle, B., Bradford, M., Brearley, F. Q., Breuer-Ndoundou Hockemba, M., Bunyavejchewin, S., Calderado Leal Matos, D., Castillo-Santiago, M., Catharino, E. L. M., Chai, S.-L., Chen, Y., Colwell, R. K., Robin, C. L., Clark, C., Clark, D. B., Clark, D. A., Culmsee, H., Damas, K., Dattaraja, H. S., Dauby, G., Davidar, P., DeWalt, S. J., Doucet, J.-L., Duque, A., Durigan, G., Eichhorn, K. A. O., Eisenlohr, P. V., Eler, E., Ewango, C., Farwig, N., Feeley, K. J., Ferreira, L., Field, R., de Oliveira Filho, A. T., Fletcher, C., Forshed, O., Franco, G., Fredriksson, G., Gillespie, T., Gillet, J.-F., Amarnath, G., Griffith, D. M., Grogan, J., Gunatilleke, N., Harris, D., Harrison, R., Hector, A., Homeier, J., Imai, N., Itoh, A., Jansen, P. A., Joly, C. A., de Jong, B. H. J., Kartawinata, K., Kearsley, E., Kelly, D. L., Kenfack, D., Kessler, M., Kitayama, K., Kooyman, R., Larney, E., Laumonier, Y., Laurance, S., Laurance, W. F., Lawes, M. J., Amaral, I. L. d., Letcher, S. G., Lindsell, J., Lu, X., Mansor, A., Marjokorpi, A., Martin, E. H., Meilby, H., Melo, F. P. L., Metcalfe, D. J., Medjibe, V. P., Metzger, J. P., Millet, J., Mohandass, D., Montero, J. C., de Morisson Valeriano, M., Mugerwa, B., Nagamasu, H., Nilus, R., Ochoa-Gaona, S., Onrizal, Page, N., Parolin, P., Parren, M., Parthasarathy, N., Paudel, E., Permana, A., Piedade, M. T. F., Pitman, N. C. A., Poorter, L., Poulsen, A. D., Poulsen, J., Powers, J., Prasad, R. C., Puyravaud, J.-P., Razafimahaimodison, J.-C., Reitsma, J., dos Santos, J. R., Roberto Spironello, W., Romero-Saltos, H., Rovero, F., Rozak, A. H., Ruokolainen, K., Rutishauser, E., Saiter, F., Saner, P., Santos, B. A., Santos, F., Sarker, S. K., Satdichanh, M., Schmitt, C. B., Schöngart, J., Schulze, M., Suganuma, M. S., Sheil, D., da Silva Pinheiro, E., Sist, P., Stevart, T., Sukumar, R., Sun, I.-F., Sunderand, T., Suresh, H. S., Suzuki, E., Tabarelli, M., Tang, J., Targhetta, N., Theilade, I., Thomas, D. W., Tchouto, P., Hurtado, J., Valencia, R., van Valkenburg, J. L. C. H., Van Do, T., Vasquez, R., Verbeeck, H., Adekunle, V., Vieira, S. A., Webb, C. O., Whitfeld, T., Wich, S. A., Williams, J., Wittmann, F., Wöll, H., Yang, X., Adou Yao, C. Y., Yap, S. L., Yoneda, T., Zahawi, R. A., Zakaria, R., Zang, R., de Assis, R. L., Garcia Luize, B., & Venticinque, E. M. (2015). An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1423147112>
- Smith, L. M., & Reynolds, H. L. (2015). Plant–soil feedbacks shift from negative to positive with decreasing light in forest understory species. *Ecology*, 96(9), 2523-2532. <https://doi.org/10.1890/14-2150.1>

- Soane, B. D. (1990). The role of organic matter in soil compactibility: A review of some practical aspects. *Soil and Tillage Research*, 16(1–2), 179–201. [https://doi.org/http://dx.doi.org/10.1016/0167-1987\(90\)90029-D](https://doi.org/http://dx.doi.org/10.1016/0167-1987(90)90029-D)
- Sohn, M. B., & Li, H. (2017). A GLM-based latent variable ordination method for microbiome samples. *Biometrics*, 74(2), 448–457. <https://doi.org/10.1111/biom.12775> PMID - 28991375
- Somarriba, E., Cerda, R., Orozco, L., Cifuentes, M., Dávila, H., Espin, T., Mavisoy, H., Ávila, G., Alvarado, E., Poveda, V., Astorga, C., Say, E., & Deheuvels, O. (2013). Carbon stocks and cocoa yields in agroforestry systems of Central America. *Agriculture, Ecosystems & Environment*, 173(0), 46–57. <https://doi.org/http://dx.doi.org/10.1016/j.agee.2013.04.013>
- Souza, H., Graaff, J., & Pulleman, M. (2012). Strategies and economics of farming systems with coffee in the Atlantic Rainforest Biome. *Agroforestry Systems*, 84(2), 227–242. <https://doi.org/10.1007/s10457-011-9452-x>
- Styger, E., Rakotondramasy, H. M., Pfeffer, M. J., Fernandes, E. C. M., & Bates, D. M. (2007). Influence of slash-and-burn farming practices on fallow succession and land degradation in the rainforest region of Madagascar. *Agriculture, Ecosystems & Environment*, 119(3–4), 257–269. <https://doi.org/10.1016/j.agee.2006.07.012>
- Taubert, F., Fischer, R., Groeneveld, J., Lehmann, S., Müller, M. S., Rödiger, E., Wiegand, T., & Huth, A. (2018). Global patterns of tropical forest fragmentation. *Nature*, 554(7693), 519–522. <https://doi.org/10.1038/nature25508>
- Taylor-Rodríguez, D., Kaufeld, K., Schliep, E. M., Clark, J. S., & Gelfand, A. E. (2016). Joint Species Distribution Modeling: Dimension Reduction Using Dirichlet Processes. *Bayesian Analysis*, 12(2), 939–967. <https://doi.org/10.1214/16-ba1031>
- Tedersoo, L., Bahram, M., & Zobel, M. (2020). How mycorrhizal associations drive plant population and community biology. *Science*, 367(6480), eaba1223. <https://doi.org/10.1126/science.aba1223>
- Tedesco, M. J., Gianello, C., Bissani, C. A., Bohnen, H., & Volkweiss, S. J. (1995). *Análise de Solo, Plantas e Outros Materiais* (2<sup>a</sup> ed.). UFRGS.
- Ter Braak, C. J. F., & Prentice, I. C. (1988). A Theory of Gradient Analysis. In M. Begon, A. H. Fitter, E. D. Ford, & A. Macfadyen (Eds.), *Advances in Ecological Research* (Vol. 18, pp. 271–317). Academic Press. [https://doi.org/https://doi.org/10.1016/S0065-2504\(08\)60183-X](https://doi.org/https://doi.org/10.1016/S0065-2504(08)60183-X)
- Ter Steege, H., Pitman, N., Sabatier, D., Castellanos, H., Van Der Hout, P., Daly, D., Silveira, M., Phillips, O., Vasquez, R., Van Andel, T., Duivenvoorden, J., De Oliveira, A., Ek, R., Lilwah, R., Thomas, R., Van Essen, J., Baider, C., Maas, P., Mori, S., Terborgh, J., Núñez Vargas, P., Mogollón, H., & Morawetz, W. (2003). A spatial model of tree  $\alpha$ -diversity and tree density for the Amazon. *Biodiversity & Conservation*, 12(11), 2255–2277. <https://doi.org/10.1023/A:1024593414624>
- Thiele-Bruhn, S., Bloem, J., de Vries, F. T., Kalbitz, K., & Wagg, C. (2012). Linking soil biodiversity and agricultural soil management. *Current Opinion in Environmental Sustainability*, 4(5), 523–528. <https://doi.org/http://dx.doi.org/10.1016/j.cosust.2012.06.004>
- Thomaz, E. L. (2013). Slash-and-burn agriculture: Establishing scenarios of runoff and soil loss for a five-year cycle. *Agriculture, Ecosystems & Environment*, 168(0), 1–6. <https://doi.org/http://dx.doi.org/10.1016/j.agee.2013.01.008>
- Tittonell, P., Muriuki, A., Klapwijk, C. J., Shepherd, K. D., Coe, R., & Vanlauwe, B. (2013). Soil Heterogeneity and Soil Fertility Gradients in Smallholder Farms of the East

## References

---

- African Highlands. *Soil Sci. Soc. Am. J.*, 77(2), 525-538. <https://doi.org/10.2136/sssaj2012.0250>
- Too, C. C., Keller, A., Sickel, W., Lee, S. M., & Yule, C. M. (2018). Microbial Community Structure in a Malaysian Tropical Peat Swamp Forest: The Influence of Tree Species and Depth [Original Research]. *Frontiers in Microbiology*, 9(2859). <https://doi.org/10.3389/fmicb.2018.02859>
- Tu, L.-H., Hu, H.-L., Hu, T.-X., Zhang, J., Liu, L., Li, R.-H., Dai, H.-Z., & Luo, S.-H. (2011). Decomposition of Different Litter Fractions in a Subtropical Bamboo Ecosystem as Affected by Experimental Nitrogen Deposition. *Pedosphere*, 21(6), 685-695. [https://doi.org/http://dx.doi.org/10.1016/S1002-0160\(11\)60171-9](https://doi.org/http://dx.doi.org/10.1016/S1002-0160(11)60171-9)
- Uriarte, M., Turner, B. L., Thompson, J., & Zimmerman, J. K. (2015). Linking spatial patterns of leaf litterfall and soil nutrients in a tropical forest: a neighborhood approach. *Ecological Applications*, 25(7), 2022-2034. <https://doi.org/10.1890/15-0112.1>
- USDA, S. S. S. (2010). *Keys to Soil Taxonomy* (11 ed.). Department of Agriculture Natural Resources Conservation. [http://soils.usda.gov/technical/classification/tax\\_keys/](http://soils.usda.gov/technical/classification/tax_keys/)
- Usinowicz, J., Chang-Yang, C.-H., Chen, Y.-Y., Clark, J. S., Fletcher, C., Garwood, N. C., Hao, Z., Johnstone, J., Lin, Y., Metz, M. R., Masaki, T., Nakashizuka, T., Sun, I. F., Valencia, R., Wang, Y., Zimmerman, J. K., Ives, A. R., & Wright, J. S. (2017). Temporal coexistence mechanisms contribute to the latitudinal gradient in forest diversity. *Nature*, 550(7674), 105-108. <https://doi.org/10.1038/nature24038>
- van der Putten, W. H., Bardgett, R. D., Bever, J. D., Bezemer, T. M., Casper, B. B., Fukami, T., Kardol, P., Klironomos, J. N., Kulmatiski, A., Schweitzer, J. A., Suding, K. N., Van de Voorde, T. F. J., & Wardle, D. A. (2013). Plant–soil feedbacks: the past, the present and future challenges. *Journal of Ecology*, 101(2), 265-276. <https://doi.org/10.1111/1365-2745.12054>
- Van Noordwijk, M., & Garrity, D. P. (1995). Nutrient use efficiency in agroforestry systems. In *Potassium in Asia, Balanced Fertilization to Increase and Sustain Agricultural Production* (pp. 245-279). International Potash Institute. <http://www.icraf.org/Sea/Publications/files/paper/PP0034-04.PDF>
- Van Wagner, C. E. (1968). The Line Intersect Method in Forest Fuel Sampling. *Forest Science*, 14(1), 20-26. <http://www.ingentaconnect.com/content/saf/fs/1968/00000014/00000001/art00007>
- Varma, A. (2003). The economics of slash and burn: a case study of the 1997–1998 Indonesian forest fires. *Ecological Economics*, 46(1), 159-171. [https://doi.org/http://dx.doi.org/10.1016/S0921-8009\(03\)00139-3](https://doi.org/http://dx.doi.org/10.1016/S0921-8009(03)00139-3)
- Velázquez, E., Martínez, I., Getzin, S., Moloney, K. A., & Wiegand, T. (2016). An evaluation of the state of spatial point pattern analysis in ecology. *Ecography*, 39(11), 1042-1055. <https://doi.org/10.1111/ecog.01579>
- Velázquez, E., & Wiegand, T. (2020). Competition for light and persistence of rare light-demanding species within tree-fall gaps in a moist tropical forest. *Ecology*, n/a(n/a), e03034. <https://doi.org/10.1002/ecy.3034>
- Verbeke, G., & Molenberghs, G. (2017). Modeling Through Latent Variables. *Annual Review of Statistics and Its Application*, 4(1), 267-282. <https://doi.org/10.1146/annurev-statistics-060116-054017>
- Walkley, A., & Black, I. A. (1934). An Examination of the Degtjareff Method for Determining Soil Organic Matter, and A Proposed Modification of the Chromic Acid Titration Method. *Soil Science*, 37(1), 29-38.

- [http://journals.lww.com/soilsci/Fulltext/1934/01000/An\\_Examination\\_of\\_the\\_Deg\\_tjareff\\_Method\\_for.3.aspx](http://journals.lww.com/soilsci/Fulltext/1934/01000/An_Examination_of_the_Deg_tjareff_Method_for.3.aspx)
- Wang, H., Liu, S.-R., Mo, J.-M., Wang, J.-X., Makeschin, F., & Wolff, M. (2010). Soil organic carbon stock and chemical composition in four plantations of indigenous tree species in subtropical China. *Ecological Research*, 25(6), 1071-1079. <https://doi.org/10.1007/s11284-010-0730-2>
- Warton, D. I. (2005). Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics*, 16(3), 275-289. <https://doi.org/10.1002/env.702>
- Warton, D. I. (2017). Why you cannot transform your way out of trouble for small counts. *Biometrics*, 74(1), 362-368. <https://doi.org/10.1111/biom.12728>
- Warton, D. I., Blanchet, G. F., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution*, 30(12), 766-779. <https://doi.org/10.1016/j.tree.2015.09.007>
- Warton, D. I., Foster, S. D., De'ath, G., Stoklosa, J., & Dunstan, P. K. (2015). Model-based thinking for community ecology [journal article]. *Plant Ecology*, 216(5), 669-682. <https://doi.org/10.1007/s11258-014-0366-3>
- Warton, D. I., & Hui, F. K. C. (2017). The central role of mean-variance relationships in the analysis of multivariate abundance data: a response to Roberts (2017). *Methods in Ecology and Evolution*, 8(11), 1408-1414. <https://doi.org/doi:10.1111/2041-210X.12843>
- Warton, D. I., Lyons, M., Stoklosa, J., & Ives, A. R. (2016). Three points to consider when choosing a LM or GLM test for count data. *Methods in Ecology and Evolution*, 7(8), 882-890. <https://doi.org/10.1111/2041-210X.12552>
- Warton, D. I., Wright, S. T., & Wang, Y. (2012). Distance-based multivariate analyses confound location and dispersion effects. *Methods in Ecology and Evolution*, 3(1), 89-101. <https://doi.org/10.1111/j.2041-210X.2011.00127.x>
- Weiss, S., Treuren, W. V., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L. C., Xu, Z. Z., Ursell, L., Alm, E. J., Birmingham, A., Cram, J. A., Fuhrman, J. A., Raes, J., Sun, F., Zhou, J., & Knight, R. (2016). Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *The ISME journal*, 10(7), 1669-1681. <https://doi.org/10.1038/ismej.2015.235>
- Werner, G. D., Cornwell, W. K., Sprent, J. I., Kattge, J., & Kiers, E. T. (2014). A single evolutionary innovation drives the deep evolution of symbiotic N<sub>2</sub>-fixation in angiosperms. *Nature Communications*, 5, 4087.
- Wiegand, T., Gunatilleke, C. V. S., Gunatilleke, I. A. U. N., & Huth, A. (2007). How individual species structure diversity in tropical forests. *Proceedings of the National Academy of Sciences*, 104(48), 19029-19033. <https://doi.org/10.1073/pnas.0705621104>
- Wiegand, T., May, F., Kazmierczak, M., & Huth, A. (2017). What drives the spatial distribution and dynamics of local species richness in tropical forest? *Proceedings of the Royal Society B: Biological Sciences*, 284(1863), 20171503. <https://doi.org/doi:10.1098/rspb.2017.1503>
- Wiegand, T., & Moloney, K. A. (2004). Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, 104(2), 209-229. <https://doi.org/10.1111/j.0030-1299.2004.12497.x>
- Wiegand, T., & Moloney, K. A. (2013). *Handbook of spatial point-pattern analysis in*

- ecology*. CRC Press.
- WinklerPrins, A. (2009). Sweep and Char and the Creation of Amazonian Dark Earths in Homegardens. In W. Woods, W. Teixeira, J. Lehmann, C. Steiner, A. WinklerPrins, & L. Rebellato (Eds.), *Amazonian Dark Earths: Wim Sombroek's Vision* (pp. 205-211). Springer Netherlands. [https://doi.org/10.1007/978-1-4020-9031-8\\_10](https://doi.org/10.1007/978-1-4020-9031-8_10)
- Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., Midgley, J. J., Navas, M.-L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V. I., Roumet, C., Thomas, S. C., Tjoelker, M. G., Veneklaas, E. J., & Villar, R. (2004). The worldwide leaf economics spectrum [10.1038/nature02403]. *Nature*, 428(6985), 821-827. [https://doi.org/http://www.nature.com/nature/journal/v428/n6985/suppinfo/nature02403\\_S1.html](https://doi.org/http://www.nature.com/nature/journal/v428/n6985/suppinfo/nature02403_S1.html)
- Xia, F., Chen, J., Fung, W. K., & Li, H. (2013). A Logistic Normal Multinomial Regression Model for Microbiome Compositional Data Analysis. *Biometrics*, 69(4), 1053-1063. <https://doi.org/10.1111/biom.12079>
- Yamada, M., & Gholz, H. L. (2002). An evaluation of agroforestry systems as a rural development option for the Brazilian Amazon [journal article]. *Agroforestry Systems*, 55(2), 81-87. <https://doi.org/10.1023/a:1020523107243>
- Yamaura, Y., K?ry, M., & Royle, A. J. (2016). Study of biological communities subject to imperfect detection: bias and precision of community N-mixture abundance models in small-sample situations. *Ecological Research*, 31(3), 289-305. <https://doi.org/10.1007/s11284-016-1340-4>
- Yan, Y., Klinkhamer, P. G., van Veen, J. A., & Kuramae, E. E. (2019). Environmental filtering: a case of bacterial community assembly in soil. *Soil Biology and Biochemistry*, 136, 107531.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. 2008, 27(8), 25. <https://doi.org/10.18637/jss.v027.i08>
- Zhang, K., Cheng, X., Dang, H., Ye, C., Zhang, Y., & Zhang, Q. (2013). Linking litter production, quality and decomposition to vegetation succession following agricultural abandonment. *Soil Biology and Biochemistry*, 57(0), 803-813. <https://doi.org/http://dx.doi.org/10.1016/j.soilbio.2012.08.005>
- Zhou, X., Leite, M. F. A., Zhang, Z., Tian, L., Chang, J., Ma, L., Li, X., van Veen, J. A., Tian, C., & Kuramae, E. E. (2021). Facilitation in the soil microbiome does not necessarily lead to niche expansion. *Environmental Microbiome*, 16(1), 4. <https://doi.org/10.1186/s40793-021-00373-2>

## Summary

Plant communities affect the soil subsystem and vice-versa. Anthropogenic management activities impact both plants and soil, thereby interfering with the co-dependence between plant, soil, and microbial components. Land management practices are known to interfere with both above and belowground components of terrestrial ecosystems, but we still know very little about how land management impacts the combination of aboveground and belowground sub-systems and their linkages with each other. In this thesis, I focus on identifying potential land uses that induce or increase the co-dependence between plants and soil microorganisms by evaluating (1) successional changes in plant community assemblages in both taxonomic and functional composition; and (2) the effects of anthropogenic management on above- and belowground interactions between plant biomass and soil communities in two types of land-use systems in the Amazon forest: natural fallow after slash-and-burn agriculture and agroforestry systems. In this thesis, I took advantage of many different approaches (plant spatial arrangement, estimation of aboveground biomass, analysis of soil physico-chemical properties, and soil metagenomics) to investigate the influence of land use and anthropogenic management on aboveground and belowground communities and their interactions.

With respect to plant-plant interactions (**Chapter 2**), I investigated successional dynamics using a chronosequence approach covering 2-25 years of regeneration after first-cycle shifting cultivation and compare these with mature rainforests. Additionally, I considered the impact of land use intensification with three second-cycle slash-and-burn forests ('degraded' sites). The results revealed that land-use intensification (2<sup>nd</sup>-cycle regrowth) strongly affected all aspects of the spatial organization of secondary vegetation, increasing clustering, co-occurrences of functionally distinct plants, and systematic increase in focal species' negative impacts on surrounding legume diversity. Thus, land-use intensification affects

spatial organization of self-regenerative vegetation far more than first-cycle secondary succession does. Therefore, a second-cycle of slash-and-burn agriculture is enough to induce a stronger clustering of leguminous plants at both univariate and bivariate interactions.

When looking into plant-soil relationships (**Chapter 3**), I evaluated three types of rainforests (young, old, and mature rainforest) and three different agroforestry systems (enriched fallow, commercial plantation, and homegardens). I found a strong impact of agroforestry systems on both soil factors and aboveground biomass. Interestingly, I observed an increasing co-dependence between soil factors and plant biomass in successional systems, but not in the agroforests. These results suggested that agricultural practices in the agroforests disrupt the co-dependence between aboveground-belowground interactions.

Finally, I added the soil microbiome to the picture of aboveground-belowground interactions by taking advantage of advanced statistical methods (reviewed in **Chapters 4**). The results revealed a fungal-driven relationship along succession and in mature forest (**Chapter 5**). On the other hand, agroforestry systems exhibited a weaker co-dependence between plant biomass and soil factors, and their soil microbiomes appear to promote more specific bacterial populations as opposed to fungal taxa.

In conclusion, integrating high-resolution genomic data with complex environmental data in observational studies is an important step towards identifying ecological differentiation and shifts in aboveground-belowground interactions. Only a combined approach that accounts for the plants' biomass, the soil factors, and the soil microbial community allows one to disentangle the relative importance of each component of the system - a central challenge for the sustainability of land use in the Amazon.

## Samenvatting (Dutch summary)

Plantengemeenschappen beïnvloeden het bodemsubstelsysteem en vice-versa. Antropogeen landbeheer heeft invloed op zowel planten als bodem, waarbij de samenhang tussen plant, bodem en microbiële componenten wordt verstoord. Het is bekend dat landbeheerpraktijken interfereren met zowel boven- als ondergrondse componenten van terrestrische ecosystemen, maar we weten nog steeds heel weinig over hoe landbeheer de combinatie van bovengrondse en ondergrondse subsystemen en hun onderlinge verbanden beïnvloedt. Mijn proefschrift richt zich op het identificeren van landbouwbeheer maatregelen die de samenhang tussen planten en bodemmicro-organismen kan behouden of vergroten door het evalueren van (1) successie in de samenstelling van plantengemeenschappen in zowel taxonomie als functie; en (2) de effecten van landbouwbeheer op boven- en ondergrondse interacties tussen plantenbiomassa en bodemgemeenschappen in twee landgebruikssystemen in het Amazonewoud: natuurlijke braaklegging na slash-and-burn landbouw en agroforestry. In dit proefschrift, heb ik gebruik gemaakt van verschillende benaderingen (ruimtelijke ordening van planten, schatting van bovengrondse biomassa, analyse van fysisch-chemische eigenschappen van de bodem, en bodemmetagenomics) om de invloed van landgebruik en landbeheer op boven- en ondergrondse gemeenschappen en hun interacties te onderzoeken.

Met betrekking tot de plant-plant interacties (**Hoofdstuk 2**), heb ik de successiedynamiek van vegetatie in een chronosequentie van velden 2-25 jaar na de eerste-cyclus van zwerfandbouw onderzocht en deze regeneratie fase vergeleken met volgroeide regenwouden. Bovendien onderzocht ik de impact van intensivering van landgebruik met drie tweede-cyclus slash-and-burn bossen ('gedegradeerde' sites). Uit de resultaten is gebleken dat intensivering van landgebruik (hergroei in de tweede-cyclus) een sterke invloed had op alle aspecten van de ruimtelijke organisatie van de secundaire vegetatie, met een toename van de clustering, het naast elkaar voorkomen van functioneel verschillende planten, en een systematische toename van

de negatieve invloed van sleutelsoorten op de omringende diversiteit van leguminosen. Intensivering van landgebruik heeft dus veel meer invloed op de ruimtelijke organisatie van regenereerde vegetatie dan secundaire successie tijdens de eerste cyclus. Daarom is een tweede cyclus van “slash-and-burn” landbouw voldoende om een sterkere clustering van vlinderbloemigen teweeg te brengen bij zowel univariate als bivariate interacties.

Bij het onderzoeken van plant-bodem relaties (**Hoofdstuk 3**), concentreerde ik op drie typen regenwoud (jong, oud en tropisch regenwoud) en drie verschillende agroforestry systemen (verrijkte braak, commerciële plantage, en huis-tuin-en-keuken). Ik vond een sterke impact van agroforestry systemen op zowel bodemfactoren als bovengrondse biomassa. Interessant is dat ik een toenemende co-dependentie tussen bodemfactoren en plantenbiomassa waarnam in successiesystemen, maar niet in de agroforestry systemen. Deze resultaten suggereren dat de landbouwpraktijken in de agroforestry systemen de co-dependentie tussen de bovengrondse en ondergrondse interacties verstoren.

Tenslotte, heb ik het bodem-microbioom opgenomen in het plaatje van bovengrondse-ondergrondse interacties door gebruik te maken van geavanceerde statistische methoden (besproken in **Hoofdstuk 4**). De resultaten lieten een schimmel-gedreven relatie zien tijdens successie en in volgroeid tropisch bos (**Hoofdstuk 5**). In tegenstelling tot de successiereeks en volgroeid tropisch bos, vertoonden agroforestry systemen een zwakkere co-dependentie tussen plantenbiomassa en bodemfactoren, en bodem-microbiomen waarin specifieke bacteriële populaties werden bevorderd.

Concluderend kan worden gesteld dat het integreren van hoge-resolutie genomische gegevens met complexe milieugegevens in observationele studies een belangrijke stap kan zijn in het identificeren van ecologische differentiatie en verschuivingen in bovengrondse-ondergrondse interacties. Een gecombineerde aanpak, die rekening houdt met de biomassa van de planten, bodemfactoren en de bodemmicrobiële gemeenschap, maakt het mogelijk om het relatieve belang van elke

component van het systeem te ontrafelen - een centrale uitdaging voor de verduurzaming van het landgebruik in het Amazone.

## Acknowledgments

In a continuous probability function, the probability at a single point is always zero. This means that the likelihood of a single event is only considered over an interval. In other words, an event in our life can be mistaken as impossible because we can only consider it impossible under the light of other similar events. This sounds like a good lesson from statistics to carry for life: only when you put several events in perspective can you see how remarkable life is. This Acknowledgments section is to thank the people that helped and supported me during the several events of my life that made this thesis happen.

I thank my promoters, Prof. George Kowalchuk and Prof. Eiko Kuramae, for giving me the chance to experience all the events that led to this thesis. An especial thanks to Prof. Eiko Kuramae, who decided to collect the poster of Prof. Tsai during the Brazilian Congress of Microbiology and decided to give her card to a tall guy with some nice pictures of soil microbes. This small event made all this story possible.

I thank my former supervisors, Prof. Christoph Gehring, Prof. Guillaume Xavier, and Prof. Flávio Henrique Reis Moraes, for the attention and dedicated effort that fostered my academic development.

To my beloved wife Raycenne Rosa Leite for her company during past and present events, especially when I become the most boring person in the world. Thanks for talking with me about things other than scientific articles and statistics, for becoming a demi-goddess of New Ether, and for being the person who best took care of my books.

An especial thanks to my paranymphs, Cristina Rotoni, and Mauricio Dimitrov, with I had the honor to share several funny events. From the moment I arrived at NIOO, Mauricio was always available to help and advise me on life and my career. Later, I was lucky enough to meet Cristina, who lived close to Spoleto without knowing that this name is famous in Brazil due to a chain of Brazilian

restaurants that developed the concept of fast food to Italian cuisine. Such unlikely event ensured several jokes and moments of fun.

An especial thanks to the Kuramae's group. Thanks to Afnan Suleiman, Ohana Costa, and Késia Lourenço, the golden trio of Microbial Ecology. Thanks to Thiago Schlempler, Adriano Reis Lucheta, Matheus Cipriano, and Marcelo Zerillo, the old guard of Kuramae's group, for the amazing stories. Thanks to Mattias de Hollander and Noriko Cassman, who though me bioinformatics and were always available to help me during my mistakes.

A big thanks to all my colleagues from the Microbial Ecology Department at NIOO, especially, Victor Carrion, Viviane Cordovez, Desalegn Etalo, Juan Jaramillo. To all my friends from São Paulo that I ended up meeting in the Netherlands: Luiz Moretti, João Bossolani, Letusa Momesso, Rafaela Bugarelli, Mariana Cia. To my Chinese friends Han Wang, Jingjing Chang, and Zhikang Wang, and all the colleagues from the Northeast Institute of Geography and Agroecology in Changchun, China, especially to Prof. Chunjie Tian, for the amazing time we spend together and fantastic collaborations.

A big thanks to all the Senior Scientists from the ME, Paul Bodelier, Paolina Garbeva, Mark Zwart, Wietse de Boer, and Jos Raaijmakers. Your brilliant science and opinions were always a source of inspiration and critical analysis of my work and career. I also thank Riks Laanbroek and Hans van Veen for giving me the opportunity to learn with them and collaborate in scientific studies.

To my friend Ronildson Lima Luz, lab mate, ideas' buddy, time traveler, ex-chemist of the perfect dodge, Doorbreaker, and orc of Gruumsh, a hero during the weekends and a friend in this world and all the others. To André Luiz Raposo Barros, the coolest friend I ever made, who of the three spirits inherited the one that yields the best stories. To my illustrious colleague Roberto Lima da Silva (*in memoriam*), for shortening long walks and speeding up long waits, friends who, with simple conversations, can change the curvature of spacetime are rare.

To all my classmates from the Agroecology Program at the Maranhão State

University: Alexandra Rocha da Piedade, Elizabeth Araújo Costa, Júlio César Bravo Medina, Leticia Moura Ramos, Hulda Rocha e Silva , Maria de Jesus de Sousa Silva, Marta Jordana Arruda Coelho, Pedro Ivo Menezes Bitú, Robson Jardellys de Souza Maciel, Rozalino Antônio Aguiar Júnior, Servio Robinson Cabezas Morillo, Stefania Pinzón Triana, Vinícius Ribamar Alencar Macedo, and Virley Gardeny Lima Sena, for their support in assignments, fieldwork, lab work, and for sharing moments of suffering, effort, and joy. To all the employees of the Agroecology Program, for all the help, smiles, and coffee, and to the other professors of the Program, for the extremely enriching classes that helped me survive the entire process that was writing this thesis. I also want to mention all the farmers who shared their knowledge and allowed this moment to come true. I hope one day I will help you as much as you helped me.

To all the warriors, barbarians, rangers, thieves, mages, wizards, druids, bards, and clerics who lived, fought, lied, died, and did not even exist but made life in the real world enjoyable.

Last and most importantly, I thank my family for yesterday, today, and forever.

## About the author

My name is Márcio Fernandes Alves Leite, I was born on 29 August 1989 in São Luís, Maranhão, Brazil. In 2007, I started to study Environmental Engineering at CEUMA University. During my studies, I met Prof. Dr. Flávio Henrique Reis Moraes, from whom I learned about the ecology of microbes and their importance for sustainability. Willing to continue learning about this topic, I asked him to participate in one of his research projects. Since then, I got involved with



microbiology for the sustainability of agriculture. I got my Environmental engineer degree in 2011 after studying the impact of biochar in the soil fungal community. After presenting this work as a poster during the 26<sup>th</sup> Brazilian Congress of Microbiology in 2011, I met Prof. Dr. Eiko Kuramae. In 2012, I was admitted in the Master Program of Agroecology at the Maranhão, under the supervision of Prof. Dr Christoph Gehring, where I studied how the agricultural practices impacted the aboveground belowground relationship between plants and soil factors. In 2014, I started my Ph.D. research at the Maranhão State University. In 2015, I was awarded with a scholarship to visit the Netherlands Institute of Ecology as part of a sandwich program for one year. In 2019, I returned to the Netherlands as part of the TKI project SMART at the Department of Microbial Ecology of the Netherlands Institute of Ecology. Later, I was invited to join the Ph.D. Program at Ecology and Biodiversity research group at Utrecht University having Prof. George Kowalchuk and Prof. Eiko Kuramae as promoters. My research focus on the impact of agriculture on the aboveground-belowground interactions between plant, soil factors, and the soil microbiome.

## Publication List

Momesso, L., Crusciol, C. A., **Leite, M. F. A.**, Bossolani, J. W., & Kuramae, E. E. (2022). Forage Grasses Steer Soil Nitrogen Processes, Microbial Populations, and Microbiome Composition in A Long-term Tropical Agriculture System. *Agriculture, Ecosystems & Environment*, 323, 107688.

**Leite, M. F. A.**, Dimitrov, M. R., Freitas-Iório, R. P., de Hollander, M., Cipriano, M. A., Andrade, S. A., ... & Kuramae, E. E. (2021). Rearranging the sugarcane holobiont via plant growth-promoting bacteria and nitrogen input. *Science of the Total Environment*, 800, 149493.

Zhou, X., **Leite, M. F. A.**, Zhang, Z., Tian, L., Chang, J., Ma, L., ... & Kuramae, E. E. (2021). Facilitation in the soil microbiome does not necessarily lead to niche expansion. *Environmental microbiome*, 16(1), 1-13.

Bossolani, J. W., Crusciol, C. A., **Leite, M. F. A.**, Merloti, L. F., Moretti, L. G., Pascoaloto, I. M., & Kuramae, E. E. (2021). Modulation of the soil microbiome by long-term Ca-based soil amendments boosts soil organic carbon and physicochemical quality in a tropical no-till crop rotation system. *Soil Biology and Biochemistry*, 156, 108188.

Oyserman, B. O., Cordovez, V., Flores, S. S., **Leite, M. F. A.**, Nijveen, H., Medema, M. H., & Raaijmakers, J. M. (2021). Extracting the GEMs: Genotype, Environment, and Microbiome Interactions Shaping Host Phenotypes. *Frontiers in microbiology*, 11, 3444.

Chang, J., Shi, S., Tian, L., **Leite, M. F. A.**, Chang, C., Ji, L., ... & Kuramae, E. E. (2021). Self-crossing leads to weak co-variation of the bacterial and fungal communities in the rice rhizosphere. *Microorganisms*, 9(1), 175.

**Leite, M. F. A.**, & Kuramae, E. E. (2020). You must choose, but choose wisely: model-based approaches for microbial community analysis. *Soil Biology and Biochemistry*, 151, 108042.

Luz, R. L., **Leite, M. F. A.**, Zelarayán, M. C., Boddey, R. M., & Gehring, C. (2020). Litter decomposition and nutrient release dynamics of leaves and roots of the babassu palm in eastern Amazonia. *Acta Amazonica*, 50(3), 213–222.

Gehring, C., Zelarayán, M. C., Luz, R. L., Almeida, R. B., Boddey, R. M., & **Leite, M. F. A.** (2020). Babassu palm (*Attalea speciosa* Mart.) super-dominance shapes its surroundings via multiple biotic, soil chemical, and physical interactions and accumulates soil carbon: a case study in eastern Amazonia. *Plant and Soil*, 454(1), 447-460.

Rocha, K. F., Kuramae, E. E., Borges, B. M. F., **Leite, M. F. A.**, & Rosolem, C. A. (2020). Microbial N-cycling gene abundance is affected by cover crop specie and development stage in an integrated cropping system. *Archives of Microbiology*, 202, 2005-2012.

Oyserman, B. O., Cordovez, V., Flores, S. S., **Leite, M. F. A.**, Nijveen, H., Medema, M. H., & Raaijmakers, J. M. (2021). Extracting the GEMs: Genotype, Environment, and Microbiome Interactions Shaping Host Phenotypes. *Frontiers in microbiology*, *11*, 3444.

Kuramae, E. E., **Leite, M. F. A.**, Suleiman, A. K., Gough, C. M., Castillo, B. T., Faller, L., ... & Syring, J. (2019). Wood Decay Characteristics and Interspecific Interactions Control Bacterial Community Succession in *Populus grandidentata* (Bigtooth Aspen). *Frontiers in microbiology*, *10*, 979.

Armada, E., **Leite, M. F. A.**, Medina, A., Azcón, R., & Kuramae, E. E. (2018). Native bacteria promote plant growth under drought stress condition without impacting the rhizomicrobiome. *FEMS microbiology ecology*, *94*(7), fiy092.

Laanbroek, H. J., Zhang, Q. F., **Leite, M. F. A.**, Verhoeven, J. T., & Whigham, D. F. (2018). Effects of *Rhizophora* mangrove leaf litter and seedlings on carbon and nitrogen cycling in salt marshes—potential consequences of climate-induced mangrove migration. *Plant and Soil*, *426*(1), 383-400.

da Silva, R. L., **Leite, M. F. A.**, Muniz, F. H., de Souza, L. A. G., de Moraes, F. H. R., & Gehring, C. (2017). Degradation impacts on riparian forests of the lower Mearim river, eastern periphery of Amazonia. *Forest Ecology and Management*, *402*, 92-101.

Kip, N., Jansen, S., **Leite, M. F. A.**, De Hollander, M., Afanasyev, M., Kuramae, E. E., & Van Veen, J. A. (2017). Methanogens predominate in natural corrosion protective layers on metal sheet piles. *Scientific reports*, *7*(1), 1-11.

Schlemper, T. R., **Leite, M. F. A.**, Lucheta, A. R., Shimels, M., Bouwmeester, H. J., van Veen, J. A., & Kuramae, E. E. (2017). Rhizobacterial community structure differences among sorghum cultivars in different growth stages and soils. *FEMS microbiology ecology*, *93*(8).

**Leite, M. F. A.**, Pan, Y., Bloem, J., Ten Berge, H., & Kuramae, E. E. (2017). Organic nitrogen rearranges both structure and activity of the soil-borne microbial seedbank. *Scientific reports*, *7*(1), 1-11.

Cassman, N. A., **Leite, M. F. A.**, Pan, Y., de Hollander, M., van Veen, J. A., & Kuramae, E. E. (2016). Plant and soil fungal but not soil bacterial communities are linked in long-term fertilized grassland. *Scientific reports*, *6*(1), 1-11.

Cardozo, E. G., Muchavisoy, H. M., Silva, H. R., Zelarayán, M. L. C., **Leite, M. F. A.**, Rousseau, G. X., & Gehring, C. (2015). Species richness increases income in agroforestry systems of eastern Amazonia. *Agroforestry systems*, *89*(5), 901-916.

**Leite, M. F. A.**, Luz, R. L., Muchavisoy, K. H. M., Zelarayán, M. L. C., Cardoso, E. G., Moraes, F. H. R., ... & Gehring, C. (2016). The effect of land use on aboveground biomass and soil quality indicators in spontaneous forests and agroforests of eastern Amazonia. *Agroforestry systems*, *90*(6), 1009-1023.







Royal Netherlands  
Academy of Arts  
and Sciences