Impact of rainforest conversion: How prokaryotic communities respond to anthropogenic land use changes

Dissertation

For the award of the degree

"Doctor rerum naturalium" (Dr.rer.nat.)

of the Georg-August-Universität Göttingen

within the doctoral program Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Dirk Berkelmann

from Bielefeld, Germany

Göttingen, 2020

Thesis committee:

Prof. Dr. Rolf Daniel, Department of Genomics and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August Universität Göttingen

PD Dr. Michael Hoppert, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August Universität Göttingen

Members of the examination board:

Reviewer: **Prof. Dr. Rolf Daniel**, Department of Genomics and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen

Second Reviewer: **PD. Dr. Michael Hoppert**, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen

Further members of the examination board:

Prof. Dr. Jörg Stülke, Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen

Prof. Dr. Andrea Polle, Department of Forest Botany and Tree Physiology, Büsgen institute, Georg-August University Göttingen

Prof. Dr. Stefan Scheu, Department of Animal Ecology, J.F. Blumenbach Institute of Zoology and Anthropology, Georg-August University Göttingen

Prof. Dr. Kai Heimel, Department of Molecular Microbiology and Genetics, Institute of Microbiology and Genetics, Georg-August University Göttingen

Date of the oral examination: 09.06.2020

Table of contents

1

Introduction

1.1 Deforestation – an overview of a growing threat for global biodiversity

Approximately 30% of the global land area is covered by forests (FAO, 2018). Forest areas occur in almost all climate zones and are very different. Also due to human activities, forest loss or deforestation has occurred for centuries. Globalization in the twentieth century lead to a massive increase in land demand worldwide. Especially for upscaling agriculture, an expansion of usable land was required in many areas to meet the global demand. This new land mostly originated from primary and secondary forests (Gibbs et al., 2010). In addition to permanent agricultural purposes, deforestation was and is also caused by logging, urbanization and wildfires (Curtis et al., 2018). The average area of cleared forest is drastically higher in tropical regions compared to temperate and boreal zones (Leblois et al., 2017). This is mainly due to cultivation of crops and the required circumstances for highly demanded crops (Curtis et al., 2018). Although the amount of cleared areas is different between climate zones, the rate of deforestation follows the same trend (Leblois et al., 2017) (Figure 1.1).

Depending on the climate zone, different drivers are responsible for ongoing deforestation. In the tropics, commodity-driven deforestation is the main reason, e.g. for cultivation of rubber, oil palm, coffee and cacao. Globally, 27 ± 5% of all forest disturbance between 2001 and 2015 was associated with commodity-driven deforestation (Curtis et al., 2018). In temperate zones, forestry and wildfires are the main reason for deforestation today (Figure 1.2). Biodiversity is severely affected by these processes regardless of forest type. Especially the zones with the highest deforestations rates
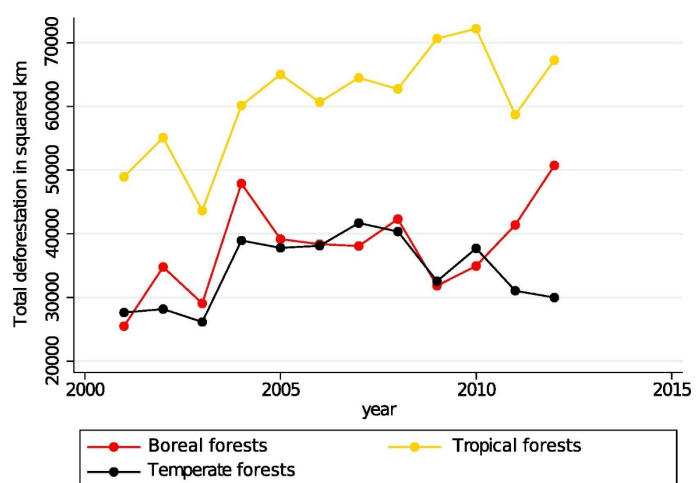
Figure 1.1: Global rates of deforestation for the three main forest types. Reprinted from Leblois et al. 2017, with permission from Elsevier (license number 4795360533577).

5

worldwide, Amazonia, Southeast Asia and Central Africa, provide evidence that biodiversity at all levels is highly at risk, foremost for larger organisms like birds or vertebrates (Barnes et al., 2014; Betts et al., 2017; Darras et al., 2018; Symes et al., 2018; Gomes et al., 2019). Therefore, the ongoing deforestation and higher demand for wood and agricultural products poses a challenge for biodiversity preservation, especially in areas where products of high global importance such as palm oil and rubber can be produced.

1.2 The expansion of rubber and oil palm cultivation in Southeast Asia

■ Commodity driven deforestation ■ Shifting Agriculture ■ Urbanization ■ Forestry ■ Wildfire ■ Zero or Minor Loss
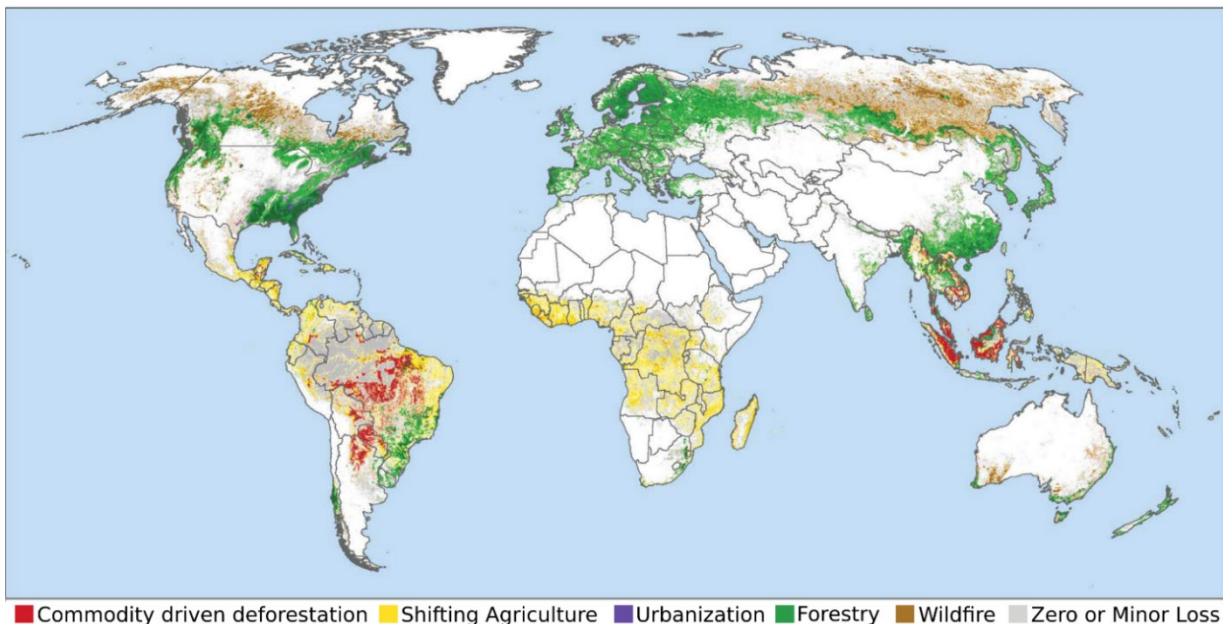
Figure 1.2: Main drivers of deforestation between 2001 to 2015. Darker color intensity indicates higher total forest loss. Adapted from Curtis et al. 2020. Reprinted with permission from AAAS (license number 4793001372025).

Palm oil is a product that is of high demand, due to its usage in a variety of products, from cooking oil, processed foods (chocolate, biscuits) and hygiene products (soaps, washing powder) to consumer products like biofuels (Koh and Wilcove, 2007; Afriyanti et al., 2016). With depleting fossil fuel reserves, biofuel in particular is of interest. Consequently, cultivation of crops suitable for biofuel production like the African oil palm (*Elaeis guineensis*) became more important. Overall, oil palm cultivation became a business of global importance, with increasing rates of production, export and consumption (Byerlee et al., 2017). Oil palm is the most productive oil crop but can only be cultivated in tropical areas, which led to the establishment of large scale oil palm plantations in Latin America, Midwest Africa and

Southeast Asia, which are continuously expanding (Furumo and Aide, 2017; Ordway et al., 2017; Taheripour et al., 2019).

Indonesia and Malaysia provide the major part of palm oil and rubber production today (Taheripour et al., 2019). After mainly cultivating rice and coffee, production shifted gradually to rubber and oil palm in the last decades, due to drastically increased demand worldwide and the possibility to cultivate these climate-zone limited crops (Gatto et al., 2015). In order to increase production, more agricultural land was needed, resulting in deforestation and conversion of already used areas for establishment of large plantations, either by big state-owned companies, or smallholder farms adjacent to the large plantations (Gibbs et al., 2010; Gatto et al., 2015). The area used for oil palm plantation has increased by 87% between the years 2000 and 2013, with varying impacts on primary forests, depending on the area and the extent of already cleared forests (Wilcove et al., 2013; Abood et al., 2015). In the Kalimantan provinces, the expansion of the plantation area came at the expense of 47% intact primary forests and 22% logged forests (Wilcove et al., 2013). These rates vary in different regions, e.g. in Jambi province, only 8% primary forest was converted to oil palm plantations between 1990 and 2015, as already logged or otherwise used areas were converted (Gatto et al., 2015; Tarigan et al., 2015). Although oil palm cultivation is often mentioned as the main reason for deforestation in Indonesia, forest clearance mostly occurred already in the past by logging or fiber production and (Gatto et al., 2015).

Oil palm cultivation has been drastically increased during the last decades, but other cash-crops like rubber are still dominant in Indonesia (Euler et al., 2016). Cultivation of rubber (*Hevea brasiliensis*) increased by 30% in South East Asia between 1990 and 2011 (Clough et al., 2016). In Jambi province, oil palm cultivation increased rapidly from 150,000 ha in 1996 to 550,000 ha in 2011, while rubber cultivation areas were expanded from 510,000 to 650,000 ha in the same time period (Gatto et al., 2015). Similar to oil palm and fiber cultivation, biodiversity is drastically reduced in rubber plantation areas with up to six times more species of vascular plants in forest areas than in rubber plantations (Beukema et al., 2007; Phommexay et al., 2011; Drescher et al., 2016a). However, traditional agroforestry variations of rubber cultivation tolerate wildlife within the cultivation area. It has been shown that rubber agroforestry (often referred to as "jungle rubber") has ecosystem services similar to primary forests and therefore might be a possibility to merge conservation and agriculture (Beukema et al., 2007; Wilcove et al., 2013).

1.3 Effects of land use conversion on soils and biodiversity

Agricultural management not only changes the appearance and structure of the habitat, but also affects soil properties by application of fertilizer, liming and use of pesticides and herbicides in order to increase productivity. Several studies reported on how deforestation and land use changes affect all trophic levels and soil properties (Allen et al., 2015; Schneider et al., 2015; Drescher et al., 2016a; Kassa et al., 2017; Rembold et al., 2017; Brinkmann et al., 2019; Krashevska et al., 2019; Schulz et al., 2019).

1.3.1 How land use management changes soil properties

Soil is an extremely versatile habitat with different microhabitats that can vary within short distances (Daniel, 2005). Agricultural management by application of similar management practices can cause general changes of soil characteristics that are independent from the cultivated crop or climate. Most practices aim at optimizing nutrient levels to maximize yields. Fertilizers are a source of nitrogen, phosphorous and potassium, often applied as combination (termed "NPK" fertilizer). Nitrogen is supplied in different forms, mostly as ammonium (NH_4^+) or less commonly as nitrate (NO_3^-) (Comte et al., 2013). Depending on N source, soil is either acidified (with ammonium) or alkalified (with nitrate). Additionally, liming, often performed by supplementation with urea or dolomite alkalifies soil to compensate acidification by mineral fertilizer (Ballard, 2000; Comte et al., 2013). Furthermore, deforestation by slash and burn practices is mostly used to clear areas in Indonesia. This results in accumulation of cations in the soil, which are traceable for several decades (Tanaka et al., 2005; Sohng et al., 2017). In addition, soil pH values increase compared to forests, which comprise acidic soils with pH values around 4 (Ballard, 2000; Tanaka et al., 2005; Comte et al., 2013; Allen et al., 2015; Sohng et al., 2017).

The applied compounds alter nutrient availability as well and thus the respective nutrient cycles within the soil. Nitrogen loss in form of nitrous oxide N_2O (a greenhouse gas) release has been shown in fertilized soils of different monoculture and agroforest cash-crop systems compared to forest (Veldkamp et al., 2008; Allen et al., 2015). To compensate for these nutrient losses, further and continuous fertilization is a necessity for intensively managed plantations (Beauchamp, 1997; Ding et al., 2013; Corre et al., 2014; Allen et al., 2015; Kielak et al., 2016; Zhou et al., 2016).

Similar to nitrogen, soil organic carbon (also referred to as SOC) and wood biomass depletes in the course of land use management and rainforest conversion, which has been observed in

several cash-crop systems including oil palm, rubber and cacao plantations (Kotowska et al., 2015, 2016; Kassa et al., 2017; Shanmugam et al., 2018). Soil organic carbon was reported to deplete by ~8% during conversion of secondary forest to plantation crops and by over 18% following primary forest conversion to plantation crops (Shanmugam et al., 2018) A crucial difference between forest an cash-crops is the lack of a litter layer that stores carbon within the ecosystem. The carbon storage drops compared to rainforest from 4.5 to 2.9 and 1.9 Mg ha^{-1} year^{-1} in rubber and oil palm plantations, respectively (Kotowska et al., 2016). The removal of litter and therefore carbon input is also represented by decreased decomposition rates, which are more pronounced in oil palm plantations (Guillaume et al., 2015). Additionally, carbon is removed by harvesting, which is tied to the productivity of the plantation. In the case of oil palm, the highest productivity is reached at a tree age of approximately 16 years (Shanmugam et al., 2018). Accordingly, carbon loss is positively correlated with the development of yield within a plantation (Shanmugam et al., 2018). Phosphorous levels decrease with land use conversion, but different rates of decline were observed for inorganic and organic P reserves (Maranguit et al., 2017).

In summary, conversion of forest to agricultural land use systems or land use intensification in general causes major changes in soil properties and correspondingly in soil nutrient stocks and cycling.

1.3.2 Aboveground biodiversity in tropical forests and converted land uses

Ongoing deforestation is a severe threat to biodiversity at different trophic levels. In case of wildfires and urbanization, it seems obvious that the drastic environmental changes lead to biodiversity decreases. However, agricultural land use in general and less management intense forms of agricultural land use (like agroforestry) potentially harbour similar niches as forest habitats. When establishing large-scale land use systems for cash-crop monocultures, the floral biodiversity is drastically reduced (Vieira et al., 2008; Flynn et al., 2009; Laliberté et al., 2010; Wilcove et al., 2013; Rembold et al., 2017). It has been shown that the reduction of plant diversity in monoculture plantations of various crops, including prominent crops such as oil palm, rubber, rice and coffee, correlates with reduced biodiversity of ground living animals, but also with that of birds and bats (Flynn et al., 2009; Almeida et al., 2016; Maas et al., 2016; Prabowo et al., 2016; Darras et al., 2018). Furthermore, smaller organisms like insects and invertebrates as well as soilborne microorganisms like protists and nematodes are negatively affected in terms of species richness or community composition (Rubiana et al., 2015; Klarner

et al., 2017; Krashevska et al., 2018, 2019; Heuss et al., 2019). These effects occur also within amphibian communities when comparing forest and riparian areas in converted land use systems, highlighting the general impact of land use intensification on different soil types (Paoletti et al., 2018). This relationship is further visible in agriculture systems that are less intensive in terms of reduced plant biodiversity or management applications. Less profitable agroforestry systems are a compromise between intensively managed monocultures and forests. These show reduced biodiversity loss due to higher overall biodiversity and the availability of more different ecological niches compared to monoculture systems (Steffan-Dewenter et al., 2007; Drescher et al., 2016a; Klarner et al., 2017).

1.3.3 Prokaryotic communities and diversity in soils and their responses to land use conversion

Microbial life is crucial for mediating nutrient cycling in soil (Fierer and Jackson, 2006). High soil heterogeneity through different particle and pore size, texture, water content and nutrient distribution provides many different ecological niches (Figure 1.3) (Vos et al., 2013). Due to the high soil heterogeneity, prokaryotic diversity is very high in soils and directly connected to soil properties and plant diversity (Brenzinger et al., 2015; Schneider et al., 2015; Kaiser et al., 2016; Fonseca et al., 2018).

In general, bulk soils in temperate and tropical regions are dominated by various members of *Acidobacteria*, *Actinobacteria* and *Proteobacteria* (Chapter 2 and 3; Lauber et al., 2009; Rousk
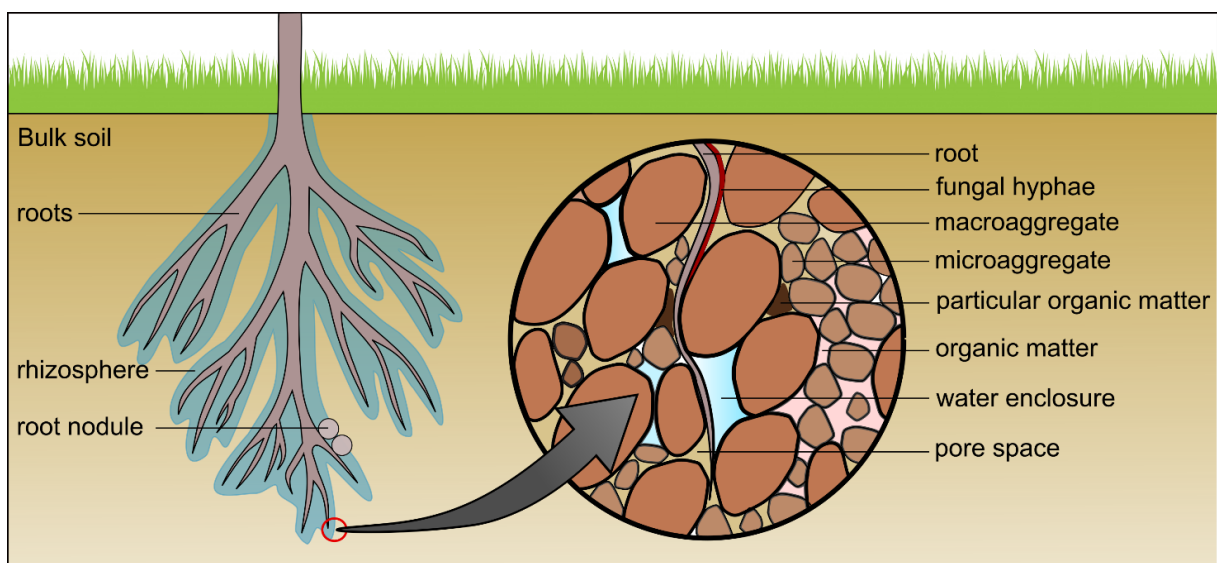
Figure 1.3: Simplified schematic of the microbial habitats in soil with emphasis on bulk soil structure.

et al., 2010; Nacke et al., 2011; Schneider et al., 2015). The community composition varies, depending on soil properties such as texture, nutrient availability or oxygen concentration, which can be different within short distances (Daniel, 2005; Fierer, 2017). One of the main drivers of prokaryotic community structure is soil pH, with a strong influence especially on *Acidobacteria* and *Actinobacteria* (Lauber et al., 2009; Rousk et al., 2010; Schneider et al., 2015; Zhalnina et al., 2015; Kaiser et al., 2016). Plant roots and their exudates further diversify the habitat and directly affect microbes in the rhizosphere (the area around roots within bulk soil) (Rennenberg et al., 2009; Fonseca et al., 2018). Besides the area around plant roots, microbes can be also located within roots as endophytes or symbionts in root nodules (Newman et al., 2016; Clúa et al., 2018). Endophytic or symbiotic lifestyles are commonly associated with specific groups of microorganisms, the most prominent members belong to the *Rhizobiales*, *Burkholderiales*, *Klebsiella* and *Rhodococcus* (De Meyer et al., 2016; Martínez-Hidalgo and Hirsch, 2017).

Agricultural land use drastically affects soil properties and nutrient levels and correspondingly microbial community structure and function. The responses of microbial communities to agricultural management in general and effects of rainforest conversion to managed systems are of high interest in order to understand and predict functional shifts and ecosystem responses. Most related studies were conducted in Brazil or central America (Navarrete et al.; Leff et al., 2012; Meyer et al., 2017). Although similar studies conducted in Southeast Asia are still rare, numbers have been growing in recent years (Chapters 2 and 3; Tripathi et al., 2012; Lee-Cruz et al., 2013; Schneider et al., 2015). Among the mentioned studies, bacterial β-diversity mostly increased with increasing land use intensity, which was accompanied with higher soil pH in managed land use systems. In these studies, the three phyla *Acidobacteria*, *Proteobacteria* and *Actinobacteria* were most abundant as it is typical for soils in general (Nacke et al., 2011; Tripathi et al., 2012; Lee-Cruz et al., 2013; Mendes et al., 2015; Schneider et al., 2015; Kaiser et al., 2016; Pedrinho et al., 2018). When considering the land use change from forest to managed land use systems, *Proteobacteria* were consistently decreasing in abundance (Chapter 3; Kerfahi et al., 2016; Schneider et al., 2015; Tripathi et al., 2012). It was suggested, that taxa like *Burkholderiales* are especially affected due to their involvement in nitrogen fixation and the influence of fertilizer input on available nitrogen sources (Chapter 3; Schneider et al., 2015). However, results for some bacterial groups are inconsistent among studies: *Acidobacteria* were negatively affected in oil palm and rubber soils in Indonesia, while increasing in abundance in Malaysian oil palm soils (Schneider et al., 2015; Kerfahi et al., 2016). Most of these studies were conducted with amplicon-based marker gene analysis. Metagenomic shotgun sequencing

studies are rare and thus microbial functionality in these soils was rarely addressed. In direct sequencing-based metagenomic studies it was shown that forest conversion to arable land reduced microbial functional diversity, but it increased after conversion to managed wheat fields (Pedrinho et al., 2018; Muñoz-Arenas et al., 2020). Still, the connection between microbial taxonomy and functionality is not fully investigated yet and is a knowledge gap in the efforts to disentangle the underlying processes of rainforest conversion to managed land use systems.

1.4 How to analyse microbial communities – An overview

Monitoring microbial communities in complex environments is still a challenge. The traditional approach is cultivation and subsequent analysis of single microorganisms from environmental samples. However, the complexity of the sample environment cannot be simulated *in situ*, creating a bottleneck (Streit and Schmitz, 2004). Therefore, complementation of cultivation-based approaches and circumventing the pitfalls of these approaches by cultivation-independent methods and direct detection of DNA or RNA in samples are necessary to analyse environmental microbial communities (Daniel, 2005; Schöler et al., 2017). One possibility for this is Fluorescence in situ hybridization (FISH), which utilizes probes binding to DNA of the target organisms. These probes are then detected by fluorescence of a bound probe, caused by either attached fluorescing compounds or enzymes that catalyse a reaction resulting in luminescence. FISH can be used to monitor specific taxonomic groups but is less practical when trying to analyse community compositions of diverse habitats, since all detected signals by a probe are indistinguishable. Therefore, FISH is suitable to detect quantities of certain taxa, but is limited by its probe design (Huber et al., 2018).

Other approaches to assess the entirety of communities are environmental DNA-based metagenomics and RNA-based metatranscriptomics. Originally, metagenomic studies relied on constructing clone libraries containing either functional genes for enzyme screenings or 16S rRNA genes for taxonomic analyses. These libraries were then function-driven or sequence-based screened or sequenced by the Sanger method to find novel genes for a product or assess community taxonomy (Daniel, 2005). With the introduction of next generation sequencing, sequencing depth was drastically increased, allowing the analysis of complex communities. Metagenomic identification of microbial communities can be separated into marker gene-focussed approaches and direct sequencing of nucleic acids also known as metagenomic shotgun sequencing (Figure 1.4) (Quince et al., 2017). However, these methods strongly rely

on databases to identify obtained sequences and available computing power to process the amount of data.
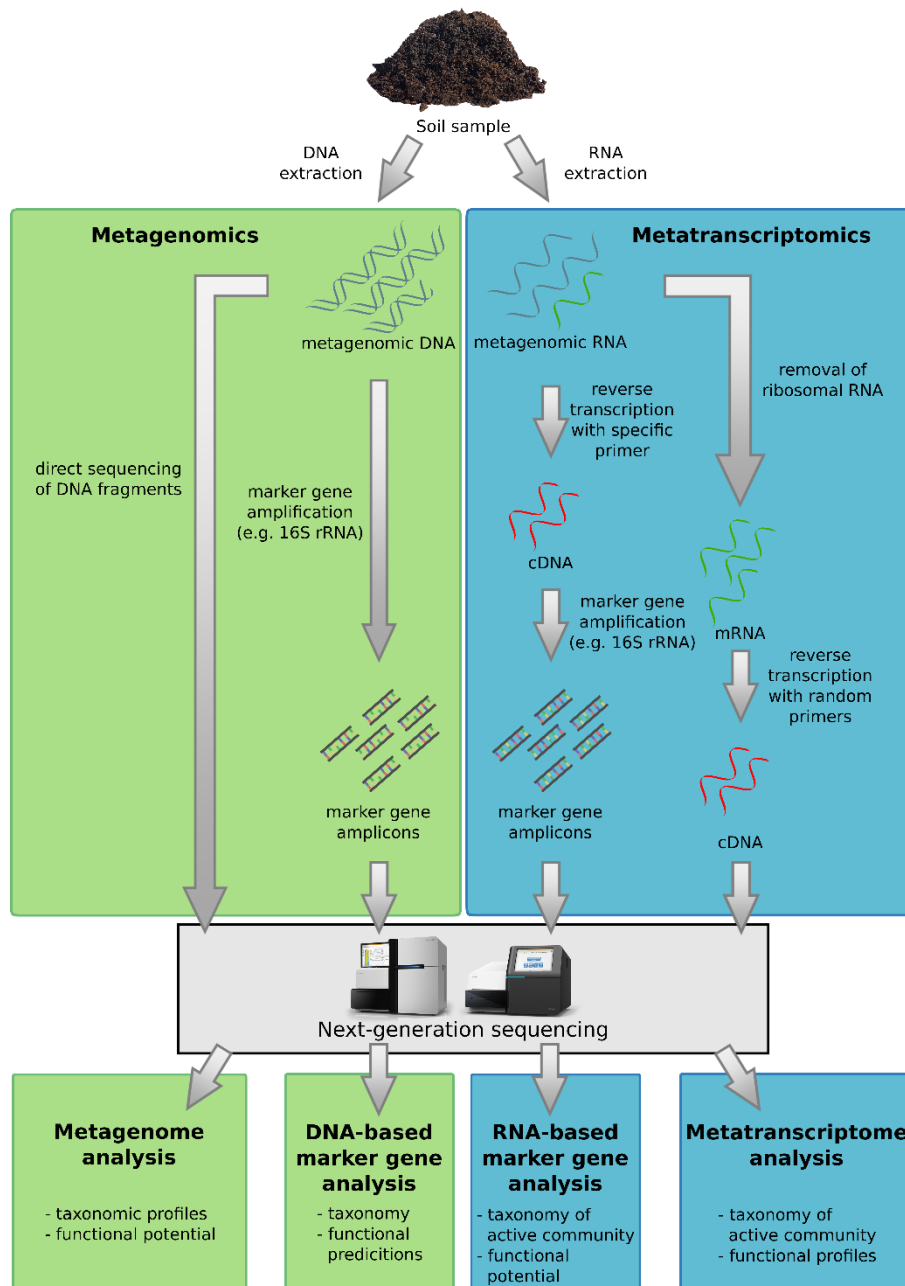
Figure 1.4: Schematic of metagenomic/metatranscriptomic analysis of microbial communities from soil.

1.4.1 Microbial community profiling by analysis of marker genes

Even before large scale direct metagenome sequencing was feasible, sequencing selected genes was a powerful tool. In order to analyse the composition of entire microbial communities, focussing on marker genes is a cost efficient and reliable option. In order to be used as a marker the respective gene needs to be ubiquitous and contain highly conserved as well as variable regions, to ensure certain flexibility and stability (Lan et al., 2016). A variety of genes is suitable

to be used as taxonomic marker, depending on the targeted taxonomic group. For bacteria, the 16S rRNA gene is considered as the gold standard, while the 18S rRNA gene and ITS (intergenic transcribed spacer) are commonly used for fungi and micro-eukaryotes (Amann et al., 1995; Lindahl et al., 2013; Lentendu et al., 2014). To analyse microbial communities, nucleic acids are extracted from environmental samples, e.g. soil. Then, the targeted gene is amplified either from DNA or from reverse transcribed environmental RNA (cDNA) by using general primers that cover as many taxa as possible. With RNA as a template, taxonomic profiles are used to analyse the active communities of a sample, excluding dormant cells or cell remains from used samples (Chapter 2; Wemheuer et al., 2014; Zhang et al., 2014). However, none of the primer pairs so far designed for this purpose is perfect as the coverage can vary between taxa. To overcome these limitations, housekeeping genes or functional genes (e.g. *nifH*, *recA*, *amoA*) are used as an alternative for specific taxonomic groups (Torriani et al., 2001; Lan et al., 2016; Oton et al., 2016; Soni et al., 2016). Although some bias is introduced by marker gene-based analysis, microbial community profiling is reliable and one of the most frequently used methods (Rausch et al., 2019). Databases are needed to identify the taxonomic background of each sequence and therefore results are dependent on size and quality of the chosen database. Further, taxonomic profiles are reliable for abundant taxa, but assessing rare taxa might face some limitations such as insufficient survey size or lack of coverage by the used primers (Klindworth et al., 2013). Even frequently used primer sets do not cover all taxonomic groups equally, which needs to be considered during analysis (Klindworth et al., 2013). After sequencing, curation of obtained sequences by quality-filtering, chimera removal and denoising is required to avoid erroneous results (Bokulich et al., 2013).

In addition to taxonomic identity, microbial activity and functionality is of high interest. Functional profiles of metagenomic communities cannot be directly investigated by marker gene analysis, however, functional predictions based on functional profiles are possible with tools like Tax4Fun2 (Wemheuer et al., 2018), PICRUSt (Langille et al., 2013) or Piphillin (Iwai et al., 2016). Since these predictions are based on functional profiles for known taxa, coverage has to be considered, especially in environments with diverse microbial communities.

1.4.2 Metagenomic shotgun sequencing

To fully investigate the impact of prokaryotic functionality on ecosystem functioning, phylogenetic characterizations by marker gene analyses have to be complemented by taking the entirety of functional genes into account. Metagenomic and metatranscriptomic shotgun

sequencing provides taxonomic information and abundances of functional genes by direct high-throughput sequencing of environmentally derived nucleic acids. In this way, the functional potential and functional profiles of active genes are assessed by DNA and RNA sequencing, respectively (Myrold et al., 2014).

High-throughput sequencing techniques mostly provide short fragments in high numbers which provides a high sequencing depth and therefore a potentially better coverage of all present nucleic acids in a sample. Depending on the employed sequencing system, several million short reads of about 150 bp to 250 bp can be obtained per sample. These sequences can be assembled into larger fragments and potentially full genomes (referred to as metagenome assembled genomes) or short reads are directly mapped on a reference dataset to assess the taxonomic and functional composition of whole communities (Quince et al., 2017).

Short fragments can be a problem though, when sequencing genomes or sequences with long repeats, since assembling these to the correct full-length sequence is very challenging. This can be solved by single-molecule sequencers like Pacific Biosciences' or Oxford Nanopore sequencing, which provides reads with up to two megabases (Istace et al., 2017; Jain et al., 2018). However, higher error rates are currently a problem which can be met by complementation with short fragments from high throughput sequencers (Goodwin et al., 2015).

Handling of sequence data is complex since data size and certain characteristics need to be considered and are still a challenge. For instance, differences in sequencing depth among samples of a dataset can cause a bias in results if not handled appropriately and in accordance to sequencing data characteristics. However, methods for sequence analyses are still under debate and have to be chosen carefully with awareness to not only data characteristics, but experimental specificities like sampling size as well (McMurdie and Holmes, 2014; Weiss et al., 2017; Quinn et al., 2018b, 2018a). Furthermore, identification of reads is again highly dependent on curated reference databases, since taxonomic classification relies entirely on successful matches with database entries.

With advancing sequencing technologies and a higher availability of data processing power, handling these enormous datasets has become easier, allowing to study microbial communities from various angles in addition to marker gene analysis (Chapter 3; Tringe et al., 2005; Fierer et al., 2012; Xu et al., 2014; Rausch et al., 2019; Nkongolo and Narendrula-Kotha, 2020; Winand et al., 2020).

1.5 Aim and structure of this study

This study was conducted as part of the "Collaborative Research Centre 990: Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems (Sumatra, Indonesia)" project within the sub-project group "B – Biota and ecosystem services". The general aim of the project is to investigate the effects of rainforest conversion to different agricultural land use systems, ranging from agroforestry with rubber (*Hevea brasiliensis*) to monoculture plantations of rubber (*Hevea brasiliensis*) and oil palm (*Elaeis guineensis*), on environmental processes, biota and ecosystems as well as human dimensions in Jambi province, Indonesia (Drescher et al., 2016a). Additionally, possibilities of alternative cultivation forms regarding management intensity and biodiversity enrichment are explored (Darras et al., 2019).

1.5.1 Study sites and design

The experimental sites are located in Jambi province on Sumatra, Indonesia (Figure 1.5). All experimental sites were duplicated in two regions of the province: the "Bukit Duabelas landscape" and the "Harapan landscape" (Figure 1.5). For each treatment, four core plots (50 x 50 m) were established and each of these core plots harboured five fixed sub plots (5 x 5 m) to streamline soil sampling throughout the project. Sub plots within the core plots were randomly distributed. In each landscape, four different land use systems were established, following a gradient of land use intensity. Sampling sites were further equipped with a meteorological station to measure climate data.

As "rainforest" control, plots were established in primary degraded forest (forests showed signs of selective logging, hence classified as degraded). Agroforestry, as a form of intermediate management intensity was included as "jungle rubber", harbouring rubber trees (*Hevea brasiliensis*) planted in forests. As highest land use intensity, plots within monocultures of rubber (*Hevea brasiliensis*) and oil palm (*Elaeis guineensis*) were established. The latter plots were established in smallholder-owned plantations in 2012. The used plantations varied in age, ranging from seven to sixteen years. All used areas consist of acrisol soils, with loam acrisol in the "Harapan" area (including fractions of sand, silt and clay) and clay acrisol in the "Bukit Duabelas" landscape with higher proportions of clay (Allen et al., 2015; Drescher et al., 2016b).

An additional experiment located in oil palm plantations within the Harapan area was established to investigate possible effects of reduced fertilizer input and herbicide use on soilborne microbial communities. This oil palm management experiment is located in a state-

owned oil palm plantation and was established in 2015 and relies on the same plot design as the core plots of the rainforest conversion experiment (Figure 1.5). Four different treatments were included: regular fertilizer and herbicide application, regular fertilizer application with mechanical weeding, reduced fertilizer application with herbicide usage and reduced fertilizer application with mechanical weeding (Darras et al., 2019).
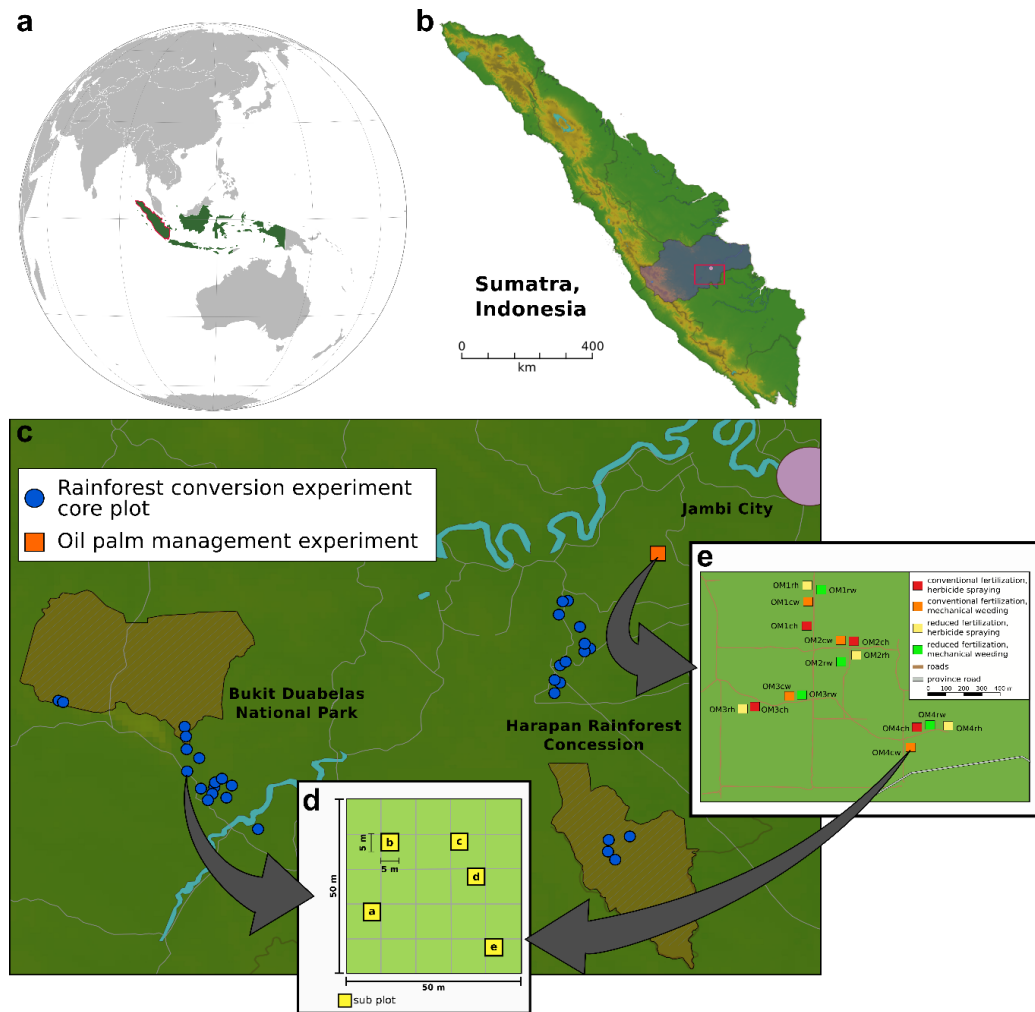
Figure 1.5: Location of the CRC990 project located in Sumatra, Indonesia (a) and Jambi province (b). All plots of the rainforest conversion experiment are shown as grey dots and the location of the oil palm management experiment as orange coloured square (c). An example of the plot design of the rainforest conversion experiment (d) shows distribution and size of subplots located within each core plot. The general plot design and distribution of the oil palm management experiment (e) shows location and treatments of all plots. Each plot of the oil palm management experiment was established with the same plot design as for the rainforest conversion experiment (d).

1.5.2 Research objectives

The prokaryotic communities in soil have been analysed in numerous experiments and habitats. Due to the importance of agriculture, communities in managed land use systems have been investigated as well. Most studies focussed on temperate but not on tropical regions. Microbial

community responses to conversion of rainforest to agricultural land use systems were hardly studied at all prior to this study. Especially the active community, excluding dormant or dead cells is of high importance to understand the effects of rainforest conversion on soilborne communities. One part of this study aimed at investigating how the active soilborne prokaryotic community is affected by rainforest conversion by analysing RNA-based 16S rRNA gene sequences. Additionally, few studies addressed functionality of soilborne communities, despite the strong connection to soil properties and ecosystem functioning, which are directly affected by land use management practices. The connections between functional genes and the taxonomic origin of these genes have rarely been addressed before. To close this gap, microbial communities and functional genes were investigated by direct metagenomic approaches. To unravel the relationship between affected functional genes and the actual community, the relationship of taxonomy and functionality was investigated by identifying the taxonomic background of detected functional genes.

Lastly, different approaches are considered to reduce land use management and herbicide usage. Therefore, the effect of reduced fertilizer input and herbicide use on soil bacterial communities in oil palm plantations was investigated by 16S rRNA gene analysis.

1.6 References

Abood, S. A., Lee, J. S. H., Burivalova, Z., Garcia-Ulloa, J., and Koh, L. P. (2015). Relative Contributions of the Logging, Fiber, Oil Palm, and Mining Industries to Forest Loss in Indonesia. *Conserv. Lett.* 8, 58–67.

Afriyanti, D., Kroeze, C., and Saad, A. (2016). Indonesia palm oil production without deforestation and peat conversion by 2050. *Sci. Total Environ.* 557–558, 562–570.

Allen, K., Corre, M. D., Tjoa, A., and Veldkamp, E. (2015). Soil nitrogen-cycling responses to conversion of lowland forests to oil palm and rubber plantations in Sumatra, Indonesia. *PLoS One* 10, e0133325.

Almeida, S. M., Silva, L. C., Cardoso, M. R., Cerqueira, P. V, Juen, L., and Santos, M. P. D. (2016). The effects of oil palm plantations on the functional diversity of Amazonian birds. *J. Trop. Ecol.* 32, 510–525.

Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev* 59.

Ballard, T. M. (2000). Impacts of forest management on northern forest soils. *For. Ecol. Manage.* 133, 37–42.

Barnes, A. D., Jochum, M., Mumme, S., Haneda, N. F., Farajallah, A., Widarto, T. H., et al. (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nat. Commun.* 5.

Beauchamp, E. G. (1997). Nitrous oxide emission from agricultural soils. *Can. J. Soil Sci.* 77, 113–123.

Betts, M. G., Wolf, C., Ripple, W. J., Phalan, B., Millers, K. A., Duarte, A., et al. (2017). Global forest loss disproportionately erodes biodiversity in intact landscapes. *Nature* 547, 441–444.

Beukema, H., Danielsen, F., Vincent, G., Hardiwinoto, S., and van Andel, J. (2007). Plant and bird diversity in rubber agroforests in the lowlands of Sumatra, Indonesia. *Agrofor. Syst.* 70, 217–242.

Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59.

Brenzinger, K., Dörsch, P., and Braker, G. (2015). pH-driven shifts in overall and transcriptionally active denitrifiers control gaseous product stoichiometry in growth experiments with extracted bacteria from soil. *Front. Microbiol.* 6, 961.

Brinkmann, N., Schneider, D., Sahner, J., Ballauff, J., Edy, N., Barus, H., et al. (2019). Intensive tropical land use massively shifts soil fungal communities. *Sci. Rep.* 9, 3403.

Byerlee, D., Falcon, W. P., and Naylor, R. (2017). The tropical oil crop revolution: food, feed, fuel, and forests. Oxford University Press.

Clough, Y., Krishna, V. V, Corre, M. D., Darras, K., Denmead, L. H., Meijide, A., et al. (2016). Land-use choices follow profitability at the expense of ecological functions in Indonesian smallholder landscapes. *Nat. Commun.* 7, 13137.

Clúa, J., Roda, C., Zanetti, M. E., and Blanco, F. A. (2018). Compatibility between Legumes and Rhizobia for the Establishment of a Successful Nitrogen-Fixing Symbiosis. *Genes (Basel).* 9, 125.

Comte, I., Colin, F., Grünberger, O., Follain, S., Whalen, J. K., and Caliman, J.-P. (2013). Landscape-scale assessment of soil response to long-term organic and mineral fertilizer application in an industrial oil palm plantation, Indonesia. *Agric. Ecosyst. Environ.* 169, 58–68.

Corre, M. D., Sueta, J. P., and Veldkamp, E. (2014). Nitrogen-oxide emissions from tropical forest soils exposed to elevated nitrogen input strongly interact with rainfall quantity and seasonality. *Biogeochemistry* 118, 103–120.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., and Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science (80-.).* 361, 1108 LP – 1111.

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

Darras, K., Batáry, P., Furnas, B., Celis-Murillo, A., Van Wilgenburg, S. L., Mulyani, Y. A., et al. (2018). Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis. *J. Appl. Ecol.* 55, 2575–2586.

Darras, K. F. A., Corre, M. D., Formaglio, G., Tjoa, A., Potapov, A., Brambach, F., et al. (2019). Reducing Fertilizer and Avoiding Herbicides in Oil Palm Plantations— Ecological and Economic Valuations. *Front. For. Glob. Chang.* 2, 65.

De Meyer, S. E., Briscoe, L., Martínez-Hidalgo, P., Agapakis, C. M., de-los Santos, P. E., Seshadri, R., et al. (2016). Symbiotic Burkholderia species show diverse arrangements of nif/fix and nod genes and lack typical high-affinity cytochrome cbb3 oxidase genes. *Mol. Plant-Microbe Interact.* 29, 609–619.

Ding, W., Luo, J., Li, J., Yu, H., Fan, J., and Liu, D. (2013). Effect of long-term compost and inorganic fertilizer application on background N2O and fertilizer-induced N2O emissions from an intensively cultivated soil. *Sci. Total Environ.* 465, 115–124.

Drescher, J., Rembold, K., Allen, K., Beckschäfer, P., Buchori, D., Clough, Y., et al. (2016a). Ecological and socio-economic functions across tropical land use systems after rainforest conversion. *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150275.

Drescher, J., Rembold, K., Allen, K., Beckschäfer, P., Buchori, D., Clough, Y., et al. (2016b). Ecological and socio-economic functions across tropical land use systems after rainforest conversion. *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150275.

Euler, M., Schwarze, S., Siregar, H., and Qaim, M. (2016). Oil Palm Expansion among Smallholder Farmers in Sumatra, Indonesia. *J. Agric. Econ.* 67, 658–676.

FAO, F. (2018). The State of the World's Forests 2018-Forest pathways to sustainable development.

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* 15, 579.

Fierer, N., and Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A* 103, 626–631.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., et al. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci.* 109, 21390 – 21395.

Flynn, D. F. B., Gogol-Prokurat, M., Nogeire, T., Molinari, N., Richers, B. T., Lin, B. B., et al. (2009). Loss of functional diversity under land use intensification across multiple taxa. *Ecol. Lett.* 12, 22–33.

Fonseca, J. P., Hoffmann, L., Cabral, B. C. A., Dias, V. H. G., Miranda, M. R., de Azevedo Martins, A. C., et al. (2018). Contrasting the microbiomes from forest rhizosphere and deeper bulk soil from an Amazon rainforest reserve. *Gene* 642, 389–397.

Furumo, P. R., and Aide, T. M. (2017). Characterizing commercial oil palm expansion in Latin America: land use change and trade. *Environ. Res. Lett.* 12, 24008.

Gatto, M., Wollni, M., and Qaim, M. (2015). Oil palm boom and land-use dynamics in Indonesia: The role of policies and socioeconomic factors. *Land use policy* 46, 292–303.

Gibbs, H. K., Ruesch, A. S., Achard, F., Clayton, M. K., Holmgren, P., Ramankutty, N., et al. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc. Natl. Acad. Sci.* 107, 16732 LP – 16737.

Gomes, V. H. F., Vieira, I. C. G., Salomão, R. P., and ter Steege, H. (2019). Amazonian tree species threatened by deforestation and climate change. *Nat. Clim. Chang.* 9, 547–553.

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., and McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* 25, 1750–1756.

Guillaume, T., Damris, M., and Kuzyakov, Y. (2015). Losses of soil carbon by converting tropical forest to plantations: erosion and decomposition estimated by δ13C. *Glob. Chang. Biol.* 21, 3548–3560.

Heuss, L., Grevé, M. E., Schäfer, D., Busch, V., and Feldhaar, H. (2019). Direct and indirect effects of land-use intensification on ant communities in temperate grasslands. *Ecol. Evol.* 9, 4013–4024.

Huber, D., Voith von Voithenberg, L., and Kaigala, G. V (2018). Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro Nano Eng.* 1, 15–24.

Istace, B., Friedrich, A., D'Agata, L., Faye, S., Payen, E., Beluche, O., et al. (2017). de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* 6, 1–13.

Iwai, S., Weinmaier, T., Schmidt, B. L., Albertson, D. G., Poloso, N. J., Dabbagh, K., et al. (2016). Piphillin: improved prediction of metagenomic content by direct inference from human microbiomes. *PLoS One* 11, e0166104.

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345.

Kaiser, K., Wemheuer, B., Korolkow, V., Wemheuer, F., Nacke, H., Schöning, I., et al. (2016). Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Sci. Rep.* 6, 33696.

Kassa, H., Dondeyne, S., Poesen, J., Frankl, A., and Nyssen, J. (2017). Impact of deforestation on soil fertility, soil carbon and nitrogen stocks: the case of the Gacheb catchment in the White Nile Basin, Ethiopia. *Agric. Ecosyst. Environ.* 247, 273–282.

Kerfahi, D., Tripathi, B. M., Dong, K., Go, R., and Adams, J. M. (2016). Rainforest conversion to rubber plantation may not result in lower soil diversity of bacteria, fungi, and nematodes. *Microb Ecol* 72, 359–371.

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., and Kuramae, E. E. (2016). The Ecology of Acidobacteria: Moving beyond Genes and Genomes. *Front Microbiol* 7, 744.

Klarner, B., Winkelmann, H., Krashevska, V., Maraun, M., Widyastuti, R., and Scheu, S. (2017). Trophic niches, diversity and community composition of invertebrate top predators (Chilopoda) as affected by conversion of tropical lowland rainforest in Sumatra (Indonesia). *PLoS One* 12.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41, 1–11.

Koh, L. P., and Wilcove, D. S. (2007). Cashing in palm oil for conservation. *Nature* 448, 993–994.

Kotowska, M. M., Leuschner, C., Triadiati, T., and Hertel, D. (2016). Conversion of tropical lowland forest reduces nutrient return through litterfall, and alters nutrient use efficiency and seasonality of net primary production. *Oecologia* 180, 601–618.

Kotowska, M. M., Leuschner, C., Triadiati, T., Meriem, S., and Hertel, D. (2015). Quantifying above- and belowground biomass carbon loss with forest conversion in tropical lowlands of Sumatra (Indonesia). *Glob Chang Biol* 21, 3620–3634.

Krashevska, V., Kudrin, A. A., Widyastuti, R., and Scheu, S. (2019). Changes in Nematode Communities and Functional Diversity With the Conversion of Rainforest Into Rubber and Oil Palm Plantations. *Front. Ecol. Evol.* 7, 487.

Krashevska, V., Malysheva, E., Klarner, B., Mazei, Y., Maraun, M., Widyastuti, R., et al. (2018). Micro-decomposer communities and decomposition processes in tropical lowlands as affected by land use and litter type. *Oecologia* 187, 255–266.

Laliberté, E., Wells, J. A., DeClerck, F., Metcalfe, D. J., Catterall, C. P., Queiroz, C., et al. (2010). Land-use intensification reduces functional redundancy and response diversity in plant communities. *Ecol. Lett.* 13, 76–86.

Lan, Y., Rosen, G., and Hershberg, R. (2016). Marker genes that are less conserved in their sequences are useful for predicting genome-wide similarity levels between closely related prokaryotic strains. *Microbiome* 4, 18.

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotech* 31, 814–821.

Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl. Environ. Microbiol.* 75, 5111 LP – 5120.

Leblois, A., Damette, O., and Wolfersberger, J. (2017). What has Driven Deforestation in Developing Countries Since the 2000s? Evidence from New Remote-Sensing Data. *World Dev.* 92, 82–102.

Lee-Cruz, L., Edwards, D. P., Tripathi, B. M., and Adams, J. M. (2013). Impact of logging and forest conversion to oil palm plantations on soil bacterial communities in Borneo. *Appl. Environ. Microbiol.* 79, 7290–7297.

Leff, J. W., Nemergut, D. R., Grandy, A. S., O'Neill, S. P., Wickings, K., Townsend, A. R., et al. (2012). The effects of soil bacterial community structure on decomposition in a tropical rain forest. *Ecosystems* 15, 284–298.

Lentendu, G., Wubet, T., Chatzinotas, A., Wilhelm, C., Buscot, F., and Schlegel, M. (2014). Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Mol. Ecol.* 23, 3341–3355.

Lindahl, B. D., Nilsson, R. H., Tedersoo, L., Abarenkov, K., Carlsen, T., Kjøller, R., et al. (2013). Fungal community analysis by high-throughput sequencing of amplified markers – a user's guide. *New Phytol.* 199, 288–299.

Maas, B., Karp, D. S., Bumrungsri, S., Darras, K., Gonthier, D., Huang, J. C.-C., et al. (2016). Bird and bat predation services in tropical forests and agroforestry landscapes. *Biol. Rev.* 91, 1081–1101.

Maranguit, D., Guillaume, T., and Kuzyakov, Y. (2017). Land-use change affects phosphorus fractions in highly weathered tropical soils. *CATENA* 149, 385–393.

Martínez-Hidalgo, P., and Hirsch, A. M. (2017). The Nodule Microbiome: N2-Fixing Rhizobia Do Not Live Alone. *Phytobiomes J.* 1, 70–82.

McMurdie, P. J., and Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Comput. Biol.* 10, 1–12.

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

Meyer, K. M., Klein, A. M., Rodrigues, J. L. M., Nüsslein, K., Tringe, S. G., Mirza, B. S., et al. (2017). Conversion of Amazon rainforest to agriculture alters community traits of methane-cycling organisms. *Mol. Ecol.* 26, 1547–1556.

Muñoz-Arenas, L. C., Fusaro, C., Hernández-Guzmán, M., Dendooven, L., Estrada-Torres, A., and Navarro-Noya, Y. E. (2020). Soil microbial diversity drops with land use change in a high mountain temperate forest: a metagenomics survey. *Environ. Microbiol. Rep.*

Myrold, D. D., Zeglin, L. H., and Jansson, J. K. (2014). The potential of metagenomic approaches for understanding soil microbial processes. *Soil Sci. Soc. Am. J.* 78, 3–10.

Nacke, H., Thürmer, A., Wollherr, A., Will, C., Hodac, L., Herold, N., et al. (2011). Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS One* 6, e17000–e17000.

Navarrete, A. A., Tsai, S. M., Mendes, L. W., Faust, K., Hollander, M., Cassman, N. A., et al. Soil microbiome responses to the short-term effects of Amazonian deforestation. *Mol. Ecol.* 24, 2433–2448.

Newman, M. M., Hoilett, N., Lorenz, N., Dick, R. P., Liles, M. R., Ramsier, C., et al. (2016). Glyphosate effects on soil rhizosphere-associated bacterial communities. *Sci. Total Environ.* 543, 155–160.

Nkongolo, K. K., and Narendrula-Kotha, R. (2020). Advances in monitoring soil microbial community dynamic and function. *J. Appl. Genet.* 61, 249–263.

Ordway, E. M., Naylor, R. L., Nkongho, R. N., and Lambin, E. F. (2017). Oil palm expansion in Cameroon: Insights into sustainability opportunities and challenges in Africa. *Glob. Environ. Chang.* 47, 190–200.

Oton, E. V., Quince, C., Nicol, G. W., Prosser, J. I., and Gubry-Rangin, C. (2016). Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeota. *ISME J.* 10, 85–96.

Paoletti, A., Darras, K., Jayanto, H., Grass, I., Kusrini, M., and Tscharntke, T. (2018). Amphibian and reptile communities of upland and riparian sites across Indonesian oil palm, rubber and forest. *Glob. Ecol. Conserv.* 16, e00492.

Pedrinho, A., Mendes, L. W., Merloti, L. F., da Fonseca, M. de C., Cannavan, F. de S., and Tsai, S. M. (2018). Forest-to-pasture conversion and recovery based on assessment of microbial communities in Eastern Amazon rainforest. *FEMS Microbiol. Ecol.* 95.

Phommexay, P., Satasook, C., Bates, P., Pearch, M., and Bumrungsri, S. (2011). The impact of rubber plantations on the diversity and activity of understorey insectivorous bats in southern Thailand. *Biodivers. Conserv.* 20, 1441–1456.

Prabowo, W. E., Darras, K., Clough, Y., Toledo-Hernandez, M., Arlettaz, R., Mulyani, Y. A., et al. (2016). Bird responses to lowland rainforest conversion in Sumatran smallholder landscapes, Indonesia. *PLoS One* 11, e0154876.

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844.

Quinn, T. P., Crowley, T. M., and Richardson, M. F. (2018a). Benchmarking differential expression analysis tools for RNA-Seq: normalization-based vs. log-ratio transformation-based methods. *BMC Bioinformatics* 19, 274.

Quinn, T. P., Erb, I., Richardson, M. F., and Crowley, T. M. (2018b). Understanding sequencing data as compositions: an outlook and review. *Bioinformatics* 34, 2870–2878.

Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., et al. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome* 7, 133.

Rembold, K., Mangopo, H., Tjitrosoedirdjo, S. S., and Kreft, H. (2017). Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. *Biol. Conserv.* 213, 234–242.

Rennenberg, H., Dannenmann, M., Gessler, A., Kreuzwieser, J., Simon, J., and Papen, H. (2009). Nitrogen balance in forest soils: nutritional limitation of plants under climate change stresses. *Plant Biol.* 11, 4–23.

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., et al. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* 4, 1340–1351.

Rubiana, R., Rizali, A., Denmead, L. H., Alamsari, W., Hidayat, P., Pudjianto, D. H., et al. (2015). Agricultural land use alters species composition but not species richness of ant communities. *Asian Myrmecology* 7, 73–85.

Schneider, D., Engelhaupt, M., Allen, K., Kurniawan, S., Krashevska, V., Heinemann, M., et al. (2015). Impact of lowland rainforest transformation on diversity and composition of soil prokaryotic communities in Sumatra (Indonesia). *Front. Microbiol.* 6.

Schöler, A., Jacquiod, S., Vestergaard, G., Schulz, S., and Schloter, M. (2017). Analysis of soil microbial communities based on amplicon sequencing of marker genes. *Biol. Fertil. Soils* 53, 485–489.

Schulz, G., Schneider, D., Brinkmann, N., Edy, N., Daniel, R., Polle, A., et al. (2019). Changes in trophic groups of protists with conversion of rainforest into rubber and oil palm plantations. *Front. Microbiol.* 10, 240.

Shanmugam, S., Dalal, R. C., Joosten, H., Raison, R. J., and Joo, G. K. (2018). SOC stock changes and greenhouse gas emissions following tropical land use conversions to plantation crops on mineral soils, with a special focus on oil palm and rubber plantations. *Agriculture* 8, 133.

Sohng, J., Singhakumara, B. M. P., and Ashton, M. S. (2017). Effects on soil chemistry of tropical deforestation for agriculture and subsequent reforestation with special reference to changes in carbon and nitrogen. *For. Ecol. Manage.* 389, 331–340.

Soni, R., Suyal, D. C., Sai, S., and Goel, R. (2016). Exploration of nifH gene through soil metagenomes of the western Indian Himalayas. *3 Biotech* 6, 25.

Steffan-Dewenter, I., Kessler, M., Barkmann, J., Bos, M. M., Buchori, D., Erasmi, S., et al. (2007). Tradeoffs between income, biodiversity, and ecosystem functioning during tropical rainforest conversion and agroforestry intensification. *Proc. Natl. Acad. Sci.* 104, 4973 LP – 4978.

Streit, W. R., and Schmitz, R. A. (2004). Metagenomics – the key to the uncultured microbes. *Curr. Opin. Microbiol.* 7, 492–498.

Symes, W. S., Edwards, D. P., Miettinen, J., Rheindt, F. E., and Carrasco, L. R. (2018). Combined impacts of deforestation and wildlife trade on tropical biodiversity are severely underestimated. *Nat. Commun.* 9, 4052.

Taheripour, F., Hertel, T. W., and Ramankutty, N. (2019). Market-mediated responses confound policies to limit deforestation from oil palm expansion in Malaysia and Indonesia. *Proc. Natl. Acad. Sci.* 116, 19193–19199.

Tanaka, S., Kendawang, J. J., Yoshida, N., Shibata, K., Jee, A., Tanaka, K., et al. (2005). Effects of Shifting Cultivation on Soil Ecosystems in Sarawak, Malaysia IV. Chemical Properties of the Soils and Runoff Water at Niah and Bakam Experimental Sites. *Soil Sci. Plant Nutr.* 51, 525–533.

Tarigan, S. D., Sunarti, and Widyaliza, S. (2015). Expansion of Oil Palm Plantations and Forest Cover Changes in Bungo and Merangin Districts, Jambi Province, Indonesia. *Procedia Environ. Sci.* 24, 199–205.

Torriani, S., Felis, G. E., and Dellaglio, F. (2001). Differentiation of Lactobacillus plantarum, L. pentosus, and L. paraplantarum by recA gene sequence analysis and multiplex PCR assay with recA gene-derived primers. *Appl. Environ. Microbiol.* 67, 3450 LP – 3454.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., et al. (2005). Comparative Metagenomics of Microbial Communities. *Science (80-.).* 308, 554 LP – 557.

Tripathi, B. M., Kim, M., Singh, D., Lee-Cruz, L., Lai-Hoe, A., Ainuddin, A. N., et al. (2012). Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. *Microb Ecol* 64, 474–484.

Veldkamp, E., Purbopuspito, J., Corre, M. D., Brumme, R., and Murdiyarso, D. (2008). Land use change effects on trace gas fluxes in the forest margins of Central Sulawesi, Indonesia. *J. Geophys. Res. Biogeosciences* 113.

Vieira, I. C. G., Toledo, P. M., Silva, J. M. C., and Higuchi, H. (2008). Deforestation and threats to the biodiversity of Amazonia. *Brazilian J. Biol.* 68, 949–956.

Vos, M., Wolf, A. B., Jennings, S. J., and Kowalchuk, G. A. (2013). Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiol. Rev.* 37, 936–954.

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27.

Wemheuer, B., Güllert, S., Billerbeck, S., Giebel, H.-A., Voget, S., Simon, M., et al. (2014). Impact of a phytoplankton bloom on the diversity of the active bacterial community in the southern North Sea as revealed by metatranscriptomic approaches. *FEMS Microbiol. Ecol.* 87, 378–389.

Wemheuer, F., Taylor, J. A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T., et al. (2018). Tax4Fun2: a R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene marker gene sequences. *bioRxiv*, 490037.

Wilcove, D. S., Giam, X., Edwards, D. P., Fisher, B., and Koh, L. P. (2013). Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. *Trends Ecol Evol* 28, 531–540.

Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Van Braekel, J., et al. (2020). Targeting the 16S rRNA Gene for Bacterial Identification in Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford Nanopore Technologies) Generation Sequencing Technologies. *Int. J. Mol. Sci.* 21.

Xu, Z., Hansen, M. A., Hansen, L. H., Jacquiod, S., and Sørensen, S. J. (2014). Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PLoS One* 9, e93445–e93445.

Zhalnina, K., Dias, R., de Quadros, P. D., Davis-Richardson, A., Camargo, F. A. O., Clark, I. M., et al. (2015). Soil pH Determines Microbial Diversity and Composition in the Park Grass Experiment. *Microb Ecol* 69, 395–406.

Zhang, Y., Zhao, Z., Dai, M., Jiao, N., and Herndl, G. J. (2014). Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. *Mol. Ecol.* 23, 2260–2274.

Zhou, W.-J., Ji, H., Zhu, J., Zhang, Y.-P., Sha, L.-Q., Liu, Y.-T., et al. (2016). The effects of nitrogen fertilization on N2O emissions from a rubber plantation. *Sci. Rep.* 6, 28230.

2

How rainforest conversion to agricultural systems in Sumatra (Indonesia) affects active soil bacterial communities

Dirk Berkelmann[1], Dominik Schneider[1], Martin Engelhaupt[1†], Melanie Heinemann[1], Stephan Christel1[‡], Marini Wijayanti[2], Anja Meryandini[2] and Rolf Daniel[1]

[1]Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University, Göttingen, Germany,

[2]Department of Biology, Faculty of Mathematics and Natural Sciences IPB, Bogor Agricultural University, Bogor, Indonesia

Check for updates

How Rainforest Conversion to Agricultural Systems in Sumatra (Indonesia) Affects Active Soil Bacterial Communities

Dirk Berkelmann[1], Dominik Schneider[1], Martin Engelhaupt[1†], Melanie Heinemann[1], Stephan Christel[1‡], Marini Wijayanti[2], Anja Meryandini[2] and Rolf Daniel[1]*

[1] Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University, Göttingen, Germany, [2] Department of Biology, Faculty of Mathematics and Natural Sciences IPB, Bogor Agricultural University, Bogor, Indonesia

Palm oil production in Indonesia increased constantly over the last decades, which led to massive deforestation, especially on Sumatra island. The ongoing conversion of rainforest to agricultural systems results in high biodiversity loss. Here, we present the first RNA-based study on the effects of rainforest transformation to rubber and oil palm plantations in Indonesia for the active soil bacterial communities. For this purpose, bacterial communities of three different converted systems (jungle rubber, rubber plantation, and oil palm plantation) were studied in two landscapes with rainforest as reference by RT-PCR amplicon-based analysis of 16S rRNA gene transcripts. Active soil bacterial communities were dominated by *Frankiales* (*Actinobacteria*), subgroup 2 of the *Acidobacteria* and *Alphaproteobacteria* (mainly *Rhizobiales* and *Rhodospirillales*). Community composition differed significantly between the converted land use systems and rainforest reference sites. *Alphaproteobacteria* decreased significantly in oil palm samples compared to rainforest samples. In contrast, relative abundances of taxa within the *Acidobacteria* increased. Most important abiotic drivers for shaping soil bacterial communities were pH, calcium concentration, base saturation and C:N ratio. Indicator species analysis showed distinct association patterns for the analyzed land use systems. Nitrogen-fixing taxa including members of *Rhizobiales* and *Rhodospirillales* were associated with rainforest soils while nitrifiers and heat-resistant taxa including members of *Actinobacteria* were associated with oil palm soils. Predicted metabolic profiles revealed that the relative abundances of genes associated with fixation of nitrogen significantly decreased in plantation soils. Furthermore, predicted gene abundances regarding motility, competition or gene transfer ability indicated rainforest conversion-induced changes as well.

Keywords: 16S rRNA gene transcripts, soil bacterial communities, rainforest conversion, active bacterial communities, oil palm plantation, Sumatra

INTRODUCTION

Palm oil and rubber production play a crucial role for the economy in several countries. Especially in Indonesia, as one of the top producers of palm oil and rubber, conversion of natural systems to agricultural systems almost doubled from 2000 to 2009 (Angelsen, 1995; McCarthy, 2010; Oosterveer, 2015; Ivancic et al., 2016). In most cases, primary and secondary rainforests were

converted to managed cash crop systems. Since the major part of the global biodiversity is inherited by tropical forests, the enormous biodiversity harbored by Indonesians rainforests was reduced drastically during this the process. Consequently, deforestation and conversion to agricultural systems in tropical regions is considered the biggest threat to global biodiversity. This affects not only animal and plant communities, but also microbial communities and tropical ecosystem functions as well (Donald, 2004; Sodhi et al., 2004; Koh et al., 2011; Wilcove et al., 2013; Barnes et al., 2014).

Microbial and, in particular, bacterial communities drive almost all biogeochemical cycles and are involved in nutrient cycling in soils (Fierer and Jackson, 2006; Delmont et al., 2011, 2012). Therefore, soil bacteria are closely connected to the lifestyles of other organisms and nutrient availability itself. Additionally, it was suggested that the community response of soil bacteria toward changes in nutrient availability and plant diversity follows predictable patterns (Waldrop et al., 2000; Leff et al., 2015). In soil, the involvement of microbes in nutrient cycling is crucial for soil fertility and therefore for plant growth and growing cash crops (Bhardwaj et al., 2014; Lynch, 2015). In the last years, several studies investigated the effects of logging and land transformation on soil bacterial communities and confirmed that rainforest conversion to oil palm or rubber plantations has severe impacts on soil prokaryotic diversity and composition (Lee-Cruz et al., 2013; Schneider et al., 2015; Kerfahi et al., 2016). Soil bacterial and archaeal diversity increased with increasing land use intensity and biomass decreased with pH and C:N ratio, which were identified as main abiotic drivers of bacterial community formation (Allen et al., 2015; Schneider et al., 2015). It was also shown that some procedures of rainforest exploitation, like logging, appear reversible, which makes further research on the topic even more crucial (Tripathi et al., 2012; Lee-Cruz et al., 2013; Kerfahi et al., 2016).

The results of most previous studies were obtained by DNA-based 16S rRNA gene analyses representing the entire community whereas effects on activity and functional distribution of prokaryotic groups have rarely been addressed. Since microorganisms can be abundant while remaining inactive or dormant and even can have different numbers of ribosomal operons, the actual impact and importance for the corresponding communities or ecosystem can be biased (Urich et al., 2008; Větrovský and Baldrian, 2013; Wemheuer et al., 2014). To avoid this issue, analyses based on RNA are required as well. Studies that aim to analyze bacterial activity by using 16S rRNA transcripts are well established for marine environments (Wemheuer et al., 2012, 2014; Rodríguez-Blanco et al., 2013; Zhang et al., 2014; Stibal et al., 2015). In soil and other terrestrial environments, however, only a limited number of RNA-based studies are available, which address the effects on land use conversion and showed that entire community analysis alone can lead to false conclusions regarding community activity (Foesel et al., 2014; Herzog et al., 2015; Mueller et al., 2016; Ragot et al., 2016).

As part of the "Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems (Sumatra, Indonesia)" (EFForTS) collaborative research center, we investigated the impact of rainforest conversion on the active soil bacterial community and deduced their functional responses. This study is one of the first that investigates the effect of rainforest transformation to rubber and oil palm plantations on active bacterial soil communities in Indonesia.

Three different agricultural systems comprising intensively managed oil palm plantations, rubber plantations and jungle rubber were studied and compared to secondary rainforest in two different landscapes in Southwest Sumatra, Indonesia. Based on large-scale amplicon-based analysis of 16S rRNA transcripts, changes in the active bacterial communities were analyzed and correlated with biotic and abiotic factors. Besides investigating the active bacterial community composition and diversity, we also investigated effects on functional traits and compared our results to previous DNA-based studies.

As a guideline we followed three hypotheses: (a) Rainforest transformation leads to no significant changes of soil bacterial diversity, whereas (b) active bacterial community structure and composition is significantly affected. Furthermore, we hypothesized that (c) for rainforest and transformed land use systems specific indicator species can be detected and predicted metabolic profiles show significant functional shifts of the active bacterial community.

MATERIALS AND METHODS

Sampling and Sample Treatment

Two landscapes in southwest Sumatra (Indonesia), the Harapan Rainforest Concession (H) and Bukit Duabelas (B), were selected for sampling (**Figure 1**). Sampling was conducted on the plots as described by Schneider et al. (2015). Soil texture differed with primarily loam Acrisol soils in Harapan and clay Acrisol soils in Bukit Duabelas. Both landscapes harbored secondary rainforest (named "BF" and "HF") and three different land use systems representing different land use intensities, which derived from rainforest conversion. The agricultural systems comprise oil palm plantations (monocultures of *Elaeis guineensis*; designated "BO" and "HO"), rubber plantations (monocultures of *Hevea brasiliensis*; designated "BR" and "HR") and rubber agroforest (designated "BJ" an "HJ"). The latter represents a traditional agroforestry system known as "jungle rubber" in which rubber trees are planted in secondary rainforest without fertilizer or liming input. Each land use system consisted of four core plots in each landscape, with three subplots (five by five meters) per core plot, resulting in 96 subplot samples out of the 32 core plots in total. A soil corer was used to take three cores at each subplot of the upper seven cm of top soil and a diameter of five cm. The three soil samples per subplot were homogenized and coarse roots and stones (> 5 mm) were removed. To prevent RNAs from degradation, RNAprotect Bacteria Reagent was applied to the samples as recommended by the manufacturer (Qiagen, Hilden, Germany). Samples were stored at −80°C until further use.
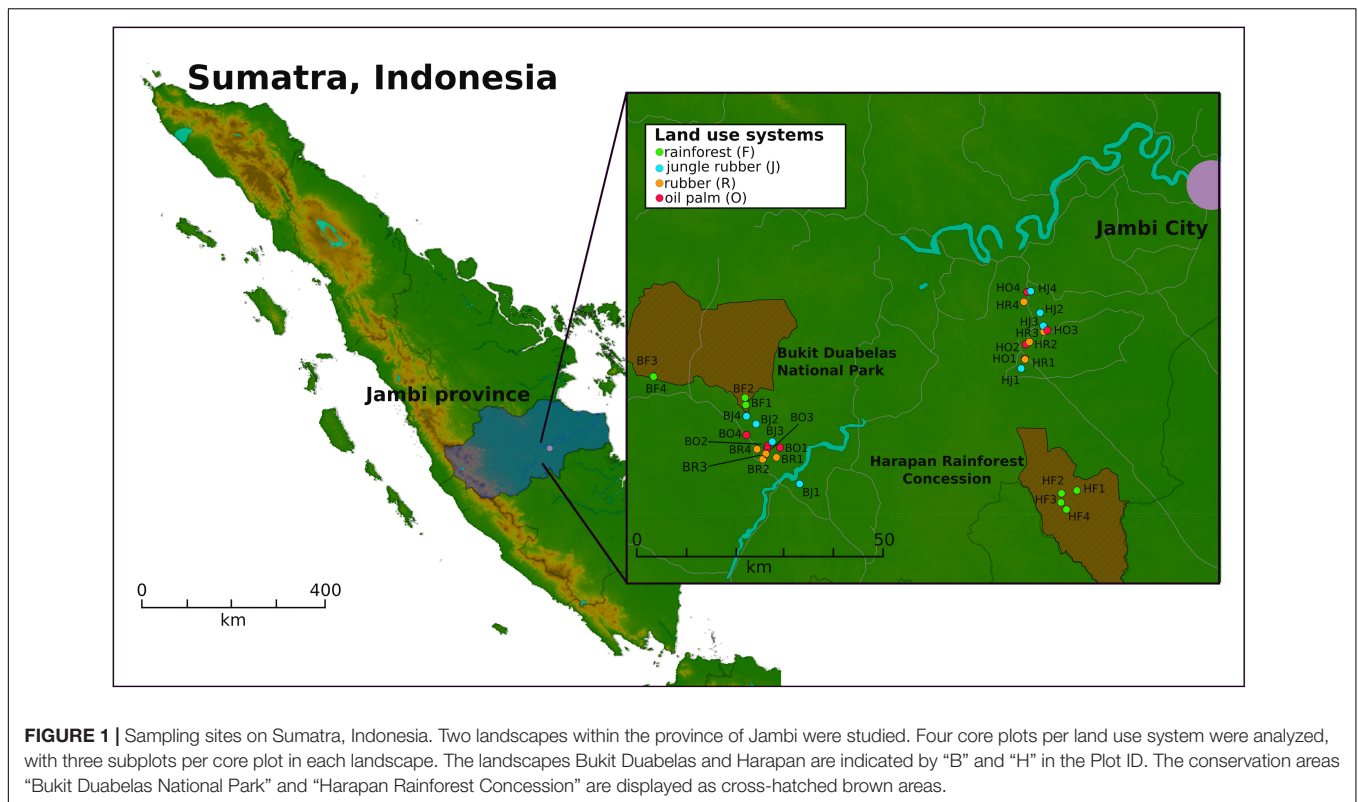
FIGURE 1 | Sampling sites on Sumatra, Indonesia. Two landscapes within the province of Jambi were studied. Four core plots per land use system were analyzed, with three subplots per core plot in each landscape. The landscapes Bukit Duabelas and Harapan are indicated by "B" and "H" in the Plot ID. The conservation areas "Bukit Duabelas National Park" and "Harapan Rainforest Concession" are displayed as cross-hatched brown areas.

Age of rubber trees ranged from 15 to 40 and 6 to 16 years in jungle rubber and rubber plantations, respectively. In oil palm plantations, the age of the tress varied between 8 to 15 years (Kotowska et al., 2015). Management for the two plantation types included application of herbicides every 6 months and addition of inorganic NPK fertilizer [100–300 kg ha^{-1} yr^{-1} in rubber plantations and 300–600 kg ha^{-1} yr^{-1} in oil palm plantations (Kotowska et al., 2015)]. All soil parameters [pH, P, N, C, C:N ratio, Al, Ca, Fe, K, Mg, Mn, Na, effective cation exchange capacity (ECEC) and base saturation] were retrieved from Allen et al. (2015).

RNA Isolation, cDNA Synthesis, and 16S rRNA Amplification

To analyze the active part of the bacterial communities, RNA was isolated from all 96 samples by using the MoBio PowerSoil RNA Isolation Kit (MO BIO Laboratories, Hilden, Germany). Isolation was initiated by using 0.5 g sample material. Residual DNA was removed by treatment with Turbo DNase as suggested by the manufacturer (Applied Biosystems, Darmstadt, Germany). The reaction mixture was subsequently purified and concentrated by using the RNeasy MinElute Cleanup Kit as recommended by the manufacturer (Qiagen). To verify the complete removal of DNA, a PCR reaction targeting the 16S rRNA gene was performed as described by Schneider et al. (2015). RNA yields were estimated by employing a Qubit® Fluorometer as recommended by the manufacturer (Thermo Fisher Scientific, Waltham, MA, United States).

RNA was converted to cDNA by using the SuperScript® III Reverse Transcriptase (Thermo Fisher Scientific, Waltham, MA, United States). As described by Wemheuer et al. (2015), a specific primer for the conserved region downstream to the targeted bacterial 16S rRNA gene region (5′-CCGTCAATTCMTTTGAGT-′) was used for cDNA synthesis. The reaction mixture (20 µl final volume) contained 4 µl of fivefold reaction buffer, 500 µM of each deoxynucleoside triphosphate, 5 mM DTT, 1 µM reverse primer, 1 U RiboLock^{TM} RNase Inhibitor (Thermo Fisher Scientific, Schwerte, Germany) and 200 U of reverse transcriptase. The solution was incubated at 55°C for 1 h and subsequently inactivated by incubation at 70°C for 15 min. To remove residual RNA in the RNA/DNA hybrids, 2.5 U RNase H (Thermo Fisher Scientific, Schwerte, Germany) was added and incubated at 37° for 15 min followed by an additional inactivation at 65°C for 10 min. Obtained cDNA was subsequently used for amplification of the hypervariable V3 to V5 regions of the 16S rRNA transcript [Forward primer: V3for_B   5′-CGTATCGCCTCCCTCGCGCCATCAG-MID-TACGGRAGGCAGCAG-3′ (Liu et al., 2007) reverse primer: V5rev_B   5′-CTATGCGCCTTGCCAGCCCGCTCAG-MID-CCGTCAATTCMTTTGAGT-3′ (Wang and Qian, 2009)]. The following thermal cycling scheme was used for amplification of partial bacterial 16S rRNA: initial denaturation at 98°C for 5 min, 25 cycles of denaturation at 98°C for 45 s, annealing for 45 s at 65°C, and extension at 72°C for 30 s, followed by a final extension period at 72°C for 5 min. All amplicon PCR reactions were performed in triplicate and pooled in equimolar amounts for sequencing. The Göttingen Genomics Laboratory determined

the sequences of the 16S rRNA amplicons by using a 454 GS-FLX sequencer and Titanium chemistry following the instructions of the manufacturer (Roche, Mannheim, Germany) for amplicon sequencing.

Analysis of 16S rRNA Transcripts

Raw sequences were processed and analyzed using QIIME 1.9.1 (Caporaso et al., 2010). Sequences with lengths below 300 and over, 1000 bp, quality scores below 25 and homopolymer stretches of more than 8 bp were removed by employing *split_libraries.py*. Denoising was performed with Acacia (default settings), chimera removal with UCHIME and reverse primer removal with cutadapt (Edgar et al., 2011; Bragg et al., 2012).

Determination of operational taxonomic units (OTUs) was performed at genetic divergence of 3% by employing the *pick_open_reference_otus.py* script with the SILVA NR SSU 128 database as reference. Taxonomic classification was performed with the same reference database and *parallel_assign_taxonomy_blast.py* script. OTU tables were created by employing *make_otu_table.py*. Further polishing including removal of singletons, chloroplast sequences, unclassified OTUs and extrinsic domain OTUs) was carried by employing *filter_otu_table.py*. Comparison of samples was performed by creating subsamples with identical sequence numbers (6,650 per subsample). Rarefaction estimates were done by *alpha_rarefaction.py*. Diversity indices, Shannon index and PD (phylogenetic diversity) index were calculated by *alpha_diversity.py*. Statistical tests were performed in R by employing standard functions and the "vegan" package (R Development Core Team, 2017). Data distribution and homogeneity of variance were determined by the Shapiro test, implemented in R (R Development Core Team, 2017). For determination of significant differences between treatments and samples, PERMANOVA analysis was performed with the "vegan" and "RVAideMemoire" packages in R. The "vegan" package was also used for calculation of distance matrices, clustering analysis and non-metric multidimensional scaling (NMDS) based on a weighted UniFrac distance matrix (Lozupone et al., 2011). For NMDS analysis, sample sequences were merged at core plot level with a resulting subsample size of 19,950.

For statistical analysis of abundance differences of single taxonomic groups between land use systems, normal distribution of values was tested first with Shapiro test in R. Depending on the result, normally distributed samples were analyzed with an ANOVA and Tukey test as *post hoc* tests. Non-normally distributed samples were tested by Kruskal–Wallis test and Pairwise Wilcoxon test as *post hoc* tests. Results were interpreted as significant with $p < 0.05$.

Identification of bacterial genera associated to the analyzed land use systems was performed by using the "Indicspecies" package in R. We calculated an abundance-based version of the phi coefficient of association, the point biserial correlation coefficient via the *multipatt* command based on abundance based OTU data (**Supplementary Table S2**). Prior to analysis, all OTUs belonging to the same genus were summarized. We visualized the associated taxa in a network that used the analyzed land uses as source nodes and the associated bacterial taxa as nodes, while the correlation coefficients were used as edges. Only taxa with significant correlation coefficients ($p < 0.05$) were included. Network generation was performed with Cytoscape version 3.5.1 by using the *edge-weighted spring embedded layout* algorithm, with edge width corresponding to the correlation coefficients and taxa abundance to node size.

Analysis of activity and metabolism was performed via functional predictions, which were calculated on version 123 of the SILVA database with the "Tax4Fun" package in R and visualized with the "gplots" package in R (Asshauer et al., 2015).

Accession Numbers

The 16S rRNA transcript sequences were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession number PRJNA278020.

RESULTS AND DISCUSSION

Impact of Different Land Use Systems on Active Soil Bacterial Community Composition

We analyzed a management gradient with increasing intensity from rainforest reference sites to jungle rubber over rubber plantations to oil palm plantations in two landscapes. The entire dataset comprised 1,333,275 high-quality 16S rRNA transcript sequences, which belonged to 32,280 different OTUs at species level (3% dissimilarity) (**Supplementary Tables S1, S2**). Species richness was highest in oil palm followed by rubber, rainforest, and jungle rubber (**Supplementary Figures S1, S2** and **Table S1**). Diversity indices Shannon and PD did not show significant differences between the analyzed land use systems ($p > 0.11$ and $p > 0.06$, respectively; **Supplementary Table S1**). Similar trends were observed in DNA-based studies targeting the conversion of rainforest to agricultural systems in which alpha and/or beta diversity increased as well, although the results were not completely consistent with respect to statistical significance (Carney et al., 2004; Tripathi et al., 2012; Schneider et al., 2015; Kerfahi et al., 2016).

Composition of the active soil bacterial community showed similar patterns in the corresponding land use systems of both analyzed landscapes Bukit Duabelas and Harapan (**Figure 2**). *Proteobacteria* decreased continuously with increasing land use intensity in Bukit Duabelas and in Harapan (46.3 and 52.2% in rainforest to 29 and 28.3% in oil palm plantation, respectively), while the abundance of *Acidobacteria* showed a maximum in jungle rubber (42.8% in Bukit Duabelas and 47.6% in Harapan) and rubber systems (47.9% in Bukit and 46.7% in Harapan) before decreasing again in oil palm (42.5% in Bukit Duabelas and 36.9% in Harapan).

The different land use systems within the landscapes showed significant changes in the active bacterial community composition ($p < 0.002$; **Supplementary Table S3**). The abundances of *Rhizobiales* within the *Proteobacteria* decreased with increasing land use intensity (18.5% in rainforest to 10.45%

FIGURE 2 | Community composition of active soil bacterial communities in three different land use systems and rainforest reference sites in two different landscapes. All plots were clustered according to their respective land use and landscape. Community compositions are displayed as relative abundances at order level based on 16S rRNA sequences obtained from extracted soil RNA. Taxa with abundances below 1% in each land use system were summarized as "rare taxa." The detected orders are grouped to corresponding phylum (for details see **Supplementary Table S2**).

in oil palm for Bukit Duabelas and 17.3% in rainforest to 10.6% in oil palm for Harapan), whereas *Frankiales* of the *Actinobacteria* increased (13% from rainforest to 15.6% in oil palm in Bukit Duabelas and from 8.6% in rainforest to 12.4% in oil palm in Harapan). Interestingly, *Acidobacteriales* and especially the *Acidobacteria* subgroup 2 showed an increase of abundance from rainforest (13.4% in Bukit Duabelas and 13% in Harapan) to rubber (22.6% in Bukit Duabelas and 19.6% in Harapan), which decreased in oil palm (17.9% in Bukit Duabelas and 18.1% in Harapan). *Proteobacteria* and *Acidobacteria* were reported to be very abundant at DNA level in the studied sites and in other locations (Rousk et al., 2010; Tripathi et al., 2012; Schneider et al., 2015). Here, at RNA level, *Proteobacteria* and *Acidobacteria* were the most abundant phyla as well. This is not a surprise, since *Acidobacteria* and *Alphaproteobacteria* were previously reported as generally high abundant in soils and important for

decomposition of soil carbon (Hansel et al., 2008; Leff et al., 2012). *Proteobacteria* and *Acidobacteria* were reported to be very abundant at DNA level in the studied sites and in other locations (Rousk et al., 2010; Tripathi et al., 2012; Schneider et al., 2015). Here, at RNA level, *Proteobacteria* and *Acidobacteria* were the most abundant phyla as well. This is not a surprise, since *Acidobacteria* and *Alphaproteobacteria* were previously reported as generally high abundant in soils and important for decomposition of soil carbon (Hansel et al., 2008; Leff et al., 2012).

As mentioned before, the abundant *Rhizobiales*, which constitute the major part of detected *Proteobacteria*, decreased with increased land use intensity and increased fertilizer application from rainforest to oil palm plantation. As *Rhizobiales* are known to be involved in plant-associated and free-living N_2 fixation, the higher availability of usable nitrogen compounds in

fertilized systems reduce the requirement for microbial nitrogen fixation and favors other phylogenetic groups (Yoneyama et al., 2017). Furthermore, denitrification might be affected as well, since several taxa within the *Rhizobiales* were also reported to be involved in *nirK*-mediated denitrification (Bremer et al., 2007; Yoshida et al., 2009). In contrast, nitrification-related taxa like the *Nitrosomonadales* or *Nitrospira* increased from rainforest to rubber and oil palm plantations (relative abundances lower than 1%; data not shown). Additionally, it is notable that while *Proteobacteria* abundance decreased, *Acidobacteria* abundances (especially subgroup 2) increased, indicating negative correlations between these groups. Despite their high abundances in several studies, the ecological role of *Acidobacteria* in soil is still poorly understood. Some studies provide contradicting results in which positive as well as negative correlations as response to high nutrient input are mentioned for this taxon (Kielak et al., 2016). Interestingly, positive correlations between *Proteobacteria* and *Acidobacteria* have been shown as well, which is contrary to our results. This could be explained by the differences in acidobacterial subgroups detected in the other studies and the so far not completely understood roles of all subgroups within the *Acidobacteria* (Kielak et al., 2016).

Effects of Abiotic Soil Parameters on Active Bacterial Communities

Shape and structure of prokaryotic communities are tightly connected with their surrounding environment and the corresponding abiotic and biotic environmental factors (Brevik et al., 2015). Environmental parameters are crucial factors for investigating soil-borne bacterial communities in agricultural systems (Rousk et al., 2010). Especially, the severe biodiversity and nutrient content alterations in agricultural land use systems compared to rainforest are of high importance for addressing and understanding the impact of rainforest conversion on microbial communities (Corre et al., 2006; Junier et al., 2010; Rousk et al., 2010; Dam et al., 2014). Based on non-metric multidimensional scaling (NMDS), differences in active community composition were tightly connected to conversion of rainforest to agricultural land use systems (**Figure 3**). Rainforest samples clustered separately from the converted systems. Additionally, clustering analysis confirmed that soil bacterial communities from rainforest core plots are distinct from that in almost all core plots of the converted systems (**Supplementary Figure S3**). Base saturation ($p < 0.001$, r^2 0.410) was one of the main drivers of active bacterial communities together with pH ($p < 0.001$, r^2 0.780), Fe content ($p < 0.018$, r^2 0.248), C:N ratio ($p < 0.011$, r^2 0.30) and Ca content ($p < 0.037$, r^2 0.2392). The two analyzed landscapes showed no significant difference in this respect. Base saturation, which displays soil fertility, exhibited a major impact on composition of active soil bacterial communities. Additionally, another study conducted on the same plots reported that base saturation was decreasing with increasing land use intensities, indicating that decreasing soil fertility has a major influence on active bacterial community structure (Allen et al., 2015). Soil pH is known as one of the major drivers for soil prokaryotic communities and pH changes
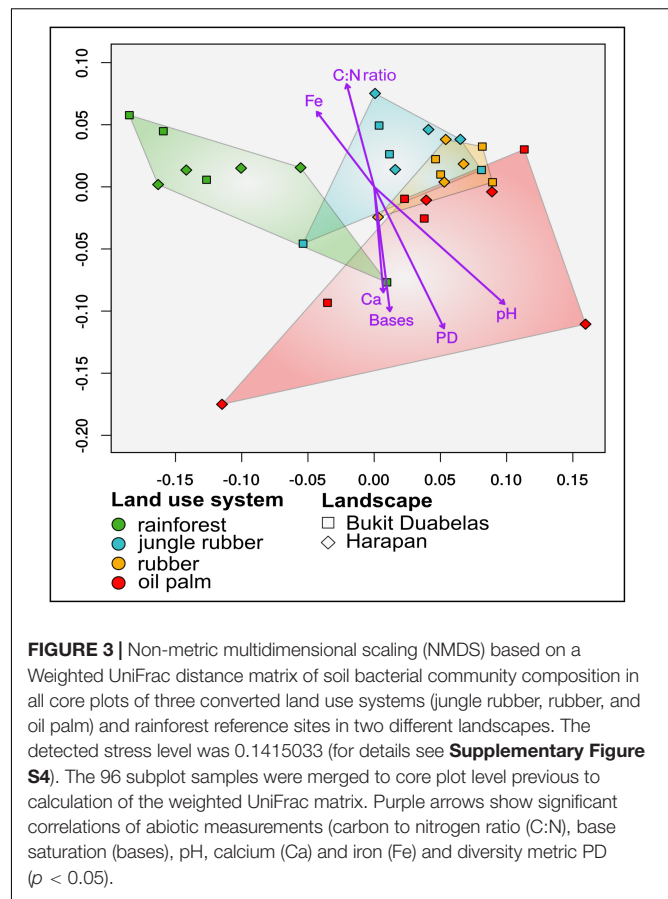
FIGURE 3 | Non-metric multidimensional scaling (NMDS) based on a Weighted UniFrac distance matrix of soil bacterial community composition in all core plots of three converted land use systems (jungle rubber, rubber, and oil palm) and rainforest reference sites in two different landscapes. The detected stress level was 0.1415033 (for details see **Supplementary Figure S4**). The 96 subplot samples were merged to core plot level previous to calculation of the weighted UniFrac matrix. Purple arrows show significant correlations of abiotic measurements (carbon to nitrogen ratio (C:N), base saturation (bases), pH, calcium (Ca) and iron (Fe) and diversity metric PD ($p < 0.05$).

were described as a common indirect consequence of fertilizer application in agricultural systems (Fierer et al., 2007; Foesel et al., 2014; Brenzinger et al., 2015; Herzog et al., 2015; Lammel et al., 2015; Stempfhuber et al., 2015; Zhalnina et al., 2015; Kaiser et al., 2016; Zhang et al., 2017). In our samples pH increased with higher land use intensity from rainforest to oil palm plantation. As biodiversity did not change significantly, pH might affect abundance of certain groups and consequently be the reason for the most prominent observed abundance changes within the *Alphaproteobacteria*, *Acidobacteriales,* and *Actinobacteria*.

Allen et al. (2015) suggested that a decreasing carbon to nitrogen ratio (C:N), which was identified as an additional significant driver for the active bacterial communities, typically indicates a shift toward a more bacteria-dominated system (Foesel et al., 2014; Allen et al., 2015). We identified C:N as a significant driver and observed for plantations compared to rainforest a decreasing C:N ratio in Harapan whereas the Bukit Duabelas samples showed similar values for all studied land use systems. Fe concentrations decreased from rainforest to the fertilized land use systems. Iron is often a limiting factor due to high demand not only of bacteria (Hibbing et al., 2009; Colombo et al., 2014). Thus, with decreasing iron content, we would expect a higher degree of bacterial competition and a community shaping effect of iron itself. Ca concentrations exhibited significant effects on active

communities and showed a positive correlation with increasing land use intensity from rainforest to oil palm plantation. This is most likely connected to liming practices and therefore to fertilizer application to counteract soil acidification (Tripathi et al., 2012). Therefore, it is indicated that soil bacterial community shifts after rainforest conversion were caused by fertilizer application, liming and reduction of plant diversity as suggested in previous studies (Rooney and Kennedy, 2013; Abdi et al., 2016; Val-Moraes et al., 2016; Liu et al., 2017). Previously observed changes of soil parameters after rainforest transformation to oil palm and rubber plantations indicated that the availability of N and other nutrients rely on continuous fertilization and liming (Allen et al., 2015). Thus, it is likely that the observed active bacterial community structure is highly dependent on ongoing treatment such as fertilizer application and liming.

Associations Between Specific Taxa of the Active Bacterial Community and Analyzed Land Uses

To identify genera, which were significantly associated with the analyzed land use systems and are suitable as indicators for one or combinations of land use systems, we calculated the point biserial correlation coefficient for each genus, which is the abundance-based counterpart of the phi coefficient. We detected for 270 (24%) of the 1,124 in total detected genera significant biserial correlation coefficients ($p < 0.05$). Most of these genera were associated to oil palm plantation (153 genera associated in total, 113 genera exclusively to oil palm) and rainforest (95 genera in total, 62 genera exclusively to rainforest). Furthermore, the point biserial correlation values and correspondingly the strength of association to the respective system were highest in oil palm plantation and rainforest compared to jungle rubber and rubber plantation (**Supplementary Figure S5**). The majority of associated genera in oil palm and rainforest belonged to the *Actinobacteria* (oil palm) and *Proteobacteria* (rainforest) (**Figure 4**). This is in accordance to our analysis of community composition in which *Proteobacteria* decreased from rainforest to oil palm plantation whereas *Actinobacteria* increased.

Strongest unique associations of rainforest were detected for genera within the *Rhizobiales* (*Beijerinckiaceae*), *Rhodospirillales* (*Acetobacteraceae*), and *Caulobacterales* (*Caulobacteraceae*). Since members of the *Rhizobiales* or *Rhodospirillales* like *Beijerinckiaceae* are associated with nitrogen fixation, the strong association to rainforest indicates a higher importance of nitrogen fixation for gaining usable nitrogen than in fertilized land use systems. Associations were also detected for *Rickettsiales,* which are also known for endosymbiotic relationships with eukaryotes (Taylor et al., 2012). This could be an indication for the higher biodiversity outside the bacterial domain in rainforests. The association of the acidobacterial subgroup 2 to rainforest is surprising, since its abundance increased from rainforest toward the studied agricultural land use systems. Except for one genus, all associations of genera within the *Clostridia* were exclusively to rainforest. Their ability to fix nitrogen in anaerobic environments might

FIGURE 4 | Association networks based on the point biserial correlation coefficient of genera related to the analyzed land use systems. Analyzed land use systems are displayed as hub nodes, while bacterial genera are displayed as nodes. Edges represent the point biserial correlation coefficient. Size of nodes and intensity of edges contribute to average abundance of genera in all land uses and weight of the respective point biserial correlation coefficient, respectively. Color of nodes contributes to prominent bacterial phyla. Hub nodes and edges are colored according to analyzed land use systems rainforest, jungle rubber, rubber, and oil palm. Only significant associations were included in the analysis ($p < 0.05$).

explain this when assuming a higher demand of nitrogen fixation in rainforest soils compared with agricultural managed soils (Hayat et al., 2010). Nitrogen-fixing clostridia were also reported as dominant in Amazonian rainforest soils, but contrary results have been reported for other tropical forest soils like the Brazilian Atlantic forest. Thus, these results indicate a specific association to soil properties rather than a general affiliation to tropical forest soils (Faoro et al., 2010).

In oil palm, the majority of associated species belonged to the *Actinobacteria*, followed by *Proteobacteria*. *Actinobacteria* are reported as tolerant to higher temperatures, as well as *Chloroflexi,* which were exclusively detected as associated taxa for oil palm (Bouskill et al., 2012). It has been proposed that higher light availability in oil palm plantations compared to rainforest, which also results in higher soil temperatures, affect the soil microbiome (Schneider et al., 2015). We detected the nitrification-related genus *Nitrospirales* and two genera of the *Nitrosomonadales* exclusively in oil palm plantations as associated taxa (Lücker et al., 2010; Ma et al., 2013). This indicates higher degrees of nitrification in oil palm soils compared to rainforest, probably caused by fertilizer application and correspondingly nitrogen input, leading to more favorable conditions for nitrifiers (Ma et al., 2013; Quan et al., 2016).

Interestingly, the detected genera with highest abundance belonged to *Acidobacteria* and were associated with combinations of land use systems. An uncultured genus within the subgroup 2 with an average abundance of 14.9%
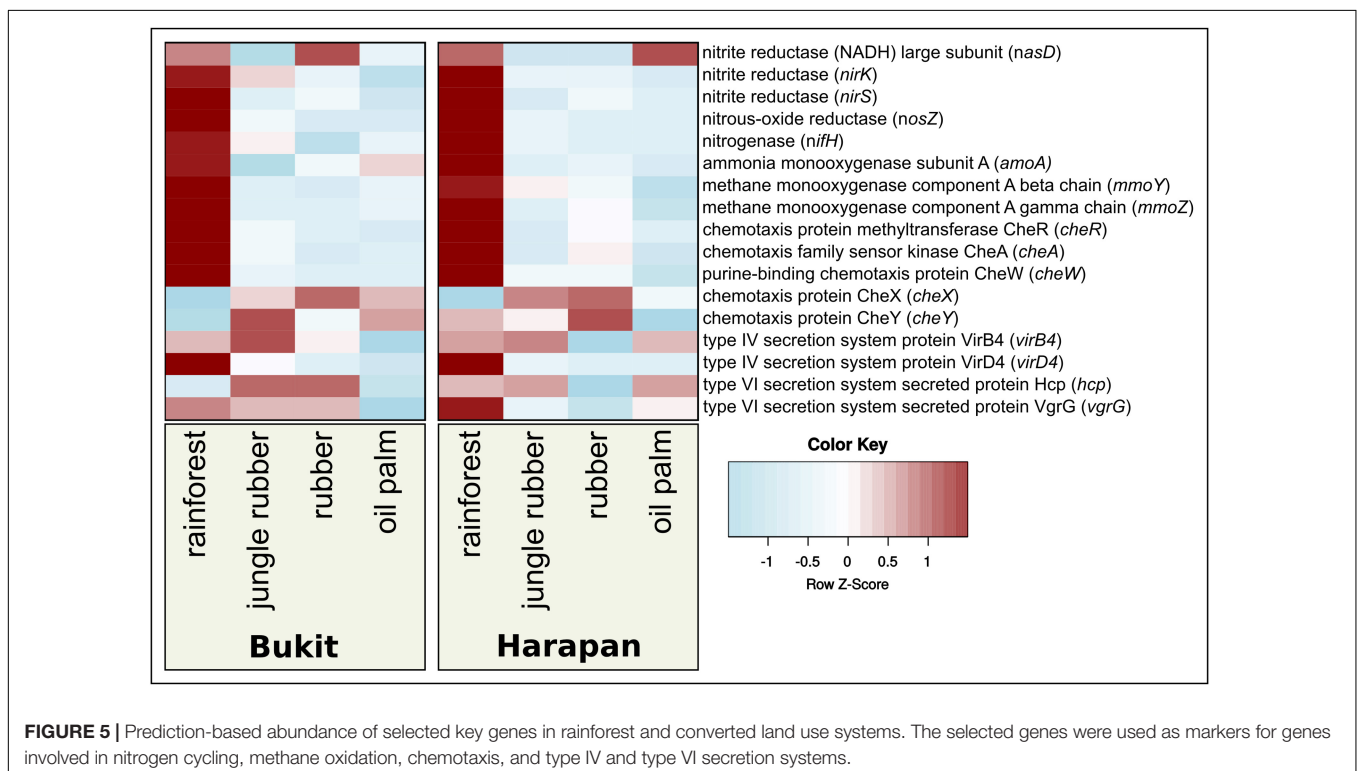
34

in all land use systems was associated with a combination of jungle rubber, rubber, and oil palm while an uncultured genus from subgroup 1 with an average abundance of 9.8% was characteristic for the combination of jungle rubber and rubber. In general, *Acidobacteriaceae* increased in abundance from rainforest to oil palm, with maxima in jungle rubber and rubber, indicating specific adaptations to these environments. The ecological role of the numerous subgroups within the *Acidobacteria* is under discussion. It was reported that the abundance of some subgroups increased during Amazonian rainforest conversion to managed soils whereas that of other subgroups decreased (Navarrete et al., 2015).

Overall, we could show that patterns of associated genera for different land use types and corresponding soil properties are distinguishable. The observed pattern corresponded to our other results, indicating a shift from higher abundances and associations of nitrogen fixation-related taxa in rainforest to conditions more favorable for groups associated with nitrification and heat-tolerance in plantations. This indicated a general shift of bacterial functions within these systems from higher importance of biological nitrogen acquisition in unfertilized systems to higher degrees of nitrification in fertilized land use systems.

Relationship Between Rainforest Conversion and Key Functions of the Bacterial Community

Besides shape and structures of soil bacterial communities, functional profiles and measures of activity are necessary to

study the bacterial responses. To obtain these profiles and activity measurements in ecological studies are often a challenge, since metatranscriptomes and high sample numbers are needed, which can be challenging in large scale projects and areas with infrastructure gaps. An alternative, especially for large sample numbers, are functional predictions based on 16S rRNA analysis using bioinformatic tools like "Tax4Fun" (Asshauer et al., 2015; Koo et al., 2017) or "PICRUSt" (Langille et al., 2013). It was shown that these tools provide a sufficient accuracy of functional profiles compared to those derived directly from metagenomic or metatranscriptomic sequence analyses. These tools are frequently used in various projects (Langille et al., 2013; Asshauer et al., 2015; Mukherjee et al., 2017; Wemheuer et al., 2017). In this study, we used Tax4Fun to investigate bacterial metabolic activity (**Figure 5**) and focused on all pathways that showed an abundance of at least 1 % (**Supplementary Table S4**). Within the thereby recovered 27 KEGG categories, relative abundances ranged from 1 to 8% (data not shown). We observed different patterns of predicted gene abundances within the analyzed land use systems in both landscapes, which were expected due to their different properties. Since previous results showed relations of bacterial community composition and fertilizer application, as well as bacterial community associations and land use changes, we analyzed the changes of relative abundances of predicted genes encoding marker enzymes for selected metabolic traits, e.g., nitrogen metabolism (**Supplementary Table S5**). We analyzed predicted gene abundances of the ammonium monooxygenase subunit A (*amoA*) for nitrification, nitrogenase (*nifH*) for nitrogen fixation, nitrous oxide reductase (*nosZ*) for denitrification

FIGURE 5 | Prediction-based abundance of selected key genes in rainforest and converted land use systems. The selected genes were used as markers for genes involved in nitrogen cycling, methane oxidation, chemotaxis, and type IV and type VI secretion systems.

and assimilatory nitrite reductase (*nasD, nirK* and *nirS*) for nitrate/nitrite assimilation (Hai et al., 2009; Pappu et al., 2017). For methane-related metabolism, we analyzed the gene encoding the methane monooxygenase (*mmoY* and *mmoZ*) (Henckel et al., 1999; Murrell et al., 2000; Sengupta and Dick, 2017). In order to address bacterial interactions, we analyzed the chemotaxis related genes *cheA*, *cheW*, *cheY*, *cheX,* and *cheR* and for secretion systems the type IV secretion system genes *virB4* and *virD4* and the type VI secretion system genes *hcp* and *vgrG* (Cascales and Cambillau, 2012; Guglielmini et al., 2014; Briegel et al., 2015; Jones and Armitage, 2015; Zinicola et al., 2015; Fedi et al., 2016).

In both landscapes, predicted abundances for nitrogenase and nitrous oxide reductase gene abundances were significantly different between rainforest and converted systems (**Figure 5**). We observed highest numbers in rainforest and lowest in converted systems ($p < 0.05$). Predicted gene abundances for *nasD, nirK* did not show significant differences ($p > 0.05$). The predicted gene abundances for *mmoA* exhibited significant higher abundances in rainforest compared to jungle rubber and plantations ($p < 0.0002$).

The gene of the nitrous oxide reductase (*nosZ*) abundances was significantly higher in rainforest samples compared to transformed systems ($p < 0.0001$). Deduced abundances for the nitrite reductase gene *nirS* showed significant decreases from rainforest to all other studied land use systems in Harapan landscape ($p < 0.02$), but no significant changes in Bukit Duabelas Landscape. In contrast to Bukit Duabelas landscape, *amoA* showed significantly higher abundance in Harapan rainforest, while analyzed methane monooxygenases did not show significant differences (**Supplementary Table S5**).

Higher abundance of nitrogenase in rainforest suggests, that biological nitrogen uptake through nitrogen fixation is decreased in converted systems, which are in accordance to the recorded suppressed biological nitrogen fixation in converted systems (Corre et al., 2006; Pajares and Bohannan, 2016). We assume that the additional input of ready to use nutrients in agricultural systems is less beneficial for soil bacterial groups with the ability to fix nitrogen, resulting in lower abundances of nitrogen fixation related taxa like *Rhizobiales* in fertilized systems (Corre et al., 2006; Barron et al., 2009; Isobe et al., 2012; Waring et al., 2014; Cong et al., 2015; Pajares and Bohannan, 2016). Interestingly, *Alphaproteobacteria*, which include the *Rhizobiales*, were negatively affected as well by rainforest conversion as shown in **Figure 2**. In Harapan landscape, predicted abundances for genes involved nitrification were lower in converted systems as well. Methane metabolism is known to be linked to nitrogen metabolism, due to similarities between the ammonia monooxygenase and the methane monooxygenase (Henckel et al., 1999; Murrell et al., 2000; Sengupta and Dick, 2017). Both enzymes catalyze similar reactions and the corresponding bacterial groups bear the potential to outcompete each other (Akiyama et al., 2014; Zheng et al., 2014). Accordingly, the methane monooxygenase gene abundance was higher in rainforest than that of ammonia monooxygenase subunit A gene. Additionally, the predicted gene abundances for denitrification were higher in rainforest

compared to the other studied systems in both landscapes. Therefore, we assume that in converted systems with higher nitrogen disposition and availability, active soil bacterial communities respond to land use management and higher nutrient input with decreased nitrogen fixation. A decrease in nitrification and denitrification seems unlikely though, since previous studies demonstrated the increase of the activity of these processes under the influence of fertilizer input (Ma et al., 2013; Quan et al., 2016; Wang et al., 2018). Additionally, we observed that taxa which are associated with the above-mentioned nutrient cycling pathways underwent the most drastic changes in relative abundance as well. Indicator species analysis showed that genera with nitrification ability such as *Nitrospira* and *Nitrosomonadales* were detected in oil palm soils (**Figure 4**). However, since functional prediction cannot assign all taxa with their respective functional potential due to a lack of detailed information about certain groups (e.g., subgroups of *Acidobacteria*), additional analysis is required in this specific case (Kielak et al., 2016).

Abundances of chemotaxis genes *cheW* and *cheR* were highest in rainforest and decreased toward higher land use intensity (all *p*-values < 0.05). However, not all observed differences of the selected chemotaxis-related genes were highest in rainforest. The *cheX* gene abundance was lowest in rainforest and highest in plantations ($p < 0.05$). No significant differences were observed for sensor kinase *cheA* and response regulator *cheY*. For all tested chemotaxis-related genes, except *cheA* and *cheY*, we detected significant differences between rainforest and plantations. The analyzed *che* genes are part of the same operon and were shown to be connected to swarming capacity and especially in pathogens (Lambert et al., 2015; Fedi et al., 2016). Since we observed significant differences between rainforest and the other studied land use systems for almost all tested genes, we assume that rainforest transformation not only affects nutrient recycling but also interactions. Additionally, it is possible that due to strong association of the analyzed *che* genes with pathogenic lifestyles, community dynamics might be altered regarding pathogenicity as well.

The abundance changes of the investigated secretion system-related functions indicated an impact of rainforest conversion on interspecific bacterial activity. Type IV secretion system genes showed significant higher abundance in rainforest compared to oil palm samples ($p < 0.05$). We conclude that exchange of nucleic acids between bacteria is decreased in nutrient-rich managed land use systems, possibly due to a lower degree of competition resulting in a lower pressure for adaptations and hence nucleic acid exchange. Type VI secretion system genes showed slightly higher abundance in rainforest compared to oil palm ($p < 0.05$) as well. These results hint less negative interaction and competition by pathogenesis via antibacterial compounds (Tripathi et al., 2016). In contrast, we previously observed a significant effect of iron content on the active bacterial community composition, which is likely caused by competition for this limiting compound. However, we did not observe similar trends for all analyzed predicted genes regarding interaction

and competition, indicating the need for further analysis. We recorded several significant changes for the studied functions, but it has to be considered that prediction-based methods only provide indications, which have to be confirmed by additional analysis such as full metagenome and metatranscriptome analyses.

CONCLUSION

We confirmed our first hypothesis (a) that the diversity of the active bacterial community was not significantly affected by rainforest conversion. We showed that rainforest transformation has a significant impact on active bacterial community composition as suggested in hypothesis (b). Furthermore, we could show that changes in soil characteristics deriving from rainforest conversion and management are a major factor in reshaping the active bacterial community. Additionally, we identified that change in pH, base saturation, Fe content and C:N ratio is significant drivers of soil bacterial community composition. This suggests a direct connection to fertilizer applications and liming, which affects composition and amount of available nutrients, i.e., nitrogen-containing compounds. Rainforest soils and converted systems revealed distinguishable patterns of associated taxa, which illustrate the changed requirements for bacterial life in the different land use systems as mentioned in hypothesis (c). Furthermore, predicted functional profiles revealed that uptake of nutrients like nitrogen through biological fixation decreases with higher land use intensity. It was also indicated that interactions in form of nucleic acid exchange as well as antagonistic or competitive behavior were reduced after rainforest conversion and it is likely that rainforest transformation leads to soil bacterial communities with severely altered nutrient cycling activity. Thus, active bacterial communities are significantly affected by rainforest transformation. In addition to the impact on active community composition, we could show that changes of soil properties introduced by management (e.g., fertilizer application) are the main drivers for adaptations and probably changes in bacterial functioning.

AUTHOR CONTRIBUTIONS

RD designed and conceived the study. Soil sampling for prokaryotic community analysis was performed by ME, MW, and AM. DS, MH, and ME carried out the field and laboratory work. DB and DS prepared and analyzed the data. All authors interpreted the results and wrote the paper.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02381/full#supplementary-material

REFERENCES

Abdi, D., Ziadi, N., Shi, Y., Gagnon, B., Lalande, R., and Hamel, C. (2016). Residual effects of paper mill biosolids and liming materials on soil microbial biomass and community structure. *Can. J. Soil. Sci.* 97, 188–199. doi: 10.1139/CJSS-2016-0063

Akiyama, H., Morimoto, S., Tago, K., Hoshino, Y. T., Nagaoka, K., Yamasaki, M., et al. (2014). Relationships between ammonia oxidizers and N2O and CH4 fluxes in agricultural fields with different soil types. *Soil. Sci. Plant. Nutr.* 60, 520–529. doi: 10.1080/00380768.2014.904206

Allen, K., Corre, M. D., Tjoa, A., and Veldkamp, E. (2015). Soil nitrogen-cycling responses to conversion of lowland forests to oil palm and rubber plantations in Sumatra, Indonesia. *PLoS One* 10:e0133325. doi: 10.1371/journal.pone.0133325

Angelsen, A. (1995). Shifting cultivation and "deforestation": a study from Indonesia. *World Dev.* 23, 1713–1729. doi: 10.1007/s13280-016-0836-z

Asshauer, K. P., Wemheuer, B., Daniel, R., and Meinicke, P. (2015). Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31, 2882–2884. doi: 10.1093/bioinformatics/btv287

Barnes, A. D., Jochum, M., Mumme, S., Haneda, N. F., Farajallah, A., Widarto, T. H., et al. (2014). Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nat. Commun.* 5:5351. doi: 10.1038/ncomms6351

Barron, A. R., Wurzburger, N., Bellenger, J. P., Wright, S. J., Kraepiel, A. M., and Hedin, L. O. (2009). Molybdenum limitation of asymbiotic nitrogen fixation in tropical forest soils. *Nat. Geosci.* 2, 42–45. doi: 10.1038/ngeo366

Bhardwaj, D., Ansari, M. W., Sahoo, R. K., and Tuteja, N. (2014). Biofertilizers function as key player in sustainable agriculture by improving soil fertility, plant tolerance and crop productivity. *Microb. Cell Fact.* 13:66. doi: 10.1186/1475-2859-13-66

Bouskill, N. J., Lim, H. C., Borglin, S., Salve, R., Wood, T. E., Silver, W. L., et al. (2012). Pre-exposure to drought increases the resistance of tropical forest soil bacterial communities to extended drought. *ISME J.* 7, 384–394. doi: 10.1038/ismej.2012.113

Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P., and Tyson, G. W. (2012). Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat. Methods* 9, 425–426. doi: 10.1038/nmeth.1990

Bremer, C., Braker, G., Matthies, D., Reuter, A., Engels, C., and Conrad, R. (2007). Impact of plant functional group, plant species, and sampling time on the composition of nirK-type denitrifier communities in soil. *Appl. Environ. Microbiol.* 73, 6876–6884. doi: 10.1128/AEM.01536-07

37

Brenzinger, K., Dörsch, P., and Braker, G. (2015). pH-driven shifts in overall and transcriptionally active denitrifiers control gaseous product stoichiometry in growth experiments with extracted bacteria from soil. *Front. Microbiol.* 6:961. doi: 10.3389/fmicb.2015.00961

Brevik, E., Cerdà, A., Mataix-Solera, J., Pereg, L., Quinton, J., Six, J., et al. (2015). The interdisciplinary nature of SOIL. *Soil* 1, 117–129. doi: 10.5194/soil-1-117-2015

Briegel, A., Ortega, D. R., Huang, A. N., Oikonomou, C. M., Gunsalus, R. P., and Jensen, G. J. (2015). Structural conservation of chemotaxis machinery across Archaea and Bacteria. *Environ. Microbiol. Rep.* 7, 414–419. doi: 10.1111/1758-2229.12265

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303

Carney, K. M., Matson, P. A., and Bohannan, B. J. (2004). Diversity and composition of tropical soil nitrifiers across a plant diversity gradient and among land-use types. *Ecol. Lett.* 7, 684–694. doi: 10.1111/j.1461-0248.2004.00628.x

Cascales, E., and Cambillau, C. (2012). Structural biology of type VI secretion systems. *Philos. Trans. R. Soc. B* 367, 1102–1111. doi: 10.1098/rstb.2011.0209

Colombo, C., Palumbo, G., He, J.-Z., Pinton, R., and Cesco, S. (2014). Review on iron availability in soil: interaction of Fe minerals, plants, and microbes. *J Soils Sediments* 14, 538–548. doi: 10.1007/s11368-013-0814-z

Cong, J., Liu, X., Lu, H., Xu, H., Li, Y., Deng, Y., et al. (2015). Available nitrogen is the key factor influencing soil microbial functional gene diversity in tropical rainforest. *BMC Microbiol.* 15:167. doi: 10.1186/s12866-015-0491-8

Corre, M. D., Dechert, G., and Veldkamp, E. (2006). Soil nitrogen cycling following montane forest conversion in central Sulawesi, Indonesia. *Soil Sci. Soc. Am. J.* 70, 359–366. doi: 10.2136/sssaj2005.0061

Dam, B., Dam, S., Kim, Y., and Liesack, W. (2014). Ammonium induces differential expression of methane and nitrogen metabolism-related genes in *Methylocystis* sp. strain SC2. *Environ. Microbiol.* 16, 3115–3127. doi: 10.1111/1462-2920.12367

Delmont, T. O., Prestat, E., Keegan, K. P., Faubladier, M., Robe, P., Clark, I. M., et al. (2012). Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J.* 6, 1677–1687. doi: 10.1038/ismej.2011.197

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., et al. (2011). Accessing the soil metagenome for studies of microbial diversity. *Appl. Environ. Microbiol.* 77, 1315–1324. doi: 10.1128/AEM.01526-10

Donald, P. F. (2004). Biodiversity impacts of some agricultural commodity production systems. *Conserv. Biol.* 18, 17–38. doi: 10.1111/j.1523-1739.2004.01803.x

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381

Faoro, H., Alves, A., Souza, E., Rigo, L., Cruz, L., Al-Janabi, S., et al. (2010). Influence of soil characteristics on the diversity of bacteria in the Southern Brazilian Atlantic Forest. *Appl. Environ. Microbiol.* 76, 4744–4749. doi: 10.1128/AEM.03025-09

Fedi, S., Barberi, T. T., Nappi, M. R., Sandri, F., Booth, S., Turner, R. J., et al. (2016). The Role of cheA Genes in swarming and swimming motility of *Pseudomonas pseudoalcaligenes* KF707. *Microbes Environ.* 31, 169–172. doi: 10.1264/jsme2.ME15164

Fierer, N., and Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* 103, 626–631. doi: 10.1073/pnas.0507535103

Fierer, N., Morse, J. L., Berthrong, S. T., Bernhardt, E. S., and Jackson, R. B. (2007). Environmental controls on the landscape-scale biogeography of stream bacterial communities. *Ecology* 88, 2162–2173. doi: 10.1890/06-1746.1

Foesel, B. U., Nägele, V., Naether, A., Wüst, P. K., Weinert, J., Bonkowski, M., et al. (2014). Determinants of Acidobacteria activity inferred from the relative abundances of 16S rRNA transcripts in German grassland and forest soils. *Environ. Microbiol.* 16, 658–675. doi: 10.1111/1462-2920.12162

Guglielmini, J., Néron, B., Abby, S. S., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. C. (2014). Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res.* 42, 5715–5727. doi: 10.1093/nar/gku194

Hai, B., Diallo, N. H., Sall, S., Haesler, F., Schauss, K., Bonzi, M., et al. (2009). Quantification of key genes steering the microbial nitrogen cycle in the rhizosphere of sorghum cultivars in tropical agroecosystems. *Appl. Environ. Microbiol.* 75, 4993–5000. doi: 10.1128/AEM.02917-08

Hansel, C. M., Fendorf, S., Jardine, P. M., and Francis, C. A. (2008). Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. *Appl. Environ. Microbiol.* 74, 1620–1633. doi: 10.1128/AEM.01787-07

Hayat, R., Ali, S., Amara, U., Khalid, R., and Ahmed, I. (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Ann. Microbiol.* 60, 579–598. doi: 10.1007/s13213-010-0117-1

Henckel, T., Friedrich, M., and Conrad, R. (1999). Molecular analyses of the methane-oxidizing microbial community in rice field soil by targeting the genes of the 16S rRNA, particulate methane monooxygenase, and methanol dehydrogenase. *Appl. Environ. Microbiol.* 65, 1980–1990.

Herzog, S., Wemheuer, F., Wemheuer, B., and Daniel, R. (2015). Effects of fertilization and sampling time on composition and diversity of entire and active bacterial communities in german grassland soilS. *PLoS One* 10:e0145575. doi: 10.1371/journal.pone.0145575

Hibbing, M. E., Fuqua, C., Parsek, M. R., and Peterson, S. B. (2009). Bacterial competition: surviving and thriving in the microbial jungle. *Nat. Rev. Microbiol.* 8, 15–25. doi: 10.1038/nrmicro2259

Isobe, K., Koba, K., Suwa, Y., Ikutani, J., Fang, Y., Yoh, M., et al. (2012). High abundance of ammonia-oxidizing archaea in acidified subtropical forest soils in southern China after long-term N deposition. *FEMS Microbiol. Ecol.* 80, 193–203. doi: 10.1111/j.1574-6941.2011.01294.x

Ivancic, H., Koh, L. P., and Wich, S. (2016). Evolution of sustainable palm oil policy in Southeast Asia. *Cogent Environ. Sci.* 2:1195032. doi: 10.1080/23311843.2016.1195032

Jones, C. W., and Armitage, J. P. (2015). Positioning of bacterial chemoreceptors. *Trends Microbiol.* 23, 247–256. doi: 10.1016/j.tim.2015.03.004

Junier, P., Molina, V., Dorador, C., Hadas, O., Kim, O.-S., Junier, T., et al. (2010). Phylogenetic and functional marker genes to study ammonia-oxidizing microorganisms (AOM) in the environment. *Appl. Microbiol. Biotechnol.* 85, 425–440. doi: 10.1007/s00253-009-2228-9

Kaiser, K., Wemheuer, B., Korolkow, V., Wemheuer, F., Nacke, H., Schöning, I., et al. (2016). Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Sci. Rep.* 6:33696. doi: 10.1038/srep33696

Kerfahi, D., Tripathi, B. M., Dong, K., Go, R., and Adams, J. M. (2016). Rainforest conversion to rubber plantation may not result in lower soil diversity of Bacteria, Fungi, and Nematodes. *Microb. Ecol.* 72, 359–371. doi: 10.1007/s00248-016-0790-0

Kielak, A. M., Barreto, C. C., Kowalchuk, G. A., van Veen, J. A., and Kuramae, E. E. (2016). The ecology of acidobacteria: moving beyond genes and genomes. *Front. Microbiol.* 7:744. doi: 10.3389/fmicb.2016.00744

Koh, L. P., Miettinen, J., Liew, S. C., and Ghazoul, J. (2011). Remotely sensed evidence of tropical peatland conversion to oil palm. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5127–5132. doi: 10.1073/pnas.1018776108

Koo, H., Hakim, J. A., Morrow, C. D., Eipers, P. G., Davila, A., Andersen, D. T., et al. (2017). Comparison of two bioinformatics tools used to characterize the microbial diversity and predictive functional attributes of microbial mats from Lake Obersee, Antarctica. *J. Microbiol. Methods* 140, 15–22. doi: 10.1016/j.mimet.2017.06.017

Kotowska, M. M., Leuschner, C., Triadiati, T., Meriem, S., and Hertel, D. (2015). Quantifying above- and belowground biomass carbon loss with forest conversion in tropical lowlands of Sumatra (Indonesia). *Glob. Change. Biol.* 21, 3620–3634. doi: 10.1111/gcb.12979

Lambert, A., Wong Ng, J., and Picardeau, M. (2015). Gene inactivation of a chemotaxis operon in the pathogen *Leptospira interrogans*. *FEMS Microbiol. Lett.* 362, 1–8. doi: 10.1093/femsle/fnu054

Lammel, D. R., Nüsslein, K., Tsai, S. M., and Cerri, C. C. (2015). Land use, soil and litter chemistry drive bacterial community structures in samples of the rainforest and Cerrado (Brazilian Savannah) biomes in Southern Amazonia. *Eur. J. Soil Biol.* 66, 32–39. doi: 10.1016/j.ejsobi.2014.11.001

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities

using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676

Lee-Cruz, L., Edwards, D. P., Tripathi, B. M., and Adams, J. M. (2013). Impact of logging and forest conversion to oil palm plantations on soil bacterial communities in Borneo. *Appl. Environ. Microbiol.* 79, 7290–7297. doi: 10.1128/AEM.02541-13

Leff, J. W., Jones, S. E., Prober, S. M., Barberán, A., Borer, E. T., Firn, J. L., et al. (2015). Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10967–10972. doi: 10.1073/pnas.1508382112

Leff, J. W., Nemergut, D. R., Grandy, A. S., O'Neill, S. P., Wickings, K., Townsend, A. R., et al. (2012). The effects of soil bacterial community structure on decomposition in a tropical rainforest. *Ecosystems* 15, 284–298. doi: 10.1007/s10021-011-9510-2

Liu, X., Chen, C., Hughes, J., Wang, W., and Lewis, T. (2017). Temporal changes rather than long-term repeated burning predominately control the shift in the abundance of soil denitrifying community in an Australian sclerophyll forest. *Microb. Ecol.* 73, 177–187. doi: 10.1007/s00248-016-0894-6

Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D., and Knight, R. (2007). Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.* 35:e120. doi: 10.1093/nar/gkm541

Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. *ISME J.* 5, 169–172. doi: 10.1038/ismej.2010.133

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., et al. (2010). A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13479–13484. doi: 10.1073/pnas.1003860107

Lynch, D. H. (2015). Nutrient cycling and soil health in organic cropping systems-Importance of management strategies and soil resilience. *Sustain. Agric. Res.* 4, 80–88. doi: 10.5539/sar.v4n3p80

Ma, J., Wang, Z., Zhu, C., Liu, S., Wang, Q., and Wu, Z. (2013). Analysis of nitrification efficiency and microbial community in a membrane bioreactor fed with low COD/N-Ratio wastewater. *PLoS One* 8:e63059. doi: 10.1371/journal.pone.0063059

McCarthy, J. F. (2010). Processes of inclusion and adverse incorporation: oil palm and agrarian change in Sumatra, Indonesia. *J. Peasant Stud.* 37, 821–850. doi: 10.1080/03066150.2010.512460

Mueller, R. C., Gallegos-Graves, L., Zak, D. R., and Kuske, C. R. (2016). Assembly of active bacterial and fungal communities along a natural environmental gradient. *Microb. Ecol.* 71, 57–67. doi: 10.1007/s00248-015-0655-y

Mukherjee, A., Chettri, B., Langpoklakpam, J. S., Basak, P., Prasad, A., Mukherjee, A. K., et al. (2017). Bioinformatic approaches including predictive metagenomic profiling reveal characteristics of bacterial response to petroleum hydrocarbon contamination in diverse environments. *Sci. Rep.* 7:1108. doi: 10.1038/s41598-017-01126-3

Murrell, J. C., Gilbert, B., and McDonald, I. R. (2000). Molecular biology and regulation of methane monooxygenase. *Arch. Microbiol.* 173, 325–332. doi: 10.1007/s002030000158

Navarrete, A. A., Venturini, A. M., Meyer, K. M., Klein, A. M., Tiedje, J. M., Bohannan, B. J. M., et al. (2015). Differential response of acidobacteria subgroups to forest-to-pasture conversion and their biogeographic patterns in the Western Brazilian Amazon. *Front. Microbiol.* 6:1443. doi: 10.3389/fmicb.2015.01443

Oosterveer, P. (2015). Promoting sustainable palm oil: viewed from a global networks and flows perspective. *J. Clean. Prod.* 107, 146–153. doi: 10.1016/j.jclepro.2014.01.019

Pajares, S., and Bohannan, B. J. (2016). Ecology of nitrogen fixing, nitrifying, and denitrifying microorganisms in tropical forest soils. *Front. Microbiol.* 7:1045. doi: 10.3389/fmicb.2016.01045

Pappu, A. R., Bhattacharjee, A. S., Dasgupta, S., and Goel, R. (2017). Nitrogen cycle in engineered and natural ecosystems—past and current. *Curr. Pollut. Rep.* 3, 120–140. doi: 10.1007/s40726-017-0051-y

Quan, Z., Huang, B., Lu, C., Shi, Y., Chen, X., Zhang, H., et al. (2016). The fate of fertilizer nitrogen in a high nitrate accumulated agricultural soil. *Sci. Rep.* 6:21539. doi: 10.1038/srep21539

R Development Core Team (2017). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Ragot, S. A., Huguenin-Elie, O., Kertesz, M. A., Frossard, E., and Bünemann, E. K. (2016). Total and active microbial communities and phoD as affected by phosphate depletion and pH in soil. *Plant Soil* 408, 15–30. doi: 10.1007/s11104-016-2902-5

Rodríguez-Blanco, A., Duval, A., Pelletier, E., Delille, D., and Ghiglione, J.-F. (2013). Effects of temperature and fertilization on the structure of total versus active bacterial communities from sub-Antarctic seawater exposed to crude oil and diesel fuel. *Polar Res.* 32:18521. doi: 10.3402/polar.v32i0.18521

Rooney, D., and Kennedy, N. M. (2013). *Bacterial Communities' Response to Nitrogen, Lime, and Plants. Sustainable Soil Management.* Waretown, NJ: Apple Academic Press.

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., et al. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* 4, 1340–1351. doi: 10.1038/ismej.2010.58

Schneider, D., Engelhaupt, M., Allen, K., Kurniawan, S., Krashevska, V., Heinemann, M., et al. (2015). Impact of lowland rainforest transformation on diversity and composition of soil prokaryotic communities in Sumatra (Indonesia). *Front. Microbiol.* 6:1339. doi: 10.3389/fmicb.2015.01339

Sengupta, A., and Dick, W. A. (2017). Methanotrophic bacterial diversity in two diverse soils under varying land-use practices as determined by high-throughput sequencing of the pmoA gene. *Appl. Soil Ecol.* 119, 35–45. doi: 10.1016/j.apsoil.2017.05.031

Sodhi, N. S., Koh, L. P., Brook, B. W., and Ng, P. K. (2004). Southeast Asian biodiversity: an impending disaster. *Trends Ecol. Evol.* 19, 654–660. doi: 10.1016/j.tree.2004.09.006

Stempfhuber, B., Engel, M., Fischer, D., Neskovic-Prit, G., Wubet, T., Schöning, I., et al. (2015). pH as a driver for ammonia-oxidizing archaea in forest soils. *Microb. Ecol.* 69, 879–883. doi: 10.1007/s00248-014-0548-5

Stibal, M., Schostag, M., Cameron, K. A., Hansen, L. H., Chandler, D. M., Wadham, J. L., et al. (2015). Different bulk and active bacterial communities in cryoconite from the margin and interior of the Greenland ice sheet. *Environ. Microbiol. Rep.* 7, 293–300. doi: 10.1111/1758-2229.12246

Taylor, M., Mediannikov, O., Raoult, D., and Greub, G. (2012). Endosymbiotic bacteria associated with nematodes, ticks and amoebae. *FEMS Immunol. Med. Microbiol.* 64, 21–31. doi: 10.1111/j.1574-695X.2011.00916.x

Tripathi, B. M., Edwards, D. P., Mendes, L. W., Kim, M., Dong, K., Kim, H., et al. (2016). The impact of tropical forest logging and oil palm agriculture on the soil microbiome. *Mol. Ecol.* 25, 2244–2257. doi: 10.1111/mec.13620

Tripathi, B. M., Kim, M., Singh, D., Lee-Cruz, L., Lai-Hoe, A., Ainuddin, A. N., et al. (2012). Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. *Microb. Ecol.* 64, 474–484. doi: 10.1007/s00248-012-0028-8

Urich, T., Lanzén, A., Qi, J., Huson, D. H., Schleper, C., and Schuster, S. C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* 3:2527. doi: 10.1371/journal.pone.0002527

Val-Moraes, S. P., de Macedo, H. S., Kishi, L. T., Pereira, R. M., Navarrete, A. A., Mendes, L. W., et al. (2016). Liming in the sugarcane burnt system and the green harvest practice affect soil bacterial community in northeastern São Paulo, Brazil. *Antonie Van Leeuwenhoek* 109, 1643–1654. doi: 10.1007/s10482-016-0764-8

Větrovský, T., and Baldrian, P. (2013). The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8:57923. doi: 10.1371/journal.pone.0057923

Waldrop, M. P., Balser, T. C., and Firestone, M. K. (2000). Linking microbial community composition to function in a tropical soil. *Soil Biol. Biochem.* 32, 1837–1846. doi: 10.1016/S0038-0717(00)00157-7

Wang, J., Chadwick, D. R., Cheng, Y., and Yan, X. (2018). Global analysis of agricultural soil denitrification in response to fertilizer nitrogen. *Sci. Total Environ.* 61, 908–917. doi: 10.1016/j.scitotenv.2017.10.229

Wang, Y., and Qian, P.-Y. (2009). Conservative fragments in bacterial 16s rRNA genes and primer design for 16s ribosomal DNA amplicons in metagenomic studies. *PLoS One* 4:e7401. doi: 10.1371/journal.pone.0007401

Waring, B. G., Weintraub, S. R., and Sinsabaugh, R. L. (2014). Ecoenzymatic stoichiometry of microbial nutrient acquisition in tropical soils. *Biogeochemistry* 117, 101–113. doi: 10.1007/s10533-013-9849-x

Wemheuer, B., Güllert, S., Billerbeck, S., Giebel, H.-A., Voget, S., Simon, M., et al. (2014). Impact of a phytoplankton bloom on the diversity of the active bacterial

39

community in the southern North Sea as revealed by metatranscriptomic approaches. *FEMS Microbiol. Ecol.* 87, 378–389. doi: 10.1111/1574-6941.12230

Wemheuer, B., Wemheuer, F., and Daniel, R. (2012). RNA-based assessment of diversity and composition of active archaeal communities in the German Bight. *Archaea* 2012:695826. doi: 10.1155/2012/695826

Wemheuer, B., Wemheuer, F., Hollensteiner, J., Meyer, F.-D., Voget, S., and Daniel, R. (2015). The green impact: bacterioplankton response toward a phytoplankton spring bloom in the southern North Sea assessed by comparative metagenomic and metatranscriptomic approaches. *Front. Microbiol.* 6:805. doi: 10.3389/fmicb.2015.00805

Wemheuer, B., Wemheuer, F., Meier, D., Billerbeck, S., Giebel, H.-A., Simon, M., et al. (2017). Linking compositional and functional predictions to decipher the biogeochemical significance in DFAA turnover of abundant bacterioplankton lineages in the North Sea. *Microorganisms* 5:E68. doi: 10.3390/microorganisms5040068

Wilcove, D. S., Giam, X., Edwards, D. P., Fisher, B., and Koh, L. P. (2013). Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. *Trends Ecol. Evol.* 28, 531–540. doi: 10.1016/j.tree.2013.04.005

Yoneyama, T., Terakado-Tonooka, J., and Minamisawa, K. (2017). Exploration of bacterial N2-fixation systems in association with soil-grown sugarcane, sweet potato, and paddy rice: a review and synthesis. *Soil Sci. Plant Nutr.* 63, 1–13. doi: 10.1080/00380768.2017.1407625

Yoshida, M., Ishii, S., Otsuka, S., and Senoo, K. (2009). Temporal shifts in diversity and quantity of nirS and nirK in a rice paddy field soil. *Soil Biol. Biochem.* 41, 2044–2051. doi: 10.1016/j.soilbio.2009.07.012

Zhalnina, K., Dias, R., de Quadros, P. D., Davis-Richardson, A., Camargo, F. A. O., Clark, I. M., et al. (2015). Soil pH determines microbial diversity and composition in the park grass experiment. *Microb. Ecol.* 69, 395–406. doi: 10.1007/s00248-014-0530-2

Zhang, Y., Shen, H., He, X., Thomas, B. W., Lupwayi, N. Z., Hao, X., et al. (2017). Fertilization shapes bacterial community structure by alteration of soil pH. *Front. Microbiol.* 8:1325. doi: 10.3389/fmicb.2017.01325

Zhang, Y., Zhao, Z., Dai, M., Jiao, N., and Herndl, G. J. (2014). Drivers shaping the diversity and biogeography of total and active bacterial communities in the South China Sea. *Mol. Ecol.* 23, 2260–2274. doi: 10.1111/mec.12739

Zheng, Y., Huang, R., Wang, B., Bodelier, P., and Jia, Z. (2014). Competitive interactions between methane-and ammonia-oxidizing bacteria modulate carbon and nitrogen cycling in paddy soil. *Biogeosciences* 11, 3353–3368. doi: 10.5194/bg-11-3353-2014

Zinicola, M., Higgins, H., Lima, S., Machado, V., Guard, C., and Bicalho, R. (2015). Shotgun metagenomic sequencing reveals functional genes and microbiome associated with bovine digital dermatitis. *PLoS One* 10:e0133674. doi: 10.1371/journal.pone.0133674

40

Supplemental Information for Chapter 2

Contents

Figure S1. Average Shannon and PD diversity indices. Values were summarized from all subplot values at land use level in the corresponding landscape.

Figure S2. Rarefaction curves for landscape level and land use systems.

Figure S3. Clustering analysis of used samples on core plot level. Cluster analysis was performed by using hclust.

Figure S4. Stressplot for calculated NMDS shown in Figure 3 based on weighted UniFrac.

Figure S5. Distribution of obtained point biserial correlation coefficients. Each dot displays one obtained value in the respective land use systems. Only significant values were included (p <0.05).

Table S1. Shannon and PD diversity indices and raw sequence data (sequence numbers and filtered sequences) at subplot level.

Table S2. OTU table at 97% sequence identity at species level. Absolute counts are shown for each subplot with a subsampling size of 6,650 per plot. The table is deposited on the enclosed CD under \Chapter2_Supplement_information\

Table S3. PERMANOVA results for community composition analysis. Analysis was performed with Supplementary Table S2 in R via the "vegan" and "RVAideMemoire" packages. The table is deposited on the enclosed CD under \Chapter2_Supplement_information\

Table S4. FTU (Fraction Taxonomy Unexplained) values obtained from Tax4Fun analysis.

Table S5. Gene abundances of selected enzymes obtained from Tax4Fun analysis. The table is deposited on the enclosed CD under \Chapter2_Supplement_information\

Figure S1. Average Shannon and PD diversity indices. Values were summarized from all subplot values at land use level in the corresponding landscape.

Figure S2. Rarefaction curves for landscape level and land use systems.

Cluster Dendrogram

Figure S3. Clustering analysis of used samples on core plot level. Cluster analysis was performed by using hclust.

Figure S4. Stressplot for calculated NMDS shown in Figure 3 based on weighted UniFrac.

Figure S5. Distribution of obtained point biserial correlation coefficients. Each dot displays one obtained value in the respective land use systems. Only significant values were included (p <0.05).
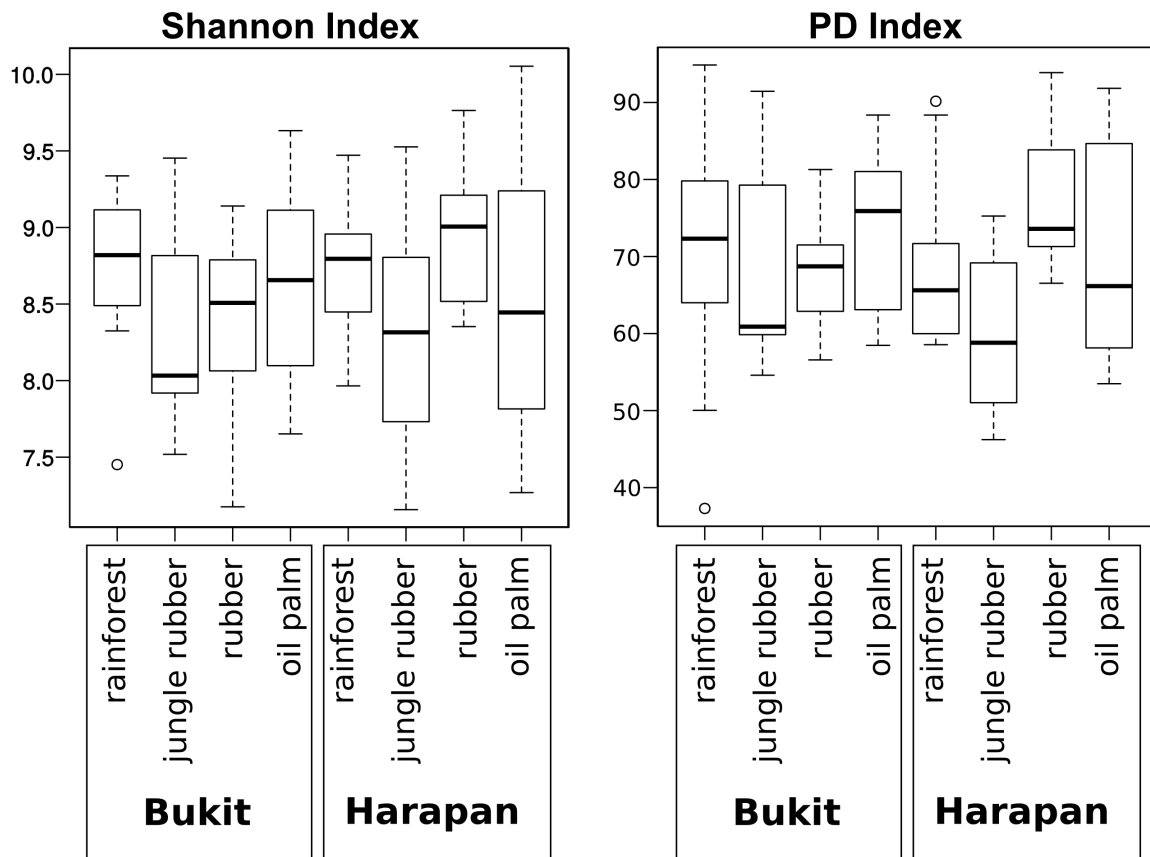
Table S1. Shannon and PD diversity indices and raw sequence data (sequence numbers and filtered sequences) at subplot level.

Subplot	treatment	PD	Shannon	Sub-sample size	raw reads	quality filtered sequences	Excluded sequences	Excluded sequences [%]
BF1a	rainforest	77.78397	8.9534092	6650	13354	13250	104	0.779
BF1b	rainforest	70.62199	8.871184833	6650	14728	14614	114	0.774
BF1c	rainforest	79.66874	9.140674788	6650	11736	11615	121	1.031
BF2a	rainforest	80.68331	8.546670769	6650	10262	10173	89	0.867
BF2b	rainforest	71.79895	8.32464721	6650	16350	16244	106	0.648
BF2c	rainforest	94.85922	9.336816238	6650	20670	20548	122	0.59
BF3a	rainforest	50.02743	8.432566931	6650	12615	12566	49	0.388
BF3b	rainforest	72.83071	9.089375278	6650	16695	16590	105	0.629
BF3c	rainforest	62.39849	8.612355429	6650	15172	15092	80	0.527
BF4a	rainforest	79.95485	9.25737844	6650	17844	17753	91	0.51
BF4b	rainforest	37.29744	7.451494867	6650	16972	16826	146	0.86
BF4c	rainforest	65.59762	8.767138693	6650	16554	16390	164	0.991
BJ1a	jungle rubber	54.58956	7.737533276	6650	14324	11276	3048	21.279
BJ1b	jungle rubber	59.82159	7.90235756	6650	18029	14007	4022	22.309
BJ1c	jungle rubber	60.83014	8.016156465	6650	20366	15822	4544	22.312
BJ2a	jungle rubber	73.39391	8.544248736	6650	16118	12575	3543	21.982
BJ2b	jungle rubber	60.84859	7.980631261	6650	13849	11087	2762	19.944
BJ2c	jungle rubber	58.41978	7.518356014	6650	17012	13205	3807	22.378
BJ3a	jungle rubber	59.88341	8.274874454	6650	16065	12726	3339	20.784
BJ3b	jungle rubber	60.94682	8.048837682	6650	16641	13001	3640	21.874
BJ3c	jungle rubber	62.50865	7.935060041	6650	15030	11769	3261	21.697
BJ4a	jungle rubber	87.22935	9.253845301	6650	17296	13465	3831	22.15
BJ4b	jungle rubber	85.14287	9.088407918	6650	20497	15918	4579	22.34
BJ4c	jungle rubber	91.44436	9.452870014	6650	17773	13741	4032	22.686

BR1a	rubber	58.25818	7.376049062	6650	19741	15155	4586	23.231
BR1b	rubber	70.13153	8.246266729	6650	15804	12505	3299	20.874
BR1c	rubber	69.54557	8.597016378	6650	16581	12923	3658	22.061
BR2a	rubber	67.89795	8.224496515	6650	16425	12745	3680	22.405
BR2b	rubber	73.49165	8.418174583	6650	19175	14950	4225	22.034
BR2c	rubber	72.87183	8.781957748	6650	16216	12678	3538	21.818
BR3a	rubber	60.42499	7.902511846	6650	19384	15212	4172	21.523
BR3b	rubber	69.70397	8.794856931	6650	14031	11093	2938	20.939
BR3c	rubber	56.57903	7.175782359	6650	18442	14375	4067	22.053
BR4a	rubber	65.34713	8.720004974	6650	16755	13215	3540	21.128
BR4b	rubber	66.485	8.874448357	6650	21075	16519	4556	21.618
BR4c	rubber	81.28779	9.140134875	6650	19332	15233	4099	21.203
BO1a	oil palm	88.36238	9.538683284	6650	14759	11777	2982	20.205
BO1b	oil palm	60.41019	7.675071379	6650	16993	13279	3714	21.856
BO1c	oil palm	86.93988	9.632170981	6650	14941	11808	3133	20.969
BO2a	oil palm	71.61516	8.417503819	6650	18549	14449	4100	22.104
BO2b	oil palm	77.70757	8.930340467	6650	16965	13168	3797	22.381
BO2c	oil palm	75.69981	8.89436097	6650	15598	12245	3353	21.496
BO3a	oil palm	76.09927	8.316459091	6650	18760	14579	4181	22.287
BO3b	oil palm	83.26894	9.110372472	6650	16292	12693	3599	22.091
BO3c	oil palm	78.79737	9.114949112	6650	16489	12865	3624	21.978
BO4a	oil palm	63.99848	8.193463238	6650	16278	12768	3510	21.563
BO4b	oil palm	62.19975	8.00173587	6650	18159	14167	3992	21.984
BO4c	oil palm	58.46591	7.65196856	6650	14905	11785	3120	20.933
HF1a	rainforest	66.80867	8.834243004	6650	14683	14602	81	0.552
HF1b	rainforest	71.30142	9.020032141	6650	17268	17125	143	0.828
HF1c	rainforest	69.85548	8.893968347	6650	16002	15841	161	1.006
HF2a	rainforest	64.42587	8.653347637	6650	13991	13912	79	0.565
HF2b	rainforest	58.55292	8.806872197	6650	12637	12550	87	0.688
HF2c	rainforest	90.15538	9.471466574	6650	23437	23269	168	0.717
HF3a	rainforest	58.62191	8.579941203	6650	15717	15550	167	1.063
HF3b	rainforest	88.36131	9.217581708	6650	23899	23774	125	0.523
HF3c	rainforest	59.3511	7.965494893	6650	15160	15079	81	0.534
HF4a	rainforest	72.05893	8.783713962	6650	16810	16730	80	0.476
HF4b	rainforest	61.69076	8.304079677	6650	16449	16355	94	0.571
HF4c	rainforest	60.62107	8.316448096	6650	15917	15797	120	0.754
HJ1a	jungle rubber	68.52989	8.577754619	6650	22053	15329	6724	30.49
HJ1b	jungle rubber	69.81603	8.867475622	6650	16819	11963	4856	28.872
HJ1c	jungle rubber	49.80552	7.554602568	6650	15109	11111	3998	26.461
HJ2a	jungle rubber	75.25357	9.526280261	6650	14713	10828	3885	26.405
HJ2b	jungle rubber	67.49928	9.026316741	6650	17317	12578	4739	27.366
HJ2c	jungle rubber	58.73849	8.656274384	6650	20122	14399	5723	28.442
HJ3a	jungle rubber	46.22859	7.752660037	6650	23433	16377	7056	30.111
HJ3b	jungle rubber	52.24632	8.045664129	6650	21688	15463	6225	28.703
HJ3c	jungle rubber	48.18154	7.157237719	6650	23108	16165	6943	30.046
HJ4a	jungle rubber	58.88131	8.052721634	6650	20413	14344	6069	29.731

HJ4b	jungle rubber	74.7635	8.741948364	6650	18561	13260	5301	28.56
HJ4c	jungle rubber	56.85882	7.710746346	6650	16714	12129	4585	27.432
HR1a	rubber	93.86729	9.763885928	6650	19957	14200	5757	28.847
HR1b	rubber	82.35294	9.010674935	6650	19398	13799	5599	28.864
HR1c	rubber	85.55128	9.580794721	6650	18416	13141	5275	28.644
HR2a	rubber	78.18748	8.877747573	6650	18892	13382	5510	29.166
HR2b	rubber	74.06955	8.527355345	6650	16254	11721	4533	27.889
HR2c	rubber	70.79938	9.02309226	6650	23050	16187	6863	29.774
HR3a	rubber	73.11822	9.033129915	6650	15193	11041	4152	27.328
HR3b	rubber	72.10467	8.352653588	6650	15702	11439	4263	27.149
HR3c	rubber	85.3364	9.388059607	6650	17884	12600	5284	29.546
HR4a	rubber	68.87694	8.507035503	6650	14898	10757	4141	27.796
HR4b	rubber	71.7911	9.000737034	6650	16667	11994	4673	28.037
HR4c	rubber	66.52854	8.456401428	6650	19435	13682	5753	29.601
HO1a	oil palm	55.3679	8.443130522	6650	18204	13265	4939	27.131
HO1b	oil palm	86.71354	9.541979978	6650	18184	13035	5149	28.316
HO1c	oil palm	91.83263	10.05300621	6650	16424	11850	4574	27.849
HO2a	oil palm	64.89689	7.87901635	6650	17408	12545	4863	27.935
HO2b	oil palm	82.61829	8.936008951	6650	19102	13468	5634	29.494
HO2c	oil palm	67.31357	7.510464017	6650	20221	13907	6314	31.225
HO3a	oil palm	72.38518	8.618475299	6650	14161	10321	3840	27.117
HO3b	oil palm	87.63691	9.554992114	6650	14392	10539	3853	26.772
HO3c	oil palm	56.00181	7.751215499	6650	17282	12556	4726	27.346
HO4a	oil palm	53.48871	8.446879469	6650	21410	15119	6291	29.383
HO4b	oil palm	60.27255	7.268500522	6650	18467	13125	5342	28.927
HO4c	oil palm	65.00442	8.20392639	6650	17543	12633	4910	27.988

Table S4. FTU (Fraction Taxonomy Unexplained) values obtained from Tax4Fun analysis.

Plot name	FTU	Site and Treatment
BF1	0.4741	Bukit rainforest
BF2	0.5553	Bukit rainforest
BF3	0.4964	Bukit rainforest
BF4	0.4935	Bukit rainforest
BJ1	0.4301	Bukit jungle rubber
BJ2	0.5631	Bukit jungle rubber
BJ3	0.4939	Bukit jungle rubber
BJ4	0.5021	Bukit jungle rubber
BO1	0.5108	Bukit rubber
BO2	0.5381	Bukit rubber
BO3	0.5506	Bukit rubber
BO4	0.6025	Bukit rubber
BR1	0.5475	Bukit oil palm
BR2	0.5332	Bukit oil palm
BR3	0.5744	Bukit oil palm

BR4	0.5366	Bukit oil palm
HF1	0.5091	Harapan rainforest
HF2	0.4702	Harapan rainforest
HF3	0.5121	Harapan rainforest
HF4	0.5207	Harapan rainforest
HJ1	0.5434	Harapan jungle rubber
HJ2	0.511	Harapan jungle rubber
HJ3	0.5095	Harapan jungle rubber
HJ4	0.4948	Harapan jungle rubber
HO1	0.4914	Harapan rubber
HO2	0.6278	Harapan rubber
HO3	0.5666	Harapan rubber
HO4	0.5655	Harapan rubber
HR1	0.5334	Harapan oil palm
HR2	0.5573	Harapan oil palm
HR3	0.6279	Harapan oil palm
HR4	0.5029	Harapan oil palm

3

Unravelling the effects of tropical land use conversion on the soil microbiome

Dirk Berkelmann[1], Dominik Schneider[1], Anja Meryandini[2] and Rolf Daniel[1]

[1]Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University, Göttingen, Germany,

[2]Department of Biology, Faculty of Mathematics and Natural Sciences IPB, Bogor Agricultural University, Bogor, Indonesia

Author contributions:

Conceived and designed the study: RD.

Management of soil sampling and transport: AM.

Performed the experiments: DB.

Analysed the data: DB and DS.

Wrote the paper: DB, DS, AM and RD.

Unravelling the effects of tropical land use conversion on the soil microbiome

Dirk Berkelmann[1], Dominik Schneider[1], Anja Meryandini[2] and Rolf Daniel[1*]

Abstract

Background: The consequences of deforestation and agricultural treatments are complex and affect all trophic levels. Changes of microbial community structure and composition associated with rainforest conversion to managed systems such as rubber and oil palm plantations have been shown by 16S rRNA gene analysis previously, but functional profile shifts have been rarely addressed. In this study, we analysed the effects of rainforest conversion to different converted land use systems, including agroforestry ("jungle rubber") and monoculture plantations comprising rubber and oil palm, on soilborne microbial communities by metagenomic shotgun sequencing in Sumatra, Indonesia.

Results: The diversity of bacteria and archaea decreased whereas diversity of fungi increased in the converted land use systems. The soil microbiome was dominated by bacteria followed by fungi. We detected negative effects of land use conversion on the abundance of *Proteobacteria* (especially on *Rhizobiales* and *Burkholderiales*) and positive effects on the abundance of *Acidobacteria* and *Actinobacteria*. These abundance changes were mainly driven by pH, C:N ratio, and Fe, C and N content. With increasing land use intensity, the functional diversity decreased for bacteria, archaea and fungi. Gene abundances of specific metabolisms such as nitrogen metabolism and carbon fixation were affected by land use management practices. The abundance of genes related to denitrification and nitrogen fixation increased in plantations while abundance of genes involved in nitrification and methane oxidation showed no significant difference. Linking taxonomic and functional assignment per read indicated that nitrogen metabolism-related genes were mostly assigned to members of the *Rhizobiales* and *Burkholderiales*. Abundances of carbon fixation genes increased also with increasing land use intensity. Motility- and interaction-related genes, especially genes involved in flagellar assembly and chemotaxis genes, decreased towards managed land use systems. This indicated a shift in mobility and interspecific interactions in bacterial communities within these soils.

Conclusions: Rainforest conversion to managed land use systems drastically affects structure and functional potential of soil microbial communities. The decrease in motility- and interaction-related functions from rainforest to converted land use systems indicated not only a shift in nutrient cycling but also in community dynamics. Fertilizer application and correspondingly higher availability of nutrients in intensively managed plantations lead to an environment in which interspecific interactions are not favoured compared to rainforest soils. We could directly link effects of land management, microbial community structure and functional potential for several metabolic processes. As our study is the first study of this size and detail on soil microbial communities in tropical systems, we provide a basis for further analyses.

Keywords: Metagenomics, Oil palm, Soil bacterial communities, Rainforest conversion, Rubber, Soil microbial community

---

* Correspondence: rdaniel@gwdg.de
[1]Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University, Grisebachstr. 8, 37077 Göttingen, Germany
Full list of author information is available at the end of the article

Background

Conversion of natural systems to agriculturally managed land use systems is constantly increasing worldwide [1, 2]. Indonesia is one of the world's largest palm oil and rubber producer and harbours a high biodiversity in tropical rainforests. Thus, effects of land use changes and rainforest conversion to agriculturally managed systems on biodiversity and ecosystem functions are of high interest including conflicts and trade-offs between conservation of biodiversity and economic revenue. In recent years, studies targeting different trophic levels as well as biogeochemical, ecological and socioeconomical effects of the conversion have been published [3, 4]. Soil microbial communities are integral components of terrestrial ecosystems. Microorganisms in soils comprise prokaryotes (archaea and bacteria), fungi and protists. Ecosystem functioning depends to a large extent on the functional diversity and activity of the belowground microbial system [5, 6]. In addition, microorganisms play a key role in decomposing soil organic matter and mineralizing nutrients in soil [7]. It has been shown that rainforest conversion to plantations has negative effects on the biodiversity of fungi [8, 9], protists [10], vertebrates [11], insects and plants [12–14], archaea [15], but not on bacterial diversity which increased [15, 16]. Additionally, the microbial community composition was severely affected by these conversion processes with a high impact on *Proteobacteria* which showed an abundance decrease and *Actinobacteria,* which showed an abundance increase with increasing land use intensity [15, 16]. Furthermore, negative effects on aboveground and belowground carbon stocks, CO_2 fluxes and leaching in these converted land use systems were shown before [17–21]. Management practices such as application of fertilizer, liming, application of herbicides (e.g. glyphosate) and harvesting in turn affect functions of microbial communities [22–24].

Previous studies targeting soil microbial communities in the tropics often relied on the analysis of markers like the 16S rRNA gene. This is cost-efficient, taxonomically accurate and relatively fast but functional profiles can only be predicted based on taxonomic information using bioinformatic tools such as Tax4Fun2 [25], Piphillin [26] and PiCrust [15, 27–31]. Prediction-based analysis of microbiome functions provides first insights into environmental microbial processes such as nutrient cycling and emission of climate gases influenced by agricultural management [16]. While these studies provided a first impression of the functional potential, direct sequencing of metagenomic DNA and identification of functional genes is needed to verify prediction-based findings [32]. However, studies covering functional analyses of rainforest conversion to oil palm monocultures based on direct sequencing of metagenomes are rare. To our knowledge, only Tripathi et al. [27] analysed microbial
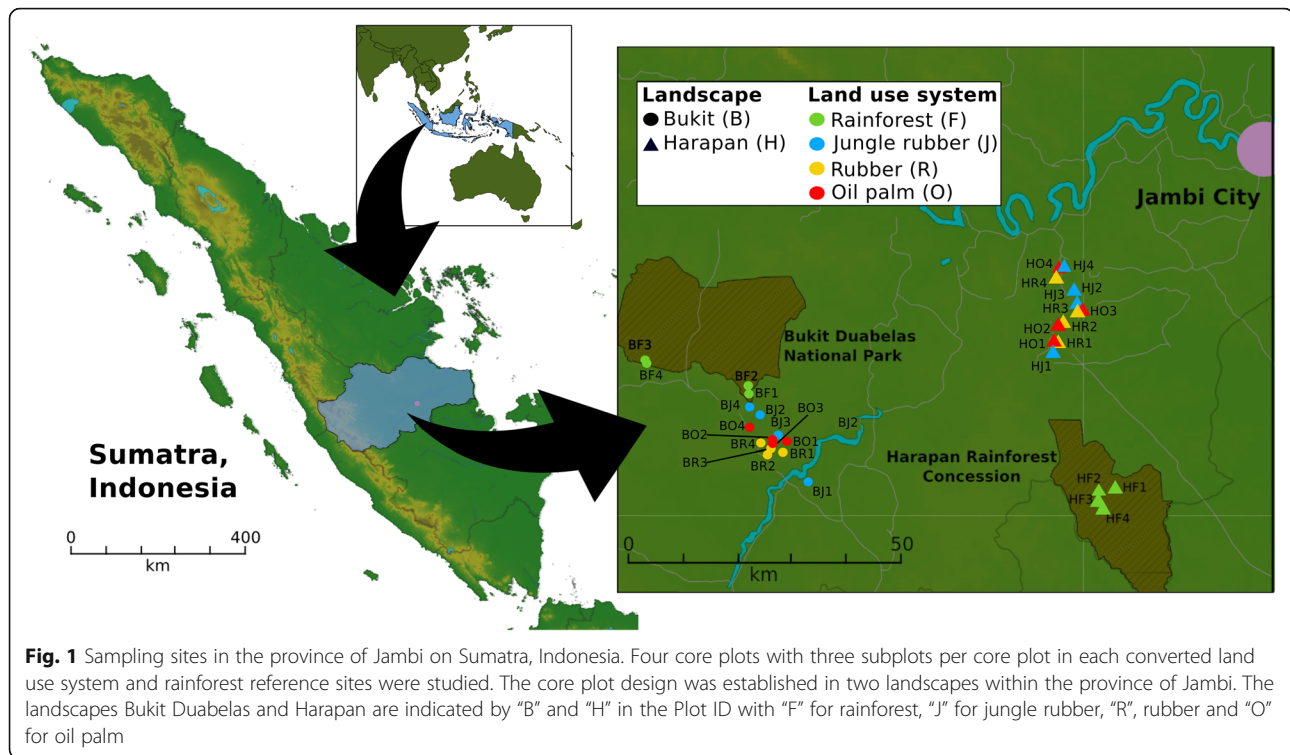
functioning in oil palm soils so far, but with a focus on logging effects. Effects of rainforest conversion to jungle rubber and rubber plantations systems on functional gene profiles, as analysed in this study, were not addressed yet. In a previous study, which was based on taxonomy-derived functional predictions, it was suggested that nutrient cycling related processes such as nitrogen fixation, denitrification and methane oxidation as well as motility- and interaction-related processes such as chemotaxis and type IV secretion systems are negatively affected by rainforest conversion to managed systems. It was hypothesized that fertilizer input and liming in intensively managed systems such as rubber and oil palm plantations reduces the need for nutrient acquisition, and thereby influencing interactions, mobility and communication of soilborne bacterial communities [16]. Even if taxonomic and functional profiles are studied, it is still not known which taxonomic group is responsible for which processes, leaving a gap that needs to be addressed as well.

To gain deeper insights into these processes, we analysed rainforest conversion to jungle rubber, rubber plantations and oil palm plantations by direct sequencing of the corresponding soil metagenomes. Subsequently, taxonomic and functional profiles in the different land use systems were analysed and compared. We linked functional results with their corresponding taxonomic background. Based on our previous studies analysing the 16S rRNA marker gene sequences in the same sampling sites and studies on similar systems in Southeast Asia [9, 10, 15, 16, 18, 33], we formulated three hypotheses. We assume that diversity will decrease for soilborne archaea, Eukaryotes and increase for bacteria with increasing land use intensity (H1). Furthermore, we expect functional profiles to differ between rainforest and plantation monocultures, in which nitrogen and methane metabolism as well as motility- and interaction-related processes are negatively affected by the reduced aboveground diversity in plantations (H2). Furthermore, we hypothesize that strongly affected taxonomic groups like *Proteobacteria* and *Actinobacteria* are linked to crucial processes in nutrient cycling such as nitrogen-fixation, nitrification and methane oxidation. In addition, some of these affected groups are indicators for motility-related processes like chemotaxis and quorum sensing (H3).

Results

General characteristics of the soil metagenome dataset

Metagenome sequencing and quality-filtering of the 32 Indonesian soil samples covering the land use types rainforest, jungle rubber, rubber plantation and oil palm plantation in two landscapes (see Methods for details, Fig. 1) resulted in more than 1.11 billion high-quality reads in total and approximately 33 million reads on average per sample (Additional file 1: Table S1). The average read length per

Fig. 1 Sampling sites in the province of Jambi on Sumatra, Indonesia. Four core plots with three subplots per core plot in each converted land use system and rainforest reference sites were studied. The core plot design was established in two landscapes within the province of Jambi. The landscapes Bukit Duabelas and Harapan are indicated by "B" and "H" in the Plot ID with "F" for rainforest, "J" for jungle rubber, "R", rubber and "O" for oil palm

paired-end read was 146 bp with an average GC content of 60%. To get an overview of the taxonomic community structure, reads were first clustered at domain level (Additional file 2: Figure S1). Approximately 40% of all reads could not be taxonomically classified. We recorded a bacterial dominance in bulk soils of all land use systems including the rainforest controls with abundances ranging from 48.5 to 52.5% of all reads (524,556,933 reads in total). Highest bacterial abundances were detected in rainforest samples (51% in Bukit and 52.5% in Harapan), which also contained the lowest number of unclassified reads of all sample types. Approximately 7% of all sequences were classified as *Eukaryota* (78,682,140 reads in total), while less than 0.6% of all reads were classified as archaea (4,251,297 reads in total) and viruses (1,147,105 reads).

Effects of rainforest conversion on microbial diversity

We analysed taxonomic and functional diversity of the entire soil community and separately of bacteria, archaea and fungi and compared the results derived from all converted land use systems with those of rainforest by using the Shannon index for each land use type (Fig. 2). Taxonomic diversity decreased for all prokaryotes from rainforest to the monoculture land uses rubber and oil palm (Fig. 2a). Lowest values for the entire community and bacteria, and archaea were detected in rubber and oil palm soils, respectively (with $p < 0.05$ compared to rainforest). In contrast, fungi showed an opposite trend with an increase in diversity from rainforest to converted

systems with the highest diversity in rubber (all $p < 0.05$). We additionally analysed functional diversity on gene level (Fig. 2b). We observed gradual decreases in functional diversity from rainforest over jungle rubber to plantations for all analysed groups. As recorded for taxonomic diversity of the entire community and bacteria, lowest values for functional diversity were obtained in rubber. For both groups, the detected decreases from rainforest to both monoculture systems were significant with $p < 0.001$. Although functional diversity decreased for archaea as well, changes were less pronounced and with high variance in rubber. Contrary to taxonomic diversity, functional diversity for fungi decreased from rainforest to converted systems with lowest values in oil palm samples ($p < 0.05$).

Effects of rainforest conversion on microbial community structures

Studying the microbial community composition at high taxonomic resolution based solely on short metagenomic reads is challenging. In order to provide a clear but still detailed analysis and avoid overinterpretation at species level, we chose order level to compare the communities. Differential abundance analysis of the soil bacterial community composition showed that significant changes between the different land use systems were more pronounced in the Harapan landscape than in the Bukit landscape (Additional file 3: Figure S2).

Fig. 2 Taxonomic (**a**) and functional diversity (**b**) of the microbial communities in the land use systems. Shannon diversity indices were calculated for the entire community as well as bacteria, archaea and fungi in each land use system. Significant changes in each land use system compared to rainforest are indicated with * (* = $p < 0.05$, ** = $p < 0.001$, *** = $p < 0.0001$)
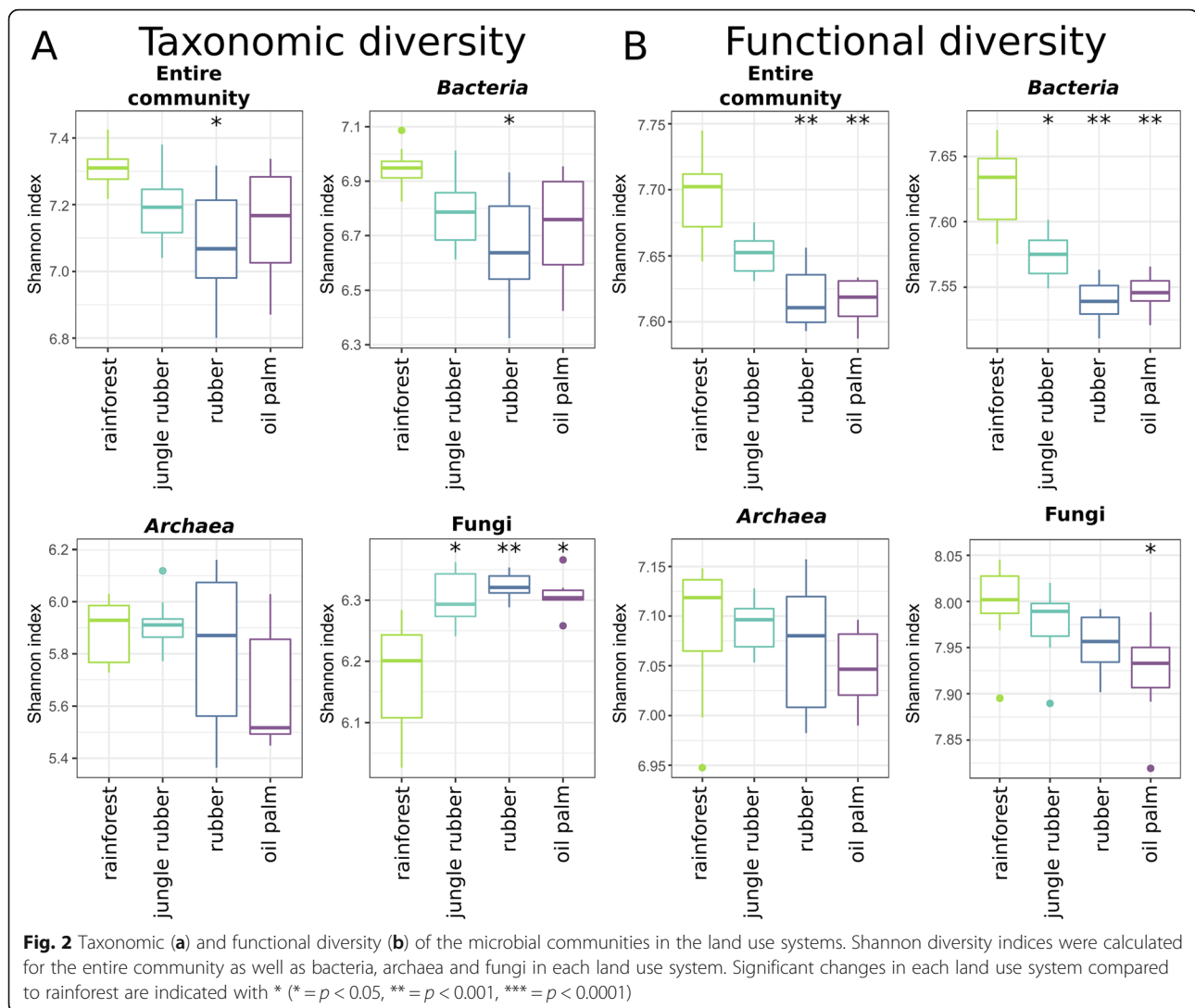
The most abundant bacterial orders were uncultured acidobacterial orders, *Rhizobiales*, *Acidobacteriales*, *Burkholderiales* and *Streptomycetales* (Fig. 3a, b). Abundances of *Rhizobiales* and *Acidobacteriales* were high in general but did not show significant changes in the converted land use systems. *Burkholderiales*, which are involved in denitrification [34] and nitrogen fixation [35], showed the largest differences in abundances between the different land use systems of all bacterial orders. Abundances of *Burkholderiales* decreased significantly in both landscapes from rainforest to the monoculture systems with generally higher abundances in Harapan samples (5.2 to 2.6% in Bukit and 8.1 to 2.5% in Harapan, $p < 0.05$).

The most abundant orders within the archaea belonged to the *Halobacteria* and *Thaumarchaeota*, with the largest fraction belonging to unclassified archaea (Fig. 3c). *Haloferacalaes*, *Halobacteriales* and *Natrialbales* of the *Halobacteria* were detected in all land use systems with similar abundances and without significant changes. In general,

significant changes were not detected for archaeal orders in all Harapan samples whereas in Bukit soils 16 of all detected 62 orders differed significantly between rainforest and converted land use systems.

The most abundant fungal orders were *Eurotiales* and *Hypocreales*, which both belong to the *Ascomycota* (Fig. 3d). *Hypocreales* decreased significantly in abundance from rainforest to all converted land use systems, except Bukit jungle rubber. *Hypocreales* decreased from 0.19% in Bukit rainforest and 0.18% in Harapan rainforest to 0.12% in rubber of both landscapes and 0.11 and 0.12% in oil palm plantations, respectively ($p < 0.05$). *Eurotiales* showed significant abundance changes compared to rainforest in rubber and oil palm plantations of both landscapes. Increasing abundances along the land use gradient from rainforest to converted systems were also observed for *Chaetothyriales* and *Pleosporales* with p.adj < 0.05 in Bukit rubber samples.

Fig. 3 Community composition displayed as relative abundances at order level of the entire community, bacteria, archaea and fungi. (**a**) The entire community including all used sequences is shown in bars in which all orders below 1% were clustered as "rare taxa". The five most abundant orders within the bacteria (**b**), archaea (**c**) and fungi (**d**) are shown seperately. Significant differences ($p < 0.05$) compared to rainforest are marked with *

We compared trophic groups of protists in each land use system (0.09% of all reads) using the same functional categories described by Schulz et al. [10] (Additional file 4: Figure S3). Phagotrophs were the most abundant group (37 to 39%) followed by photoautotrophs (29 to 30%) and animal parasites (27 to 31%). A notable increase of protistan animal parasites compared to rainforest reference sites was recorded in jungle rubber soils, whereas abundances of the phagotrophs and photoautotrophs remained rather stable. In contrast, results by Schulz et al. [10], which were based on 18S rRNA gene amplicons, showed an increase of phagotrophs and a decrease for animal parasites in converted systems compared to rainforest samples.

Additionally, microbiome composition was compared at order level based on extracted rRNA sequences (Additional file 5: Figure S4). Similar community structures were observed as in previous 16S rRNA marker gene studies [15, 16]. Similar to the phylogenetically assigned metagenome data, *Proteobacteria* and *Acidobacteria* were the most abundant bacterial phyla in extracted rRNA sequences. Notably, the fraction of unclassified bacteria at phylum level was higher in extracted rRNA sequences, ranging from 20.2 to 26.5% while remaining under 2.2% in shotgun data. Pairwise PERMANOVA analysis showed no significant differences between shotgun data and rRNA sequence controls.

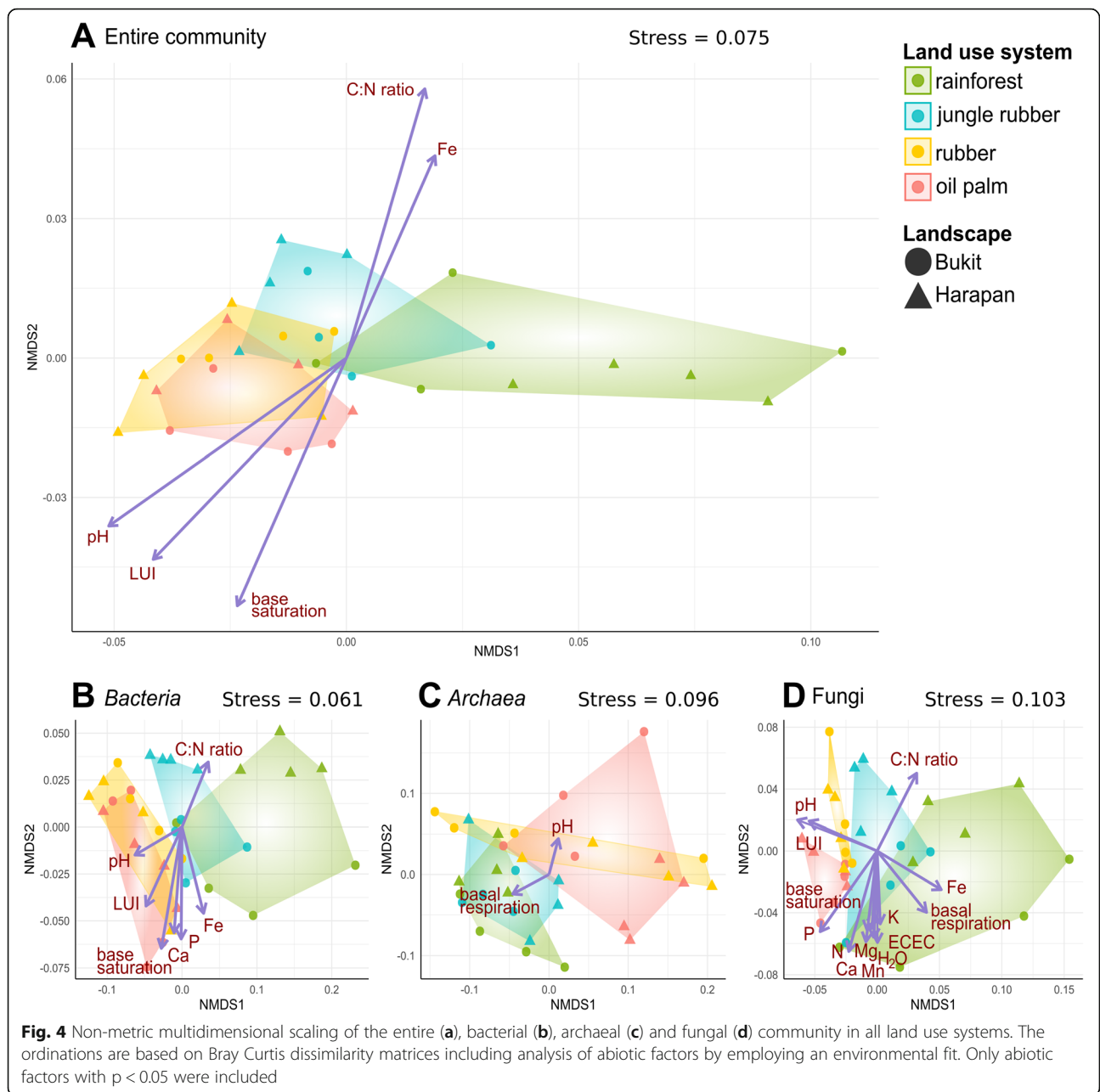### Influence of abiotic parameters on soilborne communities
The composition of microbial communities in soil is tightly connected with soil characteristics and nutrient availability. These parameters are in turn connected with

land use and management practices [36–39]. In order to investigate the impact of soil attributes on soilborne communities with respect to rainforest conversion, we employed nonparametric multidimensional scaling (NMDS) on Bray Curtis dissimilarity matrices (Fig. 4).

Multivariate analysis of the entire soil microbial community composition showed a gradient along the four different tested land use systems corresponding to increasing land use intensity (rainforest < jungle rubber < rubber plantation < oil palm plantation). NMDS stress levels ranged from 0.06 to 0.1. The parameters which correlated with the ordination of the entire community were C:N ratio, land use index (LUI), base saturation, iron content and pH ($p < 0.05$) (Additional file 6: Table S2). Similar clustering patterns were obtained for bacteria, archaea and fungi, which are all in line with the proposed land use gradient. Abiotic parameters that showed correlation with the bacterial community were LUI, pH, C:N ratio, Fe, Ca, P and base saturation (for details see, Additional file 6: Table S2). The LUI, pH value, P content and Ca content increased with higher land use intensity, reflecting the effects of agricultural liming and fertilizer application [9, 18]. These factors correlated also with the fungal community composition in addition to nitrogen, Mn, K and Mg content, water saturation, effective cation exchange (ECEC), and basal respiration. The only detected correlations to archaea were pH and basal respiration.

Impact of rainforest conversion on microbiome functions
To assess potential functional responses related to the conversion of rainforest, we used differential gene abundance analysis. Significant differences for 7294 genes in

Fig. 4 Non-metric multidimensional scaling of the entire (**a**), bacterial (**b**), archaeal (**c**) and fungal (**d**) community in all land use systems. The ordinations are based on Bray Curtis dissimilarity matrices including analysis of abiotic factors by employing an environmental fit. Only abiotic factors with p < 0.05 were included

all land use systems compared to rainforest were detected. Notably, these detected changes were stronger (positive as well as negative) in Harapan soils compared to Bukit soils (Additional file 7: Figure S5). In fact, changes in Bukit jungle rubber compared to Bukit rainforest were minor compared to all other land uses with only 7 significantly changed genes. One of these was the *nifD* gene, which is involved in nitrogen fixation and encodes a subunit of the nitrogenase enzyme. Similar to the NMDS analysis based on microbial taxonomy, the functional profile differed between rainforest and converted land use systems (Additional file 8: Figure S6).

We also analysed functional groups (at level three of the KEGG pathway hierarchy) in each converted land use system compared to rainforest as well as single genes (Figs. 5, 6, 7). The majority of functional differences were detected in Harapan soils. We detected 197 significant abundance changes of functional groups for jungle rubber, 221 for rubber plantation and 108 for oil palm plantation in Harapan soils. In Bukit soils, we detected 177 significant abundance differences of functional groups for rubber and 73 for oil palm but none for jungle rubber. Functional groups covering genes involved in nutrient cycling were analysed in detail including

Fig. 5 Functional profile of selected metabolisms based on KEGG categories (level 3) displayed as log2fold changes. Negative and positive log2fold changes indicate decreased and increased abundances in the corresponding converted land use systems compared to rainforest samples. Values with p.adj < 0.05 are marked with *
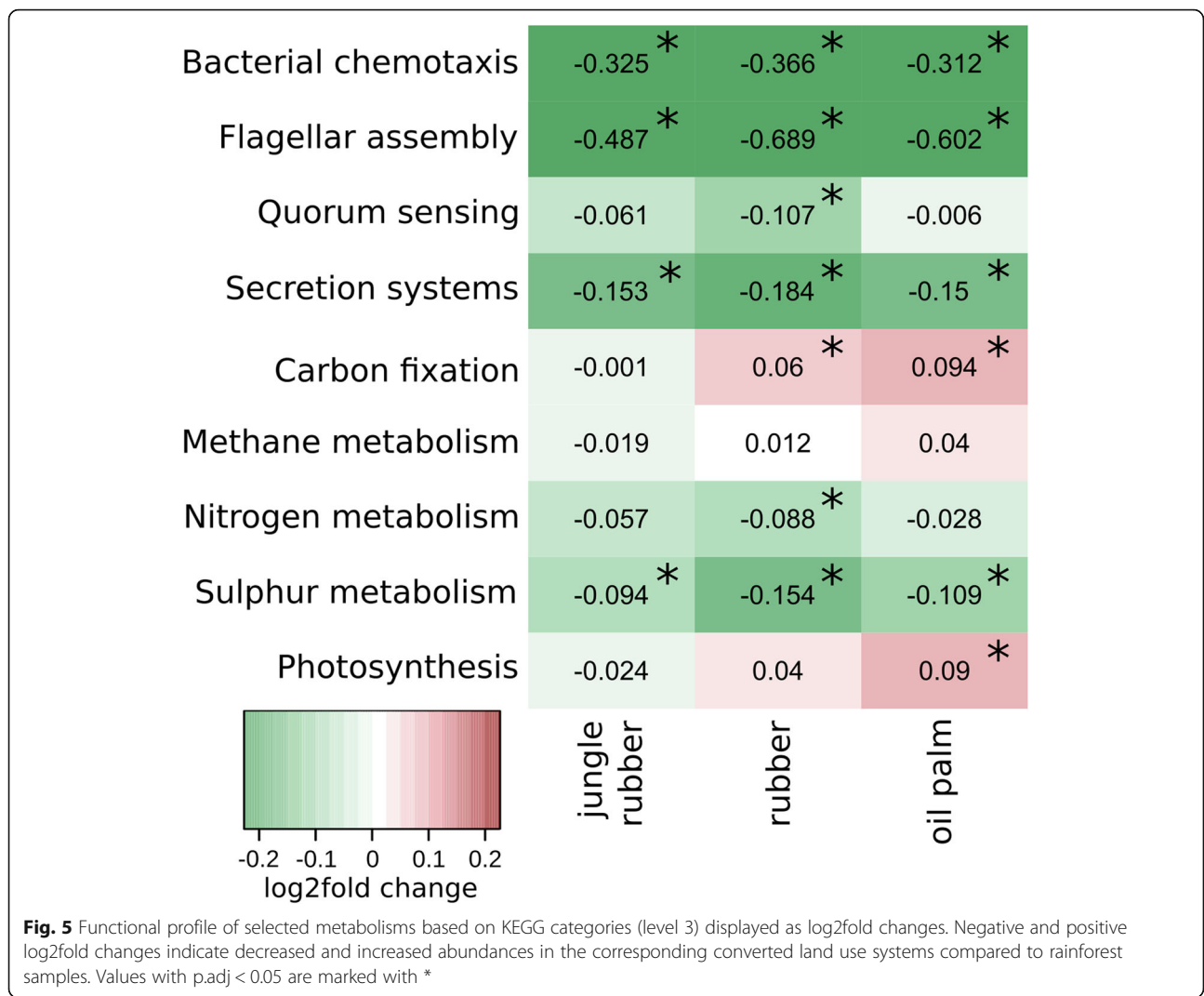
nitrogen metabolism, methane metabolism, carbon fixation and sulphur metabolism (Fig. 5). In addition, functional groups associated with interaction and competition such as chemotaxis, flagellar assembly, quorum sensing, secretion systems and photosynthesis were evaluated.

Nitrogen metabolism-related genes decreased significantly from rainforest to rubber samples with a log2fold change of − 0.08 ($p < 0.01$). Carbon fixation increased from rainforest to rubber (log2fold change 0.06 with $p < 0.05$) and oil palm samples (log2fold change 0.09 with p < 0.01). Sulphur metabolism related genes decreased from rainforest to all converted land use systems, with log2fold changes ranging from − 0.09 in jungle rubber to − 0.1 in oil palm and − 0.15 in rubber. Genes involved in photosynthesis increased from rainforest to oil palm soils with a log2fold change of 0.09 (p < 0.01). At landscape level sulphur metabolism did not change in Bukit jungle rubber and oil palm.

Additionally, we did not detect changes for nitrogen metabolism related genes in Harapan rubber samples (Additional file 9: Figure S7).

In order to analyse the potential for interactions, we selected bacterial genes involved in chemotaxis and flagellar assembly for motility, quorum sensing and secretion systems for interactions (Fig. 5 and Additional file 9: Figure S7). Bacterial chemotaxis, flagellar assembly and secretion systems showed decreases in gene abundances from rainforest to all converted land use systems. For quorum sensing, we detected a decrease in abundance only in rubber samples.

Analysis of specific gene abundances for energy metabolism and motility
Previously analysed pathways were further investigated based on abundances of characteristic genes. Selected genes were divided into two categories related to energy

Fig. 6 Log2fold changes of selected marker genes of nitrogen and methane metabolism for each analysed land use system compared to rainforest. Significant differences (p.adj. < 0.05) are marked with *. Negative and positive log2fold changes indicate decreased and increased abundances in the corresponding converted land use systems compared to rainforest samples
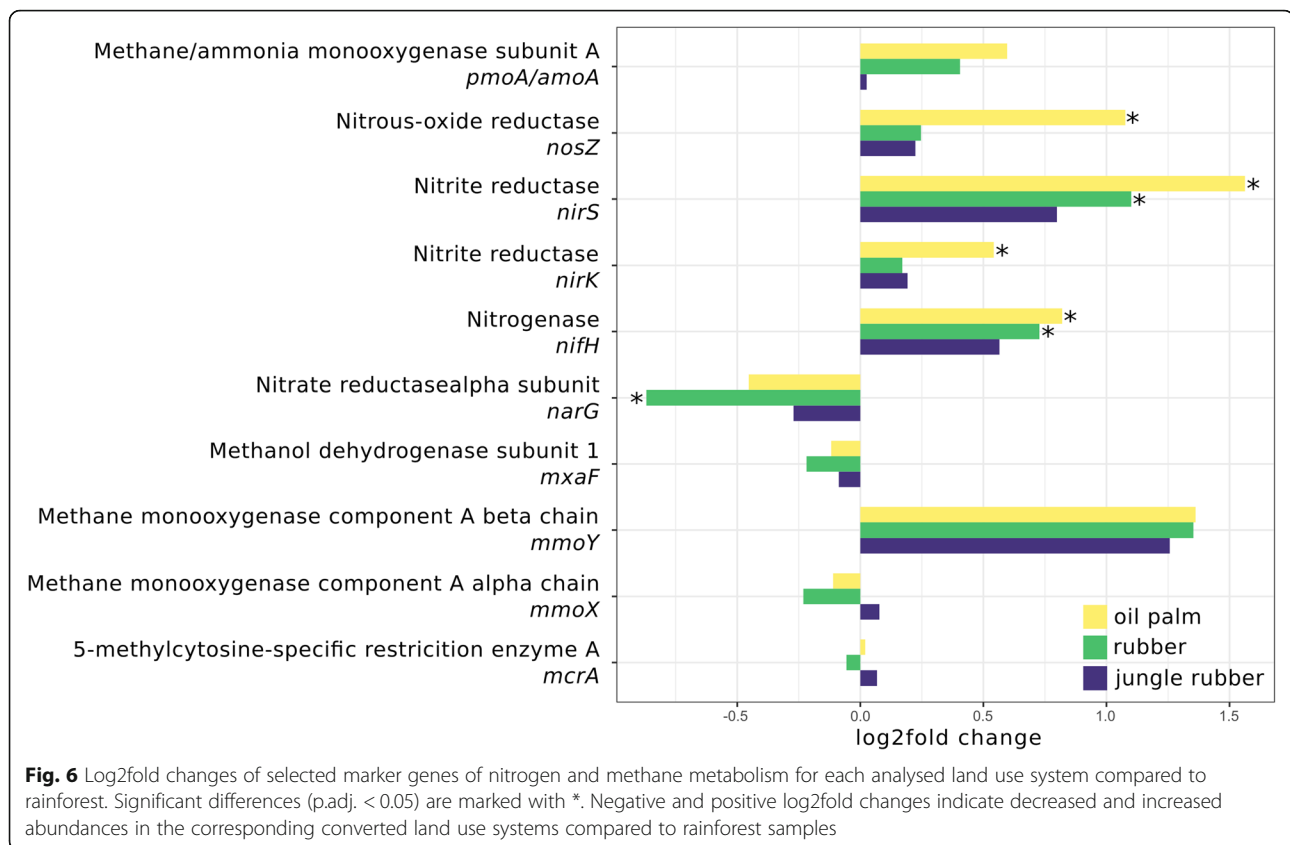
metabolism or motility (Figs. 6, 7 and Additional file 10: Figure S8 and Additional file 11: Figure S9). The denitrification related marker genes *nirK*, *nirS* and *nosZ* increased significantly from rainforest to oil palm plantation, with log2fold changes ranging from 0.5 (*nirK*) to 1.5 (*nirS*). Nitrogen fixation marker *nifH* increased from rainforest to monoculture land use systems (0.7 in rubber and 0.8 in oil palm, $p < 0.05$) and the dissimilatory nitrate reduction marker gene *narG* increased from rainforest to rubber plantations. Other marker genes related to nitrogen or methane metabolism showed no differences compared to rainforest, indicating a higher potential for denitrification and nitrogen fixation in converted land use systems. Since our results suggested that *Rhizobiales* and *Burkholderiales* are of particular importance in the studied rainforest soils due to their connection to nitrogen cycling, the question arose whether this is connected to symbiotic root nodule formation or rather to an endophytic lifestyle. We therefore analysed the abundances of *nod* genes, which are encoding nodulation factors and are crucial for the formation of root nodule symbiosis. However, evidence that rainforest conversion affects abundance of these genes was not found (Additional file 12: Figure S10).
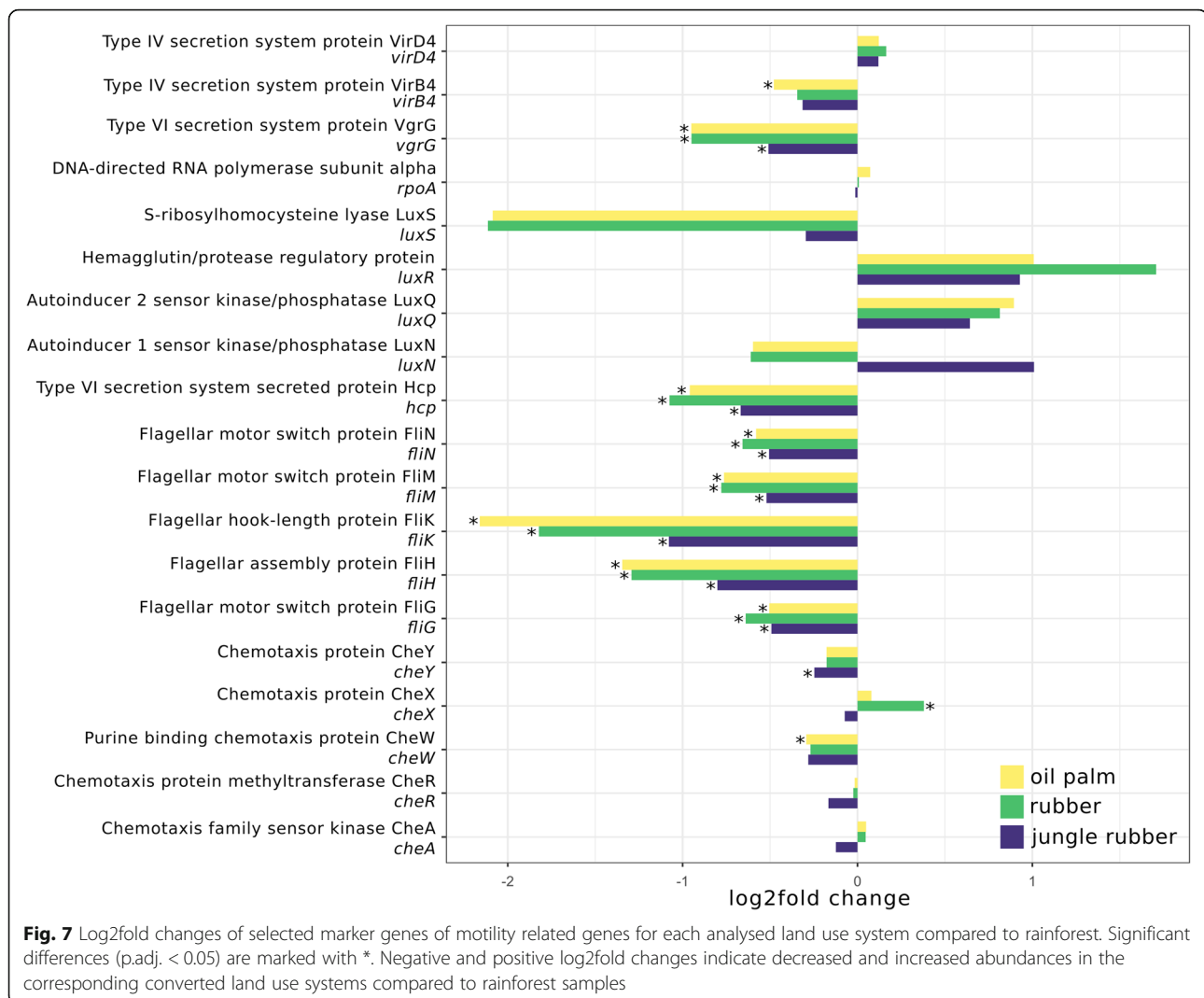
We did not see changes accompanying rainforest conversion for methane metabolism-related genes as a group.

Consequently, we analysed marker genes for methane oxidation as well (Fig. 6) but did not recorded significant changes in abundance for all analysed genes.

The abundance of flagellar assembly markers decreased significantly from rainforest to all converted systems (Fig. 7). Similar results were observed for the type VI secretion system proteins *vgrG* and *hcp*, which decreased from rainforest to all converted systems. However, when analysing the landscapes separately, most significant differences were observed in Harapan samples (Additional file 11, Figure S9). Chemotaxis markers did not show such a clear pattern. Chemotaxis gene *cheY* decreased in jungle rubber samples and *cheW* decreased from rainforest to oil palm. In contrast, *cheX* increased from rainforest to rubber samples. Type IV secretion system gene *virB4* only showed significant changes in oil palm with a minor decrease, whereas the quorum sensing marker genes showed no significant changes in abundance compared to rainforest.
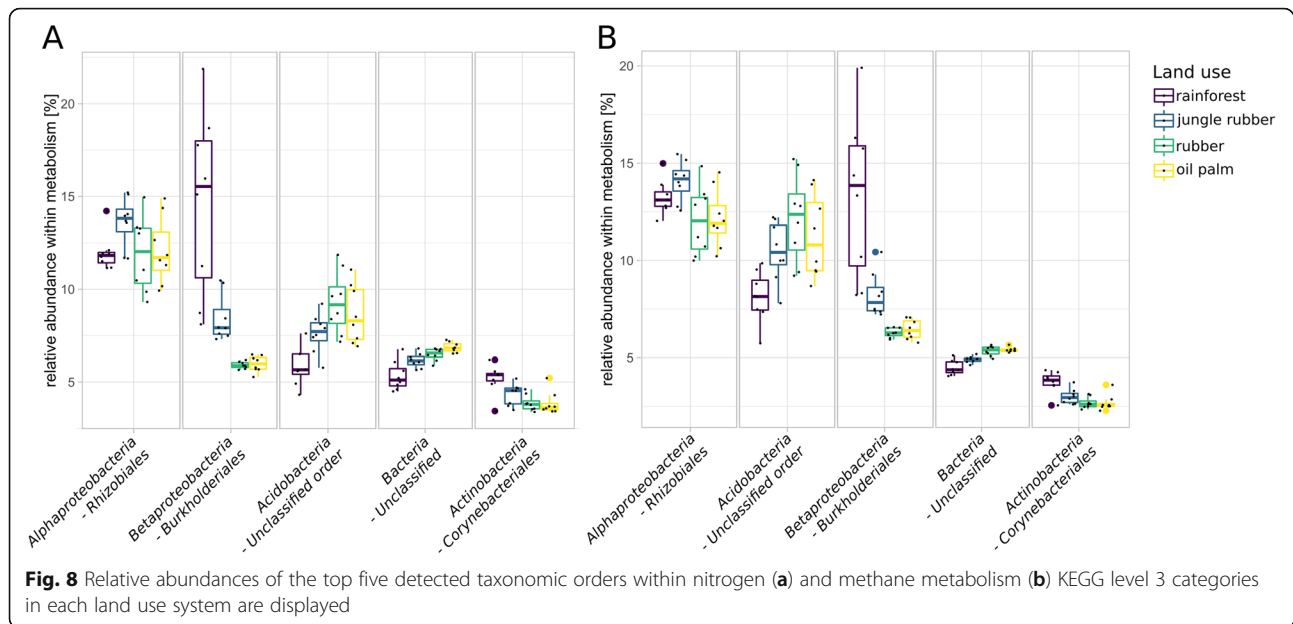
Connecting taxonomy and functions – who does what?
To unravel the full scope of effects introduced by rainforest conversion on soil microbial community structure and functional profiles, it is necessary to identify which part of the present community is involved in which

Berkelmann *et al. Environmental Microbiome*   (2020) 15:5

Page 9 of 18

Fig. 7 Log2fold changes of selected marker genes of motility related genes for each analysed land use system compared to rainforest. Significant differences (p.adj. < 0.05) are marked with *. Negative and positive log2fold changes indicate decreased and increased abundances in the corresponding converted land use systems compared to rainforest samples

processes. In context of our previous analysis, we focussed on nitrogen metabolism (Fig. 8a), methane metabolism (Fig. 8b), bacterial chemotaxis (Fig. 9a), flagellar assembly (Fig. 9b), type IV secretion systems (Fig. 9c) and type VI secretion systems (Fig. 9d). Members of the *Rhizobiales* were prominent within nitrogen metabolism-related sequences in all samples with an average abundance of 12%. *Burkholderiales* were especially abundant in rainforest samples with 14.6% with a gradual abundance decrease from jungle rubber (8.4%) to rubber and oil palm samples (5.9%) (Fig. 8a). Both orders are known to be involved in various processes related to nitrogen metabolism and were previously identified as abundant groups in rainforest soils [40]. With unclassified *Acidobacteria*, and unclassified bacteria being high in abundance as well, observed patterns broadly reflected the general community structure as described before. However, redundant genes, which are involved in several pathways, are also present within broader functional categories. This likely explains the resemblance of the categories with the general community composition, including taxa usually not associated with typical nitrogen metabolism-related processes like *Acidobacteria*, which were still abundant in our analysed samples. Consequently, we identified taxa which are connected to previously analysed functional marker genes that showed significant changes in abundance due to rainforest conversion (Additional file 13: Figure S11, and Additional file 14: Figure S12). The taxonomic profiles for the denitrification-related genes *nirS* and *nirK* were distinct. *Rhizobiales* were the dominant identified phylogenetic group in *nirK* sequences of all samples with decreasing abundance from rainforest to converted land use systems. The *nirS* gene sequences were mostly associated to unidentified bacterial taxa, with a vast increase towards managed land use systems.

Similar patterns were observed for the methane metabolism-related taxa (Fig. 8b). However, unclassified

Fig. 8 Relative abundances of the top five detected taxonomic orders within nitrogen (**a**) and methane metabolism (**b**) KEGG level 3 categories in each land use system are displayed

acidobacterial sequences were more abundant in methane metabolism-related than in nitrogen metabolism-related taxa and showed abundance increases from rainforest to converted land use systems of 8.9 to 14.3%.

Taxa related to bacterial chemotaxis (Fig. 9a) and flagellar assembly (Fig. 9b) showed similar patterns in which *Rhizobiales* and *Burkholderiales* were the most abundant taxa. Again, *Burkholderiales* showed high abundances in rainforest samples. Unclassified acidobacterial orders were among the most abundant groups with abundance increases in converted land use systems for flagellar assembly-related and chemotaxis-related genes. Dominance of *Rhizobiales* and *Burkholderiales* in rainforest samples was also detected in type IV secretion system-related taxa (Fig. 9c). A slightly different pattern was observed for type VI secretion systems-related taxa (Fig. 9d), with *Burkholderiales* showing highest mean abundances in all samples.
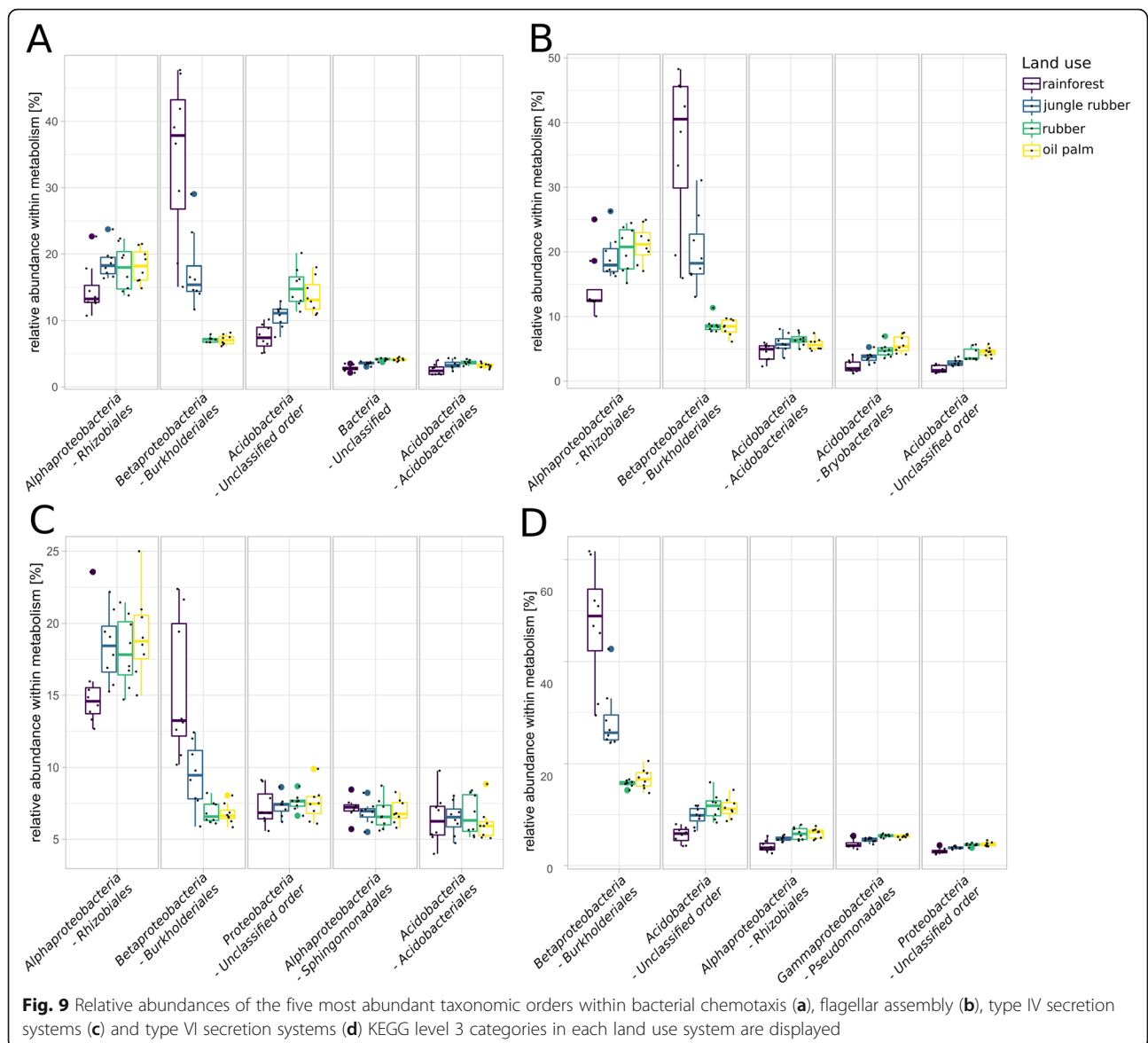
Discussion

Dominant *Proteobacteria* and *Acidobacteria* are mostly affected by rainforest conversion

Observed decreases in taxonomic biodiversity accompanying rainforest conversion to managed land use systems were reported previously for various types of organisms [10, 11, 41, 42]. Diversity of bacteria and archaea decreased in plantation soils compared to rainforest. Contrary to these results, previous studies based on 16S rRNA gene analysis of the same sampling sites observed a diversity increase for bacteria from rainforest to converted land use systems [15, 16]. A higher amount of unclassified reads in shotgun datasets compared to 16S rRNA gene datasets could be a reason for the different results. Other studies focussing on nutrient cycling in these systems

concluded that nutrient loss by leaching processes is elevated in rubber and oil palm plantation soils and affects bacterial diversity negatively [18, 43]. Fungal diversity increased with land use intensity and showed similar patterns as in a previous study [9]. The major difference between our direct metagenome sequencing-based approach and the other studies on the same sampling sites is that these were derived from 16S rRNA gene amplicon-based analysis. Since our study is based on metagenomic shotgun sequencing, we avoided possible primer bias and additionally have a higher sequencing depth, which both could lead to deviations between results derived from different approaches. Another reason for differences could be the classification of reads and the chosen tool for taxonomic binning. It has been reported that deviations in taxonomic profiles originate from false classifications rather than from sequencing errors [44]. Large-scale classifications of metagenomic short-reads are still a challenge though. To overcome possible shortcomings of single tools [44] and provide a higher classification efficiency [45], we combined the two classification tools Kraken2 and Kaiju to obtain robust results. General trends were confirmed by extracted 16S rRNA gene sequence controls and were similar to previous amplicon-based 16S rRNA gene studies [15, 16]. However, the fraction of unclassified bacterial taxa was larger in the extracted 16S rRNA gene-based control. This is the most notable differences to metagenomic shotgun data and demonstrates the advantage of a higher sequencing depth and consideration of all sequences instead of extracted marker gene sequences only.

The soil microbial community composition changed in the converted land use systems compared to rainforest

Fig. 9 Relative abundances of the five most abundant taxonomic orders within bacterial chemotaxis (**a**), flagellar assembly (**b**), type IV secretion systems (**c**) and type VI secretion systems (**d**) KEGG level 3 categories in each land use system are displayed

with similar trends as in previous studies targeting these sampling sites [9, 15, 16]. *Proteobacteria* were negatively and *Acidobacteria* positively affected by rainforest conversion. The sharp decrease in abundance of *Rhizobiales* and *Burkholderiales* in converted systems suggests that they play a key role in rainforest soils, as suggested in previous studies [15, 16]. *Streptomycetales* are known for production of bioactive secondary metabolites and are abundant in endophytic communities [46]. Considering the reduction of plant diversity in the converted systems, the decrease of *Streptomycetales* accompanying rainforest conversion corresponds to the reduction of host diversity in the converted systems. The reduction of this group indicates that the overall biodiversity changes are reflected in abundance decreases of bacterial groups known to be involved in interspecific relationships as

endophytes [47] or root nodule-forming symbionts [35], including the previously mentioned *Rhizobiales* and *Burkholderiales*.

Archaeal abundances did not seem to be affected by rainforest conversion. *Haloferacales* and *Halobateriales* were among the most abundant taxa and are known for their involvement in assimilatory nitrate reduction [48, 49]. However, we did not detect a significant involvement of these taxa in these processes, although we recorded significant changes of *narG* gene abundance. Based on our used methods, we could not identify a crucial effect of rainforest conversion on archaea. Although we did not find hints for functional importance in these systems, we cannot rule out the possibility of important archaeal activity due to the limitations of the DNA-based study design.

Our results showed that the fungal community composition in the studied land use systems was similar to an ITS-based analysis [9]. *Eurotiales* and *Hypocreales* are described as endophytes [50, 51], further supporting the assumption that the reduction of plant diversity also reduces endophyte occurrence. In addition to fungi, we analysed protists within the eukaryotes but the obtained trophic groups did not show significant differences in abundance in contrast to an 18S rRNA gene-based study [10]. This might be due to the differences between metagenomic shotgun sequencing and marker gene analysis [32]. In addition, protist sequences represented only a small portion (0.09%) in the entire metagenome dataset. Taking the number of unclassified sequences and the underrepresentation of protist genomes in the databases into account, the results for protists and other similar rare taxa should be taken with caution.

Management-related soil characteristics shape soilborne communities

Multivariate analyses showed a similar topology for all analysed taxonomic groups, reflecting the proposed management and land use intensity gradient from rainforest over jungle rubber to rubber and oil palm. The correlation of abiotic factors LUI, pH, C:N ratio and Fe to the entire community as well as to bacteria is not surprising as the communities and correspondingly our dataset were dominated by bacterial sequences. These results are in accordance with other taxonomic marker gene analysis-based studies targeting bacteria, archaea [15, 16] and fungi [9] in the same study area. Management-related applications (e.g. fertilizer, liming and herbicides) are known to have an impact on C:N ratio and soil pH. These alterations in turn affect microbial communities [15, 52], which is also reflected by our results. Allen et al. [18] analysed the soil parameters in the used sampling sites and described decreases in organic matter stocks from rainforest to converted land use systems paralleled by a decreasing C:N ratio. Furthermore, nutrient availability for microbes is improved by fertilizer input in plantations, but is heavily dependent on ongoing fertilization, as these converted land use systems are more vulnerable to nutrient losses due to changes in soil texture and soil properties introduced by plantation of rubber and oil palm monocultures [18]. Therefore, our results regarding impact of rainforest conversion on community structure seem to be direct effects of the management practices in the converted land use systems. Consequently, the modification of the soil environment is a reason for the observed functional alterations (e.g. in nitrogen metabolism) caused by land use change and applied management practices.

Functional profiles of energy and motility-related metabolisms are affected by rainforest conversion

In general, differential gene abundance analysis showed less changes in jungle rubber than in monoculture samples compared to rainforest. The gradually increase of changes reflects the land use gradient with stronger intensity in fertilized oil palm soils as proposed in previous studies [9, 15, 16].

Diversity of genes decreased with increasing land use intensity. Since most detected reads were of bacterial origin, it is not surprising that the functional diversity of the entire community shows similar results as bacterial diversity. For fungi, we detected a decrease in functional gene diversity, but an increase of taxonomic diversity, indicating an increase in less versatile species in monoculture soils or the presence of more different taxa equipped with similar gene sets.

We previously recorded that pH, carbon content, nitrogen content and C:N ratio played an important role for shaping the soil microbial communities. Our functional results support hypothesis (H2) that rainforest conversion affects metabolic reactions specific for nutrient cycling activity, including nitrogen metabolism and carbon fixation-related processes. As previously suggested, the change in nutrient availability affected soil microbial interaction and/or competition [16, 27]. Furthermore, it was hypothesized that different canopy structures in the analysed land use systems lead to higher light availability and higher ground temperatures in the converted systems, which in turn affect photosynthetic processes [15] and heat management of cells [16]. Detected higher abundance of photosynthesis related genes in oil palm further support this hypothesis.

The analysis of single functional marker genes in the metagenome dataset showed that denitrification genes *nirK* and *nirS* increased in abundance following rainforest conversion whereas nitrate reductase *narG* decreased. Allen et al. [18] described decreases in gross N mineralization, NH_4^+ immobilization and NH_4^+ pools in converted soils and generally higher N cycling rates in rainforest soils [18]. It was previously reported that *nirK* and *nirS* are positively correlated with increasing soil pH [53], which is in line with measurements conducted in our sampling sites [18] and multivariate analysis showing a significant influence of pH on microbial communities. Previous studies also described distinct distributions of the two different denitrification genes *nirK* and *nirS* among taxa, with *nirK* being associated to *Rhizobiales* [54]. Accordingly, most detected *nirK* sequences in our dataset originated from *Rhizobiales*. The higher gene abundances of denitrification-related genes in oil palm plantations correspond to previous observations in which higher N_2O emissions, a by-product of denitrification, were detected in oil palm plantations [18].

We also observed higher nitrogen fixation potential in oil palm soils, which is surprising, considering previously reported nitrogen losses in these soils [18], although higher gene abundance is not necessarily accompanied by higher gene expression rates. Decreases in abundance of marker genes for flagellar assembly, chemotaxis, and type VI and IV secretion systems indicate that a reduction of motility and interaction in soil microbial communities accompanies rainforest conversion to intensively managed land use systems. It is challenging to identify the relationship of these aspects within a soil microbial community, since specific functions/processes can be associated with specific taxonomic groups. Our results indicate a shift from a dynamic community with higher motility, interaction and communication potential in rainforest soils to a community with less interaction potential in the managed plantations. This supports previous hypotheses regarding possible impacts of rainforest conversion towards community communication traits [16, 28]. The higher availability of nutrients in the fertilized plantation soils might lead to conditions in which competition for nutrients and the resulting high level of interspecific interaction and communication is not favoured or necessary and therefore reduced. Furthermore, specific taxonomic groups, involved in nutrient cycling, possibly show different abilities in terms of motility in the nutrient-supplemented planation soil compared to rainforest soil. This supports the theory of a less mobile community in planation soils but rather due to changes in community composition than to direct effects of rainforest conversion on functionality.

Proteobacterial *Burkholderiales* are major players with respect to motility and nitrogen metabolism

We previously hypothesized that communities in rainforest soils are more dynamic and motile because of higher abundance of genes encoding flagellar assembly and chemotaxis. It is not surprising that a considerable number of reads belonging to these categories are assigned to abundant members of the community and we detected highest abundance on average for *Rhizobiales* in all samples. However, it is quite striking that the majority of flagellar assembly-related sequences in rainforest samples were assigned to members of *Burkholderiales* alone (Fig. 9). The same trend occurred for chemotaxis-related sequences. The degree of involvement in flagellar assembly leads to the suggestion that *Rhizobiales* and *Burkholderiales* are in general more capable of being motile than other groups in these soils. Additionally, significantly abundant genes connected to type IV and VI secretion systems were mostly derived from *Burkholderiales*. It was shown that members of the *Burkholderiales* are involved in co-migration with fungi in soil involving type IV pili [55]. The type VI secretion

systems are involved in prokaryotic eukaryotic interactions [55–57]. Therefore, *Burkholderiales* and *Rhizobiales* possibly play an important functional role in soil microbial communities in rainforest due to their involvement in nutrient cycling. Their decrease in abundance in the converted land use systems was accompanied by various changes of the functional potential of the entire microbial soil community, which is affected by management-induced altered soil characteristics.

Conclusion

We could show that rainforest conversion drastically affects structure and functional potential of soil microbial communities, which were dominated by bacteria. Taxonomic as well as functional diversity decreased for bacteria and archaea, whereas fungal biodiversity increased and functional diversity decreased, partly confirming hypothesis H1. Furthermore, functional profiles of the soil communities shifted along the land use gradient. Denitrification and nitrogen fixation potential increased with higher land use intensity, which is connected to abundant community members of *Burkholderiales* and *Rhizobiales*. Furthermore, we could show that pH and C: N ratio are drivers for shaping microbial community structure, which connects previously shown positive correlations of pH with land use intensity and denitrification potential. This is further supported by previous studies reporting an increase of N_2O effluxes with rainforest conversion to oil palm plantations [18].

We detected a decrease in motility- and interaction-related functions from rainforest to converted land use systems indicating not only a shift in nutrient cycling activity but also in community dynamics confirming our second hypothesis (H2). Fertilizer application and higher short-term availability of nutrients in intensively managed plantations lead to an environment in which interspecific interactions apparently are less favoured compared to rainforest soil. Microbial communities underwent a shift in composition in which *Rhizobiales*, *Burkholderiales* and other members of the *Proteobacteria* decreased from rainforest to plantations, whereas *Acidobacteria* and *Actinobacteria* increased. Additionally, the largest fraction of sequences within analysed motility processes in rainforest belonged to members of the *Burkholderiales* and *Rhizobiales*, connecting decreases in abundance with a decrease in functions related to these processes (H3).

We could show connections between agricultural management and microbial community structure and functional potential in soil. Furthermore, this study provides a basis for further analysis on functional responses of soil-borne microbial communities to rainforest conversion, which need ongoing attention due to the global impacts of large-scale land use changes in the tropics.

Berkelmann *et al. Environmental Microbiome*   (2020) 15:5

Page 14 of 18

Methods

Study design and sampling site description

Sampling was conducted in two landscapes around the Harapan Rainforest Concession and the Bukit Duabelas National Park in midwest Sumatra (Indonesia) within the framework of the "Collaborative Research Centre 990: Ecological and Socioeconomic Functions of Tropical Lowland Rainforest Transformation Systems" (EFForTS). The two landscapes differ in soil texture, with clay acrisol soils in Bukit and loam acrisol soils in the Harapan landscape. Both landscapes harbour secondary lowland rainforest as reference site and three different converted land use systems: jungle rubber comprising planted rubber trees in secondary rainforest, rubber plantations (*Hevea brasiliensis* monocultures) and oil palm plantations (*Elaeis guineensis* monocultures). Rainforest was used as reference with low anthropogenic influence, while the three converted land use systems represent different land use intensities resulting in a land use intensity gradient (rainforest < jungle rubber < rubber < oil palm). Oil palm monoculture plantations were fertilized with 300 kg to 550 kg NPK fertilizer ha^{-1} $year^{-1}$ [18]. Additionally, liming was performed in rubber and oil palm plantations with an average of 200 kg dolomite ha^{-1} $year^{-1}$ and chemical and manual weeding was done by using Gramoxone and Roundup with an average of 2 to 5 L herbicide ha^{-1} $year^{-1}$. Further information about the sampling sites and management of these sites is described by Allen et al. [18] and Brinkmann et al. [9]. Soil sampling, preparation (root removal), transport and storage was carried out as described in detail by Schneider et al. [15] and Berkelmann et al. [16]. Each of the four analysed land use systems (rainforest, jungle rubber, rubber plantation and oil palm plantation) consisted of four core plots per landscape, including three subplots (five by five meters) per core plot, resulting in 96 subplot samples and 32 core plots in total (Fig. 1). Abiotic data of the sampling sites were obtained from Allen et al. [18], Brinkmann et al. [9] (Additional file 15: Table S3).

Sample preparation, extraction of bulk soil DNA and sequencing

For direct metagenome sequencing and prior to DNA extraction, soil samples were pooled in equal amounts at core plot level, resulting in 32 soil samples in total. DNA extraction of soil samples was performed with the MoBio Powersoil DNA extraction kit (MO BIO Laboratories Inc. Carlsbad, USA) as recommended by the manufacturer. The shotgun metagenomic sequencing of all DNA samples were performed on an Illumina HiSeq 4000 system with Nextera DNA Library Prep kits and paired-end reads of 2×150 bp as recommended by the manufacturer (Illumina, San Diego, USA).

Taxonomic and functional assignment of paired-end reads

Raw sequences were quality-filtered with fastp (version 0.19.4) with a phredscore threshold of 20, overlapping base pair correction, sliding windows of 4 bp and a minimum length of 50 bp [58]. It has been shown that the combination of different sequence classifiers leads to more robust taxonomic assignments [44]. Therefore, we used Kraken2 and Kaiju in combination. Taxonomy assignments of short reads were performed by Kraken2 (v2.0.8-beta) [57] against the BLAST nt database (as of 2019-06-08). Afterwards unclassified reads were assigned with Kaiju (version 1.7.1) [59] against the BLAST nr database (as of 2019-06-08). Outputs were merged and taxonomy strings added by "addTaxonNames" by Kaiju (Additional file 16: Table S4). Protist groups were extracted from the normalized taxonomy table with "amp_subset_taxa" from the ampvis2 R package [60] according to taxonomy strings described by Schulz et al. [10]. Visualization of trophic groups was performed with ggplot2 [61]. A rRNA gene sequence-based control analysis was performed by extracting rRNA gene sequences from quality-filtered and merged reads employing sortmerna 2.1 with all by default available databases and settings [62]. Extracted rRNA gene sequences were classified as described above.

Assignment of read functions was done with previously quality-filtered reads (Additional file 17: Table S5). Functional classification was carried out by employing Kaijux [59] with default settings against the KEGG database [63] (as of 2018-10-01).

Assigning the taxonomic background of identified functional genes

To assign the taxonomic affiliation of each sequence, taxonomic and functional assignments were further merged by combining the taxonomic classification with the obtained KEGG identifier for each sequence of the beforementioned taxonomic read assignment and functional analysis, resulting in a per read taxonomy and function. This table was then filtered according to the targeted pathway, (i.e. nitrogen metabolism). All extracted hits were then normalized and displayed as relative abundances for the respective land use system with ggplot2.

Analysis of selected functional metabolisms and respective marker genes

In order to analyse microbial functionality regarding rainforest conversion, we selected metabolisms and respective genes related to agricultural management induced changes [64]. Groups and genes of the KEGG database were used (KEGG level 3 for functional groups). We analysed functional pathways harbouring genes involved in nutrient cycling, including nitrogen metabolism, methane

metabolism, carbon fixation and sulphur metabolism. For more detailed analyses, we picked marker genes and divided them into two categories regarding energy metabolism or motility. Genes selected for nitrogen metabolism were *amoA* (ammonia monooxygenase A; K10944) [65], *nifH* (nitrogenase protein; K02588) [40], *nosZ* (nitrous-oxide reductase; KK00376) [66], *nirK* and *nirS* (both encoding a nitrite reductase; K00368 and K15864) [40] and *narG* (nitrate reductase alpha subunit; K00370) [66]. For methane related processes we used *pmoA/amoA* (methane/ammonia monooxygenase subunit A; K10944) [64], *mxaF* (methanol dehydrogenase; K14028) [67], *mmoY* (methane monooxygenase component A beta chain; K16158) and *mmoX* (methane monooxygenase component A alpha chain; K16157) [68] and *mcrA* (5-methylcytosine-specific restriction enzyme A; K07451) [69].

To investigate effects on motility and interactions we analysed functional groups comprising flagellar assembly [70, 71] and chemotaxis [72], quorum sensing [73], and type IV and VI secretion systems [74, 75]. For flagellar assembly, we selected genes encoding flagellar motor switch proteins *fliN* (K02417), *fliM* (K02416) and *fliG* (K02410), the flagellar hook length protein *fliK* (K02414) [70] and a flagellar assembly protein *fliH* (K02411) [71]. Chemotaxis was covered by the sensor kinase gene *cheA* (K03407) [76], CheA response regulator *cheY* (KK03413) and scaffolding protein *cheW* (K03408) [77], CheY-phosphatase *cheX* (K03409) [78], and *cheR* (K00575), encoding a protein methyltransferase [79]. For quorum sensing we selected genes encoding cytoplasmic autoinducer receptors *luxR* (K10913), sensor kinases *luxQ* (K10909), *luxN* (K15850) and *luxS* (K07173) [80, 81]. Type IV secretion system related genes were represented by *virD4* (K03205) [82] and *virB4* (K03199) [83] and type VI secretion system related genes by *vgrG* (K11904) and *hcp* (K11903 [75].

Statistical analyses

Diversity analysis, plotting of barplots and heatmaps were done with R [84] and RStudio [85] by using the packages ampvis2 [60], vegan [86], dplyr [61], stringr and ggplot2 [61]. Shannon diversity index calculation for taxonomic diversity was performed with previously rarefied data using ampvis2 with amp_alphadiv. Functional diversity was calculated by summing all identical genes of the entire community or by summing all identical genes in the respective domain before calculating the Shannon index with ampvis2. Statistical analysis of calculated Shannon diversity results was performed with the vegan package (version 2.5–5). First, the Shapiro test was used to determine normal distribution of the data with base R and shapiro.test. All obtained values were non-normally distributed and therefore further analysed with base R by using the Kruskal-Wallis test (Kruskal.test) with subsequent pairwise Wilcoxon test (pairwise.wilcox.test).

Differential abundance analysis and count normalization of taxonomic or functional data were performed by using the DESeq2 package [87]. Default settings with the Benjamini and Hochberg correction were used after removing singletons. Normalized counts were extracted as described in the package manual and used for data visualizations. Differential abundance analysis was done for the entire dataset by the main DESeq function. Differential abundance results were obtained by the "contrast" function, in which log2fold changes of rainforest compared to each converted land use were extracted in a pairwise fashion (results_dataframe <– results(dds, contrast = c("condition", "reference", "treatment")), resulting in positive log2fold changes when gene abundances increased from reference to the respective land use and negative values when abundances decreased. The threshold for significant differences was set to $p < 0.05$. Generated tables per used condition were merged by using the dplyr package and visualized by ggplot2. Analysis was performed in a same manner for taxonomic and functional classifications. Differences in abundances between rainforest and each land use were tested at order level. Functional profiles were tested at metabolism level (level 3 of the KEGG hierarchy) and gene level.

Obtained taxonomy profiles of extracted rRNA sequences were analysed for similarity to shotgun data by calculating Bray Curtis dissimilarity matrices with the "adonis" function of the vegan package in R and subsequent tests by using pairwise PERMANOVA tests with "pairwise.perm.manova" of the RVAideMemoire package [88] in R.

Ordination analysis was performed with the ampvis2 package [60]. Raw data was rarefied for each tested taxonomic group or the entire community by ampvis2 (amp_subset_samples) before a Bray Curtis dissimilarity matrix was calculated and visualized by Nonmetric multidimensional scaling (NMDS). Data for soil characteristics of the sampling sites were obtained from Allen et al. [18] (Additional file 11: Table S3). In addition to abiotic measurements, we also included the Land Use Index (LUI) as described by Brinkmann et al. [9] (Additional file 15: Table S3). An environmental fit was calculated with the "amp_ordinate" function in ampvis2 with an envfit significance level of p < 0.05.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40793-020-0353-3.

Additional file 1: Table S1. Sequence counts of all samples before and after quality-filtering with fastp.

Additional file 2: Figure S1. Relative abundances of each domain for each analysed land use system in the respective landscape.

Additional file 3: Figure S2. Distribution of taxonomic differences compared to rainforest at order level. Differences are displayed as log2fold (treatment vs control) changes with p.adj < 0.05 for detected

---

orders in all land use systems. Bukit jungle rubber is not depicted as no significant changes of taxonomic orders were detected.

Additional file 4: Figure S3. Composition of different trophic protist groups. Trophic groups were assigned according to Schulz et al. [10]. Compositions are displayed for each land use in the respective landscape as relative abundances.

Additional file 5: Figure S4. Community composition at order level based on extracted rRNA sequences (A) and all obtained reads by shotgun sequencing (E). orders with relative abundances below 1% were clustered as "rare taxa". The five most abundant orders based on average relative abundance of extracted 16S rRNA gene sequences in all treatments are shown for bacteria (B), archaea (C) and fungi (D) as well as for all obtained reads by shotgun sequencing (F-H).

Additional file 6: Table S2. Environmental fit results for the entire community, bacteria, archaea and fungi.

Additional file 7: Figure S5. Distribution of gene changes in each converted land use system compared to rainforest. Detected changes are displayed as log2fold changes (treatment vs control; p.adj < 0.05).

Additional file 8: Figure S6. Principal Component Analysis for detected genes of all analysed land use systems and rainforest samples in the respective landscape. PCA analysis is based on transformed counts by using the regularized log function of DESeq2.

Additional file 9: Figure S7. Functional profile of selected metabolisms based on KEGG categories (level 3) displayed as log2fold changes in separated landscapes. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems. Values with p.adj < 0.05 are marked with *.

Additional file 10: Figure S8. Log2fold changes of selected marker genes of nitrogen and methane metabolism for each analysed land use system compared to rainforest for each separate landscape. Significant differences (p.adj. < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 11: Figure S9. Log2fold changes of selected marker genes motility related marker genes for each analysed land use system compared to rainforest in each landscape. Significant differences (p.adj. < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 12: Figure S10. Log2fold changes of *nod* genes in each land use compared to rainforest. Significant differences (p.adj < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 13: Figure S11. Relative abundances of the five most abundant detected taxonomic orders within all detected sequences for nitrogen related marker genes that showed significant differences between rainforest and converted land use systems. Displayed heatmaps show the five most abundant detected taxa for nitrite reductase gene *nirK*, nitrite reductase gene *nirS* and nitrate reductase alpha subunit gene *narG*.

Additional file 14: Figure S12. Relative abundances of the five most abundant detected taxonomic orders within all detected sequences for motility related marker genes that showed significant differences between rainforest and converted land use systems.

Additional file 15: Table S3. Used abiotic measurements for all samples.

Additional file 16: Table S4. Taxonomic counts of all analysed sequences. The first column shows ID and the last column the detected taxonomy.

Additional file 17: Table S5. Count matrix containing functional assignments for all samples. The first column shows the ID the last columns the function according to the KEGG database including the KO number and the full pathway description.

Availability of data and materials
All data generated or analysed during this study are included in this published article. Obtained sequences were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under accession number PRJNA562410.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University, Grisebachstr. 8, 37077 Göttingen, Germany. [2]Department of Biology, Faculty of Mathematics and Natural Sciences IPB, Bogor Agricultural University, Bogor, Indonesia.

References
1. Alexander P, Rounsevell MDA, Dislich C, Dodson JR, Engström K, Moran D. Drivers for global agricultural land use change: the nexus of diet, population, yield and bioenergy. Glob Environ Chang. 2015;35:138–47. https://doi.org/10.1016/j.gloenvcha.2015.08.011.
2. Oosterveer P. Promoting sustainable palm oil: viewed from a global networks and flows perspective. J Clean Prod. 2015;107:146–53. https://doi.org/10.1016/j.jclepro.2014.01.019.
3. Dislich C, Keyel AC, Salecker J, Kisel Y, Meyer KM, Auliya M, et al. A review of the ecosystem functions in oil palm plantations, using forests as a reference system. Biol Rev. 2017;92:1539–69.
4. Drescher J, Rembold K, Allen K, Beckschäfer P, Buchori D, Clough Y, et al. Ecological and socio-economic functions across tropical land use systems after rainforest conversion. Philos Trans R Soc B Biol Sci. 2016;371:20150275. https://doi.org/10.1098/rstb.2015.0275.
5. TO D, Robe P, Cecillon S, Clark IM, Constancias F, Simonet P, et al. Accessing the soil metagenome for studies of microbial diversity. Appl Environ Microbiol. 2011;77:1315–24.
6. Fierer N. Embracing the unknown: disentangling the complexities of the soil microbiome. Nat Rev Microbiol. 2017;15:579. https://doi.org/10.1038/nrmicro.2017.87.
7. Kroeger ME, TO D, Eren AM, Meyer KM, Guo J, Khan K, et al. New biological insights into how deforestation in Amazonia affects soil microbial

communities using metagenomics and metagenome-assembled genomes. Front Microbiol. 2018;9:1635. https://doi.org/10.3389/fmicb.2018.01635.

8. Sahner J, Budi SW, Barus H, Edy N, Meyer M, Corre MD, et al. Degradation of root community traits as indicator for transformation of tropical lowland rain forests into oil palm and rubber plantations. PLoS One. 2015;10:1–19. https://doi.org/10.1371/journal.pone.0138077.

9. Brinkmann N, Schneider D, Sahner J, Ballauff J, Edy N, Barus H, et al. Intensive tropical land use massively shifts soil fungal communities. Sci Rep. 2019;9:3403. https://doi.org/10.1038/s41598-019-39829-4.

10. Schulz G, Schneider D, Brinkmann N, Edy N, Daniel R, Polle A, et al. Changes in trophic groups of protists with conversion of rainforest into rubber and oil palm plantations. Front Microbiol. 2019;10:240 https://www.frontiersin.org/article/10.3389/fmicb.2019.00240.

11. Prabowo WE, Darras K, Clough Y, Toledo-Hernandez M, Arlettaz R, Mulyani YA, et al. Bird responses to lowland rainforest conversion in Sumatran smallholder landscapes. Indonesia. PLoS One. 2016;11:e0154876. https://doi.org/10.1371/journal.pone.0154876.

12. Böhnert T, Wenzel A, Altenhövel C, Beeretz L, Tjitrosoedirdjo SS, Meijide A, et al. Effects of land-use change on vascular epiphyte diversity in Sumatra (Indonesia). Biol Conserv. 2016;202:20–9. https://doi.org/10.1016/j.biocon.2016.08.008.

13. Potapov AM, Klarner B, Sandmann D, Widyastuti R, Scheu S. Linking size spectrum, energy flux and trophic multifunctionality in soil food webs of tropical land-use systems. J Anim Ecol. 2019;0. https://doi.org/10.1111/1365-2656.13027.

14. Mumme S, Jochum M, Brose U, Haneda NF, Barnes AD. Functional diversity and stability of litter-invertebrate communities following land-use change in Sumatra. Indonesia Biol Conserv. 2015;191:750–8. https://doi.org/10.1016/j.biocon.2015.08.033.

15. Schneider D, Engelhaupt M, Allen K, Kurniawan S, Krashevska V, Heinemann M, et al. Impact of lowland rainforest transformation on diversity and composition of soil prokaryotic communities in Sumatra (Indonesia). Front Microbiol. 2015;6. https://doi.org/10.3389/fmicb.2015.01339.

16. Berkelmann D, Schneider D, Engelhaupt M, Heinemann M, Christel S, Wijayanti M, et al. How rainforest conversion to agricultural systems in Sumatra (Indonesia) affects active soil bacterial communities. Front Microbiol. 2018;9:2381. https://doi.org/10.3389/fmicb.2018.02381.

17. Guillaume T, Damris M, Kuzyakov Y. Losses of soil carbon by converting tropical forest to plantations: erosion and decomposition estimated by δ13C. Glob Chang Biol. 2015;21:3548–60. https://doi.org/10.1111/gcb.12907.

18. Allen K, Corre MD, Tjoa A, Veldkamp E. Soil nitrogen-cycling responses to conversion of lowland forests to oil palm and rubber plantations in Sumatra. Indonesia PLoS One. 2015;10:e0133325. https://doi.org/10.1371/journal.pone.0133325.

19. Hassler E, Corre MD, Tjoa A, Damris M, Utami SR, Veldkamp E. Soil fertility controls soil–atmosphere carbon dioxide and methane fluxes in a tropical landscape converted from lowland forest to rubber and oil palm plantations. Biogeosciences. 2015;12:5831–52. https://doi.org/10.5194/bg-12-5831-2015.

20. Kotowska MM, Leuschner C, Triadiati T, Meriem S, Hertel D. Quantifying above- and belowground biomass carbon loss with forest conversion in tropical lowlands of Sumatra (Indonesia). Glob Chang Biol. 2015;21:3620–34.

21. Kotowska MM, Leuschner C, Triadiati T, Hertel D. Conversion of tropical lowland forest reduces nutrient return through litterfall, and alters nutrient use efficiency and seasonality of net primary production. Oecologia. 2016;180:601–18. https://doi.org/10.1007/s00442-015-3481-5.

22. Bhardwaj D, Ansari MW, Sahoo RK, Tuteja N. Biofertilizers function as key player in sustainable agriculture by improving soil fertility, plant tolerance and crop productivity. Microb Cell Factories. 2014;13, Article number 66.

23. Lynch DH. Nutrient cycling and soil health in organic cropping systems-importance of management strategies and soil resilience. Sustain Agric Res. 2015;4:80–8.

24. Newman MM, Hoilett N, Lorenz N, Dick RP, Liles MR, Ramsier C, et al. Glyphosate effects on soil rhizosphere-associated bacterial communities. Sci Total Environ. 2016;543:155–60. https://doi.org/10.1016/j.scitotenv.2015.11.008.

25. Wemheuer F, Taylor JA, Daniel R, Johnston E, Meinicke P, Thomas T, et al. Tax4Fun2: a R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene marker gene sequences. bioRxiv. 2018:490037. https://doi.org/10.1101/490037.

26. Iwai S, Weinmaier T, Schmidt BL, Albertson DG, Poloso NJ, Dabbagh K, et al. Piphillin: improved prediction of metagenomic content by direct inference from human microbiomes. PLoS One. 2016;11:e0166104. https://doi.org/10.1371/journal.pone.0166104.

27. Tripathi BM, Edwards DP, Mendes LW, Kim M, Dong K, Kim H, et al. The impact of tropical forest logging and oil palm agriculture on the soil microbiome. Mol Ecol. 2016;25:2244–57.

28. Lee-Cruz L, Edwards DP, Tripathi BM, Adams JM. Impact of logging and forest conversion to oil palm plantations on soil bacterial communities in Borneo. Appl Environ Microbiol. 2013;79:7290–7.

29. Leff JW, Nemergut DR, Grandy AS, O'Neill SP, Wickings K, Townsend AR, et al. The effects of soil bacterial community structure on decomposition in a tropical rain forest. Ecosystems. 2012;15:284–98. https://doi.org/10.1007/s10021-011-9510-2.

30. Zhang J, Ding X, Guan R, Zhu C, Xu C, Zhu B, et al. Evaluation of different 16S rRNA gene V regions for exploring bacterial diversity in a eutrophic freshwater lake. Sci Total Environ. 2018;618:1254–67. https://doi.org/10.1016/j.scitotenv.2017.09.228.

31. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotech. 2013;31:814–21. https://doi.org/10.1038/nbt.2676.

32. Rausch P, Rühlemann M, Hermes BM, Doms S, Dagan T, Dierking K, et al. Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. Microbiome. 2019;7:133. https://doi.org/10.1186/s40168-019-0743-1.

33. Kerfahi D, Tripathi BM, Dong K, Go R, Adams JM. Rainforest conversion to rubber plantation may not result in lower soil diversity of bacteria, fungi, and nematodes. Microb Ecol. 2016;72:359–71.

34. Tang H, Xiao X, Li C, Tang W, Cheng K, Wang K, et al. Effects of rhizosphere and long-term fertilization practices on the activity and community structure of denitrifiers under double-cropping rice field. Commun Soil Sci Plant Anal. 2019;50:682–97. https://doi.org/10.1080/00103624.2019.1589480.

35. Li Y, Pan F, Yao H. Response of symbiotic and asymbiotic nitrogen-fixing microorganisms to nitrogen fertilizer application. J Soils Sediments. 2019;19:1948–58. https://doi.org/10.1007/s11368-018-2192-z.

36. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, et al. Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J. 2010;4:1340–51.

37. Brevik EC, Cerdà A, Mataix-Solera J, Pereg L, Quinton JN, Six J, et al. The interdisciplinary nature of SOIL. Soil. 2015;1:117–29.

38. Corre MD, Dechert G, Veldkamp E. Soil nitrogen cycling following montane forest conversion in Central Sulawesi, Indonesia. Soil Sci Soc Am J. 2006;70:359–66.

39. Lan G, Li Y, Wu Z, Xie G. Soil bacterial diversity impacted by conversion of secondary forest to rubber or eucalyptus plantations: a case study of Hainan Island, South China. For Sci. 2016;63:87–93. https://doi.org/10.5849/forsci.16-012.

40. Tang Y, Yu G, Zhang X, Wang Q, Ge J, Liu S. Changes in nitrogen-cycling microbial communities with depth in temperate and subtropical forest soils. Appl Soil Ecol. 2018;124:218–28. https://doi.org/10.1016/j.apsoil.2017.10.029.

41. Kusuma YWC, Rembold K, Tjitrosoedirdjo SS, Kreft H. Tropical rainforest conversion and land use intensification reduce understorey plant phylogenetic diversity. J Appl Ecol. 2018;55:2216–26. https://doi.org/10.1111/1365-2664.13201.

42. Rembold K, Mangopo H, Tjitrosoedirdjo SS, Kreft H. Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. Biol Conserv. 2017;213:234–42. https://doi.org/10.1016/j.biocon.2017.07.020.

43. van Straaten O, Corre MD, Wolf K, Tchienkoua M, Cuellar E, Matthews RB, et al. Conversion of lowland tropical forests to tree cash crop plantations loses up to one-half of stored soil organic carbon. Proc Natl Acad Sci. 2015;112:9956 LP–9960. https://doi.org/10.1073/pnas.1504628112.

44. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking metagenomics tools for taxonomic classification. Cell. 2019;178:779–94. https://doi.org/10.1016/j.cell.2019.07.010.

45. McIntyre ABR, Ounit R, Afshinnekoo E, Prill RJ, Hénaff E, Alexander N, et al. Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. Genome Biol. 2017;18:182. https://doi.org/10.1186/s13059-017-1299-7.

46. Chen P, Zhang C, Ju X, Xiong Y, Xing K, Qin S. Community composition and metabolic potential of endophytic *Actinobacteria* from coastal salt marsh plants in Jiangsu, China. Front Microbiol. 2019;10:1063 https://www.frontiersin.org/article/10.3389/fmicb.2019.01063.

47. Matsumoto A, Takahashi Y. Endophytic actinomycetes: promising source of novel bioactive compounds. J Antibiot (Tokyo). 2017;70:514. https://doi.org/10.1038/ja.2017.20.

48. Gupta RS, Naushad S, Baker S. Phylogenomic analyses and molecular signatures for the class *Halobacteria* and its two major clades: a proposal for

Berkelmann *et al. Environmental Microbiome*   (2020) 15:5

Page 18 of 18

division of the class *Halobacteria* into an emended order *Halobacteriales* and two new orders, *Haloferacales* Ord. Nov. and *Natrialbales* Ord. N. Int J Syst Evol Microbiol. 2015;65:1050–69.

49. Ren M, Zhang Z, Wang X, Zhou Z, Chen D, Zeng H, et al. Diversity and contributions to nitrogen cycling and carbon fixation of soil salinity shaped microbial communities in Tarim Basin. Front Microbiol. 2018;9:431 https://www.frontiersin.org/article/10.3389/fmicb.2018.00431.

50. Maulana AF, Turjaman M, Sato T, Hashimoto Y, Cheng W, Tawaraya K. Isolation of endophytic fungi from tropical forest in Indonesia. Symbiosis. 2018;76:151–62. https://doi.org/10.1007/s13199-018-0542-7.

51. Pili NN, França SC, Kyndt T, Makumba BA, Skilton R, Höfte M, et al. Analysis of fungal endophytes associated with rice roots from irrigated and upland ecosystems in Kenya. Plant Soil. 2016;405:371–80. https://doi.org/10.1007/s11104-015-2590-6.

52. Tripathi BM, Kim M, Singh D, Lee-Cruz L, Lai-Hoe A, Ainuddin AN, et al. Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. Microb Ecol. 2012;64:474–84. https://doi.org/10.1007/s00248-012-0028-8.

53. Herold MB, Giles ME, Alexander CJ, Baggs EM, Daniell TJ. Variable response of *nirK* and *nirS* containing denitrifier communities to long-term pH manipulation and cultivation. FEMS Microbiol Lett. 2018;365. https://doi.org/10.1093/femsle/fny035.

54. Yin C, Fan F, Song A, Cui P, Li T, Liang Y. Denitrification potential under different fertilization regimes is closely coupled with changes in the denitrifying community in a black soil. Appl Microbiol Biotechnol. 2015;99:5719–29. https://doi.org/10.1007/s00253-015-6461-0.

55. Yang P, Zhang M, van Elsas JD. Role of flagella and type four pili in the co-migration of *Burkholderia terrae* BS001 with fungal hyphae through soil. Sci Rep. 2017;7:2997. https://doi.org/10.1038/s41598-017-02959-9.

56. Passos da Silva D, Schofield M, Parsek M, Tseng B. An update on the sociomicrobiology of quorum sensing in Gram-negative biofilm development. Pathogens. 2017;6:51.

57. Schwarz S, West TE, Boyer F, Chiang W-C, Carl MA, Hood RD, et al. *Burkholderia* type VI secretion systems have distinct roles in eukaryotic and bacterial cell interactions. PLoS Pathog. 2010;6:e1001068. https://doi.org/10.1371/journal.ppat.1001068.

58. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. https://doi.org/10.1093/bioinformatics/bty560.

59. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun. 2016;7:11257. https://doi.org/10.1038/ncomms11257.

60. Albertsen M, Karst SM, Ziegler AS, Kirkegaard RH, Nielsen PH. Back to basics – the influence of DNA extraction and primer choice on phylogenetic analysis of activated sludge communities. PLoS One. 2015;10:e0132783. https://doi.org/10.1371/journal.pone.0132783.

61. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2016. https://ggplot2.tidyverse.org.

62. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28:3211–7. https://doi.org/10.1093/bioinformatics/bts611.

63. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.

64. Lammel DR, Nüsslein K, Tsai SM, Cerri CC. Land use, soil and litter chemistry drive bacterial community structures in samples of the rainforest and Cerrado (Brazilian Savannah) biomes in southern Amazonia. Eur J Soil Biol. 2015;66:32–9.

65. Lammel DR, Feigl BJ, Cerri CC, Nüsslein K. Specific microbial gene abundances and soil parameters contribute to C, N, and greenhouse gas process rates after land use change in southern Amazonian soils. Front Microbiol. 2015;6:1057 https://www.frontiersin.org/article/10.3389/fmicb.2015.01057.

66. Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Appl Environ Microbiol. 2018;84 http://aem.asm.org/content/84/2/e01646-17.abstract.

67. Henckel T, Friedrich M, Conrad R. Molecular analyses of the methane-oxidizing microbial community in rice field soil by targeting the genes of the 16S rRNA, particulate methane monooxygenase, and methanol dehydrogenase. Appl Env Microbiol. 1999;65:1980–90.

68. Murrell JC, Gilbert B, McDonald IR. Molecular biology and regulation of methane monooxygenase. Arch Microbiol. 2000;173:325–32.

69. Speth DR, Orphan VJ. Metabolic marker gene mining provides insight in global *mcrA* diversity and, coupled with targeted genome reconstruction, sheds further light on metabolic potential of the Methanomassiliicoccales. PeerJ. 2018;6:–e5614. https://doi.org/10.7717/peerj.5614.

70. Aldridge P, Hughes KT. Regulation of flagellar assembly. Curr Opin Microbiol. 2002;5:160–5. https://doi.org/10.1016/S1369-5274(02)00302-8.

71. Paul K, Harmon JG, Blair DF. Mutational analysis of the flagellar rotor protein FliN: identification of surfaces important for flagellar assembly and switching. J Bacteriol. 2006;188:5240 LP–5248. https://doi.org/10.1128/JB.00110-06.

72. Scharf BE, Hynes MF, Alexandre GM. Chemotaxis signaling systems in model beneficial plant–bacteria associations. Plant Mol Biol. 2016;90:549–59. https://doi.org/10.1007/s11103-016-0432-4.

73. Le Bloa S, Durand L, Cueff-Gauchard V, Le Bars J, Taupin L, Marteau C, et al. Highlighting of quorum sensing *lux* genes and their expression in the hydrothermal vent shrimp *Rimicaris exoculata* ectosymbiontic community. Possible use as biogeographic markers. PLoS One. 2017;12:e0174338. https://doi.org/10.1371/journal.pone.0174338.

74. Goessweiner-Mohr N, Arends K, Keller W, Grohmann E. Conjugative type IV secretion systems in gram-positive bacteria. Plasmid. 2013;70:289–302. https://doi.org/10.1016/j.plasmid.2013.09.005.

75. Cascales E, Cambillau C. Structural biology of type VI secretion systems. Phil Trans R Soc B. 2012;367:1102–11.

76. Fedi S, Barberi TT, Nappi MR, Sandri F, Booth S, Turner RJ, et al. The role of *cheA* genes in swarming and swimming motility of *Pseudomonas pseudoalcaligenes* KF707. Microbes Environ. 2016;31:169–72. https://doi.org/10.1264/jsme2.ME15164.

77. Briegel A, Ortega DR, Huang AN, Oikonomou CM, Gunsalus RP, Jensen GJ. Structural conservation of chemotaxis machinery across *Archaea* and *Bacteria*. Env Microbiol Rep. 2015;7:414–9.

78. Muff TJ, Foster RM, Liu PJY, Ordal GW. CheX in the three-phosphatase system of bacterial chemotaxis. J Bacteriol. 2007;189:7007 LP–013. https://doi.org/10.1128/JB.00896-07.

79. Jones CW, Armitage JP. Positioning of bacterial chemoreceptors. Trends Microbiol. 2015;23:247–56.

80. Waters CM, Bassler BL. Quorum sensing: cell-to-cell communication in Bacteria. Annu Rev Cell Dev Biol. 2005;21:319–46. https://doi.org/10.1146/annurev.cellbio.21.012704.131001.

81. Papenfort K, Bassler BL. Quorum sensing signal–response systems in gram-negative bacteria. Nat Rev Microbiol. 2016;14:576. https://doi.org/10.1038/nrmicro.2016.89.

82. Guglielmini J, Néron B, Abby SS, Garcillán-Barcia MP, de la Cruz F, Rocha EPC. Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. Nucleic Acids Res. 2014;42:5715–27.

83. Walldén K, Williams R, Yan J, Lian PW, Wang L, Thalassinos K, et al. Structure of the *VirB4* ATPase, alone and bound to the core complex of a type IV secretion system. Proc Natl Acad Sci. 2012;109:11348 LP–11353. https://doi.org/10.1073/pnas.1201408109.

84. R Development Core Team. R: A language and environment for statistical computing. 2017. https://www.r-project.org/.

85. RStudio Team. RStudio: integrated development for R. 2015. http://www.rstudio.com/.

86. Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P. et al. vegan: Community Ecology Package. 2019;R package. https://cran.r-project.org/package=vegan.

87. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.

88. Hervé M. RVAideMemoire: Testing and plotting procedures for biostatistics. R package version 0.9–73. 2019. https://cran.r-project.org/package=RVAideMemoire. Accessed in July 2019.

Publisher's Note

Supplemental Information for Chapter 3

Contents

Additional file 1: Table S1. Sequence counts of all samples before and after quality-filtering with fastp.

Additional file 2: Figure S1. Relative abundances of each domain for each analysed land use system in the respective landscape.

Additional file 3: Figure S2. Distribution of taxonomic differences compared to rainforest at order level. Differences are displayed as log2fold (treatment vs control) changes with p.adj < 0.05 for detected orders in all land use systems. Bukit jungle rubber is not depicted as no significant changes of taxonomic orders were detected.

Additional file 4: Figure S3. Composition of different trophic protist groups. Trophic groups were assigned according to Schulz et al. Compositions are displayed for each land use in the respective landscape as relative abundances.

Additional file 5: Figure S4. Community composition at order level based on extracted rRNA sequences (A) and all obtained reads by shotgun sequencing (E). orders with relative abundances below 1% were clustered as "rare taxa". The five most abundant orders based on average relative abundance of extracted 16S rRNA gene sequences in all treatments are shown for bacteria (B), archaea (C) and fungi (D) as well as for all obtained reads by shotgun sequencing (F-H).

Additional file 6: Table S2. Environmental fit results for the entire community, bacteria, archaea and fungi. The table is deposited on the enclosed CD under \Chapter3_Supplement_information\

Additional file 7: Figure S5. Distribution of gene changes in each converted land use system compared to rainforest. Detected changes are displayed as log2fold changes (treatment vs control; p.adj < 0.05).

Additional file 8: Figure S6. Principal Component Analysis for detected genes of all analysed land use systems and rainforest samples in the respective landscape. PCA analysis is based on transformed counts by using the regularized log function of DESeq2.

Additional file 9: Figure S7. Functional profile of selected metabolisms based on KEGG categories (level 3) displayed as log2fold changes in separated landscapes. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems. Values with p.adj < 0.05 are marked with *.

Additional file 10: Figure S8. Log2fold changes of selected marker genes of nitrogen and methane metabolism for each analysed land use system compared to rainforest for each separate landscape. Significant differences (p.adj. < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 11: Figure S9. Log2fold changes of selected marker genes motility related marker genes for each analysed land use system compared to rainforest in each landscape. Significant differences (p.adj. < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 12: Figure S10. Log2fold changes of nod genes in each land use compared to rainforest. Significant differences (p.adj < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 13: Figure S11. Relative abundances of the five most abundant detected taxonomic orders within all detected sequences for nitrogen related marker genes that showed significant differences between rainforest and converted land use systems. Displayed heatmaps show the five most abundant detected taxa for nitrite reductase gene *nirK*, nitrite reductase gene *nirS* and nitrate reductase alpha subunit gene *narG*.

Additional file 14: Figure S12. Relative abundances of the five most abundant detected taxonomic orders within all detected sequences for motility related marker genes that showed significant differences between rainforest and converted land use systems.

Additional file 15: Table S3. Used abiotic measurements for all samples. The table is deposited on the enclosed CD under \Chapter3_Supplement_information\

Additional file 16: Table S4. Taxonomic counts of all analysed sequences. The first column shows ID and the last column the detected taxonomy. The table is deposited on the enclosed CD under \Chapter3_Supplement_information\

Additional file 17: Table S5. Count matrix containing functional assignments for all samples. The first column shows the ID the last columns the function according to the KEGG database including the KO number and the full pathway description. The table is deposited on the enclosed CD under \Chapter3_Supplement_information\

Additional file 1: Table S1. Sequence counts of all samples before and after quality-filtering with fastp.

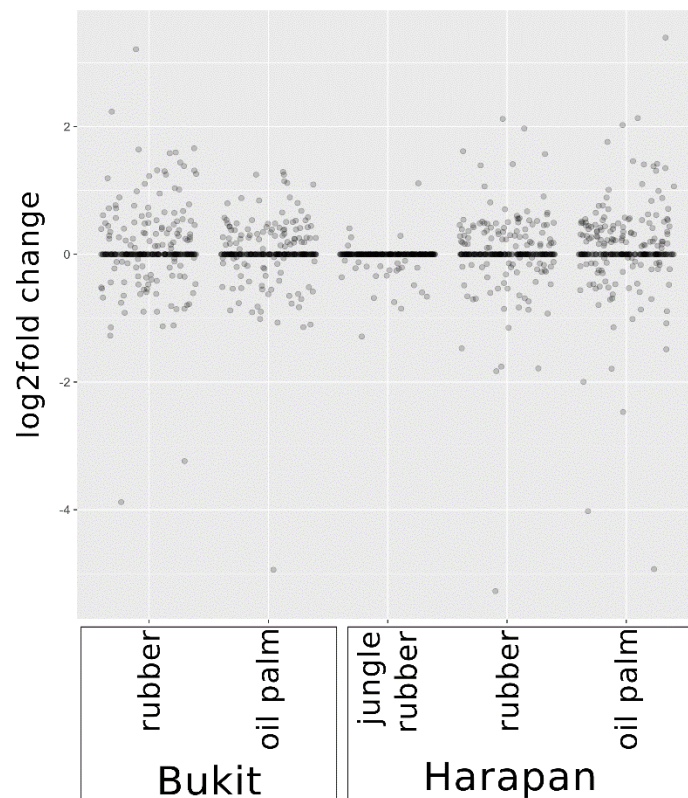| Sample	Read count (before QF)	GC content (before QF)	Read count (after QF)	GC content (after QF)	filtered [%]	remaining [%]
BF1	28225958	0,598111	26716900	0,598521	5,35	94,65
BF2	33997374	0,612474	32839798	0,612789	3,4	96,6
BF3	39824180	0,627785	38668190	0,626831	2,9	97,1
BF4	46215652	0,614638	44591542	0,614849	3,51	96,49
BJ1	28294194	0,603207	27295762	0,603591	3,53	96,47
BJ2	101699764	0,612533	98078482	0,612897	3,56	96,44
BJ3	34460134	0,608617	33426728	0,608866	3	97
BJ4	19425072	0,597543	18251772	0,597927	6,04	93,96
BO1	17259182	0,608036	16529098	0,60865	4,23	95,77
BO2	20898006	0,60109	20014412	0,601678	4,23	95,77
BO3	18954242	0,609552	18253086	0,609942	3,7	96,3
BO4	50890080	0,602904	49310800	0,603196	3,1	96,9
BR1	34871830	0,599058	33666640	0,599408	3,46	96,54
BR2	35908390	0,602978	35000026	0,603252	2,53	97,47
BR3	35317826	0,603371	34559548	0,603582	2,15	97,85
BR4	36124720	0,609015	35508276	0,609157	1,71	98,29
HF1	39877524	0,61163	38538276	0,611955	3,36	96,64
HF2	41674444	0,606638	40015006	0,606975	3,98	96,02
HF3	37939064	0,611905	36712018	0,612145	3,23	96,77
HF4	33304444	0,605236	31608576	0,605812	5,09	94,91
HJ1	47999168	0,600915	46559156	0,60119	3	97
HJ2	27810028	0,600699	26664508	0,600989	4,12	95,88
HJ3	21144922	0,589604	20035238	0,59006	5,25	94,75
HJ4	27225650	0,594439	26074146	0,594912	4,23	95,77
HO1	38716600	0,604346	37370890	0,60462	3,48	96,52
HO2	31149554	0,598075	29735514	0,598382	4,54	95,46
HO3	33041810	0,613664	31529012	0,613984	4,58	95,42
HO4	39562682	0,608788	38202384	0,608996	3,44	96,56
HR1	36957590	0,599661	35492882	0,600016	3,96	96,04
HR2	42298888	0,612423	40821394	0,612702	3,49	96,51
HR3	40017332	0,599496	38481314	0,599745	3,84	96,16
HR4	40042772	0,600104	38275410	0,600619	4,41	95,59

Additional file 2: Figure S1. Relative abundances of each domain for each analysed land use system in the respective landscape.

Additional file 3: Figure S2. Distribution of taxonomic differences compared to rainforest at order level. Differences are displayed as log2fold (treatment vs control) changes with p.adj < 0.05 for detected orders in all land use systems. Bukit jungle rubber is not depicted as no significant changes of taxonomic orders were detected.

Additional file 4: Figure S3. Composition of different trophic protist groups. Trophic groups were assigned according to Schulz et al. Compositions are displayed for each land use in the respective landscape as relative abundances.

Additional file 5: Figure S4. Community composition at order level based on extracted rRNA sequences (A) and all obtained reads by shotgun sequencing (E). orders relative abundances below 1% were clustered as "rare taxa". The five most abundant orders based on average relative abundance of extracted 16S rRNA gene sequence all treatments are shown for bacteria (B), archaea (C) and fungi (D) as well as for all obtained reads by shotgun sequencing (F-H).

Additional file 7: Figure S5. Distribution of gene changes in each converted land use system compared to rainforest. Detected changes are displayed as log2fold changes (treatment vs control; p.adj < 0.05).

Additional file 8: Figure S6. Principal Component Analysis for detected genes of all analysed land use systems and rainforest samples in the respective landscape. PCA analysis is based on transformed counts by using the regularized log function of DESeq2.

Additional file 9: Figure S7. Functional profile of selected metabolisms based on KEGG categories (level 3) displayed as log2fold changes in separated landscapes. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems. Values with p.adj < 0.05 are marked with *.

Additional file 10: Figure S8. Log2fold changes of selected marker genes of nitrogen and methane metabolism for each analysed land use system compared to rainforest for each separate landscape. Significant differences (p.adj. < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 11: Figure S9. Log2fold changes of selected marker genes motility related marker genes for each analysed land use system compared to rainforest in each landscape. Significant differences (p.adj. < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 12: Figure S10. Log2fold changes of nod genes in each land use compared to rainforest. Significant differences (p.adj < 0.05) are marked with *. Negative log2fold changes indicate higher abundances in rainforest samples, whereas positive log2fold changes indicate higher abundance in the corresponding converted land use systems.

Additional file 13: Figure S11. Relative abundances of the five most abundant detected taxonomic orders within all detected sequences for nitrogen related marker genes that showed significant differences between rainforest and converted land use systems. Displayed heatmaps show the five most abundant detected taxa for nitrite reductase gene *nirK*, nitrite reductase gene *nirS* and nitrate reductase alpha subunit gene *narG*.

Additional file 14: Figure S12. Relative abundances of the five most abundant detected taxonomic orders within all detected sequences for motility related marker genes that showed significant differences between rainforest and converted land use systems.

4

Soil bacterial community structures in relation to different oil palm management practices

Dirk Berkelmann[1], Dominik Schneider[1], Nina Hennings[2], Anja Meryandini[3] and Rolf Daniel[1]

[1]Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University, Göttingen, Germany

[2]Department of Soil Science of Temperate Ecosystems, Buesgen Institute, Georg-August-University, Göttingen, Germany

[3]Department of Biology, Faculty of Mathematics and Natural Sciences IPB, Bogor Agricultural University, Bogor, Indonesia

Author contributions:

Conceived and designed the study: RD.

Management of soil sampling and transport: AM.

Performed the experiments: DB and NH.

Analysed the data: DB and DS.

Wrote the paper: DB, DS, NH, AM and RD.

Abstract

We provide soil bacterial 16S rRNA gene amplicon and geochemical data derived from an oil palm plantation management experiment. The experimental design covered two different intensities of fertilizer application and weeding practices. We sampled the topsoil of 80 plots in total and extracted DNA and RNA. The V3-V4 regions of the 16S rRNA genes and transcripts were amplified and sequenced with an Illumina MiSeq system. Sequences were quality-filtered and taxonomy was assigned using SILVA SSU NR 138 database as reference. One year after establishing the experiment, statistically significant differences of bacterial diversity or community composition between different treatments at entire (DNA-derived) and active (RNA-derived) community level were not detected. The most dominant taxa belonged to *Acidobacteriota* and *Actinobacteriota* and were more abundant at active than at entire community level. Similar results were obtained for the abundant genera *Candidatus* Solibacter and *Haliangium*. Furthermore, clustering corresponding to the different sampling site locations was recorded. This dataset is of interest for related studies on the effect of altered management practices on soilborne communities.

Background and Summary

Palm oil is used in a variety of products, from cooking oil to biofuel with increasing global demand [1]. Thus, palm oil production has been scaled up dramatically in the last decades, leading to massive deforestation[2,3]. The biggest producer of palm oil is Indonesia, which also shows the highest rates of deforestation[4]. Due to the accompanying diversity loss, rainforest conversion and large-scale palm oil production is heavily debated and subject to research in various disciplines[5–11]. Additionally, the effects of fertilizer and herbicide applications on diversity and nutrient cycling in soil are considered as important factors for oil palm cultivation[12–15]. As bacteria mediate almost all nutrient cycling pathways in soils, several studies focussed on the impact of rainforest conversion and oil palm cultivation on soilborne bacterial communities [6,16–18]. These studies showed that certain groups, which are connected to nutrient cycling pathways, are affected by rainforest conversion to managed land use systems. These comprised the proteobacterial groups *Rhizobiales* and *Burkholderiales* as well as taxa benefitting from rainforest conversion such as *Acidobacteriales*, subgroup 2 of *Acidobacteriota* and *Streptomycetales*[5,6]. These results emphasized the effects of rainforest conversion to managed oil palm plantations, which are subjected to fertilizer, herbicide and/or mechanical weeding applications.

In this study, based on 16S rRNA gene and transcript sequencing and analysis, we provide data regarding the effects of reduced fertilizer application and mechanical weeding practices on soil bacterial communities. The analysis was performed in a state-owned oil palm plantation in Sumatra, Indonesia. Four different treatments, consisting of combinations of conventional and reduced fertilizer application as well as mechanical or herbicide-based weeding practices were analyzed (Table 1).

Table 1: Applied management practices of the four treatment types, including the provided nutrient amounts per hectare and year.

Treatment ID	Treatment	Fertilizer used	Weeding (interrow)	Liming
ch	Conventional fertilization + herbicide	260 kg N ha^{-1} yr^{-1}, 50 kg P ha^{-1} yr^{-1}, 220 kg K ha^{-1} yr^{-1}	750 cm^3 glyphosate ha^{-1} yr^{-1}	426 kg dolomite ha^{-1} yr^{-1}, 142 kg micromag ha^{-1} yr^{-1}
cw	Conventional fertilization + mechanical weeding	260 kg N ha^{-1} yr^{-1}, 50 kg P ha^{-1} yr^{-1}, 220 kg K ha^{-1} yr^{-1}	mechanical	426 kg dolomite ha^{-1} yr^{-1}, 142 kg micromag ha^{-1} yr^{-1}
rh	Reduced fertilization + herbicide	136 kg N ha^{-1} yr^{-1}, 17 kg P ha^{-1} yr^{-1}, 187 kg K ha^{-1} yr^{-1}	750 cm^3 glyphosate ha^{-1} yr^{-1}	426 kg dolomite ha^{-1} yr^{-1}, 142 kg micromag ha^{-1} yr^{-1}
rw	Reduced fertilization + mechanical weeding	136 kg N ha^{-1} yr^{-1}, 17 kg P ha^{-1} yr^{-1}, 187 kg K ha^{-1} yr^{-1}	mechanical	426 kg dolomite ha^{-1} yr^{-1}, 142 kg micromag ha^{-1} yr^{-1}

The experimental sites were established in November 2016 with four replications per treatment in short distance to each other (Figure 1). We extracted DNA and RNA from topsoil in all plots, amplified the V3-V4 region of the 16S rRNA genes and transcripts and sequenced the resulting amplicons. Additionally, soil attributes were measured for all soil samples to identify potential correlations with the corresponding soil bacterial community[19].

We obtained 6,817,019 amplicon sequences with 5,183,993 remaining sequences after quality-filtering from DNA samples and 6,412,838 raw sequences with 3,601,637 remaining sequences after quality-filtering from RNA samples[20].

We analyzed community composition, diversity and correlations to abiotic soil parameters. Most abundant phyla in the entire dataset were *Acidobacteriota* (formerly known as *Acidobacteria*), *Proteobacteria* and *Actinobacteriota* (formerly *Actinobacteria*) (Figures 2a and 2b). At order level, *Acidobacteriota* (subgroup 2), *Acidobacteriales* and *Ktedonobacterales* were the most abundant groups with no significant differences in relative abundance between the different treatments. At genus level, the most abundant taxa belonged to *Acidobaceriota* (*Candidatus* Solibacter and *Bryobacter*), *Actinobaceriota* (*Acidothermus*) and *Myxococcota* (*Haliangium*) (Figure 2b). Again, we did not detect statistically significant changes between the treatments, but notable differences between the entire (DNA-based) and potentially active

(RNA-based) community for some genera. Especially the abundant genera *Candidatus* Solibacter and *Haliangium* showed higher abundances in the active communities with relative abundance increases from 2.3 to 7.4 % (*Candidatus* Solibacter) and 0.8 to 4.8 % (*Haliangium*). In contrast, *Candidatus* Udaeobacter and HSB OF53-F07 of the *Ktedonobacterales* were among the ten most abundant genera of the entire community (approximately 2 and 1.7 %, respectively) but represented only approximately 0.2 % at active community level in all samples. In general, the detected bacterial community composition was similar to previously described communities in oil palm soils[5,6,21,22], which are mostly managed in a similar fashion with respect to conventional fertilizer and herbicide treatment[5,6,22].
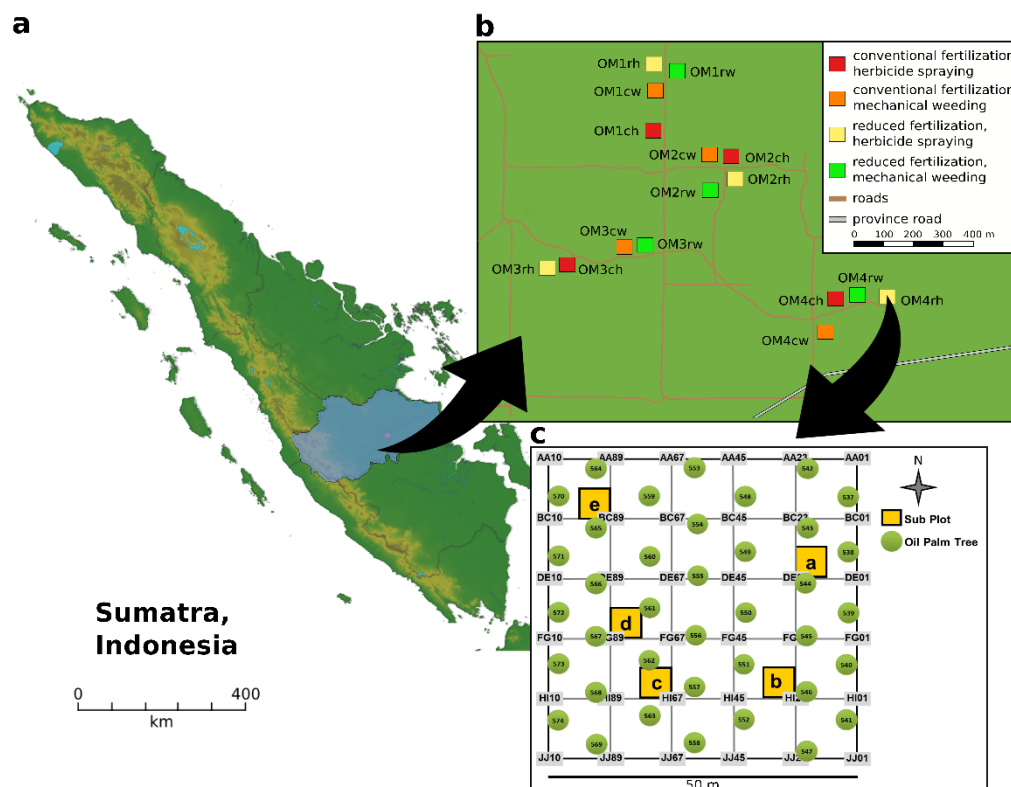
Figure 1. Sampling sites and experimental design of all sampling sites in Jambi, Indonesia. The location of the oil palm plantation in the province of Jambi is shown in respect to Jambi City (a). Squares show each plot with the respective treatment indicated by colour, with four replicates per treatment (b). An example of the experimental design is shown for each plot (c), with squares a-e indicating subplots and green circles showing planted oil palm trees.

We also analyzed the diversity at potential active and entire community level by using the Shannon index. Values ranged from 7.0 to 8.2 with no significant changes between the different treatments (Figure 3a). Ordination analysis did not show distinct clustering according to treatments but clustering due to geographical location was observed, emphasizing the importance of location over treatment. (Figure 3b). Furthermore, significant correlations

between the bacterial community and any measured abiotic factor were not recorded (all p > 0.05).
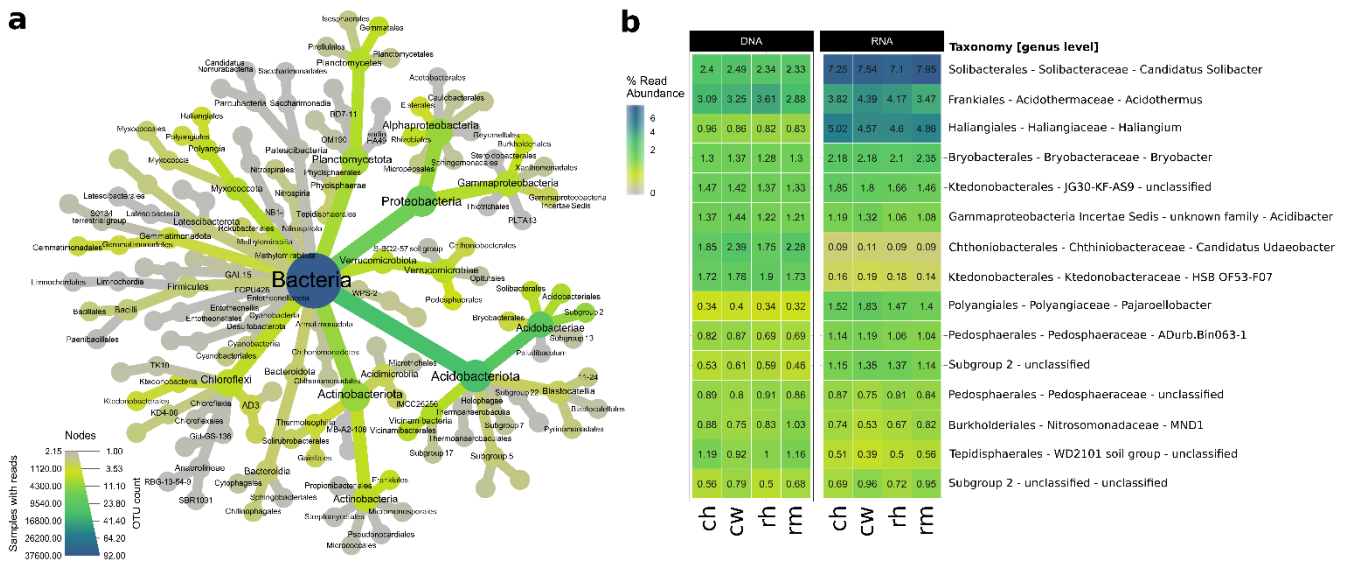
Figure 2. Community composition in all treatments at different taxonomic levels. The average community of the entire (DNA-based) and potential active (RNA-based) community at order level is shown as heat-tree (a) including all higher taxonomic levels for all used sequences. A separate visualization of the fifteen most abundant genera (b) is shown as relative abundances at genus level for the entire and the active community.
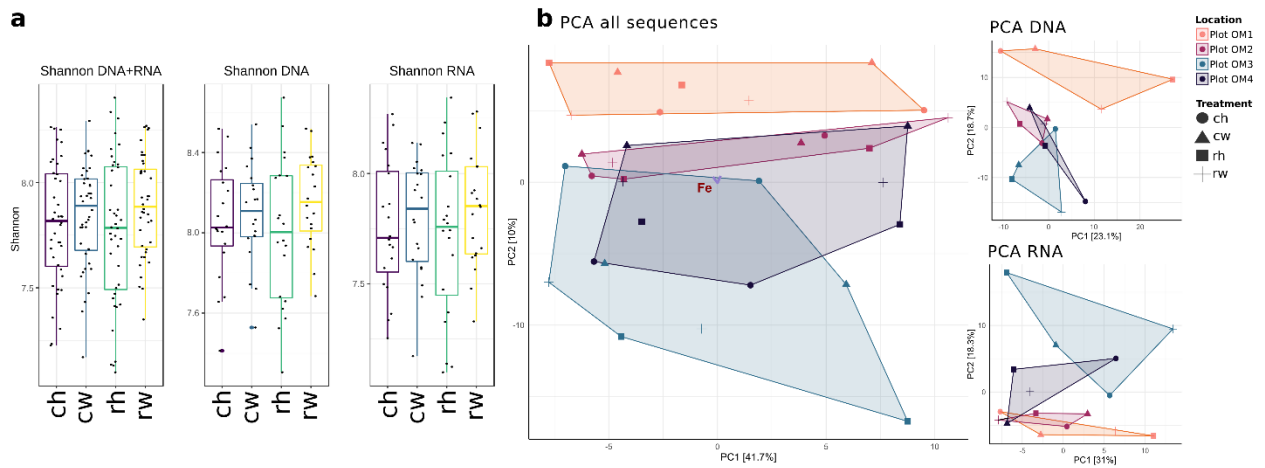
Figure 3. Detected diversity and multivariate analysis of all analyzed plots. Diversity is displayed by the Shannon index (a) for all sequences, DNA- and RNA-derived sequences. Principal Component Analysis (PCA) is shown for all sequences, DNA- and RNA-derived sequences (b). The samples were clustered at plot level, with frames and colors showing the four different plot locations and shapes for the different treatments.

Methods

Site description and soil sampling

The experiment was established as part of the EFForTS project (Ecological and socioeconomic Functions of tropical lowland rainForest Transformation Systems) in the Jambi province, located in Sumatra, Indonesia[8]. The experimental sites are located in the state-owned oil palm plantation PTPNVI, which was planted in 2002 (Figure 1). Four different locations (referred to as OM1-4) harbor four treatments, which were established in November 2016. In each of these 16 plots (50 x 50 m), five subplots were randomly established, resulting in 80 samples total.

Fertilizer treatment was conducted in two intensities: the conventional treatment usually used in the entire plantation with 260 kg nitrogen, 50 kg phosphorus and 220 kg potassium ha^{-1} yr^{-1} and reduced fertilization with 136 kg nitrogen, 17 kg phosphorous and 187 kg potassium ha^{-1} yr^{-1}. These amounts were split in half and then applied in two sessions within a year. Additionally, liming was conducted in all plots with equal amounts (426 kg dolomite and 142 kg micromag (micronutrients) ha^{-1} yr^{-1}). The herbicide treatment used 1,500 cm^3 glyphosate ha^{-1} yr^{-1} sprayed within the palm circle divided into four applications per year and 750 cm^3 glyphosate ha^{-1} yr^{-1} in inter-rows divided into two applications per year. The last application before sampling was done in April 2017. Mechanical weeding was done by cutting vegetation four times per year within the palm circle and two times per year in inter-rows with a brush cutter. The combination of these applications resulted in four different treatments: conventional fertilization with herbicide spraying (ch), conventional fertilization with mechanical weeding (cw), reduced fertilization with herbicide spraying (rh) and reduced fertilization with mechanical weeding (rw) (Table 1).

Topsoil was sampled in May 2017 from the upper seven in each subplot with a diameter of five cm. A soil corer was used to take three cores in each subplot with a distance of 1 m to each other and at least 1 m distance to trees. The three soil samples per subplot were homogenized and coarse roots and stones were removed. To prevent nucleic acids, especially RNA, from degradation RNAprotect Bacteria Reagent (Qiagen, Hilden, Germany) was applied in a ratio of 1:1. For measurements of soil parameters, we collected an additional sample, which was not supplemented with RNAprotect solution. All samples were transported in cooling boxes and stored at -80°C until further use.

Nucleic acid extraction

Frozen samples were thawed on ice. RNAprotect was removed from all samples by centrifuging for 20 min at 804.96 g and 4°C and discarding the resulting supernatant. DNA and RNA were co-extracted from 1 g of soil by using the Qiagen RNeasy PowerSoil Total RNA kit and the RNeasy PowerSoil DNA Elution kit as recommended by the manufacturer (Qiagen), except that RNA was eluted with 50 µl elution buffer instead of 100 µl. DNA contamination was removed from RNA preparations by using the TurboDNAfree kit (Applied Biosystems, Darmstadt, Germany). For this purpose, 0.1 volume DNAse buffer and 1 µl DNAse were added and incubated for 30 min at 37°C. Subsequently, a second digestion cycle was performed with 0.5 µl DNAse at 37°C for 15 min. RNA was then purified with the RNeasy MiniElute Cleanup kit (Qiagen). In order to verify complete DNA removal, a control amplification of the 16S rRNA gene was performed as described below for 16S rRNA gene amplification. Purified RNA was then reverse-transcribed into cDNA with the Superscript IV reverse transcriptase and a specific primer (5′-CCGTCAATTCMTTTGAGT-′3) as recommended by the manufacturer (Thermo Fisher Scientific, Schwerte, Germany). After cDNA synthesis, we removed residual RNA by adding 1 µl RNase H (New England Biolabs, Frankfurt am Main, Germany) to each reaction and incubation for 20 min at 37°C. Obtained DNA and cDNA were stored at -20°C until further use.

16S rRNA gene amplification and sequencing

For amplification of 16S rRNA sequences, we used 16S rRNA gene primers targeting the V3-V4 region (forward primer: 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3′, reverse primer: 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3′) as described by Klindworth[23] and added adapters for MiSeq sequencing (underlined). PCR reactions were performed in a total volume 50 µl containing 10 µl of 5-fold Phusion GC buffer, 0.2 µl 50 mM MgCl$_2$ solution, 2.5 µl DMSO, 200 µM of each of the four deoxynucleoside triphosphates and 1 U of Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific). We used 20 to 30 ng of DNA and 1 µl cDNA per reaction. The PCR reaction was started by an initial denaturation at 98°C for 1 min, followed by 25 cycles of denaturation at 98°C for 45 s, annealing at 60°C for 45 s and elongation at 72°C for 30 s. The final elongation was at 72°C for 5 minutes. Amplicons were then purified by using MagSi-NGS PREP Plus magnetic beads following the procedure recommended by the manufacturer (Steinbrenner Laborsysteme

GmbH, Wiesenbach, Germany) with the Janus Automated Workstation from Perkin Elmer (Perkin Elmer, Waltham Massachusetts, USA). Illumina MiSeq sequencing adapters were attached to the purified amplicons with the Nextera XT Index kit (Illumina, San Diego, USA). The Index PCR was done by using 5 µl of template PCR product, 2.5 µl of each index primer, 12.5 µl of 2x KAPA HiFi HotStart ReadyMix and 2.5 µl PCR grade water. Thermal cycling scheme was as follows: 95 ° C for 3 min, 8 cycles of 30 s at 95 °C, 30 s at 55 °C and 30 s at 72 °C and a final extension at 72°C for 5 min. The indexed products were purified as described before. Products were quantified by using the Quant-iT dsDNA HS assay kit and a Qubit fluorometer following the instructions of the manufacturer (Invitrogen GmbH, Karlsruhe, Germany). Purified amplicons were sequenced with an MiSeq instrument using dual indexing and reagent kit v3 (600 cycles) as recommended by the manufacturer (Illumina).

Sequence processing

Obtained paired-end sequences were first quality-filtered with fastp version 0.20[24] using a minimum phred score of 20, a minimum length of 50 bases, the default sliding window size (--cut_window_size = 4), read correction by overlap (option "correction"), adapter removal of the sequencing primers (option "adapter_fasta"), and the provided index sequences of Illumina. Quality-filtered paired-end reads were merged with PEAR version 0.9.11 and default settings[25]. Primer sequences were clipped with cutadapt version 2.5 and default settings[26]. All further steps, except mapping of sequences to ASVs (Amplicon Sequence Variant) were performed with functions implemented in vsearch version 2.1.4.1[27]. Sequences were filtered by size with "sortbylength" with a set minimum length of 300 bp. Dereplication of identical sequences was done by "derep_fulllength". Denoising and removal of low abundant sequences with less than eight replicates were done with the vsearch UNOISE3 module "cluster_unoise". Chimeric sequences were removed by employing the UCHIME module of vsearch. This included a *de novo* chimera removal ("uchime3_denovo") and a reference-based chimera removal ("uchime_ref") against the SILVA SSU 138 NR database[28]. Sequences were mapped to ASVs by vsearch ("usearch_global") with a set sequence identity threshold of 0.97. Taxonomy assignments were performed with BLASTN[29] (version 2.9.0) against the SILVA SSU 138 NR database[28] with a minimum identity threshold of 90%[30]. In addition to the taxonomy identity, we added the taxonomy id of the database, length of fragment, query percentage identity, query coverage and e-value in the taxonomy string of the table. We used identity (pident) and query coverage (qcovs) per ASV of the blast output to exclude uncertain blast hits. As recommended by the SILVA ribosomal RNA database project[31], we removed the taxonomic assignment for blast hits with

$$\frac{\text{pident} + \text{qcovs}}{2} \leq 93$$

In total, 31,987 ASVs were used for downstream analysis.

Bacterial community analysis

The bacterial community composition was further analysed in R[32] (version 3.6.1) and RStudio[33] (version 1.1.463). ASV counts were normalized by using the Geometric Mean of Pairwise Ratios (GMPR) of the GMPR package version 0.1.3[34]. Community compositions were then analyzed by the ampvis2 package version 2.4.11 and "amp_heatmap" at genus level[35]. The fifteen most abundant genera were displayed as relative abundance and clustered at treatment level. Heat-trees were displayed by the metacoder[36] package (version 0.3.2.9001).

For heat-tree calculation all counts were summed at order level and all taxa with a count-sum <1% of the lowest count-sum of all samples were excluded. The average abundance of all included taxa for all samples was calculated with:

cbind(metacoder$data$tax_abund$taxon_id, rowMeans((metacoder$data$tax_abund)[,-1])) and then added to the metacoder object with:

metacoder$data$average_taxon_abundance = average_taxon_abundance.

For diversity and ordination analysis, we used rarefaction by "amp_subset_samples" by ampvis2 as normalization of the original ASV count table. We used the Shannon diversity index as calculated by ampvis2 ("amp_alphadiv") for diversity analysis. Principal Component Analysis (PCA) was calculated with "amp_ordinate" based on Bray Curtis dissimilarity matrices. The environmental fit was calculated using the vegan package[37] with a significance threshold of $p \leq 0.05$.

Soil attribute measurements

For all abiotic measurements, soil samples were dried at 40°C for at least 10 days. We measured pH by adding the 2.25-fold volume distilled water to at least 5 g dried soil and incubate for at least 1 h prior to measurement. For C and N content determination, root fragments were manually removed, the soil was passed through a 2 mm sieve to obtain the fine soil fraction, which was ground in a ball mill (MM200, Retsch, Haan, Germany). Depending on the expected C and N contents, 5 g of soil were weighed into tin capsules. Measurements were performed by the CN analyzer vario EL cube (Elementar Analysensysteme, Hanau, Germany). Samples

were combusted at 950 °C after addition of oxygen with copper oxide as the catalyst and helium as the carrier gas. NOx gases were reduced to N_2. CO_2 and N_2 were separated by an adsorption column and were detected by a thermal conductivity detector (TCD). External certified standards of plant and soil material (IVA Analysentechnik, Meerbusch, Germany) were measured as samples for calibration validation. To account for daily variation of the room conditions and check for drifts, daily factors were determined.

Na, K, Ca, Mg, Mn, Fe, Al, S and P were measured by using an iCAP 7400 ICP-OES DUO analyser (Thermo Fisher Scientific) and standard solutions for each analyzed element (Bernd Kraft GmbH, Duisburg, Germany). Prior to measurements, 50 mg dried soil of each sample was lysed with 2 ml 65% nitric acid at 160°C for 12 h, filtered and the volume adjusted to 25 ml with water.

Data records

All obtained sequences are available at the National Center for Biotechnology Information under the Bioproject accession number PRJNA599149 and Sequence Read Archive (SRA) accession number SRP239591, containing all 160 samples (80 DNA samples and 80 RNA samples) as compressed fastq files[38]. The following files have been deposited in a figshare collection[39] and can be also accessed separately: Details regarding quality-filtering and read statistics before, during and after bioinformatic processing[20]; ASV count table with taxonomic assignments[30]; metadata information for each sample along with abiotic soil measurements[19].

Technical Validation

The sample from each subplot was derived from three different soil cores. The PCR reactions for amplification of the 16S rRNA genes transcripts and genes were performed in three technical replicates per sample. Negative controls without DNA or cDNA template were also performed. Correct amplicon size was determined on a 0.8 % agarose gel. PCR triplicates per sample were pooled in equimolar amounts for amplicon sequencing to minimize possible PCR bias.

Acknowledgements

for this study. We would like to thank the staff of the project and the local assistants who were involved in the field work. Additionally, we acknowledge the support by the Open Access Publication Funds of the University of Göttingen.

Author Contributions

RD designed and conceived the study. Management of soil sampling and field work was performed by DB and AM. DB and NH carried out the laboratory work. DB and DS prepared and analyzed the data. All authors interpreted the results, wrote the manuscript and approved submission of the manuscript.

Competing interests

The authors declare no competing financial interests.

References

1. Koh, L. P. & Wilcove, D. S. Cashing in palm oil for conservation. *Nature* **448**, 993–994 (2007).

2. Alexander, P. *et al.* Drivers for global agricultural land use change: The nexus of diet, population, yield and bioenergy. *Glob. Environ. Chang.* **35**, 138–147 (2015).

3. Oosterveer, P. Promoting sustainable palm oil: viewed from a global networks and flows perspective. *J. Clean. Prod.* **107**, 146–153 (2015).

4. Margono, B. A., Potapov, P. V, Turubanova, S., Stolle, F. & Hansen, M. C. Primary forest cover loss in Indonesia over 2000–2012. *Nat. Clim. Chang.* **4**, 730 (2014).

5. Berkelmann, D. *et al.* How rainforest conversion to agricultural systems in Sumatra (Indonesia) affects active soil bacterial communities. *Front. Microbiol.* **9**, 2381 (2018).

6. Schneider, D. *et al.* Impact of lowland rainforest transformation on diversity and composition of soil prokaryotic communities in Sumatra (Indonesia). *Front. Microbiol.* **6**, 1339 (2015).

7. Brinkmann, N. *et al.* Intensive tropical land use massively shifts soil fungal communities. *Sci. Rep.* **9**, 3403 (2019).

8. Drescher, J. *et al.* Ecological and socio-economic functions across tropical land use systems after rainforest conversion. *Philos. Trans. R. Soc. B Biol. Sci.* **19**, 371 (2016).

9. Corre, M. D., Dechert, G. & Veldkamp, E. Soil nitrogen cycling following montane forest conversion in central Sulawesi, Indonesia. *Soil Sci. Soc. Am. J.* **70**, 359–366 (2006).

10. Schulz, G. *et al.* Changes in trophic groups of protists with conversion of rainforest into rubber and oil palm plantations. *Frontiers in Microbiology* **10**, 240 (2019).

11. Krashevska, V., Klarner, B., Widyastuti, R., Maraun, M. & Scheu, S. Impact of tropical lowland rainforest conversion into rubber and oil palm plantations on soil microbial communities. *Biol. Fertil. Soils* **51**, 697–705 (2015).

12. Lynch, D. H. Nutrient cycling and soil health in organic cropping systems-Importance of management strategies and soil resilience. *Sustain. Agric. Res.* **4**, 80–88 (2015).

13. Bhardwaj, D., Ansari, M. W., Sahoo, R. K. & Tuteja, N. Biofertilizers function as key player in sustainable agriculture by improving soil fertility, plant tolerance and crop productivity. *Microb. Cell Fact.* **13** (2014).

14. Newman, M. M. *et al.* Glyphosate effects on soil rhizosphere-associated bacterial communities. *Sci. Total Environ.* **543**, 155–160 (2016).

15. Darras, K. F. A. *et al.* Reducing Fertilizer and Avoiding Herbicides in Oil Palm Plantations—Ecological and Economic Valuations. *Front. For. Glob. Change* **2**, 65 (2019).

16. Tripathi, B. M. *et al.* Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. *Microb. Ecol.* **64**, 474–484 (2012).

17. Paula, F. S. *et al.* Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities. *Mol. Ecol.* **23**, 2988–2999 (2014).

18. Kroeger, M. E. *et al.* New biological insights into how deforestation in Amazonia affects soil microbial communities using metagenomics and metagenome-assembled genomes. *Front. Microbiol.* **9**, 1635 (2018).

19. Berkelmann, D., Schneider, D., Hennings, N., Meryandini, A. & Daniel, R. Metadata, diversity and abiotic measurements of each sample. *figshare,* https://doi.org/10.6084/m9.figshare.12237707.v1 (2020).

20. Berkelmann, D., Schneider, D., Hennings, N., Meryandini, A. & Daniel, R. Sequence statistics regarding quality-filtering of obtained sequences. *figshare,* https://doi.org/10.6084/m9.figshare.12238031.v1 (2020).

21. Berkelmann, D., Schneider, D., Meryandini, A. & Daniel, R. Unravelling the effects of tropical land use conversion on the soil microbiome. *Environ. Microbiome* **15**, 5 (2020).

22. Lee-Cruz, L., Edwards, D. P., Tripathi, B. M. & Adams, J. M. Impact of logging and forest conversion to oil palm plantations on soil bacterial communities in Borneo. *Appl. Environ. Microbiol.* **79**, 7290–7297 (2013).

23. Klindworth, A. *et al.* Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **41**, 1–11 (2013).

24. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

25. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).

26. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).

27. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584–e2584 (2016).

28. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

30. Berkelmann, D., Schneider, D., Hennings, N., Meryandini, A. & Daniel, R. ASV table containing counts and taxonomic identity of all detected ASVs in each sample. *figshare,* https://doi.org/10.6084/m9.figshare.12238049.v1 (2020).

31. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).

32. R Development Core Team. R: A language and environment for statistical computing. (2017).

33. RStudio Team. RStudio: Integrated Development for R. (2015).

34. Chen, L. *et al.* GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. *PeerJ* **6**, e4600–e4600 (2018).

35. Andersen, K. S., Kirkegaard, R. H., Karst, S. M. & Albertsen, M. ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. *bioRxiv* 299537 (2018). doi:10.1101/299537

36. Foster, Z. S. L., Sharpton, T. J. & Grünwald, N. J. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLOS Comput. Biol.* **13**, e1005404 (2017).

37. Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Simpson, G. L., Solymos, P., et al. vegan: Community Ecology Package. **R package**, (2019).

38. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/SRP239591 (2020).

39. Berkelmann, D., Schneider, D., Hennings, N., Meryandini, A. & Daniel, R. Soil bacterial community structures in relation to different oil palm management practices. *figshare*, https://doi.org/10.6084/m9.figshare.c.4962227.v1 (2020).

Supplemental Information for Chapter 4

Contents

Supplementary table 1. Metadata, diversity and abiotic measurements of each sample. The table is deposited on figshare and referenced in the text (reference nr. 19). It is also deposited on the enclosed CD under \Chapter4_Supplement_information\

Supplementary table 2. Sequence statistics regarding quality-filtering of obtained sequences. The table is also deposited on figshare and referenced in the text (reference nr. 20).

Supplementary table 3. ASV table containing counts and taxonomic identity of all detected ASVs in each sample. The table is deposited on figshare and referenced in the text (reference nr. 30). It is also deposited on the enclosed CD under \Chapter4_Supplement_information\

Supplementary table 2. Sequence statistics regarding quality-filtering of obtained sequences. The table is also deposited on figshare and referenced in the text (reference nr. 20).

SampleID	Raw reads	After quality-filtering	After taxonomy filter	Discarded reads	Discarded reads [%]
OM1cha	91738	72031	71223	20515	22.36
OM1chb	81255	64114	63579	17676	21.75
OM1chc	70624	55798	55707	14917	21.12
OM1chd	90760	71917	71412	19348	21.32
OM1che	106077	84073	82896	23181	21.85
OM1cwa	111776	84036	83470	28306	25.32
OM1cwb	96856	73657	73161	23695	24.46
OM1cwc	166989	128473	127507	39482	23.64
OM1cwd	71268	55816	55422	15846	22.23
OM1cwe	44422	32784	32632	11790	26.54
OM1rha	136489	93998	93293	43196	31.65
OM1rhb	74616	56500	56206	18410	24.67
OM1rhc	129903	93075	92443	37460	28.84
OM1rhd	78737	60292	59784	18953	24.07
OM1rhe	105525	79603	78819	26706	25.31
OM1rwa	87572	63938	63493	24079	27.5
OM1rwb	90748	66524	66130	24618	27.13
OM1rwc	151753	100826	99072	52681	34.71
OM1rwd	62676	46070	45847	16829	26.85
OM1rwe	79552	58269	57949	21603	27.16
OM2cha	49064	39108	38961	10103	20.59
OM2chb	72346	55749	55503	16843	23.28
OM2chc	65267	50727	50207	15060	23.07
OM2chd	50342	38519	38312	12030	23.9
OM2che	60788	45626	45416	15372	25.29
OM2cwa	57102	43854	43646	13456	23.56
OM2cwb	90810	69726	68728	22082	24.32
OM2cwc	60166	46943	46718	13448	22.35
OM2cwd	46688	37085	36910	9778	20.94
OM2cwe	76971	56998	56624	20347	26.43
OM2rha	266197	203150	201707	64490	24.23
OM2rhb	138090	106599	105939	32151	23.28
OM2rhc	59195	46833	46593	12602	21.29
OM2rhd	54711	40807	40439	14272	26.09
OM2rhe	122473	96266	94739	27734	22.64
OM2rwa	69302	55046	54669	14633	21.11
OM2rwb	79990	63166	63032	16958	21.2
OM2rwc	94814	75679	74952	19862	20.95
OM2rwd	134025	106716	105220	28805	21.49
OM2rwe	160207	129239	128390	31817	19.86
OM3cha	55095	41212	40978	14117	25.62
OM3chb	45724	35544	35166	10558	23.09
OM3chc	57628	45502	45442	12186	21.15
OM3chd	123129	97873	97617	25512	20.72
OM3che	74900	59506	58313	16587	22.15
OM3cwa	65707	51425	50543	15164	23.08
OM3cwb	49168	38119	37584	11584	23.56
OM3cwc	58918	45212	44893	14025	23.8

OM3cwd	74557	57734	57655	16902	22.67
OM3cwe	54664	42648	42247	12417	22.72
OM3rha	71335	54629	54278	17057	23.91
OM3rhb	62537	47813	47337	15200	24.31
OM3rhc	97425	75954	75043	22382	22.97
OM3rhd	51926	41002	40771	11155	21.48
OM3rhe	69292	54564	54365	14927	21.54
OM3rwa	88271	68974	67987	20284	22.98
OM3rwb	127018	97582	97157	29861	23.51
OM3rwc	95191	71108	70433	24758	26.01
OM3rwd	128745	98097	97450	31295	24.31
OM3rwe	46805	35644	35600	11205	23.94
OM4cha	38821	29248	29190	9631	24.81
OM4chb	107244	79511	79289	27955	26.07
OM4chc	63101	49720	49354	13747	21.79
OM4chd	135609	102952	102377	33232	24.51
OM4che	110472	85138	84017	26455	23.95
OM4cwa	83381	65012	64775	18606	22.31
OM4cwb	133836	104963	103305	30531	22.81
OM4cwc	74645	58287	58070	16575	22.21
OM4cwd	63448	49553	49071	14377	22.66
OM4cwe	57280	42782	42455	14825	25.88
OM4rha	56505	44890	44740	11765	20.82
OM4rhb	131128	99137	98905	32223	24.57
OM4rhc	69144	54262	54009	15135	21.89
OM4rhd	49731	38409	38269	11462	23.05
OM4rhe	84761	64515	64192	20569	24.27
OM4rwa	61339	43572	43405	17934	29.24
OM4rwb	80144	61905	60783	19361	24.16
OM4rwc	55277	43175	42871	12406	22.44
OM4rwd	66121	51604	51326	14795	22.38
OM4rwe	59113	46173	45981	13132	22.22
r_OM1cha	52889	33898	33792	19097	36.11
r_OM1chb	82813	51360	51253	31560	38.11
r_OM1chc	102773	61029	60982	41791	40.66
r_OM1chd	51806	29370	29324	22482	43.4
r_OM1che	68001	38265	38165	29836	43.88
r_OM1cwa	29081	13141	13113	15968	54.91
r_OM1cwb	120056	69624	69535	50521	42.08
r_OM1cwc	60998	36121	36014	24984	40.96
r_OM1cwd	102424	61947	61834	40590	39.63
r_OM1cwe	60026	35369	35340	24686	41.13
r_OM1rha	85013	40690	40598	44415	52.24
r_OM1rhb	69639	38978	38838	30801	44.23
r_OM1rhc	112051	51979	51779	60272	53.79
r_OM1rhd	69202	38740	38514	30688	44.35
r_OM1rhe	69883	35623	35396	34487	49.35
r_OM1rwa	61125	30258	30119	31006	50.73
r_OM1rwb	75548	36106	35963	39585	52.4
r_OM1rwc	72044	31092	30886	41158	57.13
r_OM1rwd	83879	44495	44350	39529	47.13
r_OM1rwe	56383	32941	32869	23514	41.7
r_OM2cha	104902	57732	57589	47313	45.1
r_OM2chb	52063	26174	26131	25932	49.81

r_OM2chc	140514	80254	79996	60518	43.07
r_OM2chd	74899	42311	42197	32702	43.66
r_OM2che	90855	52410	52333	38522	42.4
r_OM2cwa	73296	44672	44572	28724	39.19
r_OM2cwb	68145	40678	40570	27575	40.47
r_OM2cwc	45040	24905	24790	20250	44.96
r_OM2cwd	110870	61146	61016	49854	44.97
r_OM2cwe	61223	32399	32317	28906	47.21
r_OM2rha	58322	33485	33368	24954	42.79
r_OM2rhb	57539	31717	31673	25866	44.95
r_OM2rhc	52559	30331	30205	22354	42.53
r_OM2rhd	44325	22581	22500	21825	49.24
r_OM2rhe	118777	68559	68170	50607	42.61
r_OM2rwa	78233	46221	46111	32122	41.06
r_OM2rwb	105304	58868	58818	46486	44.14
r_OM2rwc	85325	50463	50331	34994	41.01
r_OM2rwd	73412	49515	49295	24117	32.85
r_OM2rwe	66371	39164	38985	27386	41.26
r_OM3cha	134074	77359	77030	57044	42.55
r_OM3chb	86554	51624	51303	35251	40.73
r_OM3chc	66794	39089	39059	27735	41.52
r_OM3chd	76907	46163	46102	30805	40.05
r_OM3che	62095	36083	35903	26192	42.18
r_OM3cwa	71777	39474	39122	32655	45.5
r_OM3cwb	136339	75020	74629	61710	45.26
r_OM3cwc	70733	40166	40065	30668	43.36
r_OM3cwd	88791	49695	49622	39169	44.11
r_OM3cwe	78387	47088	46968	31419	40.08
r_OM3rha	131672	72751	72546	59126	44.9
r_OM3rhb	86173	50412	50138	36035	41.82
r_OM3rhc	78740	47617	47466	31274	39.72
r_OM3rhd	83725	50908	50735	32990	39.4
r_OM3rhe	66001	39650	39576	26425	40.04
r_OM3rwa	56543	32315	32162	24381	43.12
r_OM3rwb	79775	45526	45401	34374	43.09
r_OM3rwc	61755	33154	32912	28843	46.71
r_OM3rwd	110162	63177	62884	47278	42.92
r_OM3rwe	78715	41953	41907	36808	46.76
r_OM4cha	89880	46507	46418	43462	48.36
r_OM4chb	57719	30282	30214	27505	47.65
r_OM4chc	134088	76733	76021	58067	43.31
r_OM4chd	99667	52223	52037	47630	47.79
r_OM4che	104638	66965	66792	37846	36.17
r_OM4cwa	73975	39995	39925	34050	46.03
r_OM4cwb	61164	35431	35323	25841	42.25
r_OM4cwc	61296	31491	31469	29827	48.66
r_OM4cwd	96027	53911	53804	42223	43.97
r_OM4cwe	64202	31801	31685	32517	50.65
r_OM4rha	83377	50495	50419	32958	39.53
r_OM4rhb	70125	37824	37788	32337	46.11
r_OM4rhc	75123	43252	43086	32037	42.65
r_OM4rhd	43002	24970	24898	18104	42.1
r_OM4rhe	151822	84982	84773	67049	44.16
r_OM4rwa	78123	38890	38822	39301	50.31

r_OM4rwb	94488	55329	55138	39350	41.65
r_OM4rwc	87048	51363	51159	35889	41.23
r_OM4rwd	61270	36067	35956	25314	41.32
r_OM4rwe	72484	40834	40749	31735	43.78

5

Agroforestry management systems drive the composition, diversity, and function of fungal and bacterial endophyte communities in Theobroma cacao leaves

Franziska Wemheuer[1], Dirk Berkelmann[2], Bernd Wemheuer[2], Rolf Daniel[2], Stefan Vidal[1], Hervé Bertin Bisseleua Daghela[1,3]

[1]Section of Agricultural Entomology, Department of Crop Sciences, University of Göttingen, Grisebachstr. 6, D-37077 Göttingen, Germany

[2]Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, University of Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany

[3]Laboratory of Entomology, Institute of Agricultural Research for Development (IRAD), BP 2067, Yaoundé, Cameroon

microorganisms

Agroforestry Management Systems Drive the Composition, Diversity, and Function of Fungal and Bacterial Endophyte Communities in Theobroma Cacao Leaves

Franziska Wemheuer [1], **Dirk Berkelmann** [2], **Bernd Wemheuer** [2], **Rolf Daniel** [2],
Stefan Vidal [1,*] **and Hervé Bertin Bisseleua Daghela** [1,3]

[1] Section of Agricultural Entomology, Department of Crop Sciences, University of Göttingen, Grisebachstr. 6, D-37077 Göttingen, Germany; fwemheu@gwdg.de (F.W.); hbissel@gmail.com (H.B.B.D.)

[2] Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, University of Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany; dberkel@gwdg.de (D.B.); bwemheu@gwdg.de (B.W.); rdaniel@gwdg.de (R.D.)

[3] Laboratory of Entomology, Institute of Agricultural Research for Development (IRAD), BP 2067, Yaoundé, Cameroon

[*] Correspondence: svidal@gwdg.de; Tel.: +49-551-39-23732

Abstract: Cacao (*Theobroma cacao* L.) is one of the most economically important crops worldwide. Despite the important role of endophytes for plant growth and health, very little is known about the effect of agroforestry management systems on the endophyte communities of *T. cacao*. To close this knowledge gap, we investigated the diversity, community composition, and function of bacterial and fungal endophytes in the leaves of *T. cacao* trees growing in five major cacao-growing regions in the central region of Cameroon using DNA metabarcoding. Fungal but not bacterial alpha diversity measures differed significantly between the agroforestry management systems. Interestingly, less managed home-garden cacao forests harbored the lowest fungal richness and diversity. Our results suggest that the composition of bacterial and fungal endophyte communities is predominantly affected by agroforestry management systems and, to a lesser extent, by environmental properties. The core microbiome detected comprised important fungal phytopathogens, such as *Lasiodiplodia* species. Several predicted pathways of bacterial endophytes and functional guilds of fungal endophytes differed between the agroforest systems which might be attributed to bacteria and fungi specifically associated with a single agroforest. Our results provide the basis for future studies on foliar fungal and bacterial endophytes of *T. cacao* and their responsiveness towards agroforestry management systems.

Keywords: microbial diversity; endophytes; core microbiome; metabarcoding; agroforestry management systems; functional predictions; *Theobroma cacao*

1. Introduction

Cacao (*Theobroma cacao* L.) is one of the most economically important crops worldwide. To accommodate the increasing global demand, cocoa production has increased to approximately 5.2 million tons in 2017 (FAO Statistical Database; http://faostat.fao.org). Cacao trees have been traditionally established under thinned canopies of primary or old secondary forests [1]. To enhance their short-term income, farmers in many parts of the world have converted these shaded cacao systems into non-shaded, high intensive monocultures [1,2]. The conversion of tropical rainforests and agricultural homogenization, however, causes severe problems such as biodiversity loss as well as

an increased risk of pest outbreaks [3–5]. Cacao agroforestry management systems (AMSs), such as traditional forest gardens, may help to alleviate disease and pest problems [6,7]. Cacao AMSs include numerous cultivated plants (e.g., cocoa and bananas) and natural forest tree species. Additionally, they provide a wide range of benefits, including livelihoods for farmers, as well as the conservation of natural resources [1,8]. Consequently, the relationship among different cacao AMSs and their role in maintaining biodiversity has received more attention during the last years [1,4,9].

Endophytic fungi and bacteria have been found in all plant species investigated to date [10]. An increasing number of studies has assessed the community composition [11–14] and the diverse functional effects of endophytes on their host plants [15,16]. Beneficial endophytes have been reported to promote plant nutrition acquisition and growth [16,17]. Moreover, endophytes may enhance the resistance of their host plants to plant pathogens [18–20]. This is especially important due to the wide range of fungal pathogens attacking cacao trees [21,22]. Several of these pathogens are responsible for severe yield losses [22,23]. As a consequence, the potential of endophytic fungi and bacteria as biocontrol agents of important cacao pathogens such as *Moniliophthora roreri* or *Phytophthora capsici* has been evaluated [12,15,20]. For instance, several endospore-forming bacterial endophytes isolated from *T. cacao* inhibited the cacao pathogens *M. roreri*, *M. perniciosa*, and *P. capsica* in antagonism studies [20]. In addition, eight isolates significant inhibited *P. capsici* lesion formation in detached leaf assays compared to untreated control leaves.

Given the high ecological and economic relevance of bacterial and fungal endophytes, it is crucial to decipher endophyte communities in economically important plant species and their influencing factors. Recent studies have shown that agricultural practices such as cropping system, fertilizer, or fungicide application influenced fungal and bacterial endophyte diversity and/or community structures [24–26]. In a previous study on fungal endophyte communities of *Coffea arabica*, region and AMS significantly influenced endophytic communities [13]. To date, it is unclear whether foliar fungal and bacterial endophytes of *T. cacao* trees differ among different AMSs as well.

The goals of the current study were to fill this knowledge gap and to obtain first insights into functional and compositional changes in fungal and bacterial endophytes in leaves of *T. cacao* growing in different AMSs. Furthermore, we aimed to identify fungal and bacterial taxa that were responsive to AMSs. Our study was conducted in five major cacao-growing regions (Obala, Boumnyébel, Bakao, Talba, Kédia) in the Central Region of Cameroon. The regions differed in their AMSs ranging from less extensive to more intensively managed cacao agroforests. Foliar bacterial and fungal endophyte communities were investigated by high-throughput Illumina (MiSeq) sequencing targeting the bacterial 16S rRNA gene and the fungal internal transcribed spacer (ITS), respectively. Additionally, we evaluated the agricultural practices and environmental properties shaping bacterial and fungal communities. To better understand plant–endophyte interactions with respect to AMSs, correlation-based indicator species analyses were performed. Moreover, functional profiles were predicted from obtained 16S rRNA data, and fungal community members were classified with respect to functional guilds. We hypothesized that the AMSs and the prevailing environmental conditions impacted microbial colonization of *T. cacao* leaves and, consequently, microbial community composition.

2. Materials and Methods

2.1. Study Site and Experimental Design

The study was conducted in five major cacao-growing regions (Obala, Boumnyébel, Bakao, Talba, Kédia) in the humid dense forest region in the Central Region of Cameroon between 4°12′ and 4°30′ N and 10°6′ and 11°15′ E (Table 1, Supplementary Table S1). The altitude varied between 450 and 715 m above sea level. Within the growing regions, 20 cacao plantations in seven sites were selected due to differences in AMSs (Table 1, Supplementary Table S1). These systems were grouped as follows: (1) less managed, home-garden cacao forest (Obala; approximately 30 years old), (2) extensively managed old traditional cacao forest garden (Boumnyébel; approximately 50 years old); (3) extensively managed

young traditional cacao forest garden (Bakao; approximately 30 years old); (4) the most intensively managed mature traditional cacao forest garden (Talba; 15–20 years old); and (5) intensively managed young traditional cacao forest garden (Kédia; 8–15 years old). In Obala (sites Nkolobang and Ekabita Essélé), cocoa is grown near houses with a high variety of fruit trees species. In Boumnyébel (sites Pan Makak and Simanya), cocoa is grown under a dense cover of shade tree species. The cacao trees in Talba and Bakao are grown in larger farms and on modified savannah agroecosystems, respectively. In Kédia, cocoa is grown under full sunlight in two villages. For further details on the growing regions, see Bisseleua et al. 2009 [4] and Bisseleua et al. 2008 [7].

Table 1. Landscape characteristics of the five regions Boumnyébel (sites Pan Makak, Simanya), Talba, Bakoa, Kédia and Obala (sites Ekabita Essélé, Nkolobang). See Supplementary Table S1 for further details.

Region	Plantation	Trees	Agroforestry Management	Agricultural Land	Forest Land
Boumnyébel	Pan Makak 1 Pan Makak 2 Simanya 1 Simanya 2	4 4 4 4	extensively managed old (cocoa is grown under a dense cover of shade tree species)	20% cocoa fields, 10% annual crop	70% pristine forest, with forest reserve
Talba	Talba 1 Talba 2 Talba 3 Talba 4	4 4 4 4	intensively managed manure (cocoa is grown in larger farms)	70% cocoa fields, 5% annual field crops	25% pristine forest, no reserve
Bakoa	Bakoa 1 Bakoa 2 Bakoa 3 Bakoa 4	2 2 2 2	extensively managed young (cocoa is grown on modified savannah agroecosystems)	50% cocoa fields, 25% annual field crops, 5% patchy pasture fields	20% secondary forest, no reserve
Kédia	Kédia 1 Kédia 2 Kédia 3 Kédia 4	2 2 2 2	intensively managed young (cocoa is grown under full sun)	65% cocoa fields, 25% annual field crops, 5% pasture lands	5% secondary forest
Obala	Ekabita Essélé 1 Ekabita Essélé 2 Nkolobang 1 Nkolobang 2	4 4 4 4	home garden cacao forest (cocoa is grown nearby houses with a high variety of fruit tree species)	70% cocoa fields, 25% annual crop fields of mixed crops, agroforestry trees	5% secondary forest, no forest reserve

2.2. Sampling

At least two trees from each cacao plantation were randomly selected for sampling, resulting in eight (Bakao, Kédia) or sixteen trees (Boumnyébel, Obala, Talba) per site (Table 1, Supplementary Table S1). These trees were chosen based on the following criteria: healthy appearance, and overall good physiological and nutritional state. For each tree, three mature leaves with the same age were collected between September and October 2014. The collected leaves showed no obvious disease symptoms, including leaf spots, chlorosis, or other types of pathogen-induced lesions. They were immediately cooled down (below 4 °C) and transported to the laboratory. A total of 25 discs (50 mm^2) per leaf were stamped out from the apical, middle, and basal region of the leaf. The 75 leaf discs of each tree were pooled prior to DNA extraction. A total of 64 *T. cacao* leaf samples were investigated in this study (Supplementary Material Table S1).

2.3. Surface Sterilization and DNA Extraction

Different protocols for surface sterilization of leaf discs were tested using varying incubation times. The protocol giving the best surface sterilization success (no microbial growth for all replicates on the three different media types) was used for further analyses. The final surface sterilization protocol included the following steps: consecutive washing in 70% ethanol for 20 s, 2% sodium hypochlorite for 30 s, 70% ethanol for 20 s, followed by three times immersion in sterilized, distilled diethyl pyrocarbonate (DEPC)-treated water for 10 s. Surface-sterilized leaf discs were subsequently dried on tissue paper. Fresh solutions and separate, sterile collection tubes were used for each sample

to avoid cross-contaminations. The surface sterilization of leaf discs was controlled for effectiveness as described previously [27] by placing 5–10 sterile leaf discs per location and 50 µL aliquots of the last washing step onto common laboratory agar plates (malt extract agar (MEA), Luria-Bertani-Agar (LB) and potato dextrose agar (PDA)). The plates were incubated in the dark at approximately 25 °C for at least 3 weeks. No growth of microorganisms was observed. In addition, water from the final washing step was subjected to polymerase chain reaction (PCR) targeting the ITS region and the bacterial 16S rRNA gene. No amplification was detected. These results confirmed that the surface sterilization of all leaf discs was successful in eliminating non-cultivable and cultivable fungal and bacterial epiphytes as well as potential DNA traces from the leaf surfaces.

Total DNA was extracted, employing the peqGOLD Plant DNA Mini kit (Peqlab, Erlangen, Germany; now VWR) according to the manufacturer's instructions with two modifications described previously [27]. Briefly, all surface-sterilized leaf discs were incubated in lysis buffer at 32 °C for 12 h and subsequently homogenized using ethanol-sterilized pestles. The concentration of DNA extracts was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, USA). In total, the DNA of 64 *T. cacao* leaf samples was subjected to PCR targeting the bacterial 16S rRNA gene and the fungal ITS region.

2.4. Amplification and Sequencing of 16S rRNA Genes

Bacterial 16S rRNA genes were amplified using the forward primer S-D-Bact-0341-b-S-17 (5′-CCT ACG GGN GGC WGC AG-3′; [28]) and the reverse primer S-D-Bact-0785-a-A-21 (5′-GAC TAC HVG GGT ATC TAA TCC-3′; [28]) containing Illumina Nextera adapters for sequencing. The PCR reaction (25 µL) contained 5 µL of five-fold Phusion HF buffer, 200 µM of each of the four deoxynucleoside triphosphates, 4 µM of each primer, 1 U of Phusion high fidelity DNA polymerase (Thermo Scientific, Waltham, MA, USA), and approximately 50 ng of the extracted DNA as a template. The negative controls were performed by using the reaction mixture without a template. The following thermal cycling scheme was used: initial denaturation at 98 °C for 30 s, 30 cycles of denaturation at 98 °C for 15 s, annealing at 53 °C for 30 s, followed by extension at 72 °C for 30 s. The final extension was carried out at 72 °C for 2 min. Each sample was subjected to three independent amplifications. Obtained PCR products per sample were controlled for appropriate size, pooled in equal amounts, and purified using the peqGOLD Gel Extraction kit (Peqlab). The quantification of the PCR products was performed using the Quant-iT dsDNA HS assay kit and a Qubit fluorometer, as recommended by the manufacturer (Thermo Scientific). The DNA samples were barcoded using the Nextera XT-Index kit (Illumina, San Diego, USA) and the Kapa HIFI Hot Start polymerase (Kapa Biosystems, USA). Sequencing was performed at the Göttingen Genomics Laboratory on an Illumina MiSeq Sequencing platform (paired end 2 × 300 bp) using the MiSeq Reagent kit v3, as recommended by the manufacturer (Illumina). All bacterial samples were sequenced on the same MiSeq run.

2.5. Amplification and Sequencing of the Fungal ITS Region

The fungal endophyte community was assessed by a nested PCR approach targeting the ITS region, as described previously [24,29]. In the first PCR, the primers ITS1-F_KYO2 (5′-TAG AGG AAG TAA AAG TCG TAA-3′) [30] and ITS4 (5′- TCC TCC GCT TAT TGA TAT GC-3′) [31] were used to suppress the co-amplification of plant-derived ITS regions. Obtained PCR products were subjected to nested PCR with the primer pair ITS3_KYO2 [30] and ITS4 [31] containing the MiSeq adaptors (underlined): MiSeq-ITS3_KYO2 (5′-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GAT GAA GAA CGY AGY RAA-3′) and MiSeq-ITS4 (5′-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTC CTC CGC TTA TTG ATA TGC -3′). The PCR mixture (25 µL) contained: 5 µL of 5-fold Phusion GC buffer, 200 µM of each of the four deoxynucleoside triphosphates, 4 µM of each primer, 5% DMSO, 25 mM MgCl$_2$, 0.5 U of Phusion High Fidelity DNA polymerase (Thermo Scientific), and approximately 10 ng DNA and PCR product from the first PCR as a template, respectively. The negative controls were performed

using the reaction mixture without a template. The following thermal cycle scheme was utilized: initial denaturation at 98 °C for 30 s followed by 6 cycles of denaturation at 98 °C for 15 s, annealing at 53 °C for 30 s decreasing 0.5 °C in each cycle, followed by extension at 72 °C for 30 s and 29 cycles of denaturation at 98 °C for 15 s, annealing at 50 °C for 30 s, followed by extension at 72 °C for 30 s. The final extension was carried out at 72 °C for 2 min. Each sample was subjected to three independent amplifications. The negative controls were performed using the reaction mixture without template. Obtained PCR products were pooled in equal amounts, purified, and quantified as described for bacterial PCR products. The barcoding of purified fungal PCR products as well as sequencing were performed as described above for bacterial PCR products. All fungal samples were sequenced on the same MiSeq run.

2.6. Processing of Bacterial and Fungal Datasets

Obtained sequencing data were initially quality-filtered with the Trimmomatic tool version 0.36 [32]. Low-quality reads were truncated if the quality dropped below 12 in a sliding window of 4 bp. Subsequently, all reads shorter than 100 bp and orphan (unpaired) reads were removed. The remaining sequences were merged, quality-filtered, and further processed with USEARCH version 10.0.240 [33]. Merged bacterial reads shorter than 350 bp or longer than 550 bp were removed, while fungal reads shorter than 300 bp and longer than 500 bp were removed. Primer sequences were subsequently truncated using cutadapt (version 2.5). Reads without primer sequences as well as low-quality reads (expected error > 2) and reads with more than one ambiguous base were removed. Processed sequences of all samples were combined into a single file, and subsequently de-replicated into unique sequences. These sequences were denoised and clustered in zero-radius operational taxonomic units (zOTUs; i.e., sequences with 100% similarity) with the unoise3 algorithm [34] implemented in USEARCH [33]. Chimeric sequences were removed by the UCHIME denovo algorithm during clustering [35]. Subsequently, the remaining chimeric sequences were removed using UCHIME [35] in reference mode with the SILVA SSU Ref NR 99 132 database [36] as a reference data set for bacteria and the general release of the UNITE database version (Feb 2019) [37] for fungi.

To assign the taxonomy of bacteria and fungi, unique and chimera-free sequences were classified using the sintax algorithm against the SILVA SSU Ref NR 99 132 database [36] and the UNITE utax reference database (Feb 2019) [37], respectively. Combined sequences of all samples were mapped on the final set of unique sequences to calculate the occurrence and abundance of each zOTU in all samples. All non-bacterial and non-fungal zOTUs were removed based on their taxonomic classification. Final zOTUs tables for bacteria and fungi are provided as Supplementary Tables S2 and S3, respectively. Sequence characteristics for prokaryotic and fungal datasets are provided as Supplementary Material Tables S4 and S5, respectively.

2.7. Data Analysis

All data analyses were conducted in R version 3.6.0 [38]. Prokaryotic and fungal communities were analyzed separately. Differences were considered as statistically marginally significant with $p \leq 0.05$ and $p \leq 0.1$, respectively. Environmental properties were correlated by Spearman rank correlation using the cor.test function and grouped by hierarchical clustering using the hclust function. Each cluster contained those properties with a Spearman's rho ≥ 0.9. We selected the property with the highest correlation to all other properties from the same cluster to represent the cluster. All data were normalized prior to statistical analyses. Principal Component Analysis (PCA) was performed on resemblance matrices constructed using Euclidean distance. Differences in environmental properties among the seven sites were also evaluated by Kruskal-Wallis test, followed by Dunn's test for multiple comparisons with Benjamini–Hochberg correction using the R package FSA 0.8.25 [39]. The results of the statistical analyses are provided in Supplementary Table S4.

All alpha diversity indices were calculated 10 times and the average of all iterations was used for further statistical analyses. The zOTU tables were rarefied to 3105 (bacteria) or 2025

(fungi) sequences per sample in each iteration using the rrarefy function in vegan version 2.5–5 [40]. The diversity was calculated using the diversity function in vegan. Sample coverage was estimated using the Michaelis–Menten Fit calculated in R. For this purpose, richness and rarefaction curves were calculated using the specnumber and the rarecurve function, respectively, in picante version 1.8 [41]. The Michaelis–Menten Fit was subsequently calculated from generated rarefaction curves using the MM2 model within the drc package [42]. Good's Coverage was calculated using the R package entropart version 1.6–1 [43] using the coverage function. Final tables containing bacterial and fungal alpha diversity, richness and sample-wise coverage are provided in Supplementary Tables S5 and S6, respectively.

The alpha diversity data were tested for normal distribution with shapiro function and for homogeneity of variance with leveneTest function using the R package car version 3.0–3 [44]. As the distribution of microbial diversity and richness significantly differed from a normal distribution, differences in alpha diversity measures among the seven sites were evaluated by Kruskal-Wallis test. Statistically significant results were followed up with Dunn's test for multiple comparisons with Benjamini–Hochberg correction using the R-package FSA. We further tested for significant correlations between environmental properties/AMSs and alpha diversity measures by Spearman's rank correlation using the function cor.test.

Differences in the relative abundance of the predominant fungal ($\leq$0.5% abundance in the entire dataset) and bacterial orders ($\leq$0.5% abundance in the entire dataset) were tested by pairwise t-test with Benjamini–Hochberg correction for multiple testing. The results of the statistical analysis can be found in Supplementary Table S7. Potential differences in community composition among sites and regions were investigated by permutational multivariate analysis of variance (PERMANOVA) with 1000 random permutations using the vegdist and adonis function within the vegan package [40]. Differences in community composition between the sites were tested using pairwise PERMANOVA (https://github.com/bwemheu/pairwise.adonis; version 0.1.0). Distance-based redundancy analysis (db-RDA) with forward selection of the explanatory variables using the R package vegan was performed to analyze influences of environmental properties and agroforestry management on microbial community composition. Explanatory variables were included into the model if p was $\leq$ 0.05. Four different dissimilarity measures were calculated in R using the vegdist function [40] and tested for the bacterial and fungal datasets: unweighted as well as weighted Bray-Curtis and unweighted as well as weighted Jaccard (binary option in the vegdist function false and true, respectively). Pre-analyses revealed that weighted Bray-Curtis dissimilarities displayed a higher environmental sensitivity based on the higher coefficients of determination. Thus, only results for this distance measure are shown here.

2.8. Core Community and Correlation-Based Indicator Species Analysis

To enhance the reliability of the indicator analysis, only fungal and bacterial zOTUs detected in $\geq$75% of the trees growing in one site and with an average relative abundance of $\geq$0.01% in the entire dataset were considered, hereinafter referred to as the "core" microbiome. In addition, we applied multipattern analyses using the multipatt function from the indicspecies package [45] to identify zOTUs that are highly associated with each site. The biserial coefficients (R) with a particular site were corrected for an unequal sample size using the function r.g [46]. Associated fungal and bacterial zOTUS of each site were visualized using Cytoscape version 3.6.1 [47]. The core endophyte community and the uniquely associated fungal and bacterial zOTUs are depicted in Supplementary Table S8.

2.9. Functional Predictions

Functional information was assigned to fungal zOTUs using FUNGuild [48]. We kept guild assignments only to those zOTUs that could be assigned with the confidence ranking of "probable" and "highly probable", as recommended [48]. The sequence numbers of zOTUs assigned into the guilds were plotted as relative abundance (the number of sequences assigned to a specific guild divided by

the number of all assigned sequences; called sequence richness). In addition, the zOTU richness was determined (the number of zOTUs assigned to a specific guild per sample divided by the number of assigned zOTUs per sample). Differences in the sequence and zOTU richness were tested by pairwise *t*-test with Benjamini–Hochberg correction for multiple testing. Final tables containing functional information for fungal endophytes and results of the statistical analyses are provided as Supplementary Tables S2 and S9, respectively. Moreover, functional profiles for bacterial communities were predicted from obtained 16S rRNA data using Tax4Fun2 [49] in reference mode (Ref100NR) with copy number correction enabled. Differences in the relative abundances of putative pathways of bacterial endophytes between sites were tested by a pairwise *t*-test with Benjamini–Hochberg correction for multiple testing.

2.10. Nucleotide Sequence Accession Numbers

Sequence data were deposited in the sequence read archive (SRA) of the National Center for Biotechnology Information (NCBI) under BioProject number PRJNA594470.

3. Results

3.1. General Characteristics of the Investigated Sites

In this study, we investigated *T. cacao* leaves from different AMSs derived from five major cacao-growing regions (Obala, Boumnyébel, Bakao, Talba, Kédia) in the Central Region of Cameroon. Environmental properties measured at the different sites were correlated with each other (Supplementary Figure S1). Although some properties were significantly correlated with each other (Spearman's rho ≤ 0.9), we included all of them in the following statistical tests. Environmental and agroforestry management predictors explained more than 80% of the variation among sites ($F_{6,57}$ = 39.8, *p* = 0.001 ***, R^2 = 80.7%, 999 permutations) and more than 74% of the variation among the five growing regions representing the five different AMSs ($F_{4,59}$ = 42.9, *p* = 0.001 ***, R^2 = 74.3%, 999 permutations).

PCA ordination (Figure 1) was used to visualize the relationship of environmental data and AMSs with sampling region. The first two principal components (PC) explained 60.8% of total variation in environmental data and agroforestry practices. The PCA showed a clear clustering of Boumnyébel and Obala samples along PC1, and portioning of Boumnyébel and Talba samples from Bakoa and Obala samples along PC2. Statistical analysis further revealed that environmental properties such as annual mean temperature or humidity as well as shade tree diversity differed significantly among the seven sites (Supplementary Tables S1 and S4). For instance, rainfall was significantly higher in Simanya compared to all other sites except Pan Makak. In addition, a significantly higher temperature in Kédia and Talba than in Ekabita Essélé, Pan Makak, Simanya and Nkolobang was recorded.
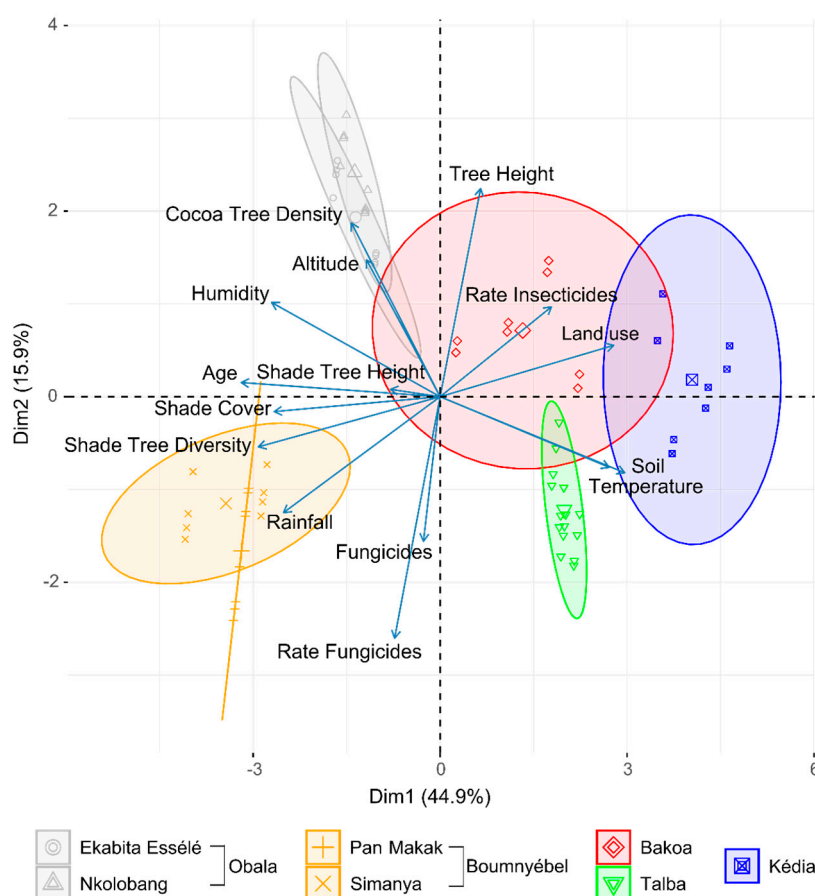
Figure 1. Biplot of the Principal Component Analysis (PCA) based on Euclidean distances. Comparison of the seven sampling sites in Cameroon by environmental characteristics and agroforestry management. The first two principal components (PC) explained 60.8% of total variation in the data. For further explanation on environmental data, see Supplementary Table S1.

3.2. Sequence Characteristics

After the removal of low-quality reads, PCR artefacts (chimeras), and plant-derived contaminations, a total of 2468,643 and 2388,301 high-quality reads were obtained for fungi and bacteria, respectively (Supplementary Tables S2 and S3). Obtained sequences were assigned to 5606 fungal and 21,902 bacterial zOTUs. Sequence numbers per sample varied between 2025 and 144,753 (average 38,573) for fungi and between 3105 and 84,048 (average 37,317) for bacteria, respectively (Supplementary Tables S5 and S6). Calculated Good's Coverage confirmed that the sampling efforts of all samples were sufficient to represent the majority of the bacterial (91.1%) and the fungal diversity (99.5%). Species accumulation curves further indicated that 82.5% of all fungal zOTUs (maximal number of zOTUs calculated = 6795) and 95.1% of all bacterial zOTUs (maximal number of zOTUs calculated = 23,036) were recovered by the surveying effort (Supplementary Figure S2). This suggests that the surveying effort was large enough to reflect the endophytic fungal and bacterial diversity in the leaves of *T. cacao*.

3.3. Foliar Endophyte Communities

Fungi were represented by three abundant phyla (≥0.5% of all sequences across all samples): Ascomycota (mean abundance across all samples: 59.5%), Basidiomycota (2.8%), and Chytridiomycota (0.7%) (Figure 2A,B, Supplementary Table S2). More than one third of the fungi (mean abundance across all samples: 36.7%) were classified as unknown fungi. The predominant fungal orders in our study were Botryosphaeriales (7.7%), Pleosporales (5.3%), Capnodiales (4.3%), Hypocreales (3.1%), Chaetothyriales (2.3%), and Glomerellales (2.0%). Lasiodiplodia (7.2%) was identified as the most

abundant fungal genus (Figure 3A). The three most abundant fungal zOTUs were two zOTUs of the genus Lasiodiplodia (Zotu1: L. brasiliensis; 5.9% and Zotu2: L. jatrophicola; 4.2%) as well as one unknown fungus (Zotu6: 5.0%).
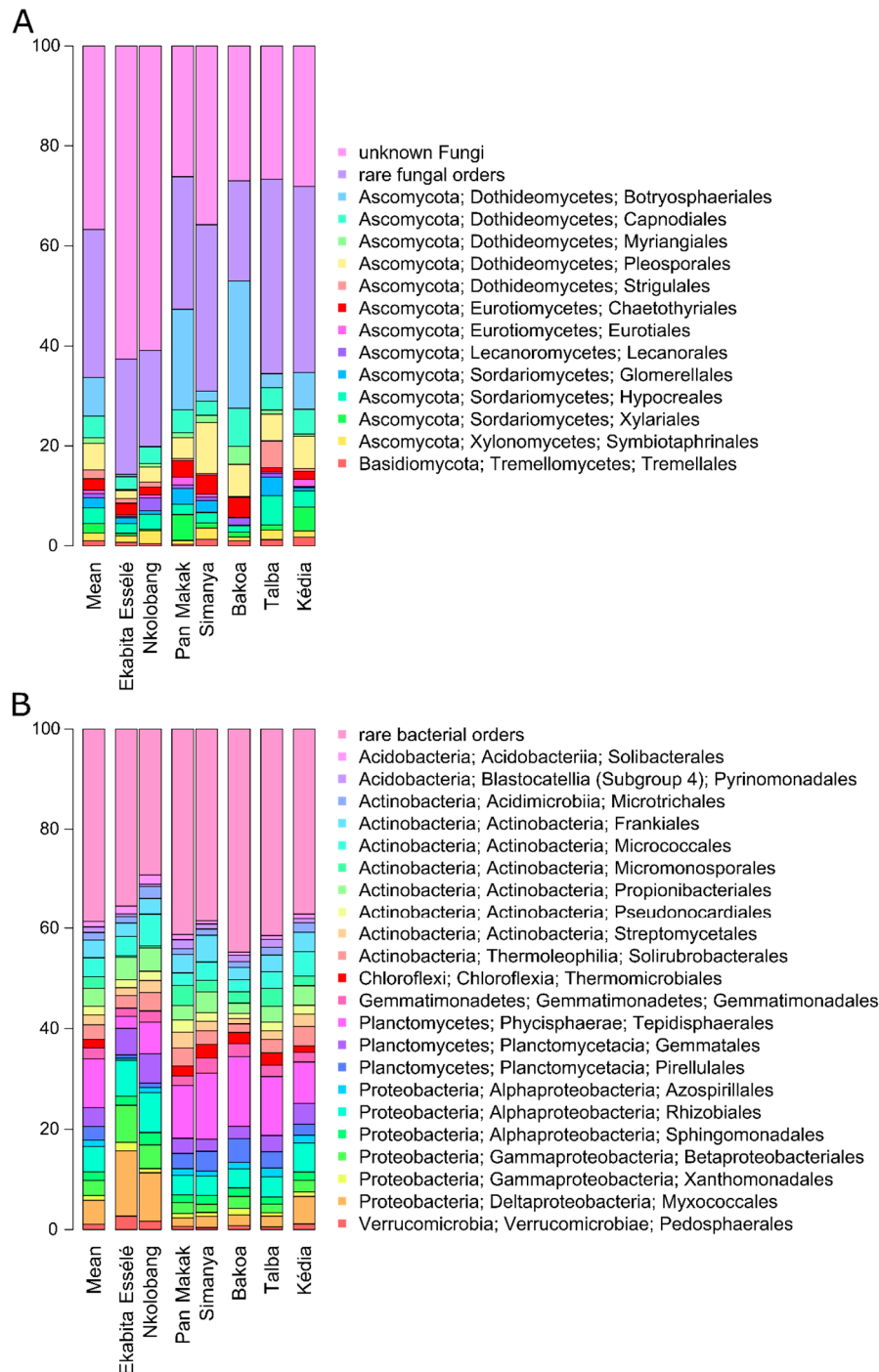
Figure 2. Abundant fungal (**A**) and bacterial (**B**) orders in *T. cacao* leaves collected in seven different agroforestry management systems in Cameroon. Only orders with an average abundance ≥1% (bacteria) or ≥0.5% (fungi) in the entire data set are shown.
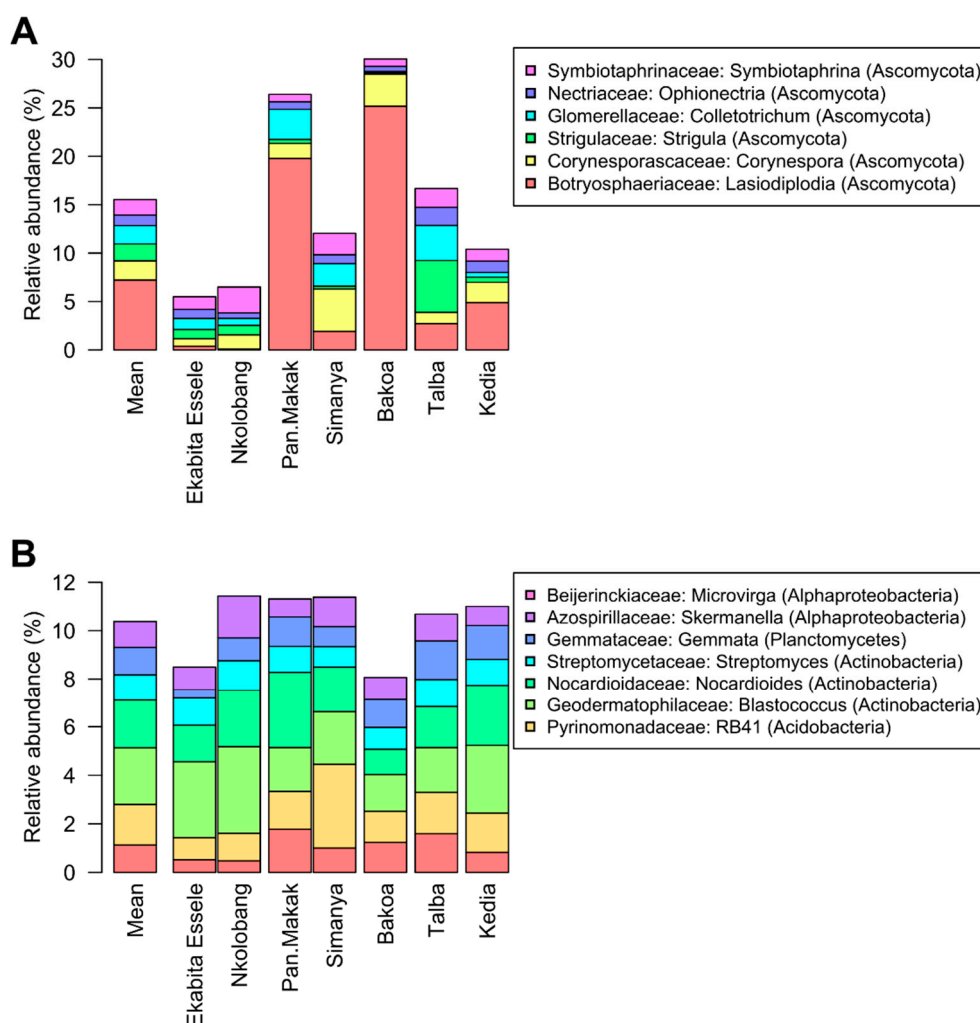
Figure 3. Abundant fungal (**A**) and bacterial (**B**) genera in *T. cacao* leaves collected in seven different agroforestry management systems in Cameroon. Only genera with an average abundance ≥1% in the entire data set are shown.

Bacterial communities were dominated by Actinobacteria (mean abundance across all samples: 30.3%), Proteobacteria (22.8%), and Planctomycetes (19.7%) (Figure 3B, Supplementary Table S3). Other abundant (≥ 1% of all sequences across all samples) bacterial phyla were Acidobacteria (9.5%), Chloroflexi (6.8%), Gemmatimonadetes (2.7%), Verrucomicrobia (2.6%), and Bacteroidetes (1.6%). Within the Proteobacteria, Alphaproteobacteria were predominant (12.3%), followed by Gammaproteobacteria (6.0%) and Deltaproteobacteria (5.3%). The dominant bacterial orders across all samples were Tepidisphaerales (9.7%), Rhizobiales (5.1%), Myxococcales (4.8%), Micrococcales (3.8%), Gemmatales (3.7%), Propionibacteriales (3.5%), and Frankiales (3.5%). Nocardioides (2.3%) Streptomyces (2.0%) were detected as the most abundant bacterial genera (Figure 3B). At zOTU level, two zOTUs of the genera *Pseudarthrobacter* (Zotu1; 0.5%) and *Streptomyces* (Zotu2; 0.4%) were predominant.

3.4. Endophyte Diversity and Community Composition Per Site

Three of the dominant fungal orders (Eurotiales, Hypocreales and Botryosphaeriales) and several predominant bacterial orders differed among the seven sites (Figure 2A, Supplementary Table S7). We detected significantly higher abundances of the Botryosphaeriales in Bakoa (25.5%) leaves compared to those from Ekabita Essélé (0.4%), Nkolobang (0.2%), and Talba (2.8%). To impact of these factors on endophyte community composition, a distance-based redundancy analysis (db-RDA) based on Bray-Curtis dissimilarities was performed. In addition, we analyzed the effect of site and

type of agroforestry system on the composition of bacterial and fungal endophyte communities by PERMANOVA.

We observed a clear clustering of fungal (Figure 4A) and, to a lesser extent, of bacterial endophyte communities (Figure 4B), by AMS and sampling site. This is supported by the results of the PERMANOVA. Sampling site ($F_{(6)}$ = 1.98, *p* = 0.001; R^2 = 17.2%) and AMS ($F_{(4)}$ = 1.95, *p* = 0.001; R^2 = 11.7%) significantly affected the composition of bacterial communities. Similarly, the composition of fungal endophytes differed significantly between sites ($F_{(6)}$ = 2.26, *p* = 0.001; R^2 = 19.2%) and agroforestry systems ($F_{(4)}$ = 2.83, *p* = 0.001; R^2 = 16.1%). The best-fit explanatory variables for fungal community composition were insecticide rates, fungicide rates, humidity, and shade tree height ($F_{5,58}$ = 2.31, *p* = 0.001; Figure 4A, Table 2). The community composition of bacterial endophytes was significantly affected by insecticide rate, fungicide rate, shade tree height, altitude and cacao tree density ($F_{6,57}$ = 2.30, *p* = 0.001; Figure 4B, Table 2). Multiple comparisons revealed that the composition of fungal communities differed significantly among all AMSs, while bacterial community composition in leaves from Obala differed from those collected in Talba, Kédia, and Bakoa (*p* ≤ 0.05).
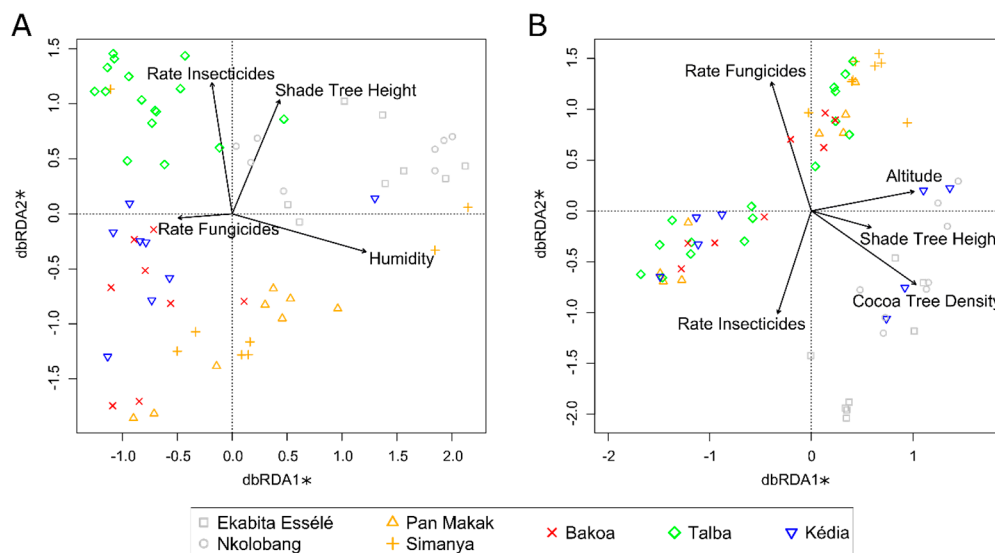
Figure 4. Distance-based redundancy analysis (db-RDA) plot showing the relationship of environmental properties and agroforestry management system to foliar fungal (**A**) and bacterial (**B**) endophyte communities of *T. cacao*. Ordination is based on weighted Bray-Curtis distances between samples and is color-coded by sampling region. Factors were chosen that significantly (*p* ≤ 0.05) contributed to the model. Axes labelled with an asterisk are significant. The first axes explained 42.5% (bacteria) or 42.2% (fungi), whereas the second axes explained 28.1% (bacteria) or 22.6% (fungi).

Table 2. Results of the db-RDA that describes the effect of environmental properties and agroforestry management systems on endophyte community composition in *T. cacao* leaves.

	DF	SumOfSqs	F	*p* Value	
Bacteria					
Entire Model	6	3.89	2.30	0.001	***
Cocoa Tree Density	1	0.49	1.72	0.024	*
Altitude	1	0.70	2.47	0.002	**
Fungicides rate/Cropping Season	1	0.51	1.82	0.022	*
Fungicides	1	0.70	2.49	0.002	**
Shade Tree Height	1	0.56	1.97	0.013	*
Insecticides rate/Cropping Season	1	0.45	1.58	0.049	*
Residual	57	16.08			
Fungi					
Entire Model	5	4.33	2.31	0.001	***
Humidity	1	1.05	2.78	0.001	***
Shade Tree Height	1	0.70	1.87	0.006	**
Fungicides rate/Cropping Season	1	0.71	1.90	0.005	**
Fungicides	1	0.58	1.54	0.019	*
Insecticides rate/Cropping Season	1	0.55	1.48	0.042	*
Residual	58	21.77			

For each model, forward selection was applied to identify which factors best described variation in community composition using an inclusion threshold of $\alpha = 0.05$ and Bray-Curtis distances. Marginal effects of terms are shown (i.e., terms were not added sequentially). The number of unrestricted permutations: 999. Significance level: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. SumOfSps: sum of squares, DF: degrees of freedom.

In addition to changes in community composition, we analyzed differences in alpha diversity measures among sites and AMSs. Richness (number of observed unique sequences) and diversity (Shannon diversity index H') for fungal communities varied between 71 and 443 and between 1.32 and 5.34, respectively (Table 3 and Supplementary Table S5). Bacterial richness and diversity ranged from to 1206 to 2096 and from 6.13 to 7.44, respectively (Supplementary Table S6). Bacterial alpha diversity measures did not differ among the sites (richness: $p = 0.79$; diversity: $p = 0.43$) and the agroforestry system types (richness: $p = 0.75$; diversity: $p = 0.32$) (Table 3). In contrast, fungal richness ($p < 0.001$) and diversity ($p = 0.001$) differed among the five agroforestry systems, with, site significantly affecting fungal richness ($p = 0.003$) and diversity ($p = 0.006$).

Table 3. Alpha diversity measures (mean ± standard deviation) for bacterial and fungal endophytes in leaves of *T. cacao*. Richness and diversity are represented by the number of observed zero-radius operational taxonomic units (zOTUs) and Shannon diversity index H', respectively.

	Fungi		Bacteria	
	Richness	**Diversity**	**Richness**	**Diversity**
Obala	178 ± 76^A	3.04 ± 0.98^A	1729 ± 218	7.02 ± 0.31
- Ekabita Essélé	174 ± 72^a	2.85 ± 0.98^a	1672 ± 255	6.92 ± 0.4
- Nkolobang	182 ± 85^a	3.24 ± 0.99	1785 ± 172	7.11 ± 0.18
Boumnyébel	279 ± 106^B	4.09 ± 1.31^B	1820 ± 85	7.17 ± 0.1
- Pan Makak	250 ± 110^b	3.96 ± 1.47	1814 ± 88	7.16 ± 0.09
- Simanya	307 ± 101^b	4.22 ± 1.23	1826 ± 87	7.19 ± 0.1
Bakoa	231 ± 82^{AB}	3.48 ± 1.11^{AB}	1789 ± 160	7.15 ± 0.19
Talba	292 ± 55^{Bb}	4.44 ± 0.41^{Bb}	1804 ± 137	7.17 ± 0.13
Kédia	330 ± 87^{Bb}	4.42 ± 0.64^B	1782 ± 101	7.12 ± 0.11

[A,B] Different superscript letters indicate significant differences ($p \leq 0.05$) between the five regions (Obala, Boumnyébel, Bakoa, Talba and Kédia). Note that fungal diversity differed marginally ($p \leq 0.1$) among Bakoa and Talba/Boumnyébel. In addition, fungal richness differed marginally between Bakoa and Kédia. [a,b] Different superscript letters indicate significant differences ($p \leq 0.05$) between the seven sites (Ekabita Essélé, Nkolobang, Pan Makak, Simanya, Bakoa, Talba and Kédia). Note that there were marginally significant differences in fungal diversity between Ekabita Essélé and Pan Makak/Simanya/Kédia, and between Nkolobang and Talba.

Multiple comparisons revealed a significantly higher fungal diversity and richness in Boumnyébel, Talba, and Kédia compared to Obala. Temperature and fungicide rate were positively correlated with fungal richness and diversity, while humidity, cacao tree density, and tree height were negatively correlated with fungal richness and/or diversity (Table 4).

Table 4. Correlation between alpha diversity measures and environmental properties/an agroforestry management system based on Spearman's rank correlation. Richness and diversity are represented by the number of observed zOTUs and Shannon diversity index H', respectively.

Tested Variable		Fungal Endophytes				Bacterial Endophytes			
		Richness		Diversity		Richness		Diversity	
	DF	rho	p	rho	p	rho	p	rho	p
Environmental properties									
Altitude	62	−0.11	0.37	−0.10	0.43	0.14	0.27	0.08	0.56
Temperature	62	0.34	**0.006**	0.26	**0.04**	−0.06	0.65	−0.01	0.92
Humidity	62	−0.34	**0.006**	−0.26	**0.04**	0.02	0.87	−0.11	0.41
Rainfall	62	0.12	0.34	0.24	<u>0.052</u>	0.04	0.77	0.04	0.73
Agroforestry management									
Age	62	−0.22	<u>0.07</u>	−0.11	0.39	0.05	0.70	0.02	0.87
Cacao Tree Density	62	−0.26	**0.04**	−0.21	<u>0.099</u>	−0.09	0.46	−0.18	0.16
Insecticides rate *	62	−0.06	0.62	−0.05	0.70	−0.08	0.51	−0.09	0.49
Fungicides rate *	62	0.29	**0.02**	0.42	**<0.001**	0.12	0.34	0.17	0.18
Shade Tree Height	62	−0.14	0.26	0.004	0.97	−0.18	0.16	−0.17	0.18
Shade Tree Diversity	62	−0.07	0.57	0.04	0.78	0.14	0.28	0.09	0.48
Shade Cover	62	−0.18	0.16	−0.07	0.58	0.02	0.86	0.02	0.88
Tree Height	62	−0.27	**0.03**	−0.29	**0.02**	−0.03	0.83	−0.02	0.85

Statistically significant ($p \leq 0.05$) and marginally significant ($p \leq 0.1$) p values are written in bold and are underlined, respectively. DF: degrees of freedom. * Fungicides and Insecticides rate/Cropping Season.

3.5. Bacterial and Fungal Taxa Associated with Agroforestry System Type

To identify bacterial and fungal core taxa associated with sites/agroforestry management systems, we performed an indicator species analysis. The core zOTUs were selected based on their relative frequency (≥75% occurrence in each of the seven sites) (Supplementary Table S8). The core endophyte community was represented by ten fungal zOTUs (0.2% of all fungal zOTUs), such as *L. brasiliensis* (Zotu1; mean abundance across all samples: 5.9%), *Colletotrichum hymenocallidis* (Zotu22; 0.8%), and *Ophionectria trichospora* (Zotu29; 0.9%). Moreover, we detected 199 bacterial zOTUs (0.9% of all bacterial zOTUs), including eleven zOTUs belonging to the genera *Streptomyces* and *Blastococcus* and nine zOTUS of the genus *Nocardioides*. A total of 20.1% of all fungal and 12.6% of all bacterial sequences were assigned to these core zOTUs.

The indicator species analysis further identified those significantly associated fungal and bacterial zOTUs using the following threshold levels: sample prevalence (≥75% in one site) and relative abundance (≥0.01% in the entire dataset). We detected a higher number of significantly associated bacterial (*n* = 523) than fungal (*n* =79) zOTUs (Figure 5, Supplementary Table S8). The lowest and highest number of significantly associated bacterial zOTUs were found in Pan Makak (*n* = 51) and Simanya (*n* = 118), respectively. Most bacterial zOTUs belonged to the genera Pir4 lineage (*n* = 19), *Nocardioides* (*n* = 12), *Actinoplanes* (*n* = 12), Candidatus *Alysiosphaera* (*n* = 11), *Marmoricola* (10), *Micromonospora* (*n* = 9), *Blastococcus* (*n* = 8), and *Krasilnikovia* (*n* = 8). For fungi, the lowest and highest number of significantly associated zOTUs were detected for Ekabita Essélé (*n* = 7) and Kédia (*n* = 22), respectively.
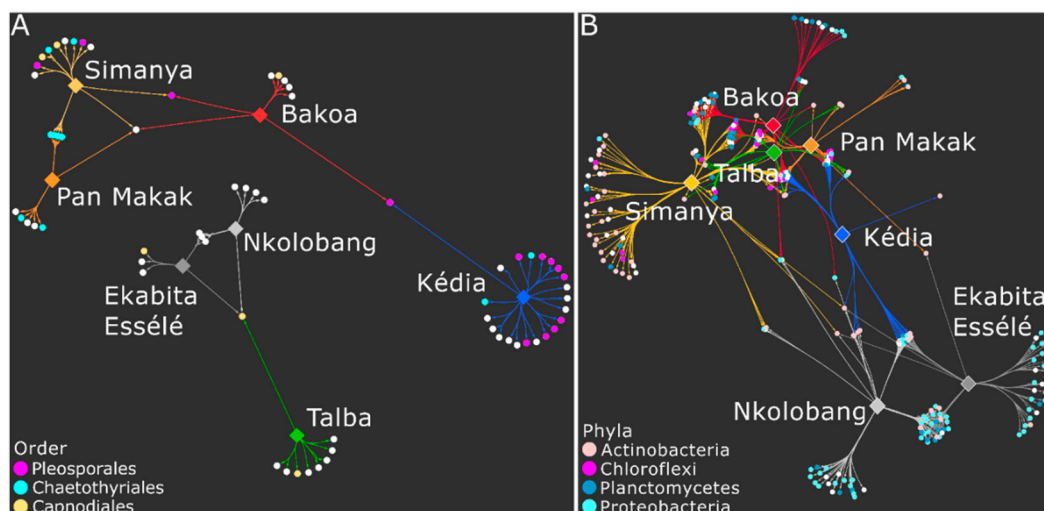
Figure 5. Bipartite association network of fungal (**A**) and bacterial (**B**) zOTUs in *T. cacao* leaves significantly associated with site. The sites are color-coded as in Figure 1. Bacterial phyla and fungal orders, which were predominant in the dataset and/or in one site, are highlighted.

3.6. Fungal Functionality Differs between the Agroforest Management Systems

We also analyzed functional changes in fungal endophytes with respect to AMS. For this purpose, functional guilds of endophytic fungi were determined using FUNGuild [48]. In total, highly probable and probable life strategies for 1766 of the 5606 zOTUs (= 31.5%) were predicted (Supplementary Table S9). We identified more abundant (≥0.5% across all samples) functional guilds by investigating the zOTU richness (*n* = 21) compared to sequence richness (*n* = 16) (Figure 6; Supplementary Table S2). Most of the fungal sequences were classified as undefined saprotrophs (sequence richness: 24.2% mean abundance across all samples; zOTU richness: 20.2.%). Other abundant functional guilds were plant pathogens (sequence richness: 23.8%; zOTU richness: 13.5%) and animal endosymbionts (sequence richness: 7.4%; zOTU richness: 14.8%). However, we observed differences in the sequence richness among sites. Undefined saprotrophs were predominant in Simanya (28.3%), Talba (20.7%), and Ekabita Essélé (29.5%), whereas lichenized fungi dominated in Nkolobang (21.7%) (Figure 6A). Plant pathogens were the dominant functional guild in Pan Makak (35.8%), Kédia (26.4%), and Bakoa (43.3%). Analysis of the zOTU richness revealed that undefined saprotrophs were the most abundant functional guild in Kédia (24.1%), Pan Makak (25.1%), Simanya (22.1%), Bakoa (24.1%), and Ekabita Essélé (18.0%) (Figure 6B). Animal endosymbionts were predominant in Nkolobang (19.3%) and lichenized fungi in Talba (17.1%).

Multiple comparisons revealed that two functional guilds (animal pathogen-plant pathogen-undefined saprotrophs and wood saprotrophs) did not differ in both sequence and zOTU richness among sites (Figure 6). Moreover, the sequence richness of most functional guilds, including animal pathogens, ericoid mycorrhizal, plant pathogens, or endophyte-plant pathogens did not differ between the sites (Figure 6A). In contrast, we observed a significantly higher sequence richness of the functional guild of arbuscular mycorrhiza in Ekabita Essélé (2.6%) and Nkolobang (3.0%) compared to the other sites (≤0.8%). The zOTU richness showed that arbuscular mycorrhiza were significantly more abundant in Ekabita Essélé (3.8%) and Nkolobang (4.2%) than in Bakoa (0.9%) (Figure 6B). We also observed a significantly lower zOTU richness for animal pathogens in Ekabita Essélé (0.9%) compared to all other sites (≥2.5%) except Nkolobang (2.1%). A significantly higher zOTU richness of plant pathogens was observed in Bakoa (17.6%) compared to Kédia (11.5%) and Talba (11.3%). Lichenized fungi were more abundant in Talba (17.1%) compared to all other sites (≤12.9%).
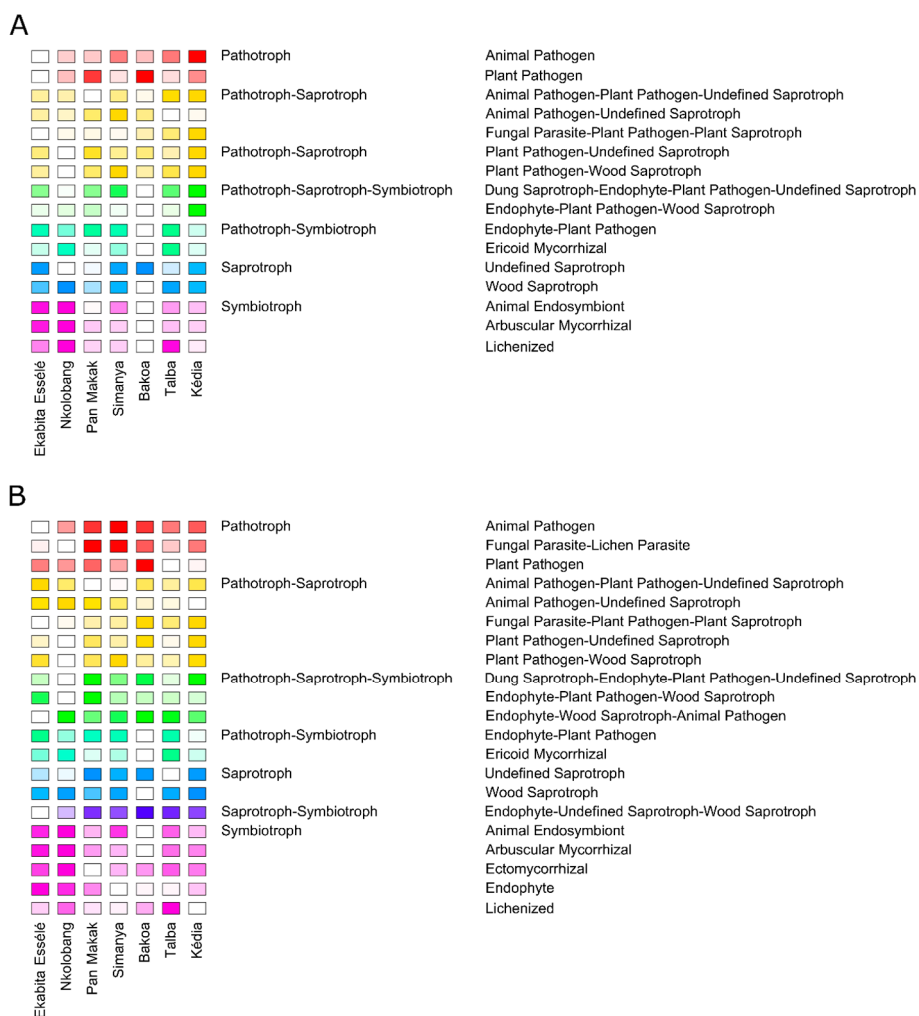
Figure 6. Functional guilds of fungal endophytes in *T. cacao* leaves. The relative abundance of fungal sequences (sequence richness) per guild (**A**) and the proportion of zOTUs (zOTU richness) per guild (**B**) is shown for the seven sites. The analyses are based on 31.5% of the zOTUs (*n* = 1766 zOTUs). The sequence richness and zOTU richness were calculated by the number of sequences assigned to a specific guild divided by the number of all assigned sequences and by the number of zOTUs assigned to a specific guild per sample divided by the number of zOTUs per sample, respectively. Following the suggestion of Nguyen et al. [48], we combined both dimensions (sequence and zOTU richness) to better reflect the relative importance of fungal life strategies in an environment. Only guilds with an average abundance ≥0.5% in the entire data set are shown.

3.7. Predicted Functional Profiles of Bacterial Endophytes

To investigate potential changes in bacterial community function between the AMSs, functional profiles for bacterial endophytes were predicted from 16S rRNA gene data using Taxa4Fun2 [49]. Approximately 82% of all zOTUs, representing 76.7% of all sequences obtained, were used in the prediction (Supplementary Table S9). We focused on important pathways involved in metabolism, environmental information processing and organismal systems, resulting in 32 abundant pathways (≥1% of all sequences across all samples) (Figure 7). The highest abundances were observed for ABC transporters (mean abundance across all samples: 11.1%), Valine, leucine, and isoleucine degradation (2.8%), propanoate metabolism, fatty acid degradation, pyruvate metabolism, and butanoate metabolism (all: 2.5%). The three pathways with the lowest mean abundance of the 32 selected pathways across all samples (1.1%) were tyrosine metabolism, porphyrin and chlorophyll metabolism, and the citrate cycle (TCA cycle).
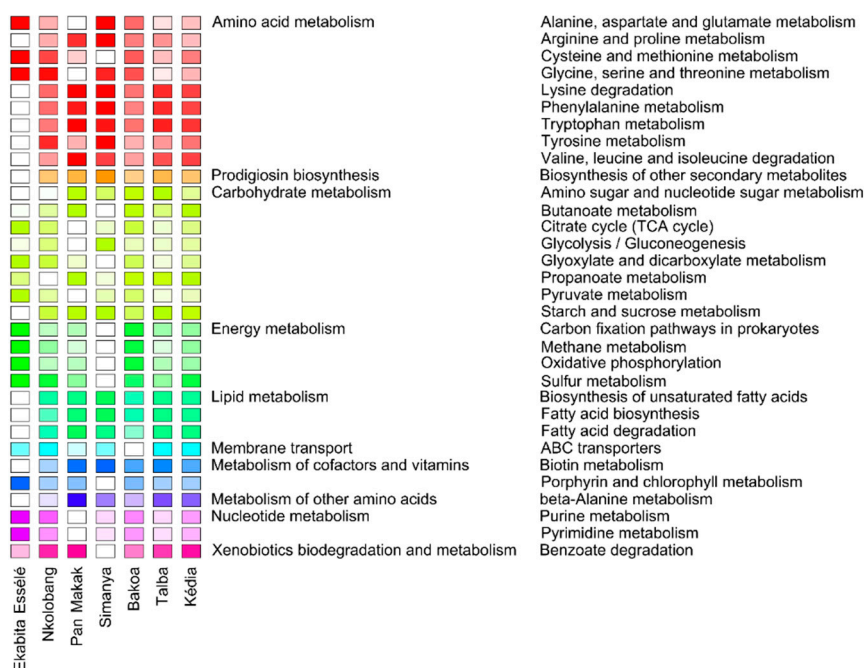
Figure 7. Predicted functional profiles of bacterial endophytes in *T. cacao* leaves. The color code refers to the relative abundance of each pathway, with high predicted abundances (dark-colored) and low predicted abundances (light-colored). The relative abundances of pathways were calculated for each site. Only pathways with an average abundance ≥1% in the entire data set are shown.

The majority (*n* = 22) of the 32 abundant pathways, such as pathways involved in the metabolism of cofactors and vitamins, pathways involved in both lipid and energy metabolism (excluding sulfur metabolism), as well as ABC transporters, did not differ in their relative abundance among the sites (Figure 7). In contrast, we observed significantly higher abundances of pyrimidine metabolism, purine metabolism, glyoxylate and dicarboxylate metabolism, as well as sulfur metabolism in Ekabita Essélé compared to Pan Makak, Simanya, and Talba. We also found significantly higher abundances of starch and sucrose metabolism in Pan Makak, Simanya, Talba, and Kédia than in Ekabita Essélé and higher abundances of the tyrosine metabolism in Simanya compared to all other sites. Significantly lower abundances of arginine and proline metabolism as well as phenylalanine metabolism were observed in Ekabita Essélé than in Simanya, whereas the opposite was detected for cysteine and methionine metabolism.

4. Discussion

To our knowledge, this is the first metabarcoding study simultaneously investigating the bacterial and fungal endophytes of *T. cacao* trees and their response towards the agroforestry system type. The leaves were colonized by a high diversity of fungal species; however, more than one third of these fungal zOTUs could not be further classified. These results indicate that the identity and ecology of many fungal endophytes in these regions remain largely uncharacterized. The large proportion of Ascomycota and comparatively few Basidiomycota generally agrees with previous studies on foliar fungal communities from tropical and temperate trees [50–53]. Two of the dominant fungal zOTUs belonged to *Lasiodiplodia*, a genus of the Botryosphaeriaceae. This family contains numerous fungal species, which are able to infect a diverse range of host plant species or known to live as saprophytes or endophytes within seeds and other living plant tissues [54,55].

The taxonomic composition of bacterial endophytes in *T. cacao* leaves differed from previous studies on bacterial endophytes of temperate tree species [51,56]. For instance, higher abundances of Proteobacteria were recorded in a recent study on Maple trees in Germany. Nonetheless, our results are in line with a previous study on bacterial communities in the roots of the tropical tree species

Eucalyptus urograndis [57] and the leaves of several woody species [58]. Two zOTUs of the bacterial genera *Pseudarthrobacter* and *Streptomyces* (both Actinobacteria) were predominant in our samples. This is noteworthy as members of the Actinobacteria, especially the genus *Streptomyces*, are known to produce a wide range of antimicrobial compounds [59,60].

The composition of fungal and bacterial endophyte communities in *T. cacao* leaves differed among the sampling sites and AMSs. The observed differences in the relative abundances of the predominant fungal and bacterial orders most likely arose from differences in management practices and environmental properties in the studied cacao agroforestry systems. This is in line with previous studies investigating endophyte communities from tree species, analyzed from different locations [13,51,61]. We further recorded that fungal but not bacterial endophyte alpha diversity measures differed among sites. Contrarily, site significantly affected bacterial but not fungal alpha diversity values in leaves of two *Acer* species [51]. In addition, AMS affected fungal richness and diversity but not bacterial alpha diversity. Köberl et al. [62] reported neither a significant impact of agroforestry type nor biogeography on the gammaproteobacterial diversity in leaves of *Musa* spp. [62]. In another study on fungal endophytes in the leaves of *B. pendula* grown in natural and managed boreal forests, silvicultural practices affected the species composition and endophyte frequencies of these fungi [63]. The authors proposed that the observed differences reflect the mode of fungal spreading and largely depend on biotic and abiotic environmental conditions determining the abundance of infection sources and the success of transmission and germination of specific fungal endophyte species. This hypothesis is supported by our study as environmental properties, such as temperature and/or humidity as well as agroforestry practices influenced alpha diversity measures and the community composition of fungal endophytes. Previous studies on the fungal endophyte communities of trees in Hawaiian landscapes [64], the leaves of Olea europaea in the Mediterranean area [65], and the leaves of Ficus tree species in the Philippines [52] also found that patterns of community similarity are strongly associated with rainfall and elevation:

Consistent with previous studies on fungicide effects on leaf-associated microbial communities [26,66,67], insecticide and fungicide application rate affected the composition of both fungal and bacterial endophytes in leaves of *T. cacao*. Interestingly, fungicide rate was positively correlated with fungal richness and diversity. In contrast, fungicide treatment had no effect on fungal endophyte diversity in *A. altissima* [26]. We speculate that the fungicide metalaxyl, used in the site sampled is highly selective against some fungi [68]. This might have created free niches in the leaf tissues subsequently colonized by other fungi.

Fungal and bacterial endophyte community composition as well as fungal richness and diversity in leaves collected in Obala significantly differed from those collected in Talba, Kédia, and Bakoa, respectively. As favorable growing conditions, such as higher humidity and temperature or lower UV radiation, can influence the rate and extent of fungal colonization and survival on leaf surfaces [65,69], we hypothesize that environmental conditions (e.g., microclimate and/or light intensity) in Obala have prevented fungal endophyte colonization. Scholtysik et al. [70] found that composition of endophytes in sun-exposed leaves from the top of full-grown *Fraxinus excelsior* trees differed considerably from leaves in the shade crowns and in the understorey. Lastly, we hypothesize that our results are related to the significantly higher cacao tree density observed in Obala or differences in shade tree height among the sites, as shade tree height and cacao tree density were main drivers of bacterial and/or fungal endophyte community composition.

Fungal and bacterial endophytes responded differently towards the factors investigated. These results are consistent with previous studies on plant-associated microbial communities [14,51] and are attributed to differences in lifestyle strategies (i.e., colonization behaviour) of fungal and bacterial endophytes [10,71]. Another possible explanation is that fungal endophytes might be more sensitive to agroforestry system type and/or environmental properties than bacteria. Due to the limitations in the study design, future studies investigating more agroforests at different cacao growing regions are needed to better understand causal factors influencing bacterial and fungal endophyte communities.

Our analysis revealed that leaves of *T. cacao* formed a core endophyte microbiota of a few fungal and 199 bacterial zOTUs, although study sites were hundreds of kilometers apart. There are several potential explanations for this results. Firstly, it is possible that the core community bacterial and fungal taxa colonized tree leaves across cacao plantations by dispersal through numerous vectors (air, rain, insect vectors), homogenizing the endophyte community as suggested for the phyllosphere communities of several tree species [72]. Another possible explanation is that some bacteria and fungi are obligate endophytes and are thus restricted to a life inside of plant tissues [73,74]. These obligate endophytes might constitute larger parts of the core community of *T. cacao* leaves. Several fungal zOTUs identified here, such as *L. brasiliensis* [75] or *C. hymenocallidis* [76], have been observed as plant pathogens, while *O. trichospora* is a relatively common species occurring on rotting wood in tropical regions [77]. It is likely that the above-mentioned fungi are latent pathogens and latent saprotrophs [78], as *T. cacao* leaves did not show any symptoms of diseases. Vega et al. [79] suggested that some endophytic fungi, such as *Colletotrichum* are either ubiquitous in coffee-growing regions because of the exchange of *Coffea* plants and seeds or because of intrinsic factors (i.e., the global distribution of the fungi themselves). These reasons might also have played a role in the current study. Lastly, it could be that the observed fungal and bacterial endophytes in the core community are seed-borne, as described for fungi such as *L. theobromae* [80] or bacteria [81]. As we did not investigate the seedling or seed endophytes of *T. cacao*, the transmission modes and colonization routes of fungal and bacterial endophytes await further research.

We further recorded that the cacao agroforests showed uniquely associated fungal and bacterial zOTUs, suggesting specific adaptions to environmental properties and agroforestry practices in the respective sites. Another explanation is that the *T. cacao* trees select for beneficial microorganisms as they provide an advantage for their host plant [73,82]. Some of the uniquely associated bacterial zOTUs, including *Streptomyces*, *Actinoplanes*, or *Micromonospora*, are known to have plant growth-promoting abilities. *Actinoplanes* spp. are a group of filamentous bacteria that can parasitize *Pythium* spp. or related fungi [83]. As mentioned above, *Streptomyces* are best-known for their wide range of produced biomolecules, which in turn might be excellent agents for controlling various fungal and bacterial phytopathogens [59]. Similarly, some species of the *Micromonospora* produce antimicrobial and antifungal compounds that act to protect plants from pathogens (reviewed in [84]). In addition, we identified animal-associated fungal species. *C. aphidis* usually grows on aphids and superficially on leaves attacked by aphids [85] and *S. buchneri* was isolated from the beetle *Stegobium paniceum* that fed on the pulverized fruits of Capsicum [86]. The authors hypothesized that the life cycle of this species includes a symbiotic phase in the gut of the beetle and an unknown sexual morph growing on plant substrates.

We applied FUNGuild [48] and Tax4Fun2 to assess the functional responses of fungal and bacterial endophytes towards agroforestry practices as well as environmental conditions. Consistent with two recent studies on root-associated fungal communities in the Bolivian Andes [87] and aerial fungal endophytes of three grass species in Germany [24], saprophytic fungi dominated fungal endophyte communities. We identified more abundant functional guilds when investigating the zOTU richness compared to sequence richness. This supports the suggestion of Nguyen et al. [48] that combining both dimensions reflect the relative importance of fungal life strategies in an environment. The high abundance of ABC transporters in the predicted functional profile of bacterial endophytes might be related to the plant-associated lifestyle of endophytes, which requires the efficient uptake of plant synthesized nutrients [88]. Similarly, Hardoim et al. [10] suggested that regulatory genes related to the stoichiometry of carbon and nitrogen metabolism and those involved in the metabolism of vitamins and nucleotides and in stress responses are of fundamental importance for a life inside plants.

We detected that several fungal functional guilds and predicted pathways of bacterial endophytes differed in their relative abundances among the sites and agroforestry system types. These results might be related to differences in agroforestry practices and environmental properties, which altered

community composition and conseqeuntly community functioning. The functional changes of fungal and bacterial endophytes towards agricultural practices have been observed previously [11,24,67].

It has been proposed that the core microbiome might be functionally significant for the host plant, while the accessory microbiome is expected to contain more dispensable functions or microorganisms whose presence is related to interactions with the surrounding environmental conditions [82]. We thus assume that functional guilds and predicted bacterial pathways that did not differ between the sites are related to the core microbiome, while the uniquely associated endophytes are responsible for observed functional changes. However, only a low number of fungal and bacterial zOTUs and sequences could be used in the analyses, as the ecological role of most microorganisms in the plant endosphere and their functions remain still unknown [10,71]. Consequently, further studies are needed to better understand how management regimes affect functional traits of bacterial and fungal endophyte communities and their functioning in leaves of the economically important tree species *T. cacao.*

5. Conclusions

To date, studies on fungal and bacterial endophytes in *T. cacao* trees growing in different agroforestry systems are lacking. In the current study, we applied large-scale metabarcoding to assess compositional and functional responses of fungal and bacterial endophytes in the leaves of *T. cacao* trees growing in five major cacao-growing regions in the Central Region of Cameroon. The diversity and richness of fungal but not bacterial endophytes differed among the five cacao regions, suggesting that fungal alpha diversity is more sensitive to agroforestry system type and/or environmental properties than bacterial alpha diversity. Our results further suggest that bacterial and fungal endophyte community composition are affected predominantly by agroforestry practices and, to a lesser extent, by environmental properties. The correlation-based indicator species indicated that the core microbial community forms stable associations with *T. cacao* across geographic scales. Functional analyses, based only a minor part of microbial zOTUs and sequences, revealed that several predicted pathways of bacterial endophytes and the functional guilds of fungal endophytes differed between the agroforests, which might be attributed to several bacteria and fungi specifically associated with a single agroforest.

References

1. Rice, R.A.; Greenberg, R. Cacao cultivation and the conservation of biological diversity. *AMBIO J. Hum. Environ.* **2000**, *29*, 167–173. [CrossRef]
2. Clough, Y.; Faust, H.; Tscharntke, T. Cacao boom and bust: Sustainability of agroforests and opportunities for biodiversity conservation. *Conserv. Lett.* **2009**, *2*, 197–205. [CrossRef]
3. Wade, A.S.; Asase, A.; Hadley, P.; Mason, J.; Ofori-Frimpong, K.; Preece, D.; Spring, N.; Norris, K. Management strategies for maximizing carbon storage and tree species diversity in cocoa-growing landscapes. *Agric. Ecosyst. Environ.* **2010**, *138*, 324–334. [CrossRef]

4. Bisseleua, D.H.B.; Missoup, A.D.; Vidal, S. Biodiversity Conservation, Ecosystem Functioning, and Economic Incentives under Cocoa Agroforestry Intensification Conservación de Biodiversidad, Funcionamiento del Ecosistema e Incentivos Económicos en la Intensificación de la Agroforestería de Cacao. *Conserv. Biol.* **2009**, *23*, 1176–1184. [CrossRef]

5. Bos, M.M.; Steffan-Dewenter, I.; Tscharntke, T. Shade tree management affects fruit abortion, insect pests and pathogens of cacao. *Agric. Ecosyst. Environ.* **2007**, *120*, 201–205. [CrossRef]

6. Bisseleua, H.B.D.; Fotio, D.; Yede, A.D.M.; Vidal, S. Shade Tree Diversity, Cocoa Pest Damage, Yield Compensating Inputs and Farmers' Net Returns in West Africa. *PLoS ONE* **2013**, *8*, e56115. [CrossRef]

7. Bisseleua, D.H.B.; Vidal, S. Plant biodiversity and vegetation structure in traditional cocoa forest gardens in southern Cameroon under different management. *Biodivers. Conserv.* **2008**, *17*, 1821–1835. [CrossRef]

8. Deheuvels, O.; Rousseau, G.X.; Quiroga, G.S.; Franco, M.D.; Cerda, R.; Mendoza, S.J.V.; Somarriba, E. Biodiversity is affected by changes in management intensity of cocoa-based agroforests. *Agrofor. Syst.* **2014**, *88*, 1081–1099. [CrossRef]

9. Clough, Y.; Barkmann, J.; Juhrbandt, J.; Kessler, M.; Wanger, T.C.; Anshary, A.; Buchori, D.; Cicuzza, D.; Darras, K.; Putra, D.D.; et al. Combining high biodiversity with high yields in tropical agroforests. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8311–8316. [CrossRef]

10. Hardoim, P.R.; van Overbeek, L.S.; Berg, G.; Pirttila, A.M.; Compant, S.; Campisano, A.; Doring, M.; Sessitsch, A. The Hidden World within Plants: Ecological and Evolutionary Considerations for Defining Functioning of Microbial Endophytes. *Microbiol. Mol. Biol. Rev.* **2015**, *79*, 293–320. [CrossRef]

11. Wemheuer, F.; Kaiser, K.; Karlovsky, P.; Daniel, R.; Vidal, S.; Wemheuer, B. Bacterial endophyte communities of three agricultural important grass species differ in their response towards management regimes. *Sci. Rep.* **2017**, *7*, 40914. [CrossRef] [PubMed]

12. Hanada, R.E.; Pomella, A.W.V.; Costa, H.S.; Bezerra, J.L.; Loguercio, L.L.; Pereira, J.O. Endophytic fungal diversity in Theobroma cacao (cacao) and T. grandiflorum (cupuaçu) trees and their potential for growth promotion and biocontrol of black-pod disease. *Fungal Biol.* **2010**, *114*, 901–910. [CrossRef] [PubMed]

13. Saucedo-García, A.; Anaya, A.L.; Espinosa-García, F.; Gonzalez, M. Diversity and Communities of Foliar Endophytic Fungi from Different Agroecosystems of Coffea arabica L. in Two Regions of Veracruz, Mexico. *PLoS ONE* **2014**, *9*, e98454. [CrossRef]

14. Granzow, S.; Kaiser, K.; Wemheuer, B.; Pfeiffer, B.; Daniel, R.; Vidal, S.; Wemheuer, F. The Effects of Cropping Regimes on Fungal and Bacterial Communities of Wheat and Faba Bean in a Greenhouse Pot Experiment Differ between Plant Species and Compartment. *Front. Microbiol.* **2017**, *8*. [CrossRef] [PubMed]

15. Mejía, L.C.; Rojas, E.I.; Maynard, Z.; Bael, S.V.; Arnold, A.E.; Hebbar, P.; Samuels, G.J.; Robbins, N.; Herre, E.A. Endophytic fungi as biocontrol agents of Theobroma cacao pathogens. *Biol. Control* **2008**, *46*, 4–14. [CrossRef]

16. Christian, N.; Herre, E.A.; Clay, K. Foliar endophytic fungi alter patterns of nitrogen uptake and distribution in Theobroma cacao. *New Phytol.* **2019**, *222*, 1573–1583. [CrossRef]

17. Leite, H.A.C.; Silva, A.B.; Gomes, F.P.; Gramacho, K.P.; Faria, J.C.; de Souza, J.T.; Loguercio, L.L. Bacillus subtilis and Enterobacter cloacae endophytes from healthy Theobroma cacao L. trees can systemically colonize seedlings and promote growth. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 2639–2651. [CrossRef]

18. Vidal, S.; Jaber, L.R. Entomopathogenic fungi as endophytes: Plant–endophyte–herbivore interactions and prospects for use in biological control. *Curr. Sci.* **2015**, *108*, 1.

19. Arnold, A.E.; Mejia, L.C.; Kyllo, D.; Rojas, E.I.; Maynard, Z.; Robbins, N.; Herre, E.A. Fungal endophytes limit pathogen damage in a tropical tree. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15649–15654. [CrossRef]

20. Melnick, R.L.; Suárez, C.; Bailey, B.A.; Backman, P.A. Isolation of endophytic endospore-forming bacteria from Theobroma cacao as potential biological control agents of cacao diseases. *Biol. Control* **2011**, *57*, 236–245. [CrossRef]

21. Vaast, P.; Somarriba, E. Trade-offs between crop intensification and ecosystem services: The role of agroforestry in cocoa cultivation. *Agrofor. Syst.* **2014**, *88*, 947–956. [CrossRef]

22. Bowers, J.H.; Bailey, B.A.; Hebbar, P.K.; Sanogo, S.; Lumsden, R.D. The impact of plant diseases on world chocolate production. *Plant Health Prog.* **2001**, *2*, 12. [CrossRef]

23. Marelli, J.-P.; Guest, D.I.; Bailey, B.A.; Evans, H.C.; Brown, J.K.; Junaid, M.; Barreto, R.W.; Lisboa, D.O.; Puig, A.S. Chocolate Under Threat from Old and New Cacao Diseases. *Phytopathology* **2019**, *109*, 1331–1343. [CrossRef] [PubMed]

24. Wemheuer, B.; Thomas, T.; Wemheuer, F. Fungal Endophyte Communities of Three Agricultural Important Grass Species Differ in Their Response Towards Management Regimes. *Microorganisms* **2019**, *7*, 37. [CrossRef]

25. Oliveira, R.; Souza, R.; Lima, T.; Cavalcanti, M. Endophytic fungal diversity in coffee leaves (Coffea arabica) cultivated using organic and conventional crop management systems. *Mycosphere* **2014**, *5*, 523–530. [CrossRef]

26. Christian, N.; Sullivan, C.; Visser, N.D.; Clay, K. Plant Host and Geographic Location Drive Endophyte Community Composition in the Face of Perturbation. *Microbial. Ecol.* **2016**, *72*, 621–632. [CrossRef]

27. Wemheuer, F.; Wemheuer, B.; Kretzschmar, D.; Pfeiffer, B.; Herzog, S.; Daniel, R.; Vidal, S. Impact of grassland management regimes on bacterial endophyte diversity differs with grass species. *Lett. Appl. Microbiol.* **2016**, *62*, 323–329. [CrossRef]

28. Klindworth, A.; Pruesse, E.; Schweer, T.; Peplies, J.; Quast, C.; Horn, M.; Glöckner, F.O. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **2013**, *41*, e1. [CrossRef]

29. Wemheuer, B.; Wemheuer, F. Assessing bacterial and fungal diversity in the plants endosphere. In *Metagenomics—Methods and Protocols*, 2nd ed.; Streit, W., Daniel, R., Eds.; Humana Press: Totowa, NJ, USA, 2016.

30. Toju, H.; Tanabe, A.S.; Yamamoto, S.; Sato, H. High-Coverage ITS Primers for the DNA-Based Identification of Ascomycetes and Basidiomycetes in Environmental Samples. *PLoS ONE* **2012**, *7*, e40863. [CrossRef]

31. White, T.J.; Bruns, T.; Lee, S.; Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protoc. Guide Methods Appl.* **1990**, *18*, 315–322.

32. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]

33. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [CrossRef] [PubMed]

34. Edgar, R.C. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. *Bioinformatics* **2018**, *34*, 2371–2375. [CrossRef] [PubMed]

35. Edgar, R.C.; Haas, B.J.; Clemente, J.C.; Quince, C.; Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **2011**, *16*. [CrossRef]

36. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2013**, *41*, D590–D596. [CrossRef]

37. Tedersoo, L.; Sánchez-Ramírez, S.; Kõljalg, U.; Bahram, M.; Döring, M.; Schigel, D.; May, T.; Ryberg, M.; Abarenkov, K. High-level classification of the Fungi and a tool for evolutionary ecological analyses. *Fungal Divers.* **2018**, *90*, 135–159. [CrossRef]

38. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018.

39. Ogle, D.H. FSA: Fisheries Stock Analysis. R Package Version 0.8.14. 2017. Available online: https://cran.r-project.org/web/packages/FSA/index.html (accessed on 25 August 2017).

40. Oksanen, J.; Blanchet, F.G.; Kindt, R.; Legendre, P.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; Stevens, M.H.H.; Wagner, H. Vegan: Community Ecology Package. R Package Version 2.4.-4. 2017. Available online: https://cran.r-project.org/web/packages/vegan/index.html (accessed on 25 August 2017).

41. Kembel, S.W.; Cowan, P.D.; Helmus, M.R.; Cornwell, W.K.; Morlon, H.; Ackerly, D.D.; Blomberg, S.P.; Webb, C.O. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **2010**, *26*, 1463–1464. [CrossRef]

42. Ritz, C.; Baty, F.; Streibig, J.C.; Gerhard, D. Dose-Response Analysis Using R. *PLoS ONE* **2016**, *10*, e0146021. [CrossRef]

43. Marcon, E.; Herault, B. Entropart: An R Package to Measure and Partition Diversity. *J. Stat. Softw.* **2015**, *67*. [CrossRef]

44. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*, 2nd ed.; Sage: Thousand Oaks, CA, USA, 2011.

45. De Cáceres, M.; Legendre, P. Associations between species and groups of sites: Indices and statistical inference. *Ecology* **2009**, *90*, 3566–3574. [CrossRef]

46. Tichy, L.; Chytry, M. Statistical determination of diagnostic species for site groups of unequal size. *J. Veg. Sci.* **2006**, *17*, 809–818. [CrossRef]

47. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]

48. Nguyen, N.H.; Song, Z.; Bates, S.T.; Branco, S.; Tedersoo, L.; Menke, J.; Schilling, J.S.; Kennedy, P.G. FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* **2016**, *20*, 241–248. [CrossRef]

49. Wemheuer, F.; Taylor, J.A.; Daniel, R.; Johnston, E.; Meinicke, P.; Thomas, T.; Wemheuer, B. Tax4Fun2: A R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene marker gene sequences. *bioRxiv* **2018**. [CrossRef]

50. Solis, M.J.L.; Dela Cruz, T.E.; Schnittler, M.; Unterseher, M. The diverse community of leaf-inhabiting fungal endophytes from Philippine natural forests reflects phylogenetic patterns of their host plant species Ficus benjamina, F. elastica and F. religiosa. *Mycoscience* **2016**, *57*, 96–106. [CrossRef]

51. Wemheuer, F.; Wemheuer, B.; Daniel, R.; Vidal, S. Deciphering bacterial and fungal endophyte communities in leaves of two maple trees with green islands. *Sci. Rep.* **2019**, *9*, 14183. [CrossRef]

52. Gazis, R.; Chaverri, P. Diversity of fungal endophytes in leaves and stems of wild rubber trees (Hevea brasiliensis) in Peru. *Fungal Ecol.* **2010**, *3*, 240–254. [CrossRef]

53. Siddique, A.B.; Khokon, A.; Unterseher, M. What do we learn from cultures in the omics age? High-throughput sequencing and cultivation of leaf- inhabiting endophytes from beech (Fagus sylvatica L.) revealed complementary community composition but similar correlations with local habitat conditions. *MycoKeys* **2017**, *20*, 1–16. [CrossRef]

54. Phillips, A.J.L.; Alves, A.; Abdollahzadeh, J.; Slippers, B.; Wingfield, M.J.; Groenewald, J.Z.; Crous, P.W. The Botryosphaeriaceae: Genera and species known from culture. *Stud. Mycol.* **2013**, *76*, 51–167. [CrossRef]

55. Sakalidis, M.L.; Ray, J.D.; Lanoiselet, V.; Hardy, G.E.S.; Burgess, T.I. Pathogenic Botryosphaeriaceae associated with Mangifera indica in the Kimberley Region of Western Australia. *Eur. J. Plant Pathol.* **2011**, *130*, 379–391. [CrossRef]

56. Jakuschkin, B.; Fievet, V.; Schwaller, L.; Fort, T.; Robin, C.; Vacher, C. Deciphering the pathobiome: Intra-and interkingdom interactions involving the pathogen Erysiphe alphitoides. *Microb. Ecol.* **2016**, *72*, 870–880. [CrossRef] [PubMed]

57. Da Siva Fonseca, E.; Peixoto, R.S.; Rosado, A.S.; de Carvalho Balieiro, F.; Tiedje, J.M.; da Costa Rachid, C.T.C. The Microbiome of Eucalyptus Roots under Different Management Conditions and Its Potential for Biological Nitrogen Fixation. *Microb. Ecol.* **2018**, *75*, 183–191. [CrossRef] [PubMed]

58. Griffin, E.A.; Harrison, J.G.; Kembel, S.W.; Carrell, A.A.; Joseph Wright, S.; Carson, W.P. Plant host identity and soil macronutrients explain little variation in sapling endophyte community composition: Is disturbance an alternative explanation? *J. Ecol.* **2019**, *107*, 1876–1889. [CrossRef]

59. Viaene, T.; Langendries, S.; Beirinckx, S.; Maes, M.; Goormachtig, S. Streptomyces as a plant's best friend? *FEMS Microb. Ecol.* **2016**, *92*. [CrossRef] [PubMed]

60. Golinska, P.; Wypij, M.; Agarkar, G.; Rathod, D.; Dahm, H.; Rai, M. Endophytic actinobacteria of medicinal plants: Diversity and bioactivity. *Antonie Van Leeuwenhoek* **2015**, *108*, 267–289. [CrossRef] [PubMed]

61. Lau, M.K.; Arnold, A.E.; Johnson, N.C. Factors influencing communities of foliar fungal endophytes in riparian woody plants. *Fungal Ecol.* **2013**, *6*, 365–378. [CrossRef]

62. Köberl, M.; Dita, M.; Martinuz, A.; Staver, C.; Berg, G. Agroforestry leads to shifts within the gammaproteobacterial microbiome of banana plants cultivated in Central America. *Front. Microbiol.* **2015**, *6*, 91.

63. Helander, M.; Wäli, P.; Kuuluvainen, T.; Saikkonen, K. Birch leaf endophytes in managed and natural boreal forests. *Can. J. For. Res.* **2006**, *36*, 3239–3245. [CrossRef]

64. Zimmerman, N.B.; Vitousek, P.M. Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 13022–13027. [CrossRef]

65.　Arnold, A.E.; Herre, E.A. Canopy cover and leaf age affect colonization by tropical fungal endophytes: Ecological pattern and process in Theobroma cacao (Malvaceae). *Mycologia* **2003**, *95*, 388–398. [CrossRef]

66.　Moulas, C.; Petsoulas, C.; Rousidou, K.; Perruchon, C.; Karas, P.; Karpouzas, D.G. Effects of systemic pesticides imidacloprid and metalaxyl on the phyllosphere of pepper plants. *Biomed. Res. Int.* **2013**, *2013*, 969750. [CrossRef] [PubMed]

67.　Karlsson, I.; Friberg, H.; Steinberg, C.; Persson, P. Fungicide effects on fungal community composition in the wheat phyllosphere. *PLoS ONE* **2014**, *9*, e111786. [CrossRef] [PubMed]

68.　Kerkenaar, A.; Sijpesteijn, A.K. Antifungal activity of metalaxyl and furalaxyl. *Pest. Biochem. Physiol.* **1981**, *15*, 71–78. [CrossRef]

69.　Wilson, D. Ecology of woody plant endophytes. In *Microbial Endophytes*; Bacon, C.W., White, J.F., Eds.; Marcel Dekker, Inc: New York, NY, USA, 2000; pp. 389–420.

70.　Scholtysik, A.; Unterseher, M.; Otto, P.; Wirth, C. Spatio-temporal dynamics of endophyte diversity in the canopy of European ash (Fraxinus excelsior). *Mycol. Prog.* **2012**, *12*. [CrossRef]

71.　Rodriguez, R.; White, J., Jr.; Arnold, A.E.; Redman, A.R.A. Fungal endophytes: Diversity and functional roles. *New Phytol.* **2009**, *182*, 314–330. [CrossRef]

72.　Laforest-Lapointe, I.; Messier, C.; Kembel, S.W. Host species identity, site and time drive temperate tree phyllosphere bacterial community structure. *Microbiome* **2016**, *4*, 27. [CrossRef]

73.　Hardoim, P.R.; van Overbeek, L.S.; Elsas, J.D.V. Properties of bacterial endophytes and their proposed role in plant growth. *Trends Microbiol.* **2008**, *16*, 463–471. [CrossRef]

74.　Card, S.; Johnson, L.; Teasdale, S.; Caradus, J. Deciphering endophyte behaviour: The link between endophyte biology and efficacious biological control agents. *FEMS Microbiol. Ecol.* **2016**, *92*, fiw114. [CrossRef]

75.　Coutinho, I.B.L.; Freire, F.C.O.; Lima, C.S.; Lima, J.S.; Gonçalves, F.J.T.; Machado, A.R.; Silva, A.M.S.; Cardoso, J.E. Diversity of genus Lasiodiplodia associated with perennial tropical fruit plants in northeastern Brazil. *Plant Pathol.* **2017**, *66*, 90–104. [CrossRef]

76.　Yang, Y.; Liu, Z.; Cai, L.; Hyde, K.; Yu, Z.; McKenzie, E. Colletotrichum anthracnose of Amaryllidaceae. *Fungal Divers.* **2009**, *39*, 123–146.

77.　Rossman, A.Y. The Genus Ophionectria (Euascomycetes, Hypocreales). *Mycologia* **1977**, *69*, 355–391. [CrossRef]

78.　Porras-Alfaro, A.; Bayman, P. Hidden Fungi, Emergent Properties: Endophytes and Microbiomes. *Ann. Rev. Phytopathol.* **2011**, *49*, 291–315. [CrossRef] [PubMed]

79.　Vega, F.E.; Simpkins, A.; Aime, M.C.; Posada, F.; Peterson, S.W.; Rehner, S.A.; Infante, F.; Castillo, A.; Arnold, A.E. Fungal endophyte diversity in coffee plants from Colombia, Hawai'i, Mexico and Puerto Rico. *Fungal Ecol.* **2010**, *3*, 122–138. [CrossRef]

80.　Cilliers, A.J.; Swart, W.J.; Wingfield, M.J. A Review of Lasiodiplodia theobromae with Particular Reference to its Occurrence on Coniferous Seeds. *S. Afr. For. J.* **1993**, *166*, 47–52. [CrossRef]

81.　Truyens, S.; Weyens, N.; Cuypers, A.; Vangronsveld, J. Bacterial seed endophytes: Genera, vertical transmission and interaction with plants. *Environ. Microbiol. Rep.* **2015**, *7*, 40–50. [CrossRef]

82.　Vandenkoornhuyse, P.; Quaiser, A.; Duhamel, M.; Le Van, A.; Dufresne, A. The importance of the microbiome of the plant holobiont. *New Phytol.* **2015**, *206*, 1196–1206. [CrossRef]

83.　Sutherland, E.D.; Lockwood, J.L. Hyperparasitism of oospores of some Peronosporales by Actinoplanes missouriensis and Humicola fuscoatra and other Actinomycetes and fungi. *Can. J. Plant Pathol.* **1984**, *6*, 139–145. [CrossRef]

84.　Hirsch, A.; Valdes, M. Micromonospora: An important microbe for biomedicine and potentially for biocontrol and biofuels. *Soil Biol. Biochem.* **2010**, *42*, 536–542. [CrossRef]

85.　Bensch, K.; Braun, U.; Groenewald, J.Z.; Crous, P.W. The genus Cladosporium. *Stud. Mycol.* **2012**, *72*, 1–401. [CrossRef]

86.　Baral, H.O.; Weber, E.; Marson, G.; Quijada, L. A new connection between wood saprobism and beetle endosymbiosis: The rarely reported saprobic discomycete Tromeropsis is congeneric with the symbiotic yeast Symbiotaphrina (Symbiotaphrinales, Xylonomycetes) and two asexual morphs misplaced in Hyphozyma. *Mycol. Prog.* **2018**, *17*, 215–254. [CrossRef]

87. Barnes, C.J.; Maldonado, C.; Frøslev, T.G.; Antonelli, A.; Rønsted, N. Unexpectedly High Beta-Diversity of Root-Associated Fungal Communities in the Bolivian Andes. *Front. Microbiol.* **2016**, *7*. [CrossRef] [PubMed]
88. Taghavi, S.; van der Lelie, D.; Hoffman, A.; Zhang, Y.-B.; Walla, M.D.; Vangronsveld, J.; Newman, L.; Monchy, S. Genome Sequence of the Plant Growth Promoting Endophytic Bacterium Enterobacter sp. 638. *PLoS Genet.* **2010**, *6*, e1000943. [CrossRef] [PubMed]

Supplemental Information for Chapter 5

Contents

Figure S1. Cluster dendrogram for environmental characteristics and agroforestry management systems based on Spearman rank correlation. For further explanation on environmental data see Supplementary Table S1.

Figure S2. Species accumulation curves for fungal (A) and bacterial (B) endophytes in *T. cacao* leaves. The maximal number of zOTUs (n_{max}) was calculated based on Michaelis-Menten Fit.

Table S1. Sample characteristics. Shade cover was measured above-ground. Shade Tree Diversity means number of shades per hectare. The table is deposited on the enclosed CD under \Chapter5_Supplement_information\

Table S2. Final zOTU table for the fungal endophyte community and putative life strategies of fungal endophytes. For further information on the Sample ID see Supplementary Table S1. The table is deposited on the enclosed CD under \Chapter5_Supplement_information\

Table S3. Final zOTU table for the bacterial endophyte community. For further information on the Sample ID see Supplementary Table S1. The table is deposited on the enclosed CD under \Chapter5_Supplement_information\

Table S4. Statistical analysis of environmental data with Kruskal-Wallis test. Multiple comparisons were performed using the Dunn test. P-values were adjusted with the Benjamini-Hochberg method. Statistically significant ($p \leq 0.05$) and marginally significant ($p \leq 0.1$) values are written in bold and are underlined, respectively.

Table S5. Sequence characteristics and alpha diversity values of the fungal dataset. Alpha diversity values were calculated at the same surveying effort (number of sequences = 2025). Every value was calculated 10 times in R. The average of all 10 iterations is provided. For further information on the Sample ID see Supplementary Table S1.

Table S6. Sequence characteristics and alpha diversity values of the prokaryotic dataset. Alpha diversity values were calculated at the same surveying effort (number of sequences = 3105). Every value was calculated 10 times in R. The average of all 10 iterations is provided. For further information on the Sample ID see Supplementary Table S1.

Table S7. Differences in the relative abundances of foliar endophytes between sites. Differences were tested by pairwise comparison using t tests with pooled SD. P value adjustment method: Benjamini-Hochberg. Statistically significant ($p \leq 0.05$) and marginally significant ($p \leq 0.1$) values are written in bold and are underlined, respectively.

Table S8. Microbial core community and significantly ($p < 0.01$) associated bacterial and fungal zOTUs with the seven sites. The table is deposited on the enclosed CD under \Chapter5_Supplement_information\

Table S9. Predicted pathways and unknown fraction for each sample. We focused on important pathways involved in metabolism, environmental information processing and organismal systems. The table is deposited on the enclosed CD under \Chapter5_Supplement_information\
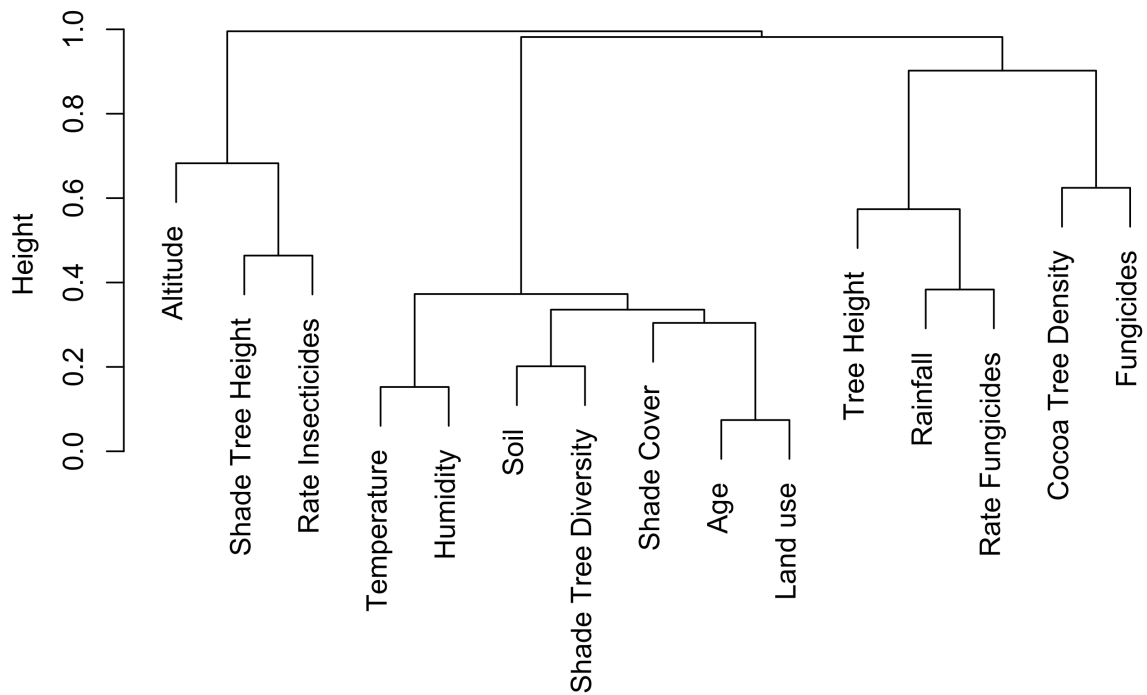
Figure S1. Cluster dendrogram for environmental characteristics and agroforestry management systems based on Spearman rank correlation. For further explanation on environmental data see Supplementary Table S1.
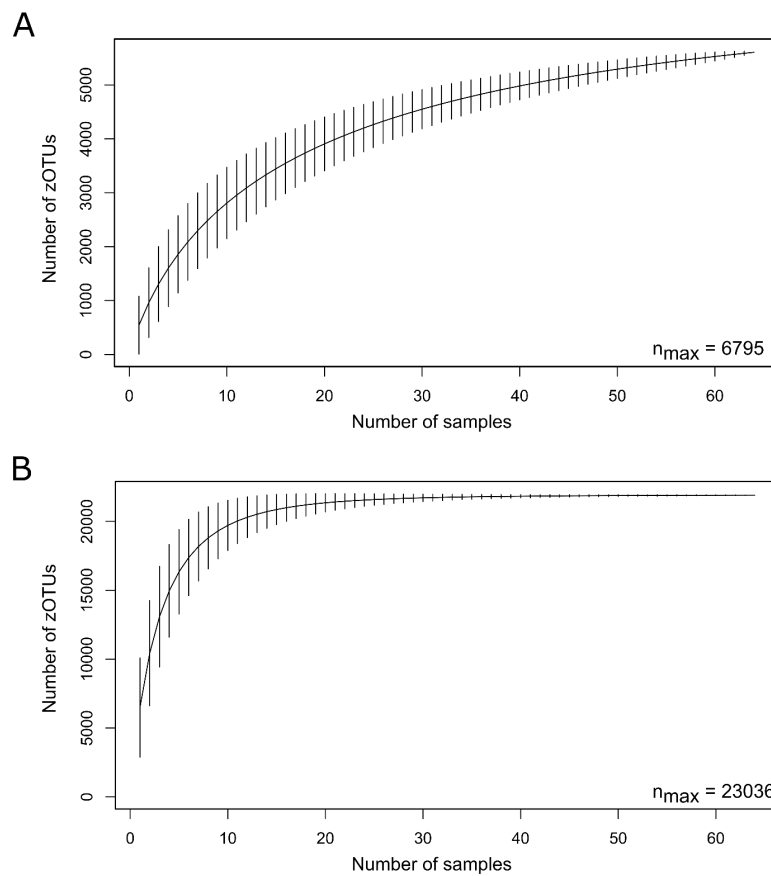
Figure S2. Species accumulation curves for fungal (A) and bacterial (B) endophytes in *T. cacao* leaves. The maximal number of zOTUs (n_{max}) was calculated based on Michaelis-Menten Fit.

Table S4. Statistical analysis of environmental data with Kruskal-Wallis test. Multiple comparison were performed using the Dunn test. P-values were adjusted with the Benjamini-Hochberg method. Statistically significant (p ≤ 0.05) and marginally significant (p ≤ 0.1) values are written in bold and are underlined, respectively.

| Multiple Comparison	Plantation Age		Tree Height		Altitude		Temperature		Humidity		Rainfall		Cacao Tree Density		Rate Insecticides		Rate Fungicides		Shade Tree Diversity		Shade Cover		Shade Tree Height	
Kruskal-Wallis test (DF = 6)	chi2	P value	chi2	P value	chi2	P value	chi2	P value	chi2	P value	chi2	P value	chi2	P value	chi2	P value	chi2	P value	chi2	P value	chi2	P value	chi2	P value
	63	0	37.24	0	48.28	0	46.76	0	48.01	0	58.56	0	34.08	0	50.35	0	38.08	0	51.68	0	43.18	0	51.3	0
	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj	Z	P.adj
Bakoa - Ekabia Essélé	-1.32	0.21	-0.15	0.93	1.83	0.13	1.72	0.15	-2.39	0.03	-3.19	0.004	-2.41	0.04	-1.75	0.14	0.91	0.54	-2.44	0.03	-1.72	0.11	-4.36	<0.001
Bakoa - Kédia	2.65	0.02	1.96	0.09	0.43	0.70	-2.58	0.02	1.78	0.13	-0.54	0.62	1.03	0.40	-0.22	0.87	0.43	0.74	0.76	0.50	2.15	0.07	-0.11	0.96
Ekabia Essélé - Kédia	3.97	<0.001	2.10	0.07	-1.40	0.26	-4.30	<0.001	4.17	<0.001	2.65	0.01	3.44	0.003	1.54	0.17	-0.48	0.73	3.20	0.004	3.87	<0.001	4.25	<0.001
Bakoa - Nkolobang	-1.32	0.22	-1.17	0.32	-1.51	0.23	1.61	0.17	-2.72	0.02	-1.67	0.12	-2.95	0.01	-1.59	0.17	1.18	0.38	-2.49	0.03	-1.72	0.12	-4.14	<0.001
Ekabia Essélé - Nkolobang	0.00	1.00	-1.02	0.36	-3.33	0.004	-0.11	0.91	-0.32	0.78	1.51	0.16	-0.54	0.73	0.16	0.87	0.27	0.83	-0.05	0.96	0.00	1.00	0.22	0.92
Kédia - Nkolobang	-3.97	<0.001	-3.12	0.01	-1.94	0.11	4.20	<0.001	-4.49	<0.001	-1.13	0.30	-3.98	<0.001	-1.38	0.22	0.75	0.56	-3.25	0.003	-3.87	0.001	-4.03	<0.001
Bakoa - Pan Makak	-3.09	0.01	3.30	0.003	0.86	0.48	2.15	0.07	-1.53	0.20	-5.02	<0.001	-0.51	0.67	1.11	0.31	-2.15	0.06	-4.12	<0.001	-2.26	0.06	-1.77	0.11
Ekabia Essélé - Pan Makak	-1.76	0.11	3.45	0.003	-0.97	0.44	0.43	0.70	0.86	0.51	-1.84	0.09	1.90	0.12	2.86	0.01	-3.06	0.01	-1.68	0.12	-0.54	0.65	2.58	0.02
Kédia - Pan Makak	-5.73	<0.001	1.34	0.27	0.43	0.74	4.73	<0.001	-3.31	0.003	-4.48	<0.001	-1.54	0.18	1.32	0.23	-2.58	0.02	-4.88	<0.001	-4.41	<0.001	-1.67	0.13
Nkolobang - Pan Makak	-1.76	0.12	4.47	<0.001	2.37	0.05	0.54	0.69	1.18	0.33	-3.35	0.002	2.44	0.04	2.70	0.02	-3.33	0.003	-1.63	0.13	-0.54	0.69	2.37	0.03
Bakoa - Simanya	-3.09	0.01	3.33	0.004	-0.32	0.75	1.18	0.35	-2.18	0.06	-5.89	<0.001	-0.51	0.71	1.64	0.16	-2.90	0.01	-4.66	<0.001	-3.44	0.002	-1.99	0.08
Ekabia Essélé - Simanya	-1.76	0.13	3.48	0.004	-2.15	0.08	-0.54	0.73	0.22	0.83	-2.70	0.01	1.90	0.14	3.40	0.003	-3.82	0.001	-2.22	0.05	-1.72	0.13	2.37	0.04
Kédia - Simanya	-5.73	<0.001	1.38	0.27	-0.75	0.53	3.77	0.001	-3.95	<0.001	-5.35	<0.001	-1.54	0.20	1.86	0.12	-3.33	0.004	-5.42	<0.001	-5.59	<0.001	-1.88	0.09
Nkolobang - Simanya	-1.76	0.14	4.50	<0.001	1.18	0.33	-0.43	0.74	0.54	0.65	-4.21	<0.001	2.44	0.05	3.24	0.004	-4.09	0.001	-2.17	0.05	-1.72	0.14	2.15	0.06
Pan Makak - Simanya	0.00	1.00	0.03	0.97	-1.18	0.36	-0.97	0.44	-0.65	0.64	-0.86	0.43	0.00	1.00	0.54	0.65	-0.75	0.59	-0.54	0.62	-1.18	0.29	-0.22	0.97
Bakoa - Talba	1.53	0.16	0.95	0.38	4.22	<0.001	-1.86	0.12	1.46	0.22	-3.80	<0.001	1.22	0.31	-3.89	0.001	-2.51	0.02	-1.03	0.35	0.06	1.00	-5.03	<0.001
Ekabia Essélé - Talba	3.06	0.005	1.13	0.32	2.11	0.08	-3.85	<0.001	4.22	<0.001	-0.12	0.90	4.00	0.001	-1.87	0.13	-3.57	0.002	1.78	0.10	2.05	0.07	0.00	1.00
Kédia - Talba	-1.53	0.17	-1.30	0.27	3.73	0.001	1.12	0.37	-0.59	0.65	-3.18	0.003	0.03	1.00	-3.64	0.001	-3.01	0.01	-1.91	0.09	-2.42	0.04	-4.91	<0.001
Nkolobang - Talba	3.06	0.01	2.30	0.04	5.96	<0.001	-3.73	0.001	4.60	<0.001	-1.87	0.09	4.63	<0.001	-2.06	0.09	-3.88	0.001	1.85	0.10	2.05	0.08	-0.25	0.99
Pan Makak - Talba	5.09	<0.001	-2.86	0.01	3.23	0.004	-4.35	<0.001	3.23	0.003	2.00	0.07	1.81	0.12	-5.17	<0.001	-0.03	0.98	3.72	0.001	2.67	0.02	-2.98	0.01
Simanya - Talba	5.09	<0.001	-2.89	0.01	4.60	<0.001	-3.23	0.003	3.98	<0.001	2.99	0.01	1.81	0.13	-5.79	<0.001	0.84	0.56	4.35	<0.001	4.04	<0.001	-2.73	0.02

Table S5. Sequence characteristics and alpha diversity values of the fungal dataset. Alpha diversity values were calculated at the same surveying effort (number of sequences = 2025). Every value was calculated 10 times in R. The average of all 10 iterations is provided. For further information on the Sample ID see Supplementary Table S1.

SeqID_Fungi	Tree ID	Raw sequences	After processing and denoising	After removal of chimeric and non-fungal sequences	Observed zOTUs (Richness)	Shannon diversity index H'	Michaelis-Menten Fit (MMF)	Good's Coverage (%)
Cocoa_Fungi_01	1	77185	63886	63145	71	1.67	99.25	99.91
Cocoa_Fungi_02	2	35580	20260	19100	314	4.61	449.51	99.36
Cocoa_Fungi_03	3	70469	30862	29547	276	4.60	357.35	99.79
Cocoa_Fungi_04	4	33026	19654	18652	256	4.77	314.57	99.68
Cocoa_Fungi_05	5	24406	13016	11944	354	4.82	493.48	99.01
Cocoa_Fungi_06	6	5590	2340	2066	314	4.75	424.29	94.58
Cocoa_Fungi_07	7	36553	18251	17292	332	4.98	443.32	99.66
Cocoa_Fungi_08	8	41988	34083	33721	87	1.53	143.40	99.56
Cocoa_Fungi_09	9	30487	13970	12537	443	5.29	651.97	99.16
Cocoa_Fungi_10	10	35320	19645	17964	425	5.34	598.69	99.40
Cocoa_Fungi_11	11	43472	28394	26053	287	4.44	400.56	99.55
Cocoa_Fungi_12	12	44110	11290	11097	133	1.72	218.96	99.37
Cocoa_Fungi_13	13	37586	25957	24622	334	4.69	474.73	99.59
Cocoa_Fungi_14	14	26182	10394	9826	224	3.06	331.84	99.25
Cocoa_Fungi_15	15	97452	71179	66326	307	4.78	396.02	99.82
Cocoa_Fungi_16	16	46481	35910	32810	307	4.42	465.44	99.45
Cocoa_Fungi_17	17	109343	86281	77684	153	3.57	187.04	99.98
Cocoa_Fungi_18	18	54324	42368	38159	272	4.19	383.14	99.69
Cocoa_Fungi_19	19	65758	46219	42067	332	4.67	480.92	99.69
Cocoa_Fungi_20	20	28564	14432	13296	318	4.83	427.15	99.24
Cocoa_Fungi_21	21	151131	106591	95436	280	4.86	346.76	99.90
Cocoa_Fungi_22	22	123161	98067	91280	286	4.27	401.08	99.86
Cocoa_Fungi_23	23	69604	52414	46610	207	3.59	299.95	99.71
Cocoa_Fungi_24	24	76345	54580	47315	307	4.45	441.85	99.74
Cocoa_Fungi_25	25	25014	18776	16905	303	4.41	428.82	99.56
Cocoa_Fungi_26	26	42032	31100	28348	348	4.68	518.98	99.40
Cocoa_Fungi_27	27	80988	44583	41512	258	4.10	353.80	99.80
Cocoa_Fungi_28	28	20804	15940	13807	316	4.57	454.91	98.79
Cocoa_Fungi_29	29	50274	34106	29197	339	4.86	475.78	99.57
Cocoa_Fungi_30	30	116067	78718	69385	380	4.84	564.87	99.87
Cocoa_Fungi_31	31	51942	38136	34419	296	4.65	387.97	99.70
Cocoa_Fungi_32	32	73219	46641	38661	271	4.43	365.65	99.74
Cocoa_Fungi_33	33	75870	63071	62259	75	1.32	121.83	99.80
Cocoa_Fungi_34	34	123636	99417	90467	263	3.61	428.27	99.82
Cocoa_Fungi_35	35	64260	45769	40128	309	4.58	442.62	99.63
Cocoa_Fungi_36	36	10164	7721	3733	189	3.70	241.35	97.62
Cocoa_Fungi_37	37	76746	52235	43640	274	4.37	375.90	99.67
Cocoa_Fungi_38	38	116458	90327	77490	298	3.82	458.47	99.77
Cocoa_Fungi_39	39	120528	94794	83207	281	4.13	403.75	99.70

Cocoa_Fungi_40	40	56492	45135	43359	161	2.30	266.18	99.65
Cocoa_Fungi_41	41	65004	46693	40212	326	4.81	452.21	99.51
Cocoa_Fungi_42	42	179592	139262	133391	208	3.78	286.61	99.93
Cocoa_Fungi_43	43	210911	154730	144753	225	3.51	345.83	99.93
Cocoa_Fungi_44	44	137396	97818	92627	322	4.35	467.78	99.88
Cocoa_Fungi_45	45	93879	66401	62594	386	4.51	624.32	99.81
Cocoa_Fungi_46	46	139086	93988	86952	442	5.21	679.71	99.85
Cocoa_Fungi_47	47	106442	67270	62540	433	5.20	661.23	99.73
Cocoa_Fungi_48	48	39302	14362	13637	300	3.96	440.04	99.65
Cocoa_Fungi_49	49	26901	8491	8273	98	1.67	144.61	99.66
Cocoa_Fungi_50	50	124200	38855	37944	197	2.41	319.89	99.89
Cocoa_Fungi_51	51	32816	6078	5957	181	2.57	256.06	99.29
Cocoa_Fungi_52	52	11118	2083	2025	106	1.92	146.19	97.98
Cocoa_Fungi_53	53	13470	3884	3800	132	2.67	174.61	98.87
Cocoa_Fungi_54	54	103144	45732	43658	315	4.11	492.81	99.84
Cocoa_Fungi_55	55	21554	2617	2494	142	2.96	175.02	98.48
Cocoa_Fungi_56	56	23858	3180	2882	223	4.47	267.77	98.54
Cocoa_Fungi_57	57	48845	31402	28436	379	4.73	575.50	99.75
Cocoa_Fungi_58	58	31595	9524	9111	178	2.93	258.56	99.24
Cocoa_Fungi_59	59	68662	15356	14837	171	2.51	251.47	99.68
Cocoa_Fungi_60	60	65023	17699	17348	173	2.33	260.42	99.82
Cocoa_Fungi_61	61	60127	23969	23697	113	2.09	148.45	99.89
Cocoa_Fungi_62	62	57903	39125	38687	188	3.81	231.80	99.88
Cocoa_Fungi_63	63	49660	23564	23360	114	3.02	132.71	99.86
Cocoa_Fungi_64	64	7390	4419	4362	141	4.47	150.22	99.54

Table S6. Sequence characteristics and alpha diversity values of the prokaryotic dataset. Alpha diversity values were calculated at the same surveying effort (number of sequences = 3105). Every value was calculated 10 times in R. The average of all 10 iterations is provided. For further information on the Sample ID see Supplementary Table S1.

SeqID_Bacteria	Tree ID	Raw sequences	After processing and denoising	After removal of chimeric and non-prokaryotic sequences	Observed zOTUs (Richness)	Shannon diversity index H'	Michaelis-Menten Fit (MMF)	Good's Coverage (%)
Cocoa_Prok_01	1	56707	30739	25913	1874	7.26	5420.95	91.07
Cocoa_Prok_02	2	80885	44769	34215	1788	7.06	5404.71	92.35
Cocoa_Prok_03	3	59005	30191	23387	1684	7.06	4440.39	90.35
Cocoa_Prok_04	4	113128	64171	50627	1785	7.14	5090.64	94.84
Cocoa_Prok_05	5	97477	49893	38514	1785	7.14	5121.87	92.77
Cocoa_Prok_06	6	90271	46306	34576	1943	7.28	6287.44	90.47
Cocoa_Prok_07	7	69147	37661	28853	1911	7.23	6026.00	89.68
Cocoa_Prok_08	8	173395	93988	69753	1739	7.07	4812.88	96.08
Cocoa_Prok_09	9	177581	95070	69455	1786	7.12	5190.02	96.57
Cocoa_Prok_10	10	109309	58064	44080	1877	7.22	5755.63	93.07
Cocoa_Prok_11	11	42104	17565	13099	1774	7.11	5150.51	82.28

Cocoa_Prok_12	12	74073	39377	29229	1737	7.04	5030.44	91.24
Cocoa_Prok_13	13	8768	4509	3105	1706	7.12	4360.25	64.19
Cocoa_Prok_14	14	70293	39918	30374	1911	7.31	5622.50	92.10
Cocoa_Prok_15	15	72960	40368	29981	1873	7.28	5317.07	91.94
Cocoa_Prok_16	16	50514	28601	22029	1944	7.31	6027.46	88.67
Cocoa_Prok_17	17	31991	17243	14188	1800	7.19	5055.14	84.19
Cocoa_Prok_18	18	62820	33934	24258	1732	7.09	4747.56	89.34
Cocoa_Prok_19	19	107706	61085	48574	1640	6.98	4287.95	94.91
Cocoa_Prok_20	20	134924	76684	61034	1723	7.14	4453.92	95.87
Cocoa_Prok_21	21	94915	50188	39614	1861	7.25	5381.13	93.89
Cocoa_Prok_22	22	105891	59495	47194	1665	7.01	4394.81	95.13
Cocoa_Prok_23	23	91331	50412	40767	1661	7.08	4151.02	94.84
Cocoa_Prok_24	24	108172	53243	39553	1822	7.19	5232.66	93.95
Cocoa_Prok_25	25	141582	73637	55043	1977	7.34	6298.06	94.51
Cocoa_Prok_26	26	153371	81307	62688	2096	7.44	7375.82	94.71
Cocoa_Prok_27	27	124230	62108	51221	1969	7.32	6297.11	94.29
Cocoa_Prok_28	28	26283	13416	11088	1935	7.29	6095.45	77.65
Cocoa_Prok_29	29	142790	74820	55368	1737	7.06	4953.77	95.93
Cocoa_Prok_30	30	40763	20531	12089	1621	7.00	4071.13	84.23
Cocoa_Prok_31	31	203439	106954	81806	1844	7.18	5594.10	97.09
Cocoa_Prok_32	32	58985	29988	24612	1779	7.13	5121.55	89.21
Cocoa_Prok_33	33	192038	99290	73909	1824	7.24	5013.10	97.05
Cocoa_Prok_34	34	172808	91027	64733	1926	7.31	5794.17	96.45
Cocoa_Prok_35	35	150678	78956	59674	2059	7.40	7133.04	94.60
Cocoa_Prok_36	36	76009	39123	28106	1628	7.03	4019.71	92.12
Cocoa_Prok_37	37	26668	14575	12390	1803	7.17	5140.23	82.56
Cocoa_Prok_38	38	152580	75577	60804	1548	6.79	4090.28	95.82
Cocoa_Prok_39	39	122524	61076	49295	1749	7.14	4736.80	94.97
Cocoa_Prok_40	40	81432	46088	37061	1772	7.14	4967.47	92.18
Cocoa_Prok_41	41	49801	27640	23323	1848	7.25	5225.25	89.37
Cocoa_Prok_42	42	62675	33964	28296	1829	7.21	5221.06	90.69
Cocoa_Prok_43	43	94321	53442	40974	1594	6.98	3945.51	94.37
Cocoa_Prok_44	44	42563	23896	18488	1784	7.12	5125.54	87.71
Cocoa_Prok_45	45	176270	100649	84048	1920	7.23	6187.97	96.98
Cocoa_Prok_46	46	78037	43859	33667	1801	7.11	5385.41	92.15
Cocoa_Prok_47	47	86545	49015	28810	1681	6.94	4763.11	92.45
Cocoa_Prok_48	48	59905	33009	18733	1797	7.08	5336.16	88.08
Cocoa_Prok_49	49	202967	109861	51877	1819	7.09	5771.31	94.61
Cocoa_Prok_50	50	113407	41713	35681	1461	6.55	3759.41	94.41
Cocoa_Prok_51	51	127576	68326	36097	1824	7.20	5171.90	93.40
Cocoa_Prok_52	52	90323	46269	25323	1719	7.03	4664.43	91.32
Cocoa_Prok_53	53	107248	62369	34057	1206	6.13	2794.66	94.51
Cocoa_Prok_54	54	164248	92162	45908	1666	7.00	4337.73	95.29
Cocoa_Prok_55	55	97065	55557	44949	1631	6.97	4253.55	95.16
Cocoa_Prok_56	56	113080	62660	41793	2051	7.38	7100.21	92.97
Cocoa_Prok_57	57	139117	75518	60419	2037	7.38	6905.53	95.27
Cocoa_Prok_58	58	101102	55254	46560	1923	7.28	6040.20	94.29
Cocoa_Prok_59	59	89443	49334	26271	1761	7.11	4860.63	91.11

Cocoa_Prok_60	60	42333	22828	11477	1814	7.13	5325.36	81.67
Cocoa_Prok_61	61	27987	13870	7593	1878	7.15	6067.57	73.01
Cocoa_Prok_62	62	61617	34004	20622	1669	7.00	4450.08	89.78
Cocoa_Prok_63	63	33168	18702	11376	1474	6.77	3579.50	84.99
Cocoa_Prok_64	64	32445	18242	9700	1724	7.05	4692.93	81.95

Table S7. Differences in the relative abundances of foliar endophytes between sites. Differences were tested by pairwise comparison using t tests with pooled SD. P value adjustment method: Benjamini-Hochberg. Statistically significant (p ≤ 0.05) and marginally significant (p ≤ 0.1) values are written in bold and are underlined, respectively.

Tremellales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.74	-	-	-	-	-
Kédia	0.56	0.39	-	-	-	-
Nkolobang	0.56	0.75	0.34	-	-	-
Pan Makak	0.56	0.74	0.34	0.91	-	-
Simanya	0.74	0.56	0.74	0.39	0.39	-
Talba	0.75	0.56	0.56	0.39	0.39	0.89

Symbiotaphrinales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.64	-	-	-	-	-
Kédia	0.64	0.97	-	-	-	-
Nkolobang	0.16	0.21	0.21	-	-	-
Pan Makak	0.99	0.64	0.64	0.16	-	-
Simanya	0.21	0.51	0.49	0.64	0.21	-
Talba	0.21	0.55	0.51	0.51	0.21	0.76

Xylariales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.99	-	-	-	-	-
Kédia	<u>0.08</u>	<u>0.07</u>	-	-	-	-
Nkolobang	0.99	0.99	<u>0.07</u>	-	-	-
Pan Makak	<u>0.07</u>	<u>0.07</u>	0.99	<u>0.07</u>	-	-
Simanya	0.99	0.99	<u>0.09</u>	0.99	<u>0.07</u>	-
Talba	0.99	0.99	<u>0.07</u>	0.99	<u>0.07</u>	0.99

Hypocreales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.80	-	-	-	-	-
Kédia	0.57	0.79	-	-	-	-
Nkolobang	0.68	0.79	0.97	-	-	-
Pan Makak	0.79	0.97	0.79	0.79	-	-
Simanya	0.79	0.97	0.79	0.79	0.97	-
Talba	**0.01**	**0.02**	0.14	<u>0.10</u>	**0.02**	**0.02**

Glomerellales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.85	-	-	-	-	-
Kédia	0.85	0.85	-	-	-	-
Nkolobang	0.85	0.85	0.94	-	-	-
Pan Makak	0.41	0.59	0.50	0.50	-	-
Simanya	0.54	0.79	0.66	0.67	0.85	-
Talba	0.38	0.41	0.38	0.38	0.85	0.67

Lecanorales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.37	-	-	-	-	-
Kédia	0.37	0.99	-	-	-	-
Nkolobang	0.57	<u>0.07</u>	<u>0.07</u>	-	-	-
Pan Makak	0.63	0.76	0.76	0.13	-	-
Simanya	0.63	0.76	0.76	0.13	0.99	-
Talba	0.62	0.76	0.76	<u>0.09</u>	0.99	0.99

Eurotiales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.74	-	-	-	-	-
Kédia	**0.04**	<u>0.09</u>	-	-	-	-
Nkolobang	0.48	0.74	0.19	-	-	-
Pan Makak	**0.04**	<u>0.08</u>	0.94	0.18	-	-
Simanya	0.48	0.74	0.19	0.97	0.18	-
Talba	0.74	0.94	**0.04**	0.65	**0.04**	0.65

Chaetothyriales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.52	-	-	-	-	-
Kédia	0.40	0.78	-	-	-	-
Nkolobang	0.40	0.78	0.95	-	-	-
Pan Makak	0.81	0.78	0.52	0.52	-	-
Simanya	0.95	0.54	0.40	0.40	0.82	-
Talba	0.23	0.52	0.78	0.78	0.37	0.23

Strigulales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.99	-	-	-	-	-
Kédia	0.99	0.99	-	-	-	-
Nkolobang	0.99	0.99	0.99	-	-	-
Pan Makak	0.99	0.99	0.99	0.99	-	-
Simanya	0.99	0.99	0.99	0.99	0.99	-
Talba	0.27	0.27	0.27	0.27	0.27	0.27

Pleosporales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.31	-	-	-	-	-

Kédia	0.98	0.31	-	-	-	-
Nkolobang	0.45	0.72	0.45	-	-	-
Pan Makak	0.58	0.58	0.58	0.74	-	-
Simanya	0.44	<u>0.07</u>	0.44	0.14	0.24	-
Talba	0.72	0.39	0.72	0.58	0.72	0.26

Myriangiales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.30	-	-	-	-	-
Kédia	0.30	0.89	-	-	-	-
Nkolobang	0.30	0.89	0.89	-	-	-
Pan Makak	0.41	0.89	0.89	0.89	-	-
Simanya	0.56	0.89	0.89	0.89	0.89	-
Talba	0.30	0.89	0.89	0.89	0.89	0.89

Capnodiales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	0.50	-	-	-	-	-
Kédia	0.81	0.81	-	-	-	-
Nkolobang	0.52	0.96	0.81	-	-	-
Pan Makak	0.81	0.81	0.96	0.85	-	-
Simanya	0.50	0.96	0.81	0.96	0.81	-
Talba	0.67	0.81	0.96	0.85	0.97	0.81

Botryosphaeriales

	Bakoa	Ekabita Essélé	Kédia	Nkolobang	Pan Makak	Simanya
Ekabita Essélé	**0.04**	-	-	-	-	-
Kédia	0.10	0.76	-	-	-	-
Nkolobang	**0.04**	0.98	0.76	-	-	-
Pan Makak	0.78	<u>0.09</u>	0.32	<u>0.09</u>	-	-
Simanya	<u>0.05</u>	0.95	0.78	0.95	0.10	-
Talba	**0.04**	0.93	0.78	0.93	<u>0.09</u>	0.96

6

Discussion

In this study, the impact of rainforest conversion and anthropogenic land use changes has been investigated. Conversions of forest to agroforestry and two monoculture plantation systems have been studied by taxonomic marker gene and direct metagenome analyses. This allows an evaluation of results from different perspectives and inclusion of taxonomic and functional potential of those microbial communities.

The importance of resource management and the impact of agriculture on the environment also leads to questions regarding established land use systems and possible impacts of the applied management practices. To assess possible effects of reduced management intensity on soilborne communities, different oil palm management intensities regarding fertilizer and herbicide usage were analysed.

6.1 Effects of rainforest conversion on microbial community composition

6.1.1 Taxonomic profiles change with increasing land use intensity

Soil is a very heterogenous habitat and harbours a high microbial diversity (Vos et al., 2013; Kaiser et al., 2016; Fonseca et al., 2018). Nevertheless, general patterns of soil community composition are regularly found. Temperate and tropical soil is mostly dominated by *Proteobacteria*, *Acidobacteria* and *Actinobacteria*. This has been described for numerous environments, including grassland, peatland, and agricultural land use systems (Lauber et al., 2009; Rousk et al., 2010; Nacke et al., 2011; Schneider et al., 2015). In this study, microbial communities mostly comprised of bacteria, followed by fungi and *Archaea* (both less than 1% relative abundance) (Chapter 3). Metagenomic shotgun data also showed large fractions of sequences that could not be classified (>30%). This could be caused by sequences belonging to novel organisms, which are not represented in the used database or sequences, which are not suitable for taxonomic classification (see 6.1.3).

Bacteria were dominated by the previously mentioned phyla, but relative abundances and proportions changed in the course of increasing land use intensity (Figure 6.1). Schneider et al. (2015) showed high abundances of *Proteobacteria* with *Rhizobiales* and *Burkholderiales* as very abundant in rainforest soils. Their abundance decreased with increasing land use intensity. *Acidobacteria* (*Acidobacteriales* and Subgroup 2 in particular) showed the opposite trend. In this study, metagenome analysis of samples from the same plots (Chapter 3) roughly confirmed these results and revealed similar community compositions. Similar patterns were observed for RNA-based 16S rRNA transcript analysis (Figure 6.1; Chapter 2). Most dominant orders within the *Proteobacteria* in all samples were *Rhizobiales* and *Burkholderiales*. Both taxa are associated with processes related to nitrogen metabolism and are involved in fixation of atmospheric nitrogen, nitrification and denitrification (Yin et al., 2015; Li et al., 2019). They exhibit very different lifestyles either as endophytes, free-living cells or symbionts in the root nodule (Hayat et al., 2010; Taylor et al., 2012). In the scenario of rainforest conversion, it seems plausible that these groups decrease in abundance upon increasing management intensity in form of fertilizer input and reduction of plant diversity. Nitrogen sources provided via fertilizer application, despite a higher susceptibility to nutrient losses, might create an environment in which microbial acquisition of nutrients (like nitrogen) is not favoured anymore. This might affect these taxa negatively or in turn support competitors of these decreasing taxa. Other groups seemed to benefit from land use conversion and accompanied increasing management intensity. Even though little is known about the ecological role of many *Acidobacteria*, members of the *Acidobacteriales* were suspected to be involved in nitrate reduction (Ward et al., 2009; Pickett et al., 2019). Some members of *Acidobacteria* and *Actinobacteria* were described as heat resistant (Bouskill et al., 2012). Less canopy density in oil palm and rubber plantations lead to more exposure to sunlight and increased topsoil temperatures (Schneider et al., 2015). *Cyanobacteria*, which are capable of photosynthesis, were another group that showed higher abundances in the active community of oil palm soils. This indeed points to higher availability of light through changes in the canopy of trees but also to a rather short-termed effect, since *Cyanobacteria* were not detected as abundant (>1% in total) in the metagenome dataset (Chapter 3) or previous DNA-based analysis in these samples, but only in the active community as shown in Chapter 2 (Schneider et al., 2015; Kvíderová and Kumar, 2020). *Cyanobacteria* are also reported for nitrogen fixation in aquatic environments and soil crusts (Patova et al., 2016; Molot, 2017). Thus, they are studied as a possible supplement for agricultural cultivation of crops like rice and peas, but have not been mentioned in oil palm related studies so far (Bidyarani et al., 2016; Chittapun et al., 2018; Peng and Bruns, 2019). Approximately 5% of all detected

nitrogenase genes in oil palm soils in the metagenome dataset originated from *Cyanobacteria*. Considering that *Cyanobacteria* only showed a notable increase of abundance in oil palm soils, it is indicated that a combination of higher light availability and fertilizer input supported cyanobacterial abundance.
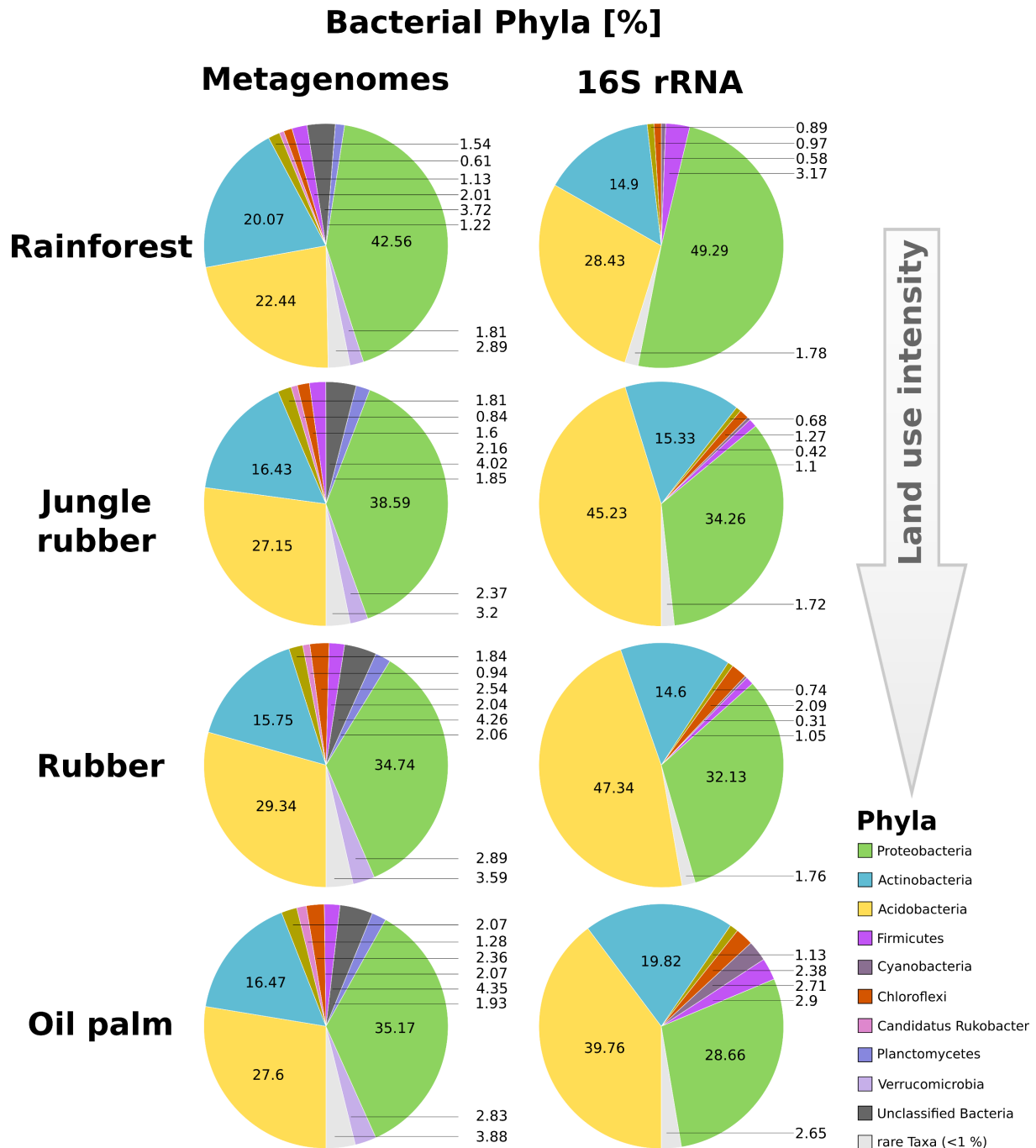
Figure 6.1: Changes of bacterial phyla abundances in rainforest samples and different land use systems obtained from metagenomic shotgun sequencing (Chapter 3) and 16S rRNA transcript derived amplicons (Chapter 2). Different phyla are shown as relative abundances of the respective dataset. Phyla with abundances <1% were clustered as "rare taxa".

In the oil palm management experiment, community compositions did not differ among different treatments, but further supported previously observed community structures in managed oil palm soils of this study and previous studies (Chapter 4) (Schneider et al., 2015; Tripathi et al., 2016).

6.1.2 Differences between RNA-based marker gene analysis of the active community and DNA-based entire community analysis by metagenomic shotgun sequencing

Most community analyses are based on DNA. These analyses are easier to handle than RNA due to the template stability but also include sequences derived from dormant or even dead cells. Thus, DNA-based studies provide no information on which members of the microbial community are in an active state (Urich et al., 2008; Wemheuer et al., 2012). The active community assessed via environmental RNA could be different in structure. The provided study on the effects on the active bacterial community (Chapter 2) is related to the study by Schneider et al. (2015), which analysed the microbial communities in same plots via DNA-based 16S rRNA gene analysis. . It should be noted, that the study by Schneider et al. (2015) employed a different sequencing technology compared to this study (454 pyrosequencing), different sets of primers and different bioinformatic data analysis approaches. Considering these limitations of comparison, similarities in community structure appear stable across different methods. Notably, some taxa which were heavily affected by rainforest conversion at DNA level, showed less response at RNA level. The abundant *Burkholderiales*, while still decreasing in abundance with higher land use intensity, showed lower abundances in Chapter 2 compared to Schneider et al. and lower decreases in abundance in the active community. *Rhizobiales* were generally more abundant in the active community (Chapter 2) than in the DNA-based metagenome dataset (Chapter 3) with higher abundance decreases (from forest to land uses), indicating a higher activity of *Rhizobiales* than previously assumed.

6.1.3 Community compositions and taxonomic assignments can differ between marker gene and metagenome analysis due to sequence characteristics and the chosen database

The two studies regarding rainforest conversion (Chapter 2 and 3) revealed some differences in taxonomic community structure. Although community structures were roughly similar, some

groups were more affected by the two different approaches than others. *Frankiales* of the *Actinobacteria* are abundant in amplicon-based studies but were very rare (less than <1%) in the direct sequencing derived dataset. Both studies were conducted using different methodological approaches and databases. The active bacterial community (Chapter 2) was analysed by using RNA-based 16S rRNA marker gene analysis with the SILVA database (Quast et al., 2012) as reference. Metagenome analysis (Chapter 3) was performed by metagenomic shotgun sequencing and taxonomy classification with the nucleotide (nt) and non-redundant (nr) NCBI BLAST databases (Federhen, 2015).

Differences in abundance between the two chapters, as mentioned for *Frankiales,* are caused by characteristics of the analysed sequences and the completeness of the chosen database. Genes used as markers are characterized by a mixture of conserved regions as well as variable regions (Nossa et al., 2010). Marker genes are amplified by PCR prior to sequencing and thus only contain sequences from taxa covered by the chosen primers. Additionally, amplification by PCR could introduce an amplification bias, as it cannot be ruled out that some sequences are less amplified than others. Metagenomic shotgun sequencing is not restricted to marker genes and sequences are not previously amplified by PCR, which poses advantages and challenges on its own. An advantage is that amplification biases or limitations of chosen primers, as in marker gene analysis, are avoided. However, since every available sequence is included, not all match the criteria for marker genes and might be classified inaccurately due to similarities among different taxonomic groups. This was evident by the high fractions of unclassified sequences in Chapter 3 during analysis of metagenomic shotgun data from the rainforest conversion experiment. Furthermore, library construction prior to sequencing possibly introduces abundance biases as well, for instance by indexing steps. These indices are attached to each sequence prior to sequencing and multiplied by PCR in order to identify the sample origin of each sequence afterwards, since many samples are processed and sequenced at the same time. Therefore, these errors might occur in a similar manner as in PCRs for amplicon generation.

Another fact to consider is the unequal distribution of gene copy numbers of the 16S rRNA gene within a genome. It has been shown that the number of gene copies within a genome differ between taxonomic groups, possibly causing higher counts of sequences for taxa with higher copy numbers (Crosby and Criddle, 2003). Copy numbers of rRNA and tRNA genes can be checked (for instance by *rrn*db (Klappenbach et al., 2001)) if complete genomes of the targeted taxa are available and then normalized accordingly to enable comparisons between taxonomic groups (Lee et al., 2008; Louca et al., 2018). The taxa *Rhodospirillales*, which were highly abundant in this study, have a higher 16S rRNA gene copy number than *Rhizobiales* (Table 6.1).

The latter were also abundant in metagenomes, whereas *Rhodospirillales* were less abundant compared to marker gene studies. However, if normalization profiles are not complete because of a lack of available genomes of some taxa present in the analysed community, an uneven normalization could not eradicate the possibly present bias. Instead, another bias would be introduced, further influencing the outcome of analysis. Thus, normalization of gene copy numbers is still a method in development that needs to be evaluated for actual effectiveness prior to usage for every individual case.

Other sources of differences originate from sequence processing and analysis. It is important to consider that different taxonomy structures are used and available among different databases. Acidobacterial subgroups are organized as orders in the SILVA database taxonomy, whereas in the NCBI database they are not assigned to a regular taxonomic level but sorted within unclassified taxa of *Acidobacteria*, causing differences at order level as seen in the presented results (Figure 6.1).

Probably the most crucial difference between the two methods is the used reference database. In this study, the nt and nr NCBI genome databases were used for taxonomic classification of metagenomic sequences (Chapter 3) and the SILVA database for marker gene analysis (Chapter 2, 4 & 5). Full genome databases like the NCBI databases harbour information not only on taxonomic identities, but on functional genes as well, unlike marker gene focussed databases like the SILVA database. Thus, these databases can be used for functional gene analysis and combinations of taxonomy and functionality. However, full genomes are harder to obtain and curate than marker gene sequences. Consequently, specialized marker gene databases like SILVA (Quast et al., 2012) or Greengenes (McDonald et al., 2012) provide less sequences in total, but more entries per gene and organism for taxonomic classification. The before mentioned *Frankiales* are well described in the SILVA database, but only 6 genomes are available in the NCBI database (Table 6.1). It is possible, that sequences in the metagenome dataset belonging to *Frankiales* simply did not match a reference sequence in the database because of no available match in the few genomes.

Advantages of metagenomic shotgun sequencing for taxonomic classification are sequencing depth, avoidance of primer and PCR bias as well as consideration of all available sequences. In contrast, marker gene analysis only considers sequences suitable as markers and is relying on more specialized databases but provides taxonomic classification with a higher resolution.

Table 6.1: Statistics of selected bacterial orders regarding available genomes in the NCBI database, 16S rRNA gene sequence entries in the SILVA database (version 128) as well as mean 16S rRNA gene copy numbers according to rrndb (accessed in March 2020).

Taxonomic Order	Genomes available	Average 16S rRNA gene copy number	Database entries
Acidobacteriales	17	1.3 ± 0.5	3,099
Frankiales	6	2.3 ± 0.5	3,098
Rhizobiales	475	2.8 ± 1.4	37,547
Burkholderiales	1,218	3.5 ± 1.7	74,711
Rhodospirillales	119	4.5 ± 2.2	11,693
Myxococcales	32	3.3 ± 0.8	4,390

It is still under debate whether metagenomic shotgun sequencing or amplicon-based community analysis is more accurate for taxonomic classification. The ongoing development and improvement of databases and classification methods did not reach a point where the complexity of soilborne communities can be unravelled completely (Tessler et al., 2017; Breitwieser et al., 2019; Rausch et al., 2019). This study emphasizes the possible differences when comparing different databases and methods but also shows that broader patterns are observed across different methods.

6.2 Effects of rainforest conversion on microbial diversity

Rainforest conversion is accompanied with drastic reductions of biodiversity at several trophic levels (Drescher et al., 2016; Rembold et al., 2017). Contrary to these observations, several studies analysing tropical forest conversion to rubber and oil palm plantations in Southeast Asia and Central America, based on 16S rRNA gene analysis, reported higher microbial diversity values or no diversity changes in converted land use systems (Carney et al., 2004; Tripathi et al., 2012; Lee-Cruz et al., 2013; Schneider et al., 2015). It has been suggested, that bacterial diversity is positively correlated with pH in soils, which has been affected by land use management practices. Furthermore, these proposed connections between pH and microbial diversity have been discussed in other studies as well (Lauber et al., 2009; Zhalnina et al., 2015).

6.2.1 Results of biodiversity analysis are influenced by the chosen method

In this study, effects of rainforest conversion to managed land use systems were investigated by analysis of the active bacterial community with RNA derived 16S rRNA gene sequences (Chapter 2) and the entire community by metagenomic shotgun sequencing (Chapter 3).

Schneider et al. (2015) previously showed a significant increase of soil bacterial diversity from rainforest to plantations, which could not be shown for the active community of the same plots in this study. These differences could be explained by dormant organisms, detected during DNA-based analysis, or possibly even relic DNA. However, metagenomic shotgun analysis showed decreasing microbial diversity from forest to rubber plantations (Chapter 3). A higher sequencing depth provided by metagenomic shotgun sequencing also covers a larger proportion of the analysed community and provides quantitative results (Jovel et al., 2016; Brumfield et al., 2020). In principle, metagenomic shotgun reads cover each gene within a sample, including genes or non-coding regions that are not well suited for taxonomic identification (as discussed in 6.1.3). For example, the presence of highly conserved gene regions in the metagenomic reads could lead to classification errors (Rausch et al., 2019). Furthermore, 16S rRNA genes work well as taxonomy marker, since they contain hypervariable regions that are flanked by conserved regions. Furthermore, marker gene databases like SILVA (Quast et al., 2012) or Greengenes (McDonald et al., 2012) contain sequences from millions of species, whereas genome databases contain genomes of tens of thousands of species (Breitwieser et al., 2019). Marker gene analysis provided superior results regarding diversity assessment and taxonomic classification, which is likely due to higher suitability of the marker gene derived sequences for taxonomic classification as well as a better curated marker gene database. In this study, a significant difference of beta diversity was not detected for the active bacterial community by comparing land use systems with rainforest samples (Chapter 2). The only significant bacterial diversity change was recorded from rainforest to rubber land use in the metagenomic shotgun dataset (Chapter 3). In total, comparing the datasets of Chapter 2 and 3 of this study and the DNA-based amplicon data from Schneider et al. (2015) the direct metagenome-based approach showed generally higher diversity, but contrary trends compared with 16S rRNA amplicon sequences (Figure 6.2). Furthermore, a correlation of diversity to increasing soil pH could not be identified as proposed by other studies, which were conducted in agriculture related studies (Lee-Cruz et al., 2013; Schneider et al., 2015), grassland (Zhalnina et al., 2015) or across multiple soil types (Lauber et al., 2009) investigating bacterial communities from soil with 16S rRNA gene analyses approaches. Therefore, a general assumption that rainforest conversion to intensively managed land use systems significantly affects prokaryotic diversity and/or is coupled to increasing soil pH cannot be supported based on the results of this study. It should be noted that soil pH differences were on a rather small scale in the mentioned Chapters (ranging from 4.09 to 4.7), which could be another reason for the lack of correlation.

Additionally, different intensity levels of land use management (Chapter 4) did not show any effect on biodiversity. It is possible that the effects of reduced land use intensity appear after a longer period of time than at the time of the management experiment presented here (Chapter 4), which was less than a year.
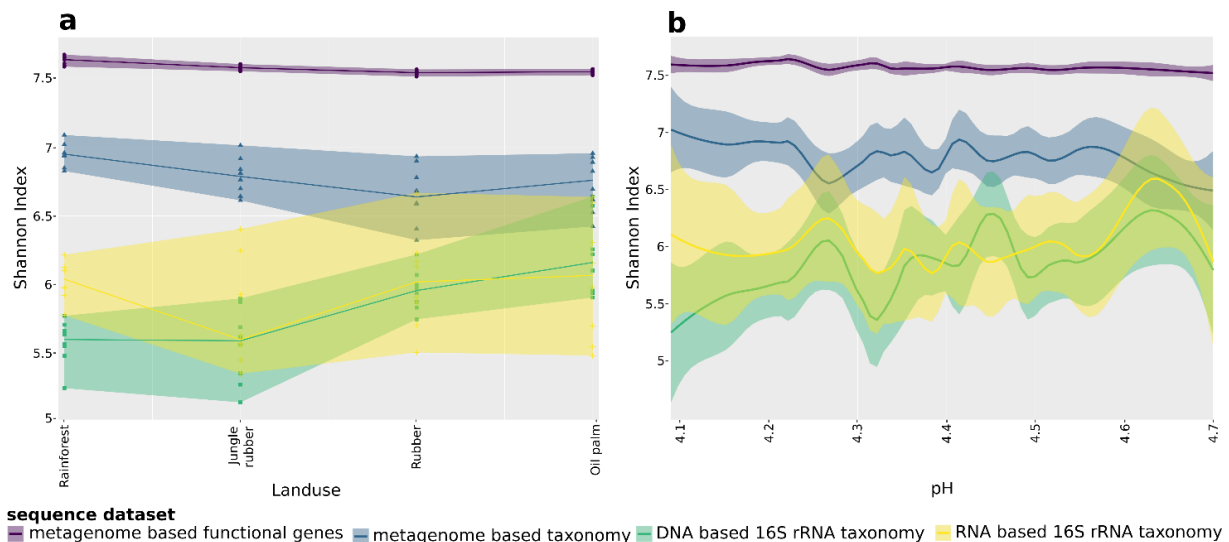
Figure 6.2: Shannon diversity indices of metagenome-based sequences (functional and taxonomic genes; Chapter 3)) and 16S rRNA amplicon sequences. RNA-based 16S rRNA gene data is taken from Chapter 2 (RNA-based) and DNA-based data from Schneider et al. (2015). All used data is visualized along the land use gradient (a) and against soil pH (b). Shaded areas indicate value of variance (a) and plot predictions based on a linear model (b) by geom-smooth (Wickham, 2016).

6.2.2 Functional diversity of bacteria and fungi is reduced by rainforest conversion

Taxonomic diversity and community composition are integral parts to describe microbial communities. However, microbial functionality and functional diversity are important factors for the analysis of community responses to changing environmental conditions as well. Functionality and functional potential can only be addressed by methods that include functional genes like metagenomic shotgun sequencing and have rarely been addressed in land use conversion related studies.

In this study, functional diversity regarding rainforest conversion was analysed by employing Shannon diversity (Chapter 3). Significant decreases were detected for bacteria in all land use systems and for fungi in oil palm plantations. The observed effects did not or not entirely correspond with taxonomic diversity, emphasizing that taxonomic and functional diversity should be addressed separately and are not necessarily correlated. Decreases in functional

diversity from rainforest to converted land use systems indicate higher redundancy in forest microbial communities with more organism harbouring the same function. This could provide a higher overall functional stability in rainforest soils and make the communities in rainforest samples less susceptible to environmental changes. It has been suggested before that functional diversity is correlated to nutrient content and, unlike taxonomic diversity, not to soil pH (Paula et al., 2014). Complementary to that suggestion, soil analysis showed a significant decrease of soil iron content and N availability from forest to managed land use systems in the same experimental plots (Allen et al., 2015). Fungal taxonomic diversity increased with increased land use intensity but fungal functional diversity decreased, indicating the presence of more fungal taxa but with redundant functional gene sets, as previously suggested in another study, which investigated fungal communities in the same experimental sites by ITS-based amplicon analysis (Brinkmann et al., 2019).

6.3 Affected taxa are tied to changes in soil properties and land use management

Changes of community structure are connected to changes in soil properties caused by land use change. Multivariate analysis of the active soil bacterial community (Chapter 2) and the entire community (Chapter 3) showed significant correlations with pH, C:N ratio, iron content, phosphorous content as well as base saturation in both studies regarding land use conversion. Increased soil pH is probably elevated in managed systems by dolomite application (liming). Decreased C:N ratios and also changes in soil pH are suspected to be remnants of initial biomass burnings during deforestation for agricultural applications, leaving measurable traces for decades after the initial burning (Tanaka et al., 2005; Sohng et al., 2017). With soil pH and C:N ratio as drivers of the active and dormant community structure, the link between land use management and altered soil properties to microbial community structure indicates that the measured community structure changes allegedly caused by rainforest conversion were introduced by changes in soil properties and partly affected by the initial deforestation via burning.

Additionally, similar clustering patterns were observed for the active bacterial community (Chapter 2) and the metagenome based community (Chapter 3) , indicating that the active bacterial community is driven or influenced by the same factors, following the proposed land use intensity gradient (rainforest < jungle rubber < rubber < oil palm). The observed clustering in multivariate analyses was similar for *Archaea* and fungi, with soil pH again as best correlation

factor. This demonstrates that rainforest conversion not only has direct effects on active and entire bacterial communities, but on other domains as well. Furthermore, land use management that alters soil properties such as soil pH and C:N ratio is an important factor for the observed community composition changes.

6.4 Impact of rainforest conversion on functional potential

6.4.1 Nitrogen metabolism and community interactions are negatively affected by rainforest conversion

It has been shown in Chapter 2 and 3 that rainforest conversion changes taxonomic profiles as well as microbial diversity and that these changes correlate with altered soil properties. Functional profiles were analysed by predicted gene abundances based on 16S rRNA gene data for the active bacterial community as well as metagenomic shotgun data for the entire microbial community. Due to strong effects of land use management by fertilizer application on related soil properties (Allen et al., 2015) and effects on abundances of microbial taxa associated to nitrogen and methane metabolism (Chapter 2; Schneider et al., 2015), these metabolisms were of particular interest for functional potential analysis (Chapter 3). Additionally, it has been hypothesized before that community dynamics in form of motility and interactions could be negatively influenced by land use management in a study that used 16S rRNA gene analysis and metagenomic shotgun sequencing to analyse effects of logging and conversion to oil palm in Borneo (Tripathi et al., 2016). Predicted functional profiles in Chapter 2 were derived from 16S rRNA transcript-based taxonomic assessment by using Tax4Fun (Asshauer et al., 2015) whereas functional potential profiles for microbial communities in Chapter 3 were directly derived from metagenomes by employing Kaiju (Menzel et al., 2016) in combination with the NCBI database (Federhen, 2015). The predicted functional abundances of nitrogen metabolism-related genes involved in nitrogen fixation and denitrification were significantly decreased in converted land use systems, suggesting a lower potential for nitrogen metabolism in plantations. Direct metagenome-derived profiles showed opposite results with higher gene abundances of nitrogen fixation and denitrification in oil palm and rubber plantations compared to rainforest. Increases in abundances of denitrification genes are supported by reports of positive correlation of pH with denitrification rates and higher nitrogen oxide emissions, a biproduct of denitrification, which were conducted in the same oil palm soils (Allen et al., 2015; Herold et al., 2018). Most of the nitrogen metabolism related sequences derived from *Rhizobiales* and

Burkholderiales, especially in rainforest samples. In the converted land use systems nitrogen metabolism related genes were mostly associated to *Rhizobiales* and unclassified *Acidobacteriales*. *Burkholderiales* are dominant in rainforest samples, suggesting a key role in nitrogen related processes. This pattern shifts in converted land use systems, with more sequences originating from *Rhizobiales* and *Acidobacteria*. This is further illustrated by detected sequences of denitrification genes *nirK* and *nirS*. Both genes encode enzymes, which are involved in the reduction of nitrite to nitric oxide as part of the denitrification process. Even though the function of both genes is similar, the distribution among taxa is different, with *nirK* sequences mostly originating from *Rhizobiales* in all samples. Similar observations were made in another study regarding denitrification potential under different fertilizer regimes in black soil (Yin et al., 2015). It was reported that monoculture plantation soils are susceptible to nutrient losses also by elevated nitrous gas releases as well (Allen et al., 2015). Nitrogen fixation gene abundances also increased and originated from symbiotic *Rhizobiales*, which is surprising as plant diversity in general is decreased in converted land use systems soils. It is indicated that for nitrogen metabolism related processes *Rhizobiales* compared to *Burkholderiales* play a bigger role in converted land use systems than in rainforest soils, which are dominated by members of the *Burkholderiales*.

Another focus of the metagenome-based analysis of soilborne microbial communities (Chapter 3) were motility and interaction related genes, represented by flagellar assembly, quorum sensing, chemotaxis and secretion system related genes. Type IV and Type VI secretion systems genes were predicted to be higher in abundance in rainforest soils of the active bacterial community. Most chemotaxis genes showed the same pattern. Functional potential analysed by metagenomic shotgun sequencing (Chapter 3) provided corresponding results in which the functional categories bacterial chemotaxis, flagellar assembly, quorum sensing and secretion systems showed significant abundance decreases with increasing land use intensity from rainforest to oil palm plantations. Identified flagellar assembly genes mostly originated from *Burkholderiales* and *Rhizobiales,* which decreased in abundance as well. Although a significant decrease in functional genes can be tied to specific taxa, it seems unlikely that a change in these genes is driven only by these two orders, since *Acidobacteria*, increased with land use intensity and carry flagellar assembly genes. The reduction of present flagellar assembly and quorum sensing genes in land use samples compared to rainforest points towards a reduction of motility and swarming potential within the present microbial community in soil. Similar connections of quorum sensing gene abundance and interaction potential were made in a study targeting microbiome responses towards infections in bovines but were not made in soil before (Zinicola

et al., 2015). The reduced motility is accompanied by reduced potential for exchange of nucleic acids or pathogenic behaviour, indicated by reductions of type IV and VI secretion system genes (Cascales and Cambillau, 2012; Walldén et al., 2012; Guglielmini et al., 2014). Type IV secretion systems are used to exchange nucleic acids and genes within a community. A decreased capability suggests higher susceptibility to environmental changes, which is amplified when assuming less capabilities of motility. Type VI secretion system related genes mostly derived from *Burkholderiales*. Originally, this group has been known for comprising pathogens but in recent years also as root nodule symbionts. These symbionts were regrouped as *Paraburkholderia*, apart from the previously known pathogens (Dobritsa and Samadpour, 2016). This study showed that type VI secretion system genes were negatively affected by rainforest conversion and originate from *Burkholderiales* but the main detected family within that order was *Paraburkholderia*. It is possible that the *Paraburkholderia* harbour members, which are heavily involved in nitrogen metabolism related processes in rainforest soils while also being able of pathogenic behaviour through type VI secretion systems (Loong et al., 2019).

This study provides evidence that functional potential is affected by management practices in plantations, leading to higher denitrification potential and drastic reduction of motility and interaction potential, mostly associated to members of *Proteobacteria*.

6.4.2 Comparing functional predictions based on taxonomy and functional gene analysis

Functional predictions derived from taxonomic community compositions bear some limitations. Especially, when a considerable part of the detected community consists of taxa with no or only a few available genomes, since available genomes are the basis for used functional profiles, which are the basis for calculating the prediction of the tested community (see 6.1.3). As mentioned before, soils are very heterogeneous, resulting in highly diverse microbial communities (Vos et al., 2013; Kaiser et al., 2016; Fonseca et al., 2018). Environments with highly diverse microbial communities are probably more susceptible to biases by functional predictions. As a result, functional predictions based on marker gene analyses need to be considered with caution with respect to the type of sample.

6.5 Effects of varying land use intensity in oil palm and cacao

Rainforest conversion and the establishment of new agricultural areas introduce drastic changes to the present ecosystem. Especially deforestation and land use management practices are frequently discussed with respect to damages of biodiversity and the environment (Wilcove et al., 2013; Azhar et al., 2017; Darras et al., 2019). Consequently, it is of high interest whether established, potentially harmful practices are replaceable by alternative practices. Another aspect is whether the stability of the environmental changes and their effects on microbial communities are detectable or even reversible upon changes in management.

In Chapter 4, effects of reduced management practices in oil palm plantations on the soilborne bacterial community were analysed. In addition to reduced fertilizer application, mechanical weeding instead of herbicide application was tested. DNA-based 16S rRNA gene analysis showed no significant differences of bacterial community composition, diversity or measured soil properties accompanying the different management practices. In the active (RNA-based) community, *Candidatus* Solibacter and *Haliangium* showed higher relative abundances compared to the entire (DNA-based) community composition. It has been suspected, that *Candidatus* Solibacter might be involved in nitrate reduction, which converts nitrate to nitrite (Ward et al., 2009; Pickett et al., 2019). Further details regarding the ecological role of *Candidatus* Solibacter are still rare though, making further evaluation very challenging.

These results raised the question why no effect of the introduced changes was recorded. It is possible that the applied reduced fertilizer amounts (roughly half the amount than regular fertilization) were too conservatively chosen to show any effect on the soilborne communities. The detected missing effect of a complete lack of herbicide application on the bacterial community seems unlikely as an effect on microbial community composition was detected in other studies (Guijarro et al., 2018). The oil palm plots for the management study were established in November 2016 and soil sampling was done in May 2017. Possibly, the introduced changes do not cause measurable differences within that timeframe, since the previous management practices were performed for roughly 15 years while the experimental treatments were applied for less than a year. Assuming that a reduction of land use intensity initiates any sort of change of microbial community composition or diversity, it might take more time to produce detectable differences, since the regularly applied management caused stable long termed effects. This is further hinted at by detected differences between the active and entire community, as RNA-based results can be seen as short termed responses, compared to DNA-based results. Another example for how long introduced changes or events are detectable

are changes in soil properties after rainforest conversion to plantations by burning. These effects, for instance drastic increase in pH, are traceable for many years (Tanaka et al., 2005; Allen et al., 2015; Sohng et al., 2017).

In addition to oil palm management variation, different management intensities in terms of herbicide and pesticide usage regarding endophytic bacterial and fungal communities were studied in agroforestry cacao cultures (Chapter 5). Similar to the oil palm management experiment, bacterial diversity did not change with altered management. However, fungal diversity was positively affected by increased management intensity. Instead of management intensity, the location of the sampling site was a crucial factor for differences in bacterial community composition, possibly connected to differences in canopy cover of the different sampling sites. It was hypothesized that shade intensity was a factor that influenced endophyte communities, which was suggested before for fungal endophytic communities (Scholtysik et al., 2013). This is in accordance with previous assumptions that canopy cover and therefore light availability might have a general effect on microbial communities in soil.

6.6 Conclusions

This study provided insights into how prokaryotic communities are affected by rainforest conversion to managed land use systems. It was shown that soil microbial community compositions and functional potential are significantly affected by the conversions. These effects diverge for different taxa. Members of the *Proteobacteria* are more dominant in rainforest soils, whereas *Acidobacteria* are of higher abundance in converted soils. Furthermore, it was shown that the active (RNA-based) and dormant (DNA-based) bacterial communities are differently affected. In addition to taxonomic profiles, relations between bacterial taxa such as *Proteobacteria* and *Acidobacteria* and functional gene occurrence were unravelled. Management practices and the accompanied changes in soil properties affected not only bacterial community composition but also their functional potential. Especially the functional potential of metabolisms which are connected to land use management (e.g. nitrogen metabolism to applied fertilizer) was affected by rainforest conversion. The taxonomic groups who were mostly influenced by these processes, are involved in these metabolic processes. Overall, this study demonstrates not only strong effects of rainforest conversion on microbial community structures and their functional potential, but the direct connection of land use management practices to specific taxa. With these results as a basis, further research could

focus on potential effects of rainforest conversion on functional profiles by employing metatranscriptomic approaches.

6.7 References

Allen, K., Corre, M. D., Tjoa, A., and Veldkamp, E. (2015). Soil nitrogen-cycling responses to conversion of lowland forests to oil palm and rubber plantations in Sumatra, Indonesia. *PLoS One* 10.

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

Azhar, B., Saadun, N., Prideaux, M., and Lindenmayer, D. B. (2017). The global palm oil sector must change to save biodiversity and improve food security in the tropics. *J. Environ. Manage.* 203, 457–466.

Bidyarani, N., Prasanna, R., Babu, S., Hossain, F., and Saxena, A. K. (2016). Enhancement of plant growth and yields in Chickpea (Cicer arietinum L.) through novel cyanobacterial and biofilmed inoculants. *Microbiol. Res.* 188–189, 97–105.

Bouskill, N. J., Lim, H. C., Borglin, S., Salve, R., Wood, T. E., Silver, W. L., et al. (2012). Pre-exposure to drought increases the resistance of tropical forest soil bacterial communities to extended drought. *ISME J.* 7, 384.

Breitwieser, F. P., Lu, J., and Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Brief. Bioinform.* 20, 1125–1136.

Brinkmann, N., Schneider, D., Sahner, J., Ballauff, J., Edy, N., Barus, H., et al. (2019). Intensive tropical land use massively shifts soil fungal communities. *Sci. Rep.* 9, 3403.

Brumfield, K. D., Huq, A., Colwell, R. R., Olds, J. L., and Leddy, M. B. (2020). Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLoS One* 15.

Carney, K. M., Matson, P. A., and Bohannan, B. J. M. (2004). Diversity and composition of tropical soil nitrifiers across a plant diversity gradient and among land-use types. *Ecol. Lett.* 7, 684–694.

Cascales, E., and Cambillau, C. (2012). Structural biology of type VI secretion systems. *Phil. Trans. R. Soc. B* 367, 1102–1111.

Chittapun, S., Limbipichai, S., Amnuaysin, N., Boonkerd, R., and Charoensook, M. (2018). Effects of using cyanobacteria and fertilizer on growth and yield of rice, Pathum Thani I: a pot experiment. *J. Appl. Phycol.* 30, 79–85.

Crosby, L. D., and Criddle, C. S. (2003). Understanding bias in microbial community analysis techniques due to rrn operon copy number heterogeneity. *Biotechniques* 34, 790–802.

Darras, K. F. A., Corre, M. D., Formaglio, G., Tjoa, A., Potapov, A., Brambach, F., et al. (2019). Reducing Fertilizer and Avoiding Herbicides in Oil Palm Plantations— Ecological and Economic Valuations. *Front. For. Glob. Chang.* 2, 65.

Dobritsa, A. P., and Samadpour, M. (2016). Transfer of eleven species of the genus Burkholderia to the genus Paraburkholderia and proposal of Caballeronia gen. nov. to accommodate twelve species of the genera Burkholderia and Paraburkholderia. *Int. J. Syst. Evol. Microbiol.* 66, 2836–2846.

Drescher, J., Rembold, K., Allen, K., Beckschäfer, P., Buchori, D., Clough, Y., et al. (2016). Ecological and socio-economic functions across tropical land use systems after rainforest conversion. *Philos. Trans. R. Soc. B Biol. Sci.* 371.

Federhen, S. (2015). Type material in the NCBI Taxonomy Database. *Nucleic Acids Res.* 43, D1086–D1098.

Fonseca, J. P., Hoffmann, L., Cabral, B. C. A., Dias, V. H. G., Miranda, M. R., de Azevedo Martins, A. C., et al. (2018). Contrasting the microbiomes from forest rhizosphere and deeper bulk soil from an Amazon rainforest reserve. *Gene* 642, 389–397.

Guglielmini, J., Néron, B., Abby, S. S., Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. C. (2014). Key components of the eight classes of type IV secretion systems involved in bacterial conjugation or protein secretion. *Nucleic Acids Res* 42, 5715–5727.

Guijarro, K. H., Aparicio, V., De Gerónimo, E., Castellote, M., Figuerola, E. L., Costa, J. L., et al. (2018). Soil microbial communities and glyphosate decay in soils with different herbicide application history. *Sci. Total Environ.* 634, 974–982.

Hayat, R., Ali, S., Amara, U., Khalid, R., and Ahmed, I. (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Ann. Microbiol.* 60, 579–598.

Herold, M. B., Giles, M. E., Alexander, C. J., Baggs, E. M., and Daniell, T. J. (2018). Variable response of nirK and nirS containing denitrifier communities to long-term pH manipulation and cultivation. *FEMS Microbiol. Lett.* 365.

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., et al. (2016). Characterization of the gut microbiome using 16S or shotgun metagenomics. *Front. Microbiol.* 7, 459.

Kaiser, K., Wemheuer, B., Korolkow, V., Wemheuer, F., Nacke, H., Schöning, I., et al. (2016). Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. *Sci. Rep.* 6, 33696.

Klappenbach, J. A., Saxman, P. R., Cole, J. R., and Schmidt, T. M. (2001). rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Res.* 29, 181–184.

Kvíderová, J., and Kumar, D. (2020). Response of short-term heat shock on photosynthetic activity of soil crust cyanobacteria. *Protoplasma* 257, 61–73.

Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl. Environ. Microbiol.* 75, 5111 LP – 5120.

Lee-Cruz, L., Edwards, D. P., Tripathi, B. M., and Adams, J. M. (2013). Impact of logging and forest conversion to oil palm plantations on soil bacterial communities in Borneo. *Appl. Environ. Microbiol.* 79, 7290–7297.

Lee, Z. M.-P., Bussema III, C., and Schmidt, T. M. (2008). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res.* 37, D489–D493.

Li, Y., Pan, F., and Yao, H. (2019). Response of symbiotic and asymbiotic nitrogen-fixing microorganisms to nitrogen fertilizer application. *J. Soils Sediments* 19, 1948–1958.

Loong, S. K., Tan, K.-K., Zulkifle, N.-I., and AbuBakar, S. (2019). Draft genome of Paraburkholderia fungorum sequence type 868 recovered from human synovial tissues. *Data Br.* 25, 104159.

Louca, S., Doebeli, M., and Parfrey, L. W. (2018). Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome* 6, 41.

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610.

Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7, 11257.

Molot, L. A. (2017). The effectiveness of cyanobacteria nitrogen fixation: Review of bench top and pilot scale nitrogen removal studies and implications for nitrogen removal programs. *Environ. Rev.* 25, 292–295.

Nacke, H., Thürmer, A., Wollherr, A., Will, C., Hodac, L., Herold, N., et al. (2011). Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS One* 6, e17000–e17000.

Nossa, C. W., Oberdorf, W. E., Yang, L., Aas, J. A., Paster, B. J., DeSantis, T. Z., et al. (2010). Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J. Gastroenterol. WJG* 16, 4135.

Patova, E., Sivkov, M., and Patova, A. (2016). Nitrogen fixation activity in biological soil crusts dominated by cyanobacteria in the Subpolar Urals (European North-East Russia). *FEMS Microbiol. Ecol.* 92.

Paula, F. S., Rodrigues, J. L. M., Zhou, J., Wu, L., Mueller, R. C., Mirza, B. S., et al. (2014). Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities. *Mol. Ecol.* 23, 2988–2999.

Peng, X., and Bruns, M. A. (2019). Development of a nitrogen-fixing cyanobacterial consortium for surface stabilization of agricultural soils. *J. Appl. Phycol.* 31, 1047–1056.

Pickett, B., Irvine, I. C., Bullock, E., Arogyaswamy, K., and Aronson, E. (2019). Legacy effects of invasive grass impact soil microbes and native shrub growth. *Invasive Plant Sci. Manag.* 12, 22–35.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596.

Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., et al. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome* 7, 133.

Rembold, K., Mangopo, H., Tjitrosoedirdjo, S. S., and Kreft, H. (2017). Plant diversity, forest dependency, and alien plant invasions in tropical agricultural landscapes. *Biol. Conserv.* 213, 234–242.

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., et al. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* 4, 1340–1351.

Schneider, D., Engelhaupt, M., Allen, K., Kurniawan, S., Krashevska, V., Heinemann, M., et al. (2015). Impact of lowland rainforest transformation on diversity and composition of soil prokaryotic communities in Sumatra (Indonesia). *Front. Microbiol.* 6.

Scholtysik, A., Unterseher, M., Otto, P., and Wirth, C. (2013). Spatio-temporal dynamics of endophyte diversity in the canopy of European ash (Fraxinus excelsior). *Mycol. Prog.* 12, 291–304.

Sohng, J., Singhakumara, B. M. P., and Ashton, M. S. (2017). Effects on soil chemistry of tropical deforestation for agriculture and subsequent reforestation with special reference to changes in carbon and nitrogen. *For. Ecol. Manage.* 389, 331–340.

Tanaka, S., Kendawang, J. J., Yoshida, N., Shibata, K., Jee, A., Tanaka, K., et al. (2005). Effects of Shifting Cultivation on Soil Ecosystems in Sarawak, Malaysia IV. Chemical Properties of the Soils and Runoff Water at Niah and Bakam Experimental Sites. *Soil Sci. Plant Nutr.* 51, 525–533.

Taylor, M., Mediannikov, O., Raoult, D., and Greub, G. (2012). Endosymbiotic bacteria associated with nematodes, ticks and amoebae. *FEMS Immunol. Med. Microbiol.* 64, 21–31.

Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., et al. (2017). Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci. Rep.* 7, 6589.

Tripathi, B. M., Edwards, D. P., Mendes, L. W., Kim, M., Dong, K., Kim, H., et al. (2016). The impact of tropical forest logging and oil palm agriculture on the soil microbiome. *Mol. Ecol.* 25, 2244–2257.

Tripathi, B. M., Kim, M., Singh, D., Lee-Cruz, L., Lai-Hoe, A., Ainuddin, A. N., et al. (2012). Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. *Microb Ecol* 64, 474–484.

Urich, T., Lanzén, A., Qi, J., Huson, D. H., Schleper, C., and Schuster, S. C. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* 3.

Vos, M., Wolf, A. B., Jennings, S. J., and Kowalchuk, G. A. (2013). Micro-scale determinants of bacterial diversity in soil. *FEMS Microbiol. Rev.* 37, 936–954.

Walldén, K., Williams, R., Yan, J., Lian, P. W., Wang, L., Thalassinos, K., et al. (2012). Structure of the VirB4 ATPase, alone and bound to the core complex of a type IV secretion system. *Proc. Natl. Acad. Sci.* 109, 11348 LP – 11353.

Ward, N. L., Challacombe, J. F., Janssen, P. H., Henrissat, B., Coutinho, P. M., Wu, M., et al. (2009). Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Appl. Environ. Microbiol.* 75, 2046–2056.

Wemheuer, B., Wemheuer, F., and Daniel, R. (2012). RNA-based assessment of diversity and composition of active archaeal communities in the German Bight. *Archaea* 2012.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org.

Wilcove, D. S., Giam, X., Edwards, D. P., Fisher, B., and Koh, L. P. (2013). Navjot's nightmare revisited: logging, agriculture, and biodiversity in Southeast Asia. *Trends Ecol Evol* 28, 531–540.

Yin, C., Fan, F., Song, A., Cui, P., Li, T., and Liang, Y. (2015). Denitrification potential under different fertilization regimes is closely coupled with changes in the denitrifying community in a black soil. *Appl. Microbiol. Biotechnol.* 99, 5719–5729.

Zhalnina, K., Dias, R., de Quadros, P. D., Davis-Richardson, A., Camargo, F. A. O., Clark, I. M., et al. (2015). Soil pH Determines Microbial Diversity and Composition in the Park Grass Experiment. *Microb Ecol* 69, 395–406.

Zinicola, M., Higgins, H., Lima, S., Machado, V., Guard, C., and Bicalho, R. (2015). Shotgun metagenomic sequencing reveals functional genes and microbiome associated with bovine digital dermatitis. *PLoS One* 10.

7
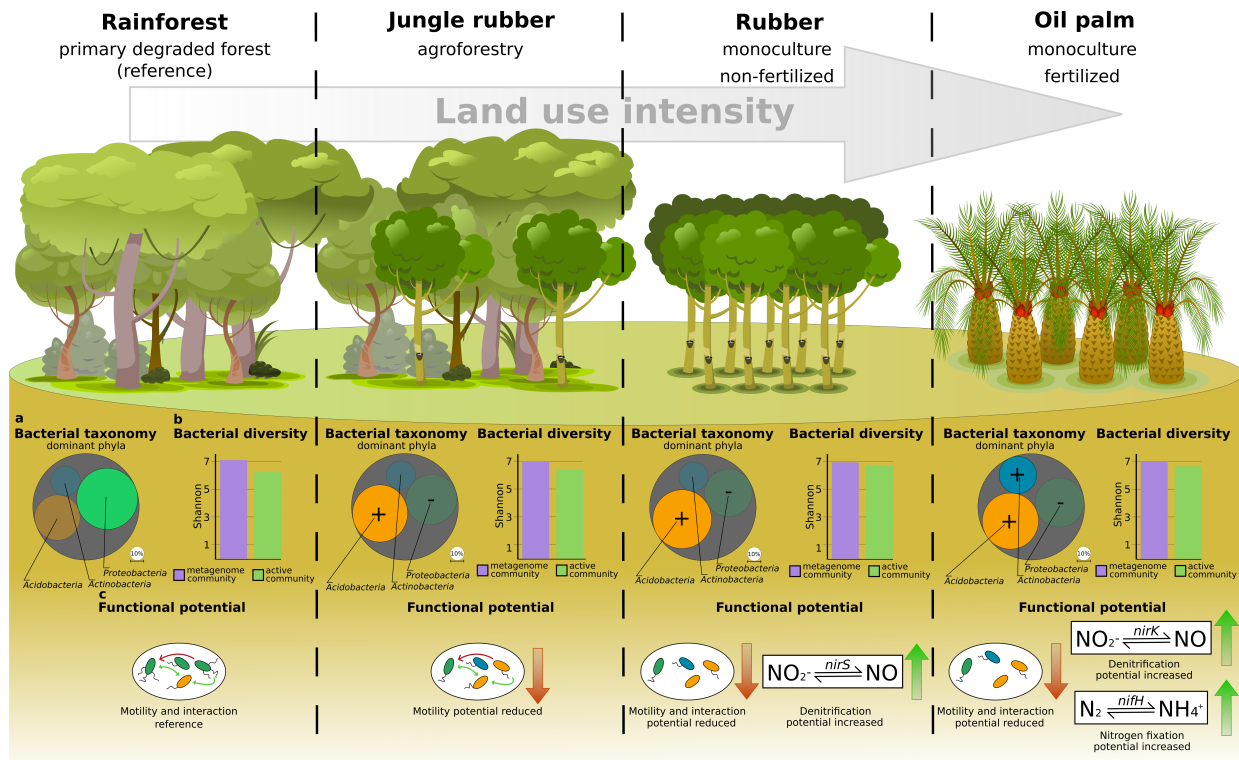
Summary

Figure 7.1: Schematic summary of main results of the rainforest conversion experiment on bacterial communities (Chapter 2 & 3). Land use intensity is shown from left to right with rainforest as reference. For rainforest and each land use bacterial taxonomy, diversity and functional potential is shown: Bacterial taxonomy is shown for the three most abundant phyla (a) with bubble diameter showing relative abundance and +/- indicating abundance changes compared to rainforest. Bacterial diversity is displayed as Shannon diversity index (b) for the active community (purple) and entire community (green). Functional potential of motility and interaction related genes (c) are displayed by number of flagella and arrows indicating interactions, with red arrows indicating gene abundance decreases and green arrows indicating gene abundance increases compared to rainforest. Differences of nitrogen metabolism related processes (c) are shown as boxes containing the reaction and gene name, with green arrows indicating gene abundance increases compared to rainforest.

In this thesis, the impact of rainforest conversion to the managed land use systems jungle rubber, rubber plantation and oil palm plantation on structure and function of soil microbial communities was investigated (Figure 7.1). The active bacterial community (RNA-based) was investigated by 16S rRNA marker gene analysis (Chapter 2) and the entire community (DNA-based) by direct metagenome sequencing (Chapter 3). Additionally, different management practices were tested for their influence on soilborne bacterial communities in oil palm (Chapter 4) and endophytic microbial communities in cacao plantations (Chapter 5).

The main focus of this work was to determine the effects of rainforest conversion to agricultural land use systems on soilborne microbial communities by using metagenomic approaches (Chapter 2 & 3). The active bacterial community was analysed by 16S rRNA gene amplicon analysis in combination with the SILVA database. The communities in all soils were dominated by *Proteobacteria*, *Acidobacteria* and *Actinobacteria*. Compared to rainforest reference soils, the relative abundance of *Acidobacteria* and *Actinobacteria* increased in managed land use systems, while *Proteobacteria* decreased. Ordination analysis showed correlations between the detected communities and measured pH and C:N ratios in soil. Functional community predicted was predicted from the taxonomic composition and showed a relative abundance decrease of nitrogen metabolism and chemotaxis related marker genes from rainforest to the converted land use systems.

In the same experimental setup, direct metagenome sequence analysis was performed in order to further investigate microbial community compositions and their functional potential. Bacterial diversity did significantly decrease from rainforest to rubber systems and functional diversity decreased in both plantation systems. Fungal diversity decreased from rainforest to converted land use systems significantly, while fungal functional diversity increased. In rainforest soils, *Rhizobiales* and *Burkholderiales* were the most dominant bacterial taxa. Towards higher land use intensity from rainforest to plantations, the communities shifted to *Acidobacteria*-dominated communities. With increasing land use intensity from rainforest to plantations accompanied by higher soil pH values, denitrification and nitrogen fixation potential increased and mostly originated from members of the bacterial orders *Burkholderiales* and *Rhizobiales*. The potential for motility and interaction, in form of flagellar assembly, chemotaxis and secretion systems, decreased towards converted land use systems. Genes involved in flagellar assembly were mostly connected to *Burkholderiales* and chemotaxis related genes to *Acidobacteria*. This study showed that community composition and functional potential is significantly affected by rainforest conversion. These changes are strongly tied to effects of land use management practices such as fertilizer input and alteration of plant diversity on soil properties and influence taxa involved in nitrogen metabolism.

In addition to rainforest conversion, effects of different management practices on bacterial communities in oil palm and cacao plantations were analysed. The reduced fertilizer and herbicide use in oil palm plantations was investigated based on RNA and DNA-driven 16S rRNA gene amplicon analysis. Although differences in bacterial diversity or community composition between treatments were not recorded, notable differences in abundance were detected between the entire and active community, emphasizing rather short-term effects.

Effects of different management intensities on endophytic bacterial and fungal communities in cacao leaves were tested by employing marker gene amplicon analysis. No differences in bacterial community composition or diversity were detected between the treatments, whereas fungal diversity was positively affected by increased management intensity. The communities were more effected by the different sampling site locations than by the treatment

This study provides new insights into the effects of rainforest conversion to managed land use systems on microbial community composition and functional potential in tropical soils. It furthermore connected taxonomic profiles with functional potential in this scenario for the first time. The results presented here contributed to close the knowledge gap between soil microbial community structure and functions that are affected by land use change-induced altered soil properties. However, DNA-based functional profiles might differ from gene activity in situ. Community activity could be further investigated by combining the results presented here with RNA-based metatranscriptomic analyses.

8

Appendix

8.1. Thesis Declaration

Declaration of independent work

I hereby confirm that I have written this doctoral thesis independently. I have not used other sources or facilities others than the ones mentioned in the Chapters. Moreover, I have not used unauthorized assistance and have not submitted this thesis previously in any form for another degree at any institution or university.

Dirk Berkelmann

Göttingen, May 2020

8.2 Acknowledgements

After starting the PhD, I also met my fellow PhD students of the CRC990 project who became dear friends over the years. Thank you for all the fun and companionship while we were in the field, on very fancy meetings on Bali or just enjoying the evenings in the skylounge. Terima kasih banyak untuk semuanya!

Further thanks go to the entire staff of the CRC990 project in Göttingen, Jambi and Bogor. Organizing the field trips went smoothly because of all your work and no one could ever find a way through the jungle of all the permit applications without your help. I would also like to thank my field assistants Zulfi Kamal, Winda Januarista and Pak Jon. Working with you was a great experience and I am very grateful for showing me around Jambi, visiting the best coffee places in town or even inviting me to your home for lunch.

Many many thanks go to Felix, Daniel, Andi, Marina, Benno, Timmer and Philipp who made days always better in very different but equally encouraging ways. I am unbelievably lucky and grateful to call you friends.

A special mention goes to Marlene, who is the big sister I never had. I learned a lot from you and I would not be the same person today if it would not be for that A-Team postcard.

Then, I want to thank my family, Marcus, Christlieb and Christine for always believing in me, being interested in what I do and unconditional support at all times. The most important person to mention here is my mother, who gave everything time and time again to make this possible and I cannot admire her enough for all the things she did for me. I would not have come this far without you and dad. This is your achievement as well and I will be forever thankful for all your sacrifices.

Finally, I want to thank Tatiana for basically everything. We met under the most unlikely circumstances and managed to create a home despite distances and all uncertainties. I cannot thank you enough for all your help and love during the last years and especially the last months. You did a lot to keep my spirits up from constant coffee supply to listening to my worries whenever I was in a bad mood. Having you around makes everything better.

What an adventure.