Doctoral Thesis

Ecological Patterns in Prokaryotic Diversity and Community Composition in Tropical Soils of Southeast Asia

February 2015

Graduate School of Seoul National University

School of Biological Sciences

Binu Mani Tripathi

Ecological Patterns in Prokaryotic Diversity and Community Composition in Tropical Soils of Southeast Asia

by

Binu Mani Tripathi

under the supervision of

Professor Jonathan Miles Adams, Ph.D.

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

February 2015

Graduate School of Seoul National University
School of Biological Sciences

ABSTRACT

The dominant factors controlling soil microbial community composition and diversity variation within the tropics are poorly known. In this study, the extent of soil microbial diversity was investigated in tropical soils of Malaysia and the way microbial communities are affected by land use (forest to agriculture), spatial scaling (at local and regional scales) and biome (tropics vs. temperate), as well as the extent to which ecological processes or other environmental factors contribute to structuring the soil microbial communities.

The effect of land use changes (i.e. conversation of forest to agriculture) on bacteria, archaea and ammonia oxidizing archaea (AOA) was studied in tropical soils of Malaysia. *Acidobacteria* and *Proteobacteria* were the most dominant bacterial phyla across all soil samples. *Thaumarchaeota* groups 1.1b and 1.1c appear to be the dominant archaeal lineages in tropical lowland soils, whereas, *Nitrososphaera* and *Nitrosotalea* clusters were recovered as the most dominant AOA taxa. It is found that land use in itself has a weak but significant effect on the microbial community composition. However, for all microbial taxa the community composition and diversity was strongly correlated with soil properties, especially soil pH. Soil pH was the best predictor of microbial community composition and diversity across the various land use types. Total bacterial and AOA diversity showed a hump-shaped correlation with soil pH, diversity peaking at ~ pH 6.0. However, total archaeal diversity showed a negative correlation with soil pH. In addition, variation in relative abundance of dominant lineages of bacteria, archaea and AOA was also significantly correlated

with soil pH. Together, these results confirm the importance of soil pH in structuring soil microbial communities in Southeast Asia.

How spatial scaling affects bacterial communities in tropical rainforest soils was further analyzed at two spatial scales: a local scale with samples spaced every 5 m over a 150 m transect, and a regional scale with samples 1 to 1,800 km apart. A weak correlation was found between spatial distance and whole bacterial community dissimilarity, but only at the local scale. In contrast, environmental distance was highly correlated with community dissimilarity at both spatial scales, stressing the greater role of environmental variables rather than spatial distance in determining bacterial community variation at different spatial scales. Soil pH was the only environmental parameter that significantly explained the variance in bacterial community at the local scale, whereas total nitrogen and elevation were additional important factors at the regional scale. Similar results were observed at both scales when only the most abundant OTUs were analyzed. A variance partitioning analysis showed that environmental variables contributed more to bacterial community variation than spatial distance at both scales. In total, these results support a strong influence of the environment, rather than neutral or dispersal effects, in determining bacterial community composition in the rainforests of Malaysia. It is possible that the remaining spatial distance effect is due to some of the myriad of other environmental factors which were not considered here, rather than dispersal limitation.

The community of both total archaea and ammonia oxidizing archaea (AOA) was compared using 16S rRNA and *amoA* gene pyrosequencing respectively, on a geographical scale in two different

biomes: tropical Malaysia and temperate Korea and Japan - to investigate to what extent archaeal and AOA diversity and community structure change between them. Total archaeal diversity showed a negative correlation with soil pH, and in contrast a hump-shaped curve for AOA diversity that peaked at ~ pH 6.0. Within each biome, soil pH also emerged as the dominant factor determining variation in community composition of both 16S rRNA and *amoA* genes. However, between biomes climate was also important in differentiating tropical from temperate archaea. For both 16S rRNA and *amoA* genes, significant phylogenetic signals were consistently detected across short phylogenetic distances. However, for 16S rRNA genes significant signals were detected also at intermediate phylogenetic distances. Quantifying phylogenetic turnover (as the deviation from a null expectation) suggested that deterministic factors imposed strong selection on total archaea and AOA communities, in both biomes. This study reveals that strong and predictable relationships exist between archaeal community structure and the environment, and that both climate and soil pH are major factors in archaeal community ecology.

Keywords: ammonia-oxidizing archaea, archaea, bacteria, land use, microbial communities, pyrosequencing, Southeast Asia, soil pH, 16S rRNA gene, tropical rain forest

TABLE OF CONTENTS

ABBREVIATIONS

OTU: Operational taxonomic unit

PCR: Polymerase chain reaction

NGS: Next generation sequencing

SRA: Short read archive

rRNA: Ribosomal ribonucleic acid

NMDS: Non-metric multidimensional scaling

BLAST: Basic local alignment search tool

PCoA: Principal coordinate analysis

RDP: Ribosomal database project

DNA: Deoxyribonucleic acid

MNTD: Mean nearest taxon distance

ses: Standardized effect size

AOA: Ammonia-oxidizing archaea

PERMANOVA: Permutational multivariate analysis of variance

RDA: Redundancy analysis

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. Tropical Soil Microbial Diversity and Recent Trends in Microbial Ecology: An Introduction

1.1. Tropical regions are exceptionally rich in biodiversity

Around the world, tropical regions are valued by biologists and conservationists for their remarkable diversity of life. Tropical rainforests comprise only 7% of the Earth's land surface, yet they support more than 60% of all known plant and animal species (Dirzo & Raven 2003). Tropical regions contain a majority of the world's endemic plant species (Joppa *et al.* 2011), old-growth tropical forests are the most species-rich in the world (Joppa *et al.* 2011), and this biodiversity enhances the forest productivity that sustains many people (Cardinale *et al.* 2012). There are noteworthy differences in rainforest species diversity among the continents. For example, Neotropics has higher bird and bat species and many epiphytic Bromeliads, while Southeast Asian tropical forests are more diverse in gliding animals and dipterocarp trees. The estimated number of rainforest plant species also varies, with 93,500, 61,700, and 20,000 species in the Neotropics, the Asia-Pacific region and Africa (including Madagascar), respectively (Corlett & Primack 2011). These differences are related to continental drift, differences in rainfall and its seasonal distribution, and extinctions caused by past natural and anthropogenic environmental changes (Corlett & Primack 2011; Gouvenain & Silander 2003; Parmentier *et al.* 2007). However, rapid anthropogenic changes have severally affected the great bio-diversity of tropical regions. For instance, due to deforestation, many large forest-dwelling mammals, half of the large primates, and nearly 9% of all known tree species are at risk (Rosen 2000), while Pitman and Jorgensen (2002) estimated that the fraction of

the tropical flora threatened with extinction may well be much higher. Furthermore, the loss of animals that produce ecological services, such as seed dispersal, nutrient recycling and pollination, might further impede forest regeneration in the disturbed areas.

1.2. Metagenomics and soil microbial community analysis

1.2.1. Metagenomics approach

The term metagenomics (Handelsman *et al.* 1998), refers to the study of the metagenome, which is total genomic DNA from environmental samples. The metagenomic techniques, which bypass the need for isolation and laboratory cultivation of individual species, have fundamentally changed studies in environmental microbiology. Metagenomics approaches allows for a more in-depth understanding of the structural (gene/species richness and distribution) and functional (metabolic) potential of environmental microbial communities. Metagenomics has made exceptional contributions to microbial ecology; among them, one of the most prominent findings of metagenomics is the first description of proteorhodopsin in marine bacteria (Béja *et al.* 2000). Metagenomics approach is mainly grouped into two categories, unselective (shotgun metagenomics) and targeted (sequence-driven metagenomics) metagenomics based on their random and directed sequencing strategies respectively.

The use of shotgun metagenomics was first reported by Tyson et al. (2004) for understanding microbial community in a low-complexity environment i.e. an acidic mine drainage habitat. They have

reconstructed nearly complete genomes of the two dominant species and the partial genomes of three less dominant species, which lead the discovery of microbial metabolic pathways in this particular ecosystem. Subsequently, several shotgun metagenomic studies provided much useful information about potential metabolic pathways of uncultivated environmental microorganisms (DeLong *et al.* 2006; Gilbert *et al.* 2008; Gill *et al.* 2006). However, in predicting the function of more diverse ecosystems like soil, the power of the shotgun metagenomics is rather limited because of inherent problems associated with limited accessibility to the genomes of less abundant members of the community. The recent development of next-generation sequencing technologies has revolutionized the field of metagenomics, enabled researchers to obtain much more DNA information from highly complex microbial communities (Mardis 2008). These NGS platforms such as 454 pyrosequencing, Illumina, Solid™ systems, and Ion Torrent™ are much faster and cheaper than the traditional Sanger method in DNA sequencing. However, short sequence reads (varying between 20 and 700 bp depending on the NGS platform used) and lower-quality sequences lack sufficient information to understand their function and ecological relevance. Furthermore, another mounting problem is the handling and management of the massive amount of metagenomic sequence data, which requires computational infrastructure and accessible tools for storage and further analysis.

In targeted metagenomics approach, a particular genes or genomic regions of interest is subjected to sequencing to reduce genetic complexity (Suenaga 2012). There are two different ways of targeted metagenomics (i) sequence-driven targeted metagenomics and (ii)

function-driven targeted metagenomics. Sequence-driven targeted metagenomics is normally conducted by PCR-based directed sequencing of metagenomic DNA. Using the ribosomal RNA genes (e.g., small subunit [SSU] and large subunit [LSU]) as phylogenetic marker gene is one of the most common methods of identifying microorganisms that have not yet been cultured or that play an important role in the environment (Acinas *et al.* 2004). Although taxonomic resolution of the 16S rRNA gene is not adequate for delineating taxa at the species or strain level in some cases (Fox *et al.* 1992), the SSU of ribosomal RNA (16S rRNA) has been used widely as a standard marker gene in prokaryotes. Microbial community analysis using taxonomic marker genes (e.g., 16S rRNA gene) commonly uses an operational taxonomic unit (OTU)-based approach, as the definition of microbial species is still vague and despite the massive sequencing efforts, currently available public databases are still devoid of the full extent of microbial diversity. As an alternative to sequence-driven targeted metagenomics, function-driven targeted metagenomics is a more direct route to the discovery of gene clusters with related metabolic roles in microbial communities. Function-driven targeted metagenomics has characterized the diversity of the major rumen enzymes approach (Ferrer *et al.* 2005). Function-driven targeted metagenomics can be used to obtain novel findings of targeted biological functions.

1.2.2. The extent of prokaryotic diversity in soil

Soil is known to be one of the most diverse habitats of microorganisms (Gans *et al.* 2005; Torsvik *et al.* 2002). The estimates

of the number of species of bacteria per gram of soil vary between 2000 and 8.3 million (Gans *et al.* 2005; Schloss & Handelsman 2006), of which fewer than 1% are culturable (Amann *et al.* 1995). It is almost impossible to estimate the true extent of microbial diversity in soil at present due to the huge number of unseen microbial cells on earth. Though, there is some consensus on the patterns of microbial diversity in soil. The overall bacterial diversity and abundance decrease with soil depth (Goberna *et al.* 2005; Will *et al.* 2010), while archaea are more abundant in deeper soil layers (Eilers *et al.* 2012; Kemnitz *et al.* 2007). The consequence of land use change on soil prokaryotic diversity is somewhat confounding. Many studies have shown that grassland and agricultural soils are diverse than forest counterparts (Roesch *et al.* 2007), however, a contrasting pattern was also observed (Nacke *et al.* 2011). Geographic distribution patterns of soil microbial communities have been shown to vary across various spatial gradients. Several studies have shown that environmental factors, and among them generally pH, are responsible for these spatial changes at local (Rousk *et al.* 2010), regional (Griffiths *et al.* 2011; Tripathi *et al.* 2012), continental (Lauber *et al.* 2009) and even cross-continental scales (Fierer & Jackson 2006). However, there are also studies that show that stochastic factors also play an important role in structuring microbial community composition (Chytrý *et al.* 2012; Green & Bohannan 2006; Telford *et al.* 2006).

More broadly, microbial species have been perceived as being ubiquitous and are often assumed to be functionally redundant, leading some researchers to suspect that microbes follow different ecological rules to higher organisms (Martiny *et al.* 2006). There is therefore

growing wider interest in whether microbial biogeographic patterns differ fundamentally from those of larger organisms (Horner-Devine *et al.* 2004; Martiny *et al.* 2006; Prosser *et al.* 2007). Several studies have revealed that there is no striking difference in the levels of bacterial diversity between different biomes (Chu *et al.* 2010a) and the diversity gets even lower toward the tropics (Lauber *et al.* 2009). A similar pattern was also observed in this study, showing that tropical biomes have a low ammonia oxidizing archaea (AOA) diversity compared to temperate biomes in surface soil (Chapter 4). However, this trend is not consistent across all microbial lineages. For instance, soil nematode species are found more diverse in tropical rainforest rather than in temperate forests (Porazinska *et al.* 2012) and marine phytoplankton's diversity declined with increasing latitude similar to the 'latitudinal diversity gradient' generally seen in macroorganisms (Barton *et al.* 2010). However, till now it is not clear that the diversity pattern in microbes parallels to those observed in macro-organisms. Microbe-specific traits (i.e. small body size, metabolic diversity, and dormancy) may make their own patterns different from macroscopic taxa.

1.2.3. The status of prokaryotic diversity in tropical soil

Compared to above-ground diversity relatively little is known about below-ground diversity in the tropical regions. Microorganisms constitute a major portion of the biodiversity in soils (Roesch *et al.* 2007) and play an essential role in soil processes (Bardgett *et al.* 2008), which ultimately affect the functioning of terrestrial ecosystems. It has been shown already that the diversity of above- and below-ground biotas are related, perhaps due to interactions between plants and soil

microorganisms and/or through material cycling (Wardle *et al.* 2004). If the interaction between above- and below-ground biotas existed in tropical regions, it might be expected that soil microorganisms in tropical regions would be more diverse and abundant than those in other regions.

There have been some studies investigating the microbial community composition and diversity in tropical and subtropical soils. Borneman and Triplett (1997) used a culture-independent cloning and Sanger sequencing approach based on 16S rRNA sequences to analyze the microbial diversity of eastern Amazonia soils; they identified previously unreported novel sequences, and suggested that "immense" microbial diversity existed in the tropical regions. Also, Wang et al. (1999) studied the diversity of a specific soil bacterial group known as *Actinobacteria* using the culture-based approach in a Singapore forest and found high diversity at the genus level. Nusslein and Tiedje (1999) reported significant changes in soil bacterial community composition due to change in vegetation cover of a Hawaiian soil from forest to pasture, and it has been also reported that conversion of forest to agriculture decreased microbial biomass and produced compositionally and functionally distinct microbial communities in Tahiti (Waldrop *et al.* 2000). Bossio et al. (2005) found that the soil bacterial community at a regenerating secondary forest on one site was more similar to an indigenous forest at another site than it was to nearby agricultural sites. Jesus et al. (2009) used terminal restriction fragment length polymorphism (T-RFLP) analysis on 16S rRNA and found that the bacterial community composition and structure in western Amazon soils were related to land use, likely through the effects of soil

attributes. The authors suggested that ecosystem conversion in the Amazon rainforest did not "deplete bacterial diversity". This observation is fundamentally different from what is reported for plant and animal diversity, which tends to decrease, both above and belowground, after ecosystem conversion (Bierregaard 2001; Lavelle & Pashanasi 1989; Soares-Filho *et al.* 2006). A practical implication of this result is that bacterial biodiversity need not be considered when assessing the impact of large-scale conversion of rainforest to agriculture. However, most of these studies were based on techniques that offer little detail on microbial community composition.

There are also some studies which used NGS methods to evaluate the microbial community composition and diversity in tropical soils. Roesch et al. (2007) used 454 pyrosequencing of 16S rRNA genes and found that estimates of bacterial diversity (the number of OTUs and Chao1 index) were similar in subtropical sites in Brazil and Florida and temperate sites in Illinois and Canada. Their study questioned the idea that high biodiversity is more prevalent in microbial communities in tropical forests. In a global-scale survey, Lauber et al. (2009), found that tropical soils are less diverse compared to soils form other biomes and soil pH emerged as the best predictor of bacterial community composition and diversity. Chu et al. (2010b) also found the similar results when comparing the Arctic tundra soil microbial diversity to other soils of other biomes including tropical soils, reinforcing the view that in contrast with the well-established latitudinal gradient in animal and plant diversity the controls on bacterial community distributions are fundamentally different from those observed for macro-organisms and that our biome definitions are

not useful for predicting variability in soil bacterial communities across the globe. It has been also found recently that land use changes from forest to pasture increase local taxonomic and phylogenetic diversity of soil bacteria, but communities become more similar across space. High-throughput NGS methods have provided us with a better idea of the immense bacterial diversity in tropical soils.

1.3. Microbial community analysis procedure

1.3.1. Initial processing and sequence quality control

Initial processing and sequence quality control analysis begins from raw sequence files (sff format) obtained from 454 pyrosequencing and metadata, which contains barcode/primer information and sample identification code. Initially, each sample is de-multiplexed according to the unique barcode for each sample, followed by the removal of adaptor/primer/linker sequences. Also, shorter (<200bp) and low-quality sequences (i.e. maximal homopolymer at 9bp, minimum ambiguous base of 1, and minimum quality score of 25) are trimmed off. This step normally allows for the minimum level of nucleotide degeneracy and sequencing errors when sorting out barcode and primer sequences (i.e. allowing one mismatch in barcode and two mismatches in each primer sequence). After performing the initial processing and sequence quality control steps, further analysis steps are performed (Figure 1).

1.3.2. Sequence alignment, pre-clustering, chimera removal and taxonomic classification

Next, an alignment of the sequence data is generated by aligning the sequence data against the reference alignment database. There are several reference databases which were aligned by their own standards. Sequences can be aligned against aligned reference sequences of SILVA (51,000 columns) and EzTaxon-e (1,457 columns) using a combination of k-mer search and pair-wise alignment in mothur. Once the sequence alignment is completed, sequences are denoised using the 'pre.cluster' command in mothur, which applies a pseudo-single linkage algorithm with the goal of removing sequences that are likely due to pyrosequencing errors (Huse *et al.* 2010). After that, putative chimeric sequences are detected and removed via the Chimera Uchime algorithm contained within mothur (Schloss *et al.* 2009). Chimeras are artificially generated recombinants between two or more parental sequences during PCR amplification (Bradley & Hillis 1997). They are normally formed when prematurely terminated fragments reanneal to template DNAs originated from different parental molecules due primarily to homologous gene regions between templates. These artificial molecules make it difficult to differentiate the original sequence from recombinants, resulting in overestimating the level of microbial diversity in environmental samples (Hugenholtz & Huber 2003; von Wintzingerode *et al.* 1997). Chimera formation rate during PCR can be reduced by setting PCR cycle as a minimum and increasing extension times (Nolte *et al.* 2010). Then taxonomic classification of each sequencing read is obtained by classifying alignments against EzTaxon-extended reference taxonomy and non-redundant nucleotide databases files using the Naïve Bayesian algorithm (Wang *et al.* 2007). EzTaxon-e database provides a

representative sequence information (type sequence) among closely related but unidentified 16S rRNA gene sequences as well as type strain sequence information from validly published bacterial species (Kim *et al.* 2012b). Closely related sequences are clustered together on the basis of a combination of phylogeny and pair-wise sequence similarity. Hence, it provides taxonomically more meaningful information than that of just similarity-based OTU clustering and the taxonomic resolution and coverage of the database will be better as more sequences are available.

1.3.3. OTU based analysis and further analysis

Once the sequences are aligned and most of the erroneous sequences are removed, alignment-based distance matrix is calculated and clustering process generates OTU at varying cutoffs. There are generally three types of clustering methods: complete-, average-, and single-linkage clustering. Basically, all linkage clustering methods calculate distances between clusters in a hierarchical way and types of methods vary depending on how to define cluster boundary. This OTU-based approach is now generally accepted in microbial community analysis.

Once OTU table or species/sample matrix are generated, versatile diversity analyses can be performed such as measuring diversity indices, individual (sequence)- or sample-based rarefaction curve, and rank abundance curve. When comparing the level of bacterial diversity between samples, using the standardized number of sequences per sample is recommended (cf. cell density or microbial biomass should be considered for more accurate diversity

measurements). If the sequences are already accurately aligned to each other, it can be used for building phylogenic tree, which makes it possible to analyze many things such as measuring phylogenetic diversity, the extent of clumping or overdispersion of certain lineages in phylogeny, and generating distance matrix between communities based on their evolutionary information. Additionally, distance or dissimilarity matrices generated from phylogeny and classical species/sample table can be used for community-level comparison between samples or relating environmental variables to each community to see what factors contribute more to structuring bacterial community within a given sample.

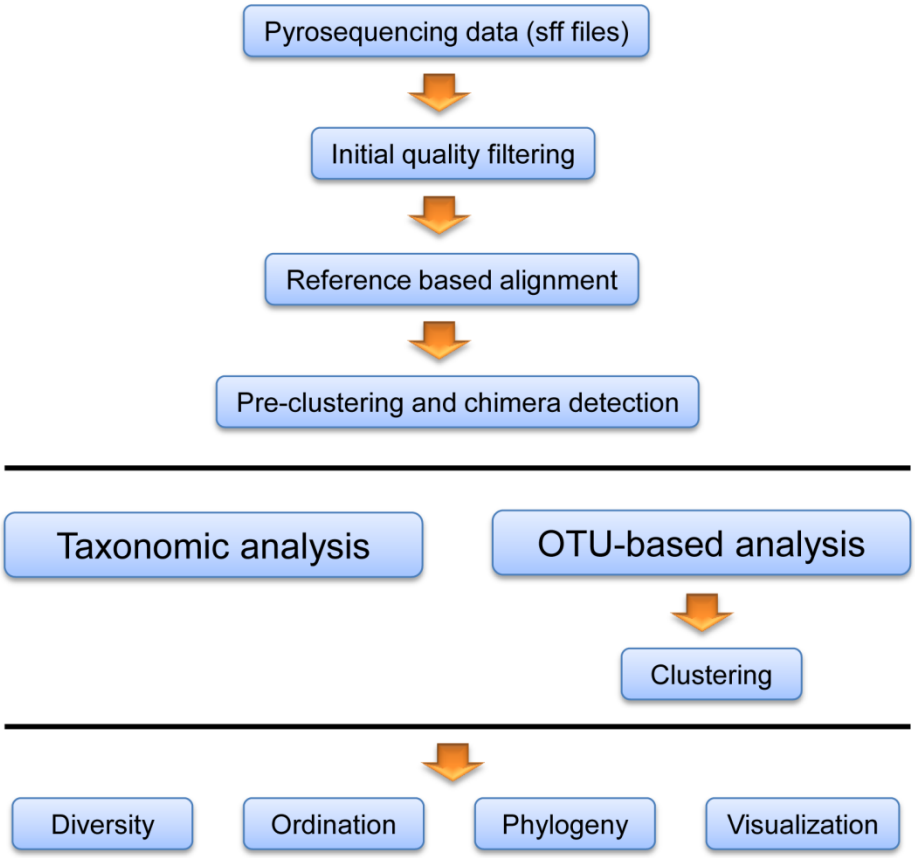

Figure 1. A suggested procedure of microbial community analysis using 454-pyrosequencing data.

1.4. Objectives of this study

There have been many theoretical and empirical studies on biodiversity and biogeography of macro-organisms from Southeast Asian tropical rainforest. Southeast Asia has the highest deforestation rate (Hansen *et al.* 2008; Sodhi *et al.* 2004), which has impacted its rich and unique biodiversity (Brook *et al.* 2003; Sodhi *et al.* 2004). Despite constituting a major portion of the biodiversity in soils (Fulthorpe *et al.* 2008; Roesch *et al.* 2007) and playing an essential role in soil processes (Bardgett *et al.* 2008), very little is known about below-ground diversity in the tropics of Southeast Asia. Therefore, there is a need to study the ecological patterns in microbial diversity and community composition in tropical soils of Southeast Asia.

The present study provides the most thorough research to date on large-scale variation in soil microbial diversity and community composition in tropical soils of Southeast Asia. Soil microbial community structure and diversity was analyzed using 454-pyrosequencing. Hence, this study essentially sets out to answer the following questions:

1) What are the dominant microbial taxa (Bacteria, Archaea and Ammonia Oxidizing Archaea) in tropical soils and how does land use (forest vs. agriculture) influence the structure and diversity of soil microbial communities? What are the major environmental factors linked to differences in the structure and diversity of these communities?
2) Are bacterial communities more similar in samples spatially located closer together than in samples further apart? If so, are

differences in community composition better explained by variation in environmental factors or by spatial distance between samples?

3) What are the dominant archaeal and AOA taxa in tropical and temperate soils? Within each biome, what are the main environmental structuring patterns in the soil archaeal community? Do key soil archaea and AOA inhabit a particular niche, or set of niches, which are predictable, being based on soil edaphic or site characteristics? Are the community patterns and levels of diversity in soil archaea/AOA the same in tropical and temperate environments?

CHAPTER 2. Understanding the Effect of Land Use on Microbial Communities in Tropical Soils of Malaysia

2.1. Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too

2.1.1. Introduction

Southeast Asia is one of the major hot spots of biodiversity (Myers *et al.* 2000). It has been reported that in this region the above-ground diversity has been severely affected by land use changes (Hoffmann *et al.* 2010). Deforestation and agricultural intensification are the most pervasive land use changes in Southeast Asia. In comparison to other tropical regions, Southeast Asia has the highest deforestation rate (Hansen *et al.* 2008; Sodhi *et al.* 2004), which has impacted its rich and unique biodiversity (Brook *et al.* 2003; Sodhi *et al.* 2004). On the other hand, conversion of land to agricultural use such as oil palm plantations has even more detrimental impacts (Edwards *et al.* 2010; Koh & Wilcove 2008). For larger organisms in the tropics, such as plants, insects, birds, or amphibians, there is clear differentiation in species composition and diversity between agricultural and nonagricultural forest environments (Edwards *et al.* 2011; Fitzherbert *et al.* 2008), although, there is evidence that a good proportion of forest species can survive in secondary forests, logged forests, and even exotic tree plantations (Edwards *et al.* 2011; Sheldon *et al.* 2010).

However, very little is known about below-ground diversity in the tropics of Southeast Asia, and the impact of land use upon it. Bacteria constitute a major portion of the biodiversity in soils (Fulthorpe *et al.* 2008; Roesch *et al.* 2007) and play an essential role in soil processes (Bardgett *et al.* 2008), which ultimately affect the

functioning of terrestrial ecosystems. It is important to know the factors that influence the biodiversity of soil bacterial communities, to understand how these communities are structured, and also to predict ecosystem responses to a changing environment.

There have been some studies that have investigated the effects of land use change on the structure of microbial communities in the tropics. Borneman and Triplett (1997) detected significant differences between soil microbial community structure in a mature forest soil and an adjacent pasture soil in eastern Amazonia. Nusslein and Tiedje (1999) reported significant changes in soil bacterial community composition due to change in vegetation cover of a Hawaiian soil from forest to pasture. Bossio et al. (2005) also found similar results in eastern Kenya. In addition, they found that the soil bacterial community at a regenerating secondary forest on one site was more similar to an indigenous forest at another site than it was to nearby agricultural sites. Jesus et al. (2009) found that the bacterial community composition and structure in western Amazon soils were significantly more correlated to changes in soil attributes than land use. However, all these studies used traditional molecular methods such as denaturing gradient gel electrophoresis, terminal restriction fragment length polymorphism, cloning, and Sanger sequencing. These approaches are often limited to the analysis of a relatively small number of clones and a few different soil samples. Taking into account the large bacterial community size and the heterogeneity of soils, only a tiny fraction of the bacterial diversity was unraveled by these studies. With the recent development of high-throughput pyrosequencing of 16S rRNA gene (Roesch *et al.* 2007), in-depth analysis of soil bacterial communities has now become

possible.

The present study provides the most thorough research to date on large-scale variation in soil bacterial diversity across different land use types in Malaysia, one of the major hotspots of biodiversity in Southeast Asia. Bacterial community structure across four land use types was analyzed using 454-pyrosequencing. The main objectives were to (1) identify whether and how the land use (forest vs. agriculture) influences the structure of bacterial communities and (2) identify environmental factors linked to differences in the structure and diversity of those communities.

2.1.2. Materials and Methods

Site description

Samples were taken from forest and non-forest sites within the lowland equatorial tropical rainforest biome (Adams 2009) at sites scattered across central and southern Malay Peninsula and Northern Borneo (Table 1). All sites sampled have an equatorial–wet climate, with precipitation equaling or exceeding potential evapotranspiration in all months of the year but two distinct peaks of rainfall in April–May and October-November in peninsular Malaysia (McGregor & Nieuwolt 1998), where as in May-June and October-January in northern Borneo (Walsh & Newbery 1999). The mean annual temperature is approximately 26.5°C throughout Malaysia with almost no variability in mean monthly temperature (Walsh & Newbery 1999).

In late February 2009, 28 composite samples were collected from primary forest (no record of logging or tree planting in the last

100 years), logged forest (records of logging or planting with native species in the last 100 years), and crop and pasture sites (number of samples per land use type are shown in Table 1). Species cultivated at the crop sites were banana, lemongrass, oil palm, papaya, sugarcane, and tapioca.

Sampling distribution was nonrandom due to a combination of factors. Sampling was determined partly by the logistics of travel time available during fieldwork on other studies. However, samples were deliberately chosen to represent a range of terra firme forest types, soils, and land use types, in lowland Malaysia-while at the same time spaced to avoid spatial clustering. Agricultural sites were sampled during travel between forest sites, their sampling determined by availability of access roads, and ability to secure sampling permission from the farmer or landowner. Agricultural samples were also taken based on crop type (the intention was to sample a range of common Malaysian crops) and pH based on a preliminary pH sample taken at the field site before sampling. The samples were chosen deliberately to represent a range of soil pH levels. The localized, nonrandom distribution of areas of particular pH ranges also prevented random sampling from being a time-effective method. In agricultural sites, only fields with crops close to maturity were sampled, rather than bare fields or stubble. Since liming and fertilizer application tend to take place during earlier growth stages, this will avoid spurious effects of recent application of these chemicals. Fields in which freshly applied fertilizer pellets were visible at the ground surface were not sampled.

Table 1. Sites sampled in this study.

Site names*	Latitude	Longitude	Land use type	Land use grouping
Pasoh forest reserve	N02 57′ 02″	E102 30′ 01″	Primary forest	Forest
Tawau hills reserve	N04 26′ 11″	E117 58′ 44″	Primary forest	Forest
Cape Racado reserve	N02 24′ 30″	E101 51′ 18″	Primary forest	Forest
Batu caves reserve	N03 14′ 37″	E101 41′ 19″	Logged forest	Forest
FRIM Kepong reserve	N03 15′ 23″	E101 37′ 24″	Logged forest	Forest
Meranti forest reserve	N02 30′ 00″	E101 52′ 00″	Logged forest	Forest
Ayer Hitam reserve	N03 01′ 24″	E101 38′ 12″	Logged forest	Forest
Oil Palm- 1	N02 52′ 25″	E101 34′ 28″	Crop field	Agriculture
Oil Palm- 2	N03 12′ 24″	E101 23′ 28″	Crop field	Agriculture
Oil Palm- 3	N02 43′ 10″	E101 38′ 42″	Crop field	Agriculture
Oil Palm- 4	N02 57′ 17″	E102 16′ 46″	Crop field	Agriculture
Papaya	N02 47′ 09″	E102 20′ 19″	Crop field	Agriculture
Sugarcane	N02 38′ 25″	E101 41′ 38″	Crop field	Agriculture
Banana	N03 10′ 09″	E101 33′ 37″	Crop field	Agriculture
Lemongrass	N03 11′ 55″	E102 14′ 58″	Crop field	Agriculture
Tapioca	N02 02′ 10″	E102 43′ 10″	Crop field	Agriculture
Vegetable garden	N02 59′ 15″	E101 36′ 51″	Crop field	Agriculture
Pasture- 1	N02 59′ 31″	E101 29′ 25″	Pasture land	Agriculture
Pasture- 2	N03 16′ 35″	E102 10′ 11″	Pasture land	Agriculture
Pasture- 3	N02 47′ 09″	E102 19′ 93″	Pasture land	Agriculture
Pasture- 4	N02 57′ 49″	E101 43′ 54″	Pasture land	Agriculture
Puchong grassy field	N03 00′ 40″	E101 36′ 16″	Pasture land	Agriculture

*All sites are located in Peninsular Malaysia, except for Tawau hills reserve which is located in Northern Borneo.

Soil collection and DNA extraction

Samples were taken at least 1 km apart. Each sampling point consisted of 1 ha and consisted of five pooled samples. This method, used by Fierer and Jackson (2006), is intended to factor out very local and transient effects (e.g., a single newly fallen leaf releasing hydrogen ions) which might confuse a picture discernible on a larger scale. The intention here was to focus on a large scale rather than highly localized patterns, which would require a separate study. At each hectare sampling point, a scoop of approximately 200 g of soil was taken from a depth of 0-5 cm (upper-layer soil) from each of the four corners of the hectare. An additional sample was also taken in the center of this hectare, and the five samples were then thoroughly homogenized in the same sterile bag. For sample collection, a sterilized trowel was used and cleaned thoroughly between successive samples. Soil samples were composited, and stored at 4℃ for up to 12 h before the samples were sieved through 4-mm mesh and simultaneously stored at −80℃ prior to DNA extraction.

DNA was extracted from each of the collected soil samples using the Power Soil DNA extraction kit (Mo Bio Laboratories, Carlsbad, CA, USA) according to the manufacturer's protocol, with 0.25 g of soil. The purified DNA was resuspended in 50 μL of solution S6 (MoBio Laboratories) and stored at −80℃ until PCR amplification.

Amplification of 16S rRNA genes and pyrosequencing

The extracted DNA was amplified using primers targeting the V1 to V3 hypervariable regions of the bacterial 16S rRNA gene (Unno

et al. 2010). The primers used for bacteria were V1-9F: 5′-X-AC-GAGTTTGATCMTGGCTCAG-3′ and V3-541R: 5′-X-AC-WTTACCGCGGCTGCTGG-3′ (where X barcode is uniquely designed for each soil sample, followed by a common linker AC). Polymerase chain reactions were carried out under the following conditions: initial denaturation at 94℃ for 5 min, followed by 10 cycles of denaturation at 94℃ for 30 s, annealing at 60℃ to 55℃ with a touchdown program for 45 s, and elongation at 72℃ for 90 s. This was followed by an additional 20 cycles of denaturation at 94℃ for 30 s, annealing at 55℃ for 45 s, and elongation at 72℃ for 90 s. The amplified products were purified using the QIAquick PCR purification kit (Qiagen, CA, USA). Amplicon pyrosequencing was performed by Macrogen Incorporation (Seoul, Korea) using 454/Roche GS-FLX Titanium Instrument (Roche, NJ, USA).

Environmental variables

To measure environmental variables, the remainder of each soil samples after DNA extraction was analyzed. Soil samples were oven dried at 60℃ until constant weight. Soil pH was measured in water at the soil to solution ratio of 1:2 using a pH meter. Total nitrogen was determined by sulfuric acid digestion using Se, $CuSO_4$, and K_2SO_4 as catalysts, with 1 g of soil. Total N in the digest was determined by the regular Kjeldahl distillation method (Bremner & Mulvaney 1982). Total carbon was determined by the Carbon Analyzer Leco CR-412 (Leco Corporation, St. Joseph, MI, USA), with 1 g of soil. Exchangeable potassium was estimated using 1 M ammonium acetate buffered at pH 7 (Thomas 1982) and determined by using atomic

absorption spectroscopy using 3 g of soil. Available phosphorus was determined by the method of Bray and Kurtz (1945) by autoanalyzer with 3 g of soil.

Processing of pyrosequencing data and taxonomic analysis

All the sequences were processed and analyzed following the procedures described previously (Unno *et al.* 2010). The total sequencing reads were divided and assigned to each sample by recognition of the unique barcode, followed by trimming sequences by removing barcode, linker, and primer sequences at both sides. The resultant sequences were subjected to a filtering process where only reads containing 0-1 ambiguous base calls (Ns) and 300 or more base pairs were selected for the further analysis. Nonspecific PCR amplicons that showed no match with the 16S rRNA gene database upon BLASTn search (expectation value of $>10^{-5}$) were also removed from the subsequent analyses. Putative chimeric sequences were detected and screened using a similarity-based approach, which splits each query sequence into two even-length fragments and then assigns each fragment to a taxon using BLAST search against EzTaxon extended database (http://eztaxon-e.ezbiocloud.net/) (Kim *et al.* 2012b), followed by removal of the sequences when two fragments differ at the order level or percent identities are greater than 95% for both fragments despite assigned to different taxonomies. All sequences were classified using EzTaxon extended database.

Phylogenetic analysis

The Mothur platform (http://www.mothur.org) was used to compare the community-level bacterial diversity across all 28 soils (Schloss *et al.* 2009). The number of phylotypes (richness) was calculated with a 97% sequence similarity cutoff based on sequence alignment against EzTaxon-aligned bacterial reference sequences. The Faith's index of phylogenetic diversity (Faith's PD) was also estimated (Faith 1992), to avoid the single level of taxonomic resolution. The number of sequences across samples was standardized to avoid incomparability of measurements resulting from different-sized samples. To do this, minimum number of sequences present in a sample was picked that was 555 sequences and randomly select this number of sequences from each of samples. The richness and Faith's PD values was calculated from this subset of 555 sequences per sample. The richness and Faith's PD value was also obtained for specific lineages of bacteria (*Alphaproteobacteria*, *Beta*/*Gammaproteobacteria*, *Acidobacteria*, and *Actinobacteria*). For this lineage-specific richness and Faith's PD estimation analyses, the number of sequences was standardized to 150, 90, 50, and 50 randomly selected sequences per soil for *Acidobacteria*, *Alphaproteobacteria*, *Beta*/*Gammaproteobacteria*, and *Actinobacteria*, respectively. The beta- and gamma-proteobacterial groups were not analyzed separately, as these groups are often combined in certain taxonomic schemes.

Statistical analysis

All statistical analyses were performed on a random subset of

555 sequences per soil sample to avoid effects on diversity metrics due to different number of reads among samples. The difference in overall community composition between each pair of samples was determined using the Bray–Curtis distance (Magurran 2004), which estimates the community distance between two samples. Relationships between Bray–Curtis distance of bacterial community, land use (forest vs. agriculture), and soil properties were assessed using PRIMER v6 (Clarke & Gorley 2006). To look at the effect of land use on bacterial community, an analysis of similarity (ANOSIM) was performed with pairwise Bray–Curtis distance as response variable and land use (forest vs. agriculture) as factor. The RELATE function (a Mantel type test) was used to determine if community-level phylogenetic distances were significantly correlated to soil properties. ANOSIM and RELATE analyses here involve distance matrices (i.e., pairwise distances between samples); pairwise distances are non-independent and therefore statistical significance needs to be evaluated by random permutation. The samples were permuted 999 times and for each time the statistic was calculated; significance is evaluated by comparing the statistic obtained from the real data against the distribution of the statistic obtained by the 999 random permutations. Statistical significance was considered at $P < 0.05$. Nonmetric multidimensional scaling (NMDS) was used to visualize the Bray-Curtis distance of bacterial community between each pair of samples.

Difference on soil properties between forest and agricultural land was determined by using Student's t-test for soil pH, total C, total N, C:N ratio, exchangeable K, and available P, which were normally distributed, and Wilcoxon test for K because its distribution was non-

normal. Correlations between soil properties were analyzed by using VARCLUS procedure in the Hmisc R package. A significant correlation was found only between total C and total N (Figure A1); therefore total N was removed from the analysis. To test whether soil properties that are significantly correlated to Bray–Curtis distance of bacterial community are also correlated to richness and Faith's PD values. The effect of soil pH on the proportion of dominant bacterial taxa was analyzed using compositional analysis (Aebischer *et al.* 1993). This approach is based on the log-ratio analysis of compositions to overcome problems of proportional data (Aebischer *et al.* 1993), for instance the proportions are dependent on each other in a group and the sum of proportions is one (Aebischer *et al.* 1993). In the whole set of data, the four most abundant taxa were chosen and then, the proportion of each chosen taxon was divided by the sum of the proportions of all the remaining taxa (i.e. all taxa besides the four major ones). Then the proportion values were log-transformed. The log-ratio values were analyzed using the regression analysis. The regression analysis was performed using linear and polynomial functions (quadratic and cubic), and the one with better fit was chosen. Regression lines were drawn by using loess function implemented in R software. A t-test was performed to evaluate if the residuals of phylotype richness and phylogenetic diversity after accounting the effect of soil pH differ between forest and agricultural land. Student's t test, Wilcoxon test, VARCLUS procedure, regression analysis, rarefaction curves, and heat map were performed using R software package 2.13.1 (http://www.R-project.org).

2.1.3. Results

General analysis of the pyrosequencing-derived dataset

Across all 28 soil samples, a total of 74,802 quality sequences were obtained, which were classified into 27,318 operational taxonomic units (OTUs) at 97% similarity level. On average, each individual sample was represented by 2,671 classifiable sequences, with a range of 555 to 8,184 sequences per sample. The dominant taxa across all soil samples were *Acidobacteria* (35.8%), *Alphaproteobacteria* (18.1%), *Beta*/*Gammaproteobacteria* (9.4%), *Chloroflexi* (6.8%), *Deltaproteobacteria* (6.0%), and *Actinobacteria* (6.3%) respectively (Figure A2 and Table 2). Most samples showed no sign of reaching an asymptote in OTU richness among the total number of reads (often over 1,000) available in the rarefaction analysis.

Land use and bacterial community

The effect of land use was tested on pairwise Bray–Curtis distances of total bacterial community as well as of four major taxa (*Acidobacteria*, *Alphaproteobacteria*, *Beta*/*Gammaproteobacteria*, and *Actinobacteria*). Significant differences between bacterial community composition for the different land use types were observed for total bacterial community ($r = 0.14$), *Acidobacteria* ($r = 0.11$), *Alphaproteobacteria* ($r = 0.11$), and *Beta*/*Gammaproteobacteria* ($r = 0.08$) ($P < 0.05$ in all cases), but not for *Actinobacteria* ($r = 0.03$, $P = 0.1$). NMDS also indicated some differentiation according to the different land use categories (Figure 2a).

Table 2. Relative average abundances of different taxa classified against EzTaxon database, in different soil types classified into various land-use categories and in all soils combined**.**

Name	Forest		Agriculture		All soils
	Primary forest	Logged forest	Pasture	Crop fields	
Acidobacteria	40.54	30.54	39.46	32.64	35.80
Alphaproteobacteria	21.87	23.17	14.52	12.98	18.13
Beta/Gammaproteobacteria	11.28	10.49	7.25	8.90	9.48
Chloroflexi	1.75	4.56	11.78	9.81	6.97
Actinobacteria	7.20	8.17	5.66	4.30	6.33
Deltaproteobacteria	4.81	7.27	5.56	6.62	6.07
Bacteroidetes	1.20	3.55	2.77	6.21	3.43
Nitrospirae	0.12	1.54	0.76	3.18	1.40
Planctomycetes	3.46	3.03	2.60	2.54	2.90
Cyanobacteria	3.67	1.37	1.74	2.46	2.31
Firmicutes	1.03	0.78	2.59	1.19	1.40
Gemmatimonadetes	0.36	0.94	0.81	2.05	1.04
TM7	0.25	0.51	1.16	1.51	0.86
Verrucomicrobia	1.43	0.65	0.44	0.59	0.78
Elusimicrobia	0.16	0.57	0.47	0.53	0.43
WS3	0.00	0.09	0.16	0.70	0.24
Armatimonadetes	0.20	0.00	0.06	0.56	0.20
Chlorobi	0.03	0.20	0.09	0.45	0.19
AD3	0.12	0.39	0.60	0.47	0.39
OP10	0.03	0.41	0.32	0.50	0.31
OMAN	0.06	0.29	0.19	0.27	0.20
OP3	0.04	0.18	0.18	0.25	0.16
TM6	0.20	0.07	0.17	0.21	0.16
WS5	0.06	0.01	0.03	0.17	0.07
others	0.15	1.23	0.64	0.93	0.74

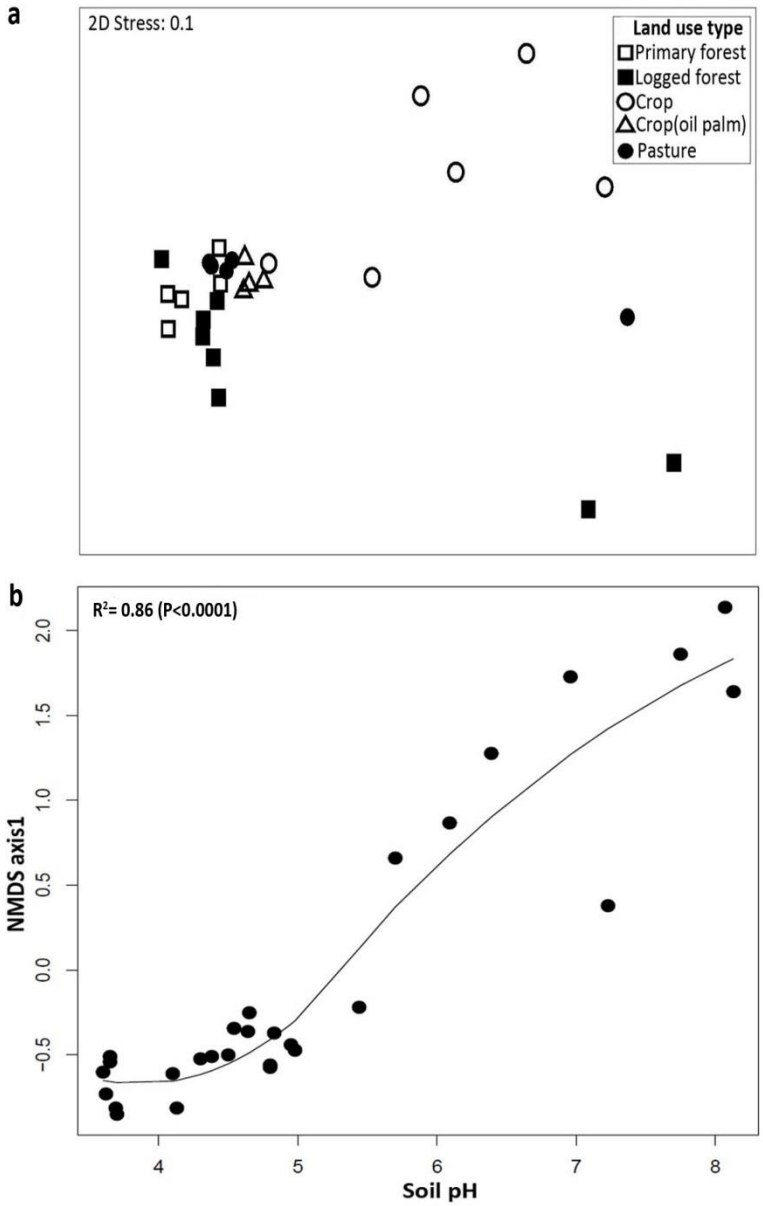

Figure 2. (a) Nonmetric multidimensional scaling plot of the bacterial community using the pairwise Bray–Curtis distances, with symbols coded by general ecosystem type and (b) the first axis of NMDS analysis regressed against soil pH and lines represent the best-fit linear model to the data.

Soil properties and bacterial community

Soil pH ($P < 0.05$) and total carbon ($P < 0.05$) varied significantly between forest and non-forest sites, while total nitrogen, and C:N ratio, available phosphorus, and exchangeable potassium did not vary significantly across the landscape (Table 3). The RELATE test indicated that soil pH, total C, and C:N ratio were significantly correlated with pairwise Bray–Curtis distances of total bacterial community as well as with Bray–Curtis distances of the four major taxa (Table 4). Among the soil parameters measured, soil pH was the most strongly correlated to the pairwise Bray–Curtis distances. Regression between the scores on the first NMDS axis and soil pH also indicated a strong relationship between the bacterial community composition and soil pH (Figure 2b). However, particular agricultural land use categories clustered strongly together, presumably partly as a result of similar pH values. For example, pasture lands form one cluster (except for one sample), and oil palm plantations form another, despite the fact that they are geographically scattered (Figure 2a).

Regression analysis results showed that soil pH, total C, and C:N ratio were significantly correlated to both phylotype richness and phylogenetic diversity of the total bacterial community, as well as of the four major taxa (Table 5). Of these, soil pH gave the most significant correlation ($P < 0.0001$) with both phylotype richness and phylogenetic diversity, when samples from distinct land use types are combined to encompass a wide pH range (Figure 3). For all bacterial groups combined, there is a peak of diversity close to neutral values (pH 7.0) (Figure 3). Soil pH was also significantly correlated ($P < 0.05$) to relative abundances of the dominant bacterial taxa (Figure 4). These

results indicate that pH is the best predictor of overall bacterial diversity across many soils. It was also found that both phylotype richness (P = 0.26) and phylogenetic diversity (P = 0.12) were not affected by land-use.

In terms of OTU level, comparison of relative abundances also revealed many apparent relationships to pH (Figure 5). The heat map of 30 most abundant OTUs shows that among the dominant OTUs in these samples, no single OTU is abundant at all pH levels, although each shows its own pattern peaking at low, intermediate or high pH levels (Figure 5). It thus seems that individual bacterial "species" are specialized in their niche to certain pH levels. Of the 30 most dominant OTUs, only two OTUs (OTU9 and OTU20) showed >97% sequence similarity with cultured bacteria, whilst other dominant OTUs are yet to be cultured (Figure 5).

The soil bacterial community composition and diversity of tropical soils was compared with samples from other biomes: subtropical (Singh *et al.* 2013), temperate (Singh *et al.* 2012b), semiarid (Kim *et al.* 2013) and subarctic (Kim *et al.* 2014). Visualization of the NMDS using pairwise distances (Bray-Curtis) clearly indicated that bacterial communities were very distinct in each biome (Figure 6). However, phylotype richness was not significantly different among tropics, subtropics and temperate biomes (Figure 7a); although semiarid and subarctic biomes were significantly different from each other and also they were different from other biomes (Figure 7a). Similarly, phylogenetic diversity in semiarid and subarctic biomes was significantly different between the two biomes (Figure 7b). Furthermore, they were significantly different from other biomes

(Figure 7b), interestingly the tropic biome was different the subtropics (Figure 7b).

Table 3. Sites sampled in this study and soil chemical properties of samples with different land-use type.

Name	Short name	Land use grouping	pH*	Total C* (%)	Total N (%)	C:N	P (μg/g)	K (μg/g)
Pasoh forest reserve- 1	PAS1	Forest	3.62	4.48	0.43	10.42	14.8	80.1
Pasoh forest reserve- 2	PAS2	Forest	3.69	4.36	0.23	18.96	16.0	68.9
Tawau hills reserve- 1	SAB1	Forest	4.38	5.92	0.52	11.38	13.8	105.5
Tawau hills reserve- 2	SAB2	Forest	4.50	3.51	0.31	11.31	17.2	70.9
Cape Racado reserve	CRA1	Forest	4.13	2.88	0.25	11.52	16.5	45.6
Batu caves reserve- 1	BCA1	Forest	8.13	11.58	0.62	18.68	17.2	117.5
Batu caves reserve- 2	BCA2	Forest	8.07	5.61	0.25	22.44	34.4	50.5
FRIM Kepong reserve- 1	FRM1	Forest	4.10	1.88	0.20	9.40	23.3	53.2
FRIM Kepong reserve- 2	FRM2	Forest	4.30	1.82	0.18	10.11	16.2	60.8
Meranti forest reserve- 1	PDM1	Forest	3.60	3.23	0.37	8.73	7.4	97.5
Meranti forest reserve- 2	PDM2	Forest	3.70	3.23	0.37	8.73	7.4	97.5
Ayer Hitam reserve- 1	AHI1	Forest	3.65	2.21	0.12	18.42	26.3	24.2
Ayer Hitam reserve- 2	AHI2	Forest	3.65	2.21	0.12	18.42	26.3	24.2
Oil Palm- 1	OIL1	Agriculture	4.54	2.65	0.22	12.05	47.0	29.4
Oil Palm- 3	OIL3	Agriculture	4.83	2.44	0.22	11.09	21.8	19.4
Papaya	PAP1	Agriculture	7.23	1.84	0.11	16.73	96.6	79.5

Table 3. Continued.

Sugarcane	SUG1	Agriculture	5.70	1.34	0.14	9.57	42.9	56.7
Banana	BAN1	Agriculture	6.96	1.28	0.14	9.14	20.9	48.4
Lemongrass	LEM1	Agriculture	6.09	1.52	0.20	7.60	27.9	106.3
Tapioca	TAP1	Agriculture	5.44	1.61	0.17	9.47	18.5	49.4
Dragon fruit	DRF1	Agriculture	6.61	2.92	0.30	9.73	0.1	123
Pasture- 1	UPA1	Agriculture	4.98	4.27	0.38	11.24	16.3	59.5
Pasture- 2	UPA2	Agriculture	4.80	4.16	0.35	11.89	16.8	42.6
Pasture- 3	UPA3	Agriculture	4.80	3.20	0.31	10.32	16.2	43.2
Pasture- 4	UPA4	Agriculture	4.95	2.74	0.18	15.22	15.2	36.1
Pasture- 5	UPA5	Agriculture	4.55	3.08	0.36	8.55	17.4	40.5
Puchong grassy field	PGF1	Agriculture	7.75	1.75	0.23	7.61	76.3	57.8

*Variables that were significantly ($P < 0.05$) differed between land-use types.

Table 4. Pearson correlations between Bray-Curtis community distance and soil properties.

Soil properties	Correlation				
	Full community	*Alphaproteobacteria*	*Beta/Gammaproteobacteria*	*Acidobacteria*	*Actinobacteria*
pH	**0.83**	**0.67**	**0.76**	**0.76**	**0.61**
Total C (%)	**0.39**	**0.41**	**0.35**	**0.40**	**0.36**
Total N (%)	0.08	0.05	0.06	0.10	0.06
C:N	**0.46**	**0.43**	**0.37**	**0.48**	**0.29**
P (µg/g)	0.15	0.04	0.19	0.16	0.11
K (µg/g)	0.04	0.02	0.11	0.06	-0.02

Significant correlation coefficients ($P < 0.05$) are denoted in bold.

Table 5. Regression between phylotype richness (OTUs) and phylogenetic diversity (Faith's PD) with total C and C:N ratio for the full community set and the four most abundant phyla[a].

Phylum	Correlation			
	Total C (%)		C:N ratio	
	OTUs	PD	OTUs	PD
Full community	**0.35*****	**0.50*****	**0.10***	**0.20****
Alphaproteobacteria	**0.18*****	**0.24*****	**0.18***	**0.18****
Beta/Gammaproteobacteria	**0.29*****	**0.33*****	**0.11***	**0.13***
Acidobacteria	**0.40*****	**0.55*****	**0.13***	**0.22***
Actinobacteria	0.08	**0.22*****	**0.14***	**0.14***

[a]relationships that are significant ($P < 0.05$) are indicated in bold. The relationships between phylotype richness and phylogenetic diversity with soil pH for the full community and the individual taxa are shown in Figure 6.

*Linear, **quadratic, and ***cubic

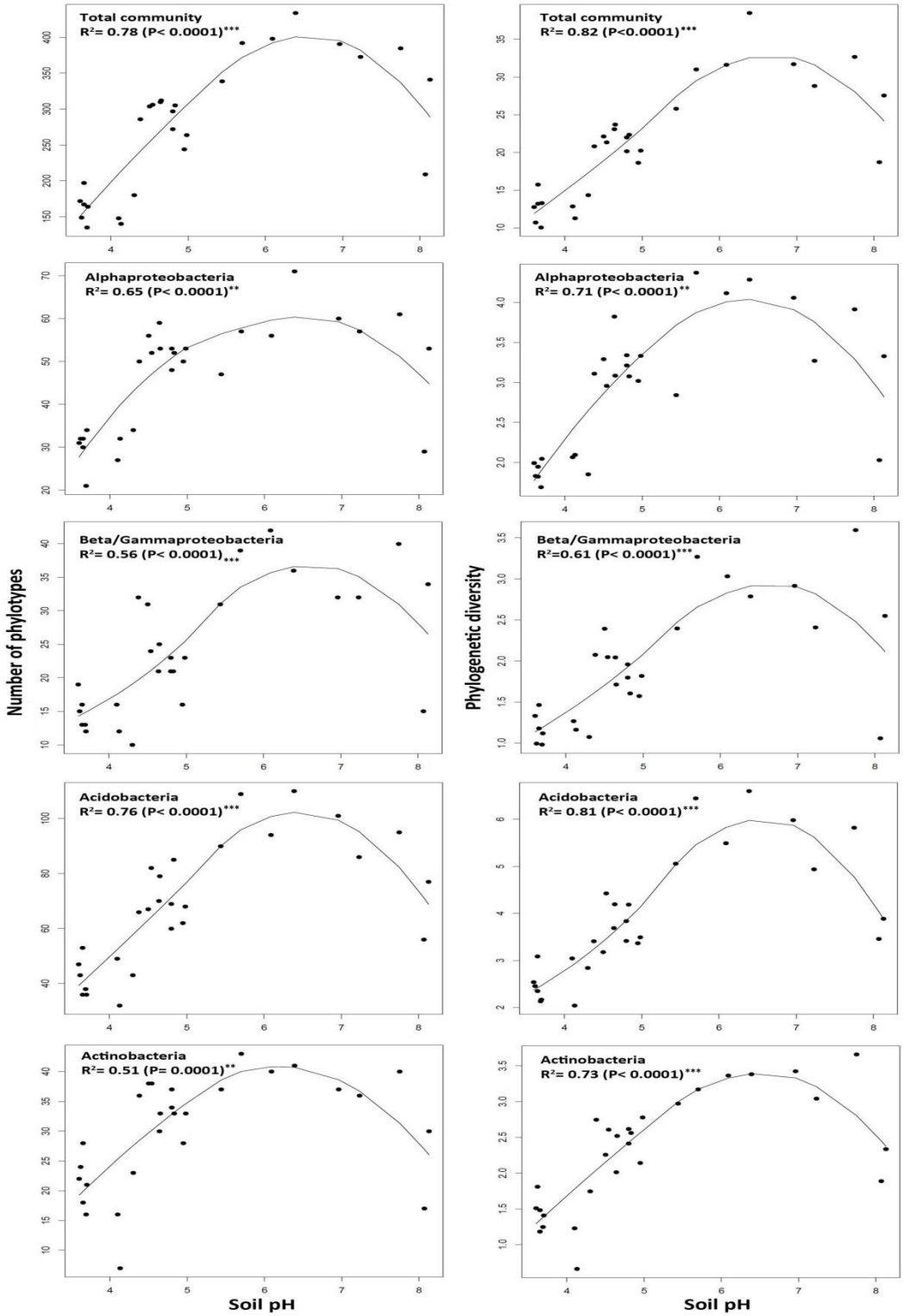

Figure 3. Relationship between soil pH and the number of phylotypes (left) and phylogenetic diversity (right) of total bacterial community and four dominant bacterial taxa. Adjusted R^2 values with the associated *P* values are shown for each taxonomic group. Lines represent the best-fit model to the data (single asterisk linear; double asterisks quadratic; triple asterisks cubic).

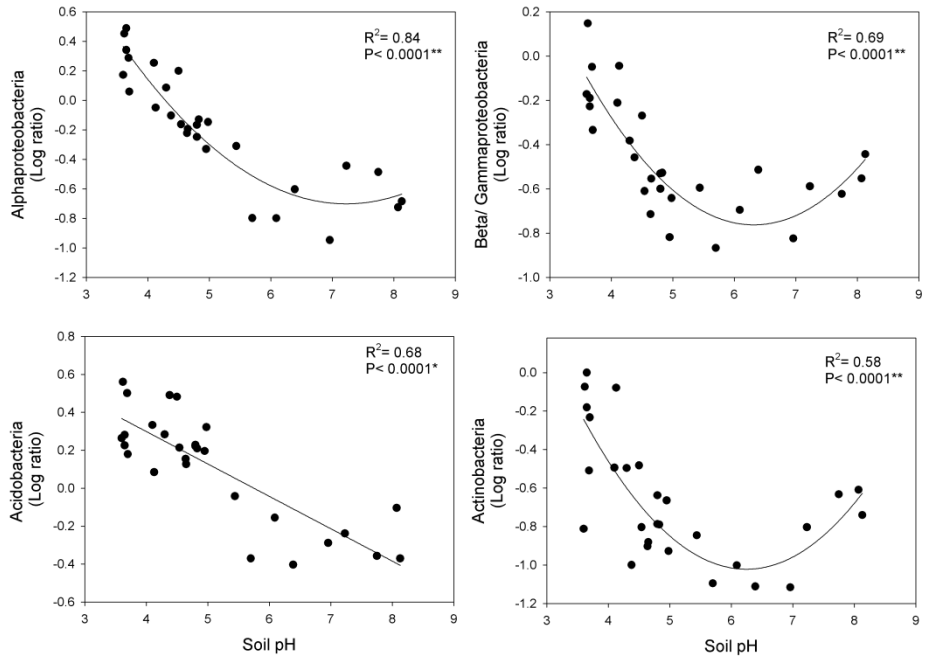

Figure 4. Relationship between log ratio of proportions of the four dominant bacterial taxa and soil pH (Log ratio was calculated using the compositional analysis, which is described in materials and methods section). Adjusted R^2 values with the associated *P* values are shown for each taxonomic group. Lines represent the bestfit model to the data (single asterisk linear; double asterisks quadratic; triple asterisks cubic).

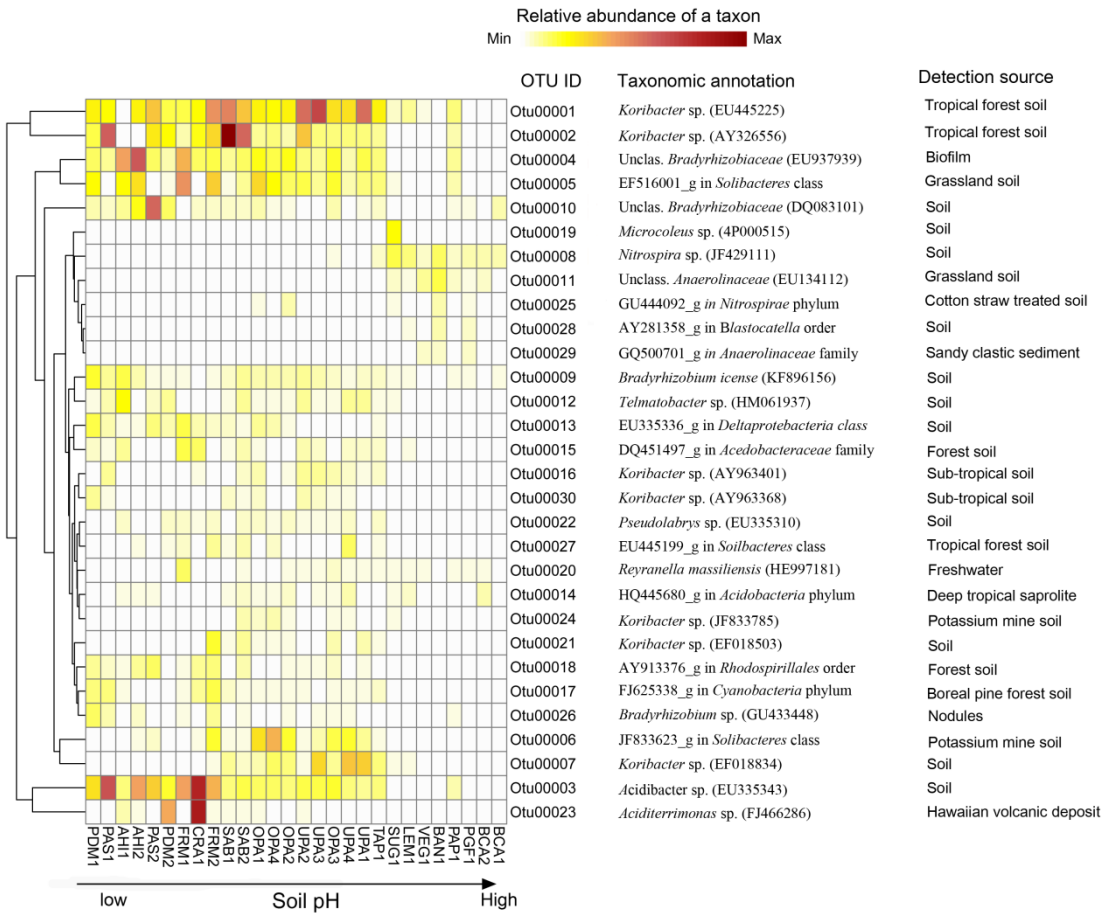
Relative abundance of a taxon
Min
Max
OTU ID
Taxonomic annotation
Detection source
Otu00001 Koribacter sp. (EU445225) Tropical forest soil
Otu00002 Koribacter sp. (AY326556) Tropical forest soil
Otu00004 Unclas. Bradyrhizobiaceae (EU937939) Biofilm
Otu00005 EF516001_g in Solibacteres class Grassland soil
Otu00010 Unclas. Bradyrhizobiaceae (DQ083101) Soil
Otu00019 Microcoleus sp. (4P000515) Soil
Otu00008 Nitrospira sp. (JF429111) Soil
Otu00011 Unclass. Anaerolinaceae (EU134112) Grassland soil
Otu00025 GU444092_g in Nitrospirae phylum Cotton straw treated soil
Otu00028 AY281358_g in Blastocatella order Soil
Otu00029 GQ500701_g in Anaerolinaceae family Sandy clastic sediment
Otu00009 Bradyrhizobium icense (KF896156) Soil
Otu00012 Telmatobacter sp. (HM061937) Soil
Otu00013 EU335336_g in Deltaprotebacteria class Soil
Otu00015 DQ451497_g in Acedobacteraceae family Forest soil
Otu00016 Koribacter sp. (AY963401) Sub-tropical soil
Otu00030 Koribacter sp. (AY963368) Sub-tropical soil
Otu00022 Pseudolabrys sp. (EU335310) Soil
Otu00027 EU445199_g in Soilbacteres class Tropical forest soil
Otu00020 Reyranella massiliensis (HE997181) Freshwater
Otu00014 HQ445680_g in Acidobacteria phylum Deep tropical saprolite
Otu00024 Koribacter sp. (JF833785) Potassium mine soil
Otu00021 Koribacter sp. (EF018503) Soil
Otu00018 AY913376_g in Rhodospirillales order Forest soil
Otu00017 FJ625338_g in Cyanobacteria phylum Boreal pine forest soil
Otu00026 Bradyrhizobium sp. (GU433448) Nodules
Otu00006 JF833623_g in Solibacteres class Potassium mine soil
Otu00007 Koribacter sp. (EF018834) Soil
Otu00003 Acidibacter sp. (EU335343) Soil
Otu00023 Aciditerrimonas sp. (FJ466286) Hawaiian volcanic deposit
PDM1
PAS1
AHI1
AHI2
PAS2
PDM2
FRM1
CRA1
FRM2
SAB1
SAB2
OPA1
OPA4
OPA2
UPA2
UPA3
OPA3
UPA4
UPA1
TAP1
SUG1
LEM1
VEG1
BAN1
PAP1
PGF1
BCA2
BCA1
low
Soil pH
High

Figure 5. The heat map shows the relative abundances of the 30 most abundant OTUs with their taxonomy at different soil pH levels.

2.1.4. Discussion

The results revealed that bacterial community structure was significantly correlated to soil parameters (especially soil pH). The relationship to soil pH in particular is so striking that it seems this is the most dominant factor which influences the bacterial ecology in tropical Southeast Asia.

The results are in agreement with other studies in the tropics (Jesus *et al.* 2009) and elsewhere (Hartman *et al.* 2008; Wakelin *et al.* 2008), where both land use and soil parameters were found to be key factors in determining the diversity and structure of bacterial communities. In the present study, soil pH was found to be the best predictor of bacterial community composition and diversity across all land use types: primary forest, logged forest, and crop and pasture land. Whatever the ecological and evolutionary factors behind these relationships, they appear to be as important in equatorial tropics of Southeast Asia as they are in any other biome or region of the world (Chu *et al.* 2010b; Fierer & Jackson 2006; Griffiths *et al.* 2011; Jesus *et al.* 2009; Lauber *et al.* 2009; Lauber *et al.* 2008).

It is also clear from the data that all four dominant taxa in these samples (*Alphaproteobacteria*, *Beta*/*Gammaproteobacteria*, *Acidobacteria*, and *Actinobacteria*) show a similar trend with pH to that seen for the total bacterial community, in general. It is interesting that although the abundance of *Acidobacteria* shows the "expected"

relationship seen in many other parts of the world (Lauber *et al.* 2009), peaking at low pH values, their OTU richness as well as their diversity shows the same pattern as the bacterial community as a whole - peaking around neutral pH. In fact, none of the taxa studied in greater detail shows a pattern of greater diversity at low pH: where they show any significant trend, they each peak in diversity around neutral.

The relative abundance of the 30 most abundant bacterial OTUs also responded strongly to soil pH and suggests that different OTUs are niche-specialized for growth at different pH levels (Figure 5). The higher sequencing depth of 454-pyrosequencing allows us to look into the information on the potential ecological roles of bacteria in tropical soils. At the OTUs level, most of the dominant OTUs were grouped in to *Korebacter* genus. *Korebacter* belongs to family *Acidobacteriaceae* (*Acidobacteria*) and they are capable of decomposing complex substrates (e.g. xylan, hemicellulose, cellulose, pectin), a property that would support an active intervention in the degradation of plant litter in soils (Eichorst *et al.* 2011; Ward *et al.* 2009). In addition, they are capable of feeding on readily oxidizable carbon (Eichorst *et al.* 2011).

In terms of land use, it is clear that the predictors of diversity in larger organisms such as trees or birds (Berry *et al.* 2008; Edwards *et al.* 2011; Fitzherbert *et al.* 2008) do not hold true for soil bacteria. Generally, from this study, it seems that land use has not significantly affected the richness and diversity of bacteria. However it is important to note that the rarefaction curves for OTU richness did not approach an asymptote, indicating that many more sequences need to be retrieved to census the entire microbial communities.

The pattern seen here is despite the supposition that greater diversity of plant species might be able to give more diverse bacteria specialized to the roots zone of particular plant species or to the decomposing litter beneath each species of plant. It is apparent that the greater bacterial diversity of the agricultural environments is due to their higher soil pH, which partly relates to liming by farmers, and partly to deliberate choice of soils which have naturally higher pH due to their better crop yields. It is interesting that limestone forest, with its high pH, has among the highest soil diversity values for the forest sites, despite its generally low tree species diversity (Whitmore *et al.* 1982).

Attention should now focus on why the striking diversity relationship with pH occurs, and why it is so general, both within and outside the tropics. Lauber et al. (2009) suggested that near-neutral pH might be regarded as a proxy for physiological availability of a variety of nutrients, such as phosphorus, whose chemistry varies with pH value. Though there is no correlation between bacterial diversity and available P, the trends observed could well be the result of precise and distinct chemical forms of these elements at different pH values. Additionally, Lauber et al. (2009) note that the internal pH of bacterial cells is normally close to neutral. An external pH environment similar to this intracellular value may mean less energy expenditure on maintaining internal pH and fewer specialized adaptations. A low pH environment in the tropics could be regarded as a "stressful" environment, sensu Grime (Grime 1987), demanding specialized adaptations that relatively few taxa have been able to acquire. However, part of the usual explanation among ecologists for why stressful/extreme environments are often poor in species is that these environments are rare and

ephemeral in time and space, preventing evolutionary adaptation. This does not seem a convincing explanation in the tropics, where low pH soils with low bacterial diversity predominate. By offering a stable, widespread environment, these acidic soils would be expected to have accumulated the greatest diversity of bacteria.

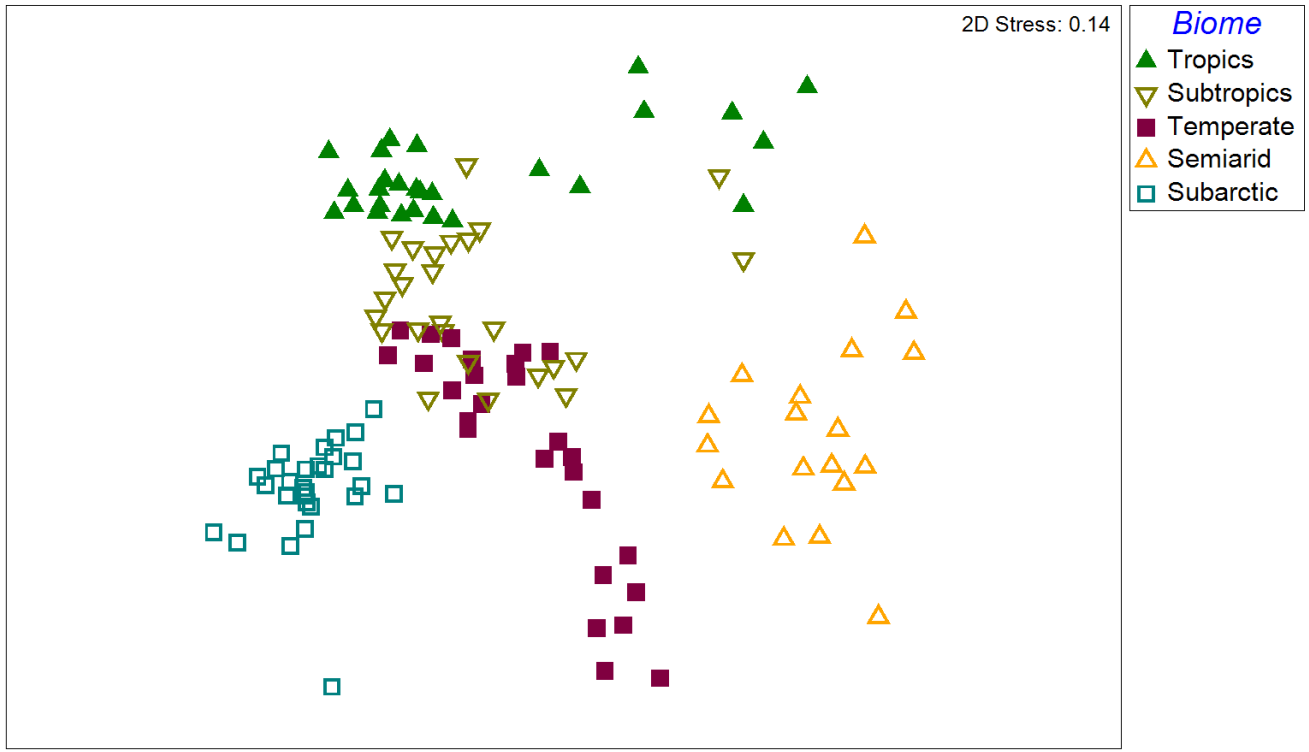

Figure 6. Bacterial community compositional structure in tropical soils and in soils from subtropics, temperate, semiarid and subarctic biomes as indicated by NMDS plot of the Bray-Curtis pairwise distances between sites.

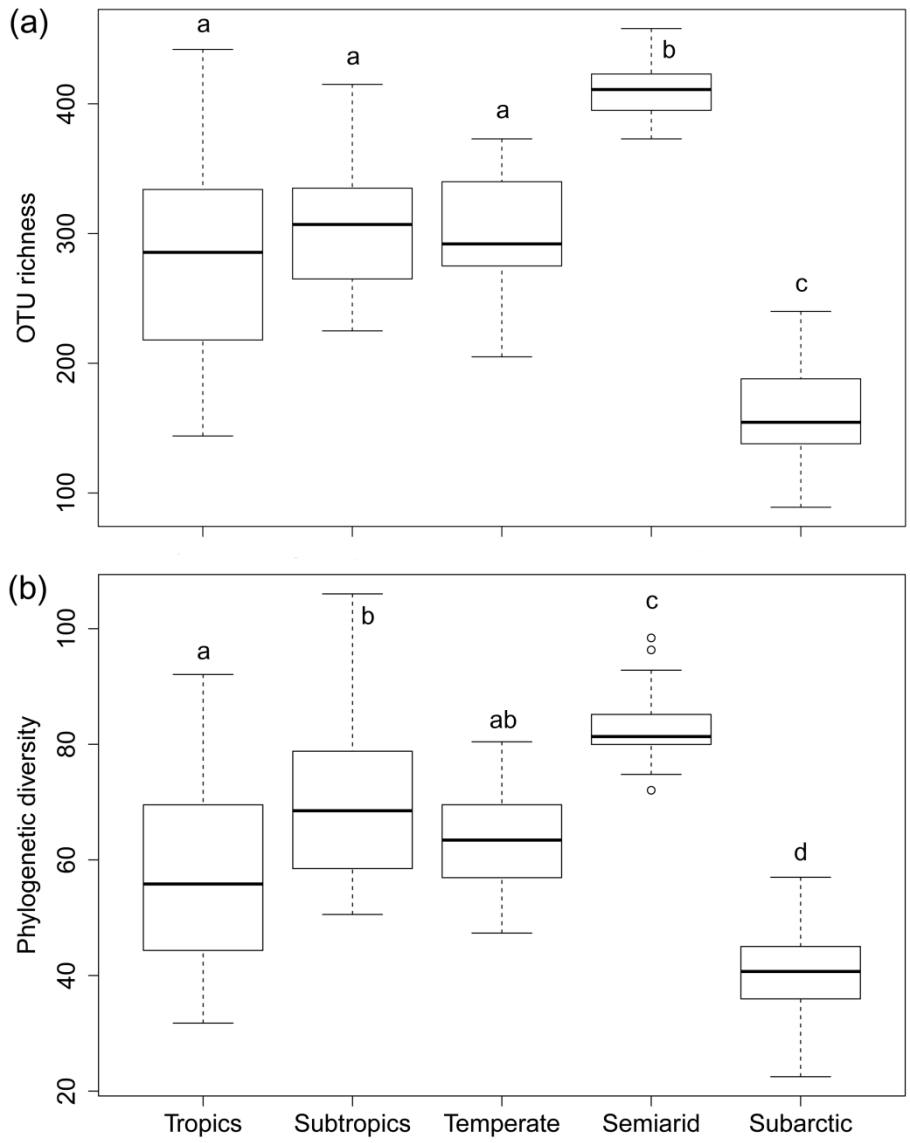

Figure 7. Bacterial OTU richness (a) and phylogenetic diversity (b) in tropical biome as compared with subtropics, temperate, semiarid and subarctic biomes. Diversity indices were calculated using random selections of 500 sequences per soil sample.

Why then might neutral environments in the tropics have accumulated more species of bacteria? A possible explanation is that the specialized adaptations to maintain the intracellular pH level around neutral are much more easily lost than gained. A lineage that speciates into a new niche can easily do so by losing specialized low or high pH adaptations but is much less likely to regain such adaptations. Hence, lineages may be able to "fall into" relatively neutral pH environments but be less likely to escape from them in evolutionary time. The result is a net buildup of bacterial taxa in niches closer to neutral soil pH. Evidence for this pattern might ultimately be found in tracing phylogenetic clades of bacteria within the data. Hypothetically, species differentiation towards the neutral middle of the pH gradient would be more frequent than movement outwards towards its extremes, with "neutral pH" clades remaining more conservative in terms of producing branches which move along the pH gradient.

Bacterial community similarities were much more closely related to differences in biome than differences in soil pH (Figure 6). Bacterial diversity was significantly lower in subarctic biome compared to other biomes (Figure 7). These results are contradictory with Chu et al. (2012) study, which showed that bacterial community composition and diversity were as variable within arctic soils as across the soils from a wide range of lower latitude biomes. These results suggest that the controls on bacterial community distributions are not only regulated by soils pH.

Mid-domain effect (MDE) could be one of the causes for this unimodal bacterial diversity pattern along the soil pH gradient. A MDE was widely used in the interpretation of unimodal diversity pattern in

macro ecology, which emphasizes the geometric constriction upon species distribution, and proposes that species usually accumulate in the middle areas along environmental gradient (Colwell & Lees 2000). However, applicability of MDE in explaining microbial distribution was doubtable when the measurement of dispersal limits for microbes is still ambiguous (Yang *et al.* 2010). As Lauber et al. (2009) have suggested, there is a need for experimental study of the relationships between soil bacterial diversity and pH. Future studies should manipulate individual factors that are known to vary with pH, to better understand their role. Such studies should consider tropical soils, as much as soils from any other part of the world.

2.2. pH dominates variation in tropical soil archaeal diversity and community structure

2.2.1. Introduction

Since the discovery of the Archaea as the third domain of life (Woese *et al.* 1990), much has changed in terms of understanding of their biology and ecology. Initially, Archaea were considered to be restricted to extreme habitats (DeLong 1998; Woese *et al.* 1990). However, new molecular methodologies have revealed the presence of Archaea in a wide variety of habitats (Auguet & Casamayor 2008; Chaban *et al.* 2006; Ehrhardt *et al.* 2007; Lliros *et al.* 2008; Ochsenreiter *et al.* 2003; Oline *et al.* 2006; Youssef *et al.* 2012). Ribosomal gene sequence analysis from uncultivated organisms and the isolation of certain key reference organisms have both expanded our knowledge and led to the proposal of various new candidate phyla. The most prominent addition to this is the candidate phylum *Thaumarchaeota* (formerly described as mesophilic *Crenarchaeota*) (Brochier-Armanet *et al.* 2008). *Thaumarchaeota* dominate archaeal communities in most soils and contain all the currently known ammonia oxidizing Archaea (AOA) (Stahl & de la Torre 2012).

Although they are found in all or most soils, the ecology of terrestrial Archaea remains poorly understood. In a global-scale survey, Auguet et al. (2010) showed that there are certain well-defined archaeal community patterns along broad environmental gradients and habitat types. Similarly, Bates et al. (2011) studied the dominant archaeal populations in 146 soils covering different biomes such as forest, grasslands, deserts, and agricultural soil across North and South

America and Antarctica. These global surveys, together with other local- and regional- scale studies, have revealed a number of factors that may affect archaeal communities, including salinity (Auguet *et al.* 2010), pH (Bengtson *et al.* 2012; Cao *et al.* 2012; Nicol *et al.* 2008), elevation (Zhang *et al.* 2009), climate and vegetation cover (Angel *et al.* 2010), and C/N ratio (Bates *et al.* 2011).

However, these studies used low-resolution molecular fingerprinting tools, which lacked the coverage and depth of high-throughput sequencing methods, like 454-pyrosequencing except for some previous broad-scale survey of terrestrial Archaea (Bates *et al.* 2011; Hu *et al.* 2013). Bates et al. (2011) used general prokaryote primers that gave far fewer reads for Archaea, due to the much lower abundance of Archaea in soils compared with Bacteria and Hu et al. (2013) studied the *Thaumarchaeota* and AOA community from a wide range of soils and ecosystem types in China, though largely from the temperate zone. Although these studies have contributed to a better understanding of Archaea, there is still a need for studies using Archaea-specific primers, which focus on variation in soil archaeal communities on a more regional scale, in environments where some factors (e.g. either climate or soil parameters) are relatively constant, but certain other factors vary strongly. This type of approach will improve understanding of the controls on diversity and composition of archaeal communities. For instance, a study of Mt. Fuji in Japan showed distinct communities of Archaea specific to each elevational zone (Singh *et al.* 2012a).

The present study concentrates on variation in soil community structure of Archaea in lowland tropical moist environments. At present,

tropical forests are being rapidly converted to crop tree plantations or other forms of agriculture (Gibbs *et al.* 2010). These land use changes have resulted in declining species richness and population density of various taxonomic groups of macro-organisms (Gibson *et al.* 2011). Several previous studies have also shown the effect of land use changes on soil bacterial communities in various tropical regions (Borneman & Triplett 1997; Cenciani *et al.* 2009; Jesus *et al.* 2009; Rodrigues *et al.* 2013; Tripathi *et al.* 2012). However, little is known about the potential impact of such changes on soil archaeal communities in the tropics (Navarrete *et al.* 2011; Taketani & Tsai 2010). Furthermore, as few studies have so far used Archaea-specific primers in pyrosequencing, there have been no detailed studies of diversity trends and the niche breadth of Archaea along pH and land use gradients in tropical soils.

Hence, this study essentially sets out to answer the following questions: (1) What are the dominant archaeal and AOA taxa in tropical soils, and how does their relative abundance vary with respect to soil pH? (2) How does the overall archaeal and AOA diversity vary along pH gradients? (3) How does the land use (forest vs. non-forest) influence the structure and diversity of archaeal and AOA communities?

2.2.2. Materials and Methods

Collection site characterization, sampling, and isolation of soil DNA

A total of 27 surface soil samples were collected from forest (13 samples) and non-forest (14 samples) sites within the lowland equatorial tropical rainforest biome at sites scattered across central and

southern Malay Peninsula and Northern Borneo, as described in previous section 2.1.2. Forest sites include primary forest (no record of logging or tree planting in the last 100 years) and logged forest (records of logging or planting with native species in the last 100 years); non-forest sites include crop and pasturelands. Species cultivated at the crop sites were banana, lemongrass, oil palm, papaya, sugarcane, and tapioca. At each sampling site, the top 5 cm of mineral (B horizon) soil was collected from five selected locations within an area of 1 ha. Soil samples were composited, stored, and shipped at 4℃ for 12 h before they were sieved through a 4-mm mesh to thoroughly homogenize and remove roots and plant detritus from the samples. Soil samples were stored at -80℃ until DNA extraction.

The following soil and site characteristics were determined from each sample and used in the subsequent statistical analyzes: pH, total nitrogen, total carbon, available phosphorus, and exchangeable potassium. Details on these soil and site characteristics, the methods used to determine these characteristics, and the procedure utilized for the soil DNA extraction have been described in section 2.1.2. The purified and resuspended DNA was stored at -20℃ and was later used for PCR amplification.

PCR amplification and pyrosequencing

For PCR amplification, bar-coded primers were used targeting the V1 to V3 region of the archaeal 16S rRNA gene, with primers and reaction conditions, as described earlier by Hur et al. (2011). PCR amplification of the archaeal *amoA* gene was performed using the

primers C*amoA*-19f /C*amoA*-616r, as described earlier (Pester *et al.* 2012). The amplified products were purified using the QIAquick PCR purification kit (Qiagen, CA, USA). Amplicon pyrosequencing was performed by Macrogen Incorporation (Seoul, Korea) using 454/Roche GS-FLX Titanium Instrument (Roche, NJ, USA).

Processing of pyrosequencing data and taxonomic analysis

The sequence data obtained after pyrosequencing were processed using mothur (Schloss *et al.* 2009). Sequences shorter than 200 nt with homopolymers longer than 8 nt and all reads containing ambiguous base calls or incorrect primer sequences were removed. Next, the archaeal 16S rRNA and *amoA* gene sequences were aligned against the EzTaxon-aligned and archaeal *amoA* aligned databases, respectively. The *amoA* reference sequences were obtained from the ARB database created by Pester et al. (2012). Putative chimeric sequences were detected and removed via the Chimera Uchime algorithm contained within mothur (Edgar *et al.* 2011).

All the16S rRNA and *amoA* gene sequences were classified against EzTaxon-extended (Kim *et al.* 2012b) and *amoA* reference databases (Pester *et al.* 2012), respectively, using the naïve Bayesian classifier implemented in mothur (at ≥ 80% bootstrap cutoff with 1,000 iterations).

Statistical processing and analysis of results

To correct for differences in the number of reads, which can

bias diversity estimates, a subset of 520 and 1,126 sequences were randomly selected of 16S rRNA and *amoA* genes, respectively, from each sample of the tropical and temperate biomes. Operational taxonomic units (OTUs) of the 16S rRNA and *amoA* genes were defined as sequence groups in which sequences differed by 3 and 15%, respectively. For cultured AOA in which both genes can be compared in the same organism, researchers have found that a 3% 16S rRNA distance equates to a 15% *amoA* gene distance (Pester *et al.* 2012). Nonmetric multidimensional scaling (NMDS) was used to visualize the Bray–Curtis distances of archaeal community across all samples. Taxonomic community composition was compared between forest and non-forest sites using ANOSIM with 999 random permutations (Clarke & Gorley 2006).

The vectors of environmental variables were fitted onto ordination space (Bray–Curtis NMDS) to detect possible associations between patterns of community structure and environmental variables using the 'envfit' function of the Vegan package in R and statistical significance were evaluated by 999 random permutations. Significant soil properties in envfit result were then correlated to the log ratio of abundance values of the most dominant taxa and diversity measures (OTU richness and Shannon index values) in SigmaPlot using linear, polynomial (quadratic) and power (cubic) law functions, and model selection was carried out based on adjusted R^2 and root-mean square error values (RMSE). The compositional analysis as mentioned in section 2.1 was used to obtain the log ratio of abundance values. Student's t-tests was used to evaluate if the residuals of OTU richness and shannon index after accounting the effect of soil pH differ between

forest and agricultural soils. Student's t-tests, Wilcoxon rank-sum tests, rarefaction curve, 'envfit' analysis, and heatmap analysis were performed using R software package 2.15.2. A bootstrapped neighbor-joining phylogenetic tree was constructed for inferring phylogeny after aligning representative phylotypes with reference sequences (J-PHYDIT software) downloaded from NCBI and EMBL in the MEGA 4 software package (Tamura *et al.* 2007).

2.2.3. Results

A total of 90,271 good quality 16S rRNA gene sequences were obtained, with coverage ranging from 557 to 6,683 reads per sample. *Thaumarchaeota* was the most abundant archaeal phylum (95.9% of all archaeal sequences) followed by *Euryarchaeota* (1.6% of all archaeal sequences), the only other phylum present across all the samples (Figure 8a). The majority of the archaeal phylotypes (50.8% of all archaeal sequences) were confined to group 1.1b of *Thaumarchaeota*, whose abundance was higher in non-forest soils (Figure 8b). However, the other two abundant groups of *Thaumarchaeota*, group1.1c and group 1.1a (36.4% and 8.6% of all archaeal sequences, respectively), were abundant in forest soils than the non-forest ones (Figure 8b). For the *amoA* gene, a total of 140,525 high quality sequences with an average of 5,621 sequences per soil were obtained. All of the obtained sequences grouped into four different clusters (Figure 8c), of which *Nitrososphaera* was the most abundant cluster (86.6% of all *amoA* gene sequences). The remaining sequences belonged to the *Nitrosotalea*, *Nitrososphaera* sistercluster and *Nitrosopumilus* cluster (9.6, 3.0, and

0.8% of all *amoA* gene sequences) (Figure 8c). The relative abundance of *Nitrososphaera* was similar across the land use types. However, the other two abundant AOA clusters relative abundance varied across the land use types; *Nitrosotalea* cluster was abundant in forest soils, whereas *Nitrososphaera* sistercluster was abundant in non-forest soils. The most abundant 16S rRNA and *amoA* gene OTUs were together partitioned by soil pH, with no single OTU abundant at all pH levels (Figure 9).

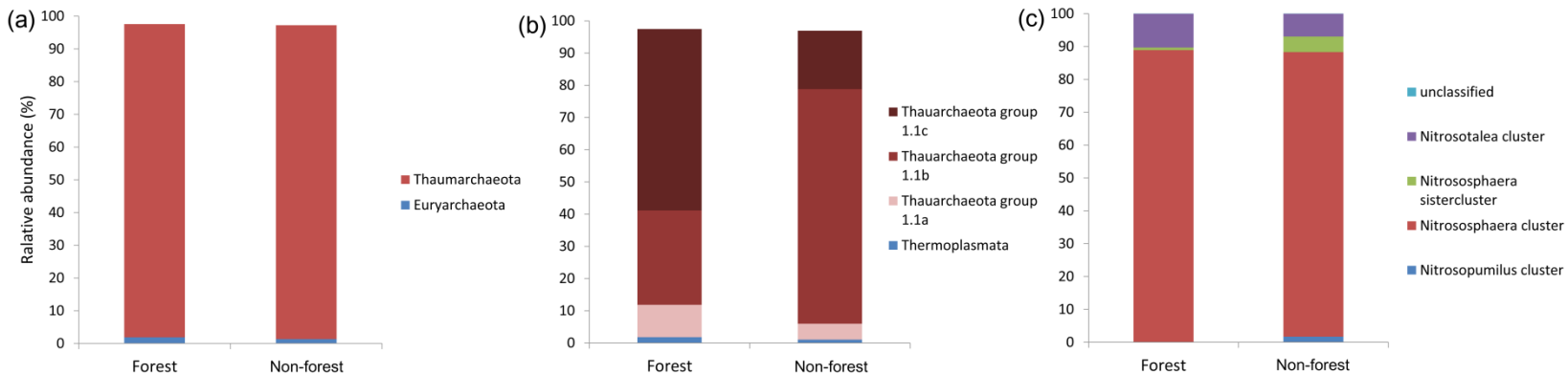

Figure 8. Relative abundances of archaeal taxa at the phylum level (a) and at the subphylum level (b), and AOA taxa (c).

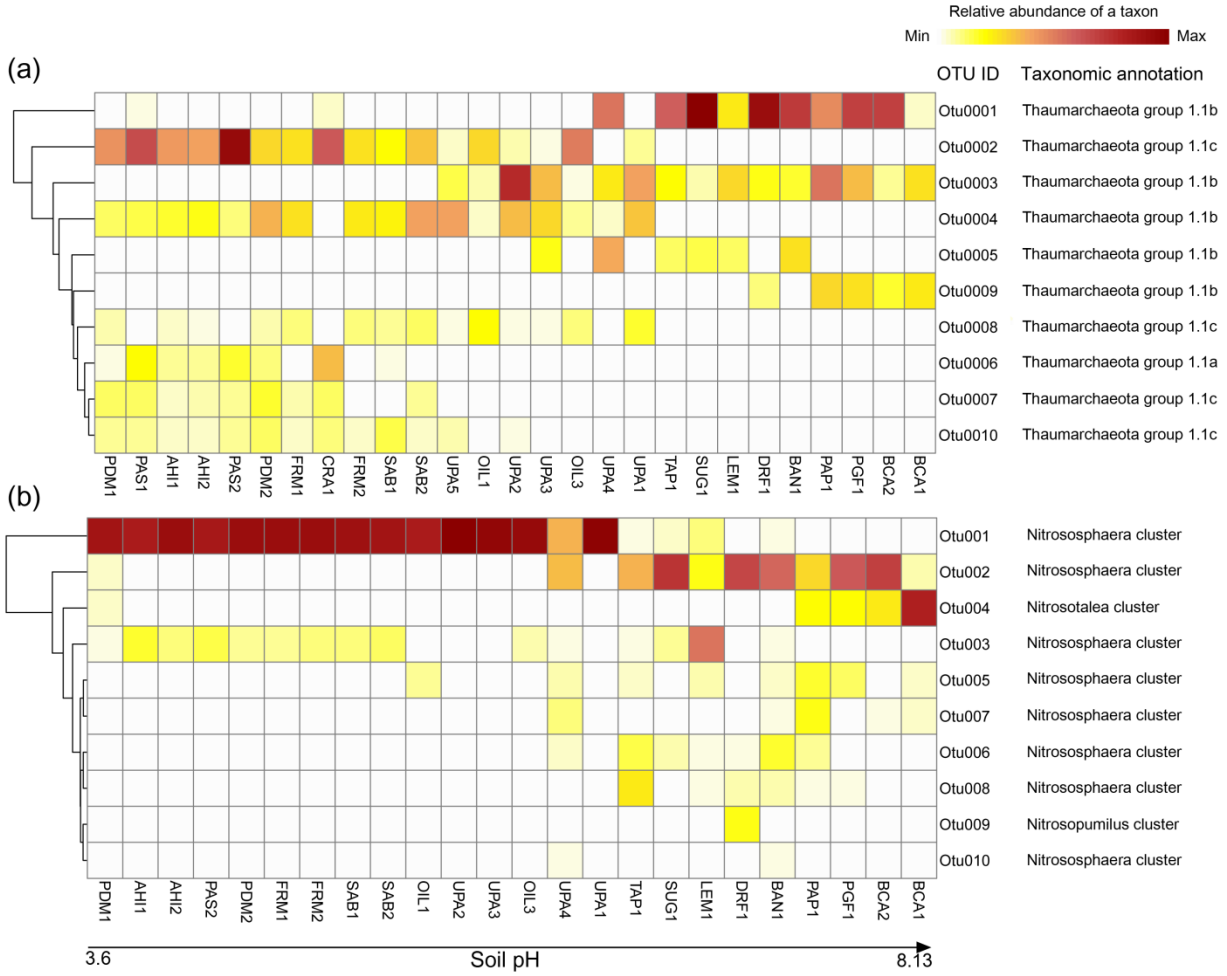

Figure 9. The heat map shows the relative abundances of the 10 most abundant (a) archaeal OTUs and (b) AOA OTUs in tropical soils against soil pH gradient.

An environmental fitting analysis, using six soil variables, showed that only soil pH was highly correlated with total archaea (R^2 = 0.91, $P < 0.001$) and AOA community (R^2 = 0.97, $P < 0.001$) structure on the ordination. pH was strongly correlated to the relative abundance of *Thaumarchaeota* group 1.1b (R^2 = 0.70, $P < 0.0001$), group 1.1c (R^2 = 0.68, $P < 0.0001$), and less strongly to *Thermoplasmata* (R^2 = 0.19, $P < 0.05$; Figure 10a). Soil pH also correlated strongly with the dominant AOA clusters (*Nitrososphaera* sub-cluster 1 and sub-cluster 7 and *Nitrososphaera* sister cluster; Figure 10b). Regression analysis results showed that soil pH was significantly correlated to both OTU richness and Shannon index of the total archaeal community, as well as of the AOA (Figure 11). Total archaeal OTU richness and Shannon index negatively correlated to soil pH (Figure 11a and Figure 11b), whereas AOA OTU richness and Shannon index positively correlated to soil pH (Figure 11c and Figure 11d).

An NMDS plot of Bray-Curtis distance showed that soil pH was a strong structuring factor in the total archaeal and AOA assemblages with samples belonging to different pH zones harboring relatively distinct communities (Figure 12). Similarly, both total archaea and AOA community structure was significantly different between forest and non-forest soils in Bray-Curtis distance-based result (ANOSIM, $P < 0.01$). For total archaea and AOA, OTU richness and Shannon index did not vary significantly with land use ($P > 0.05$).

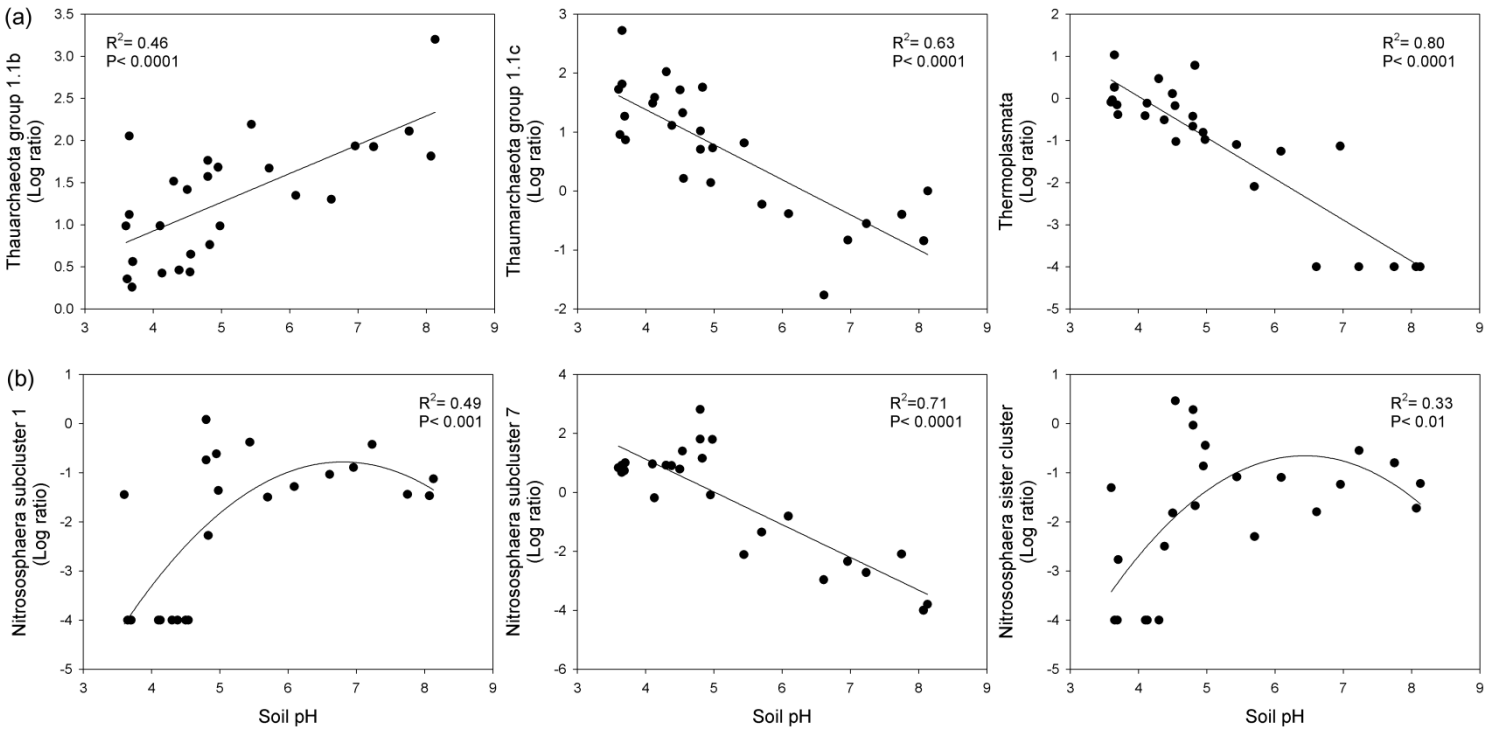

Figure 10. Relationship between soil pH and log ratio of proportion data (Obtained using compositional analysis mentioned in materials and method section) of the three dominant archaeal taxa (a), and three dominant AOA taxa (b), in tropical Malaysian soils.

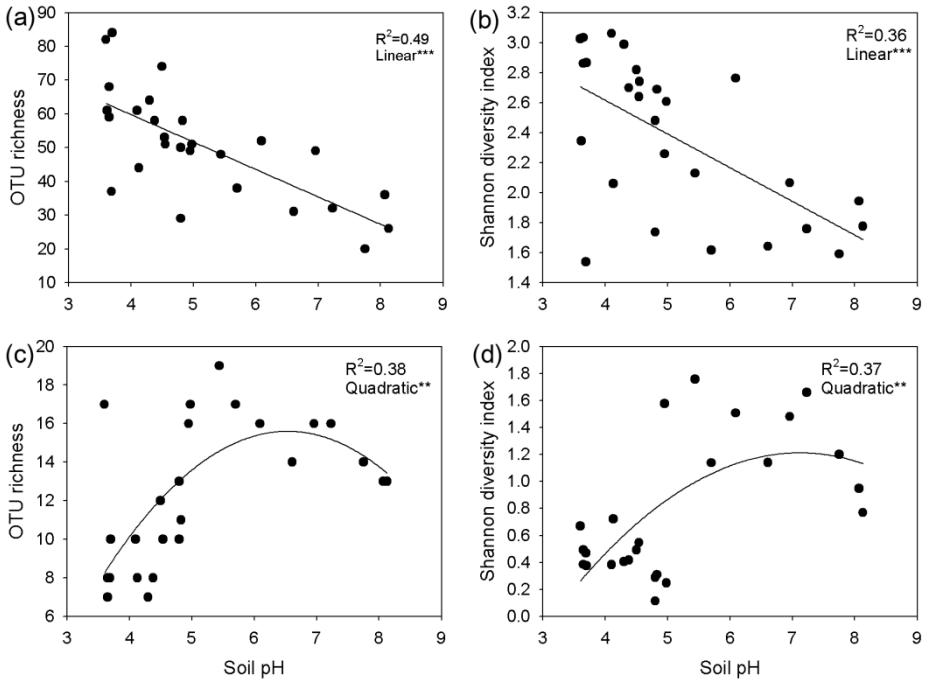

Figure 11. Relationship between soil pH, OTU richness and diversity of total archaea (a and b) and AOA (c and d), in tropical Malaysian soils.

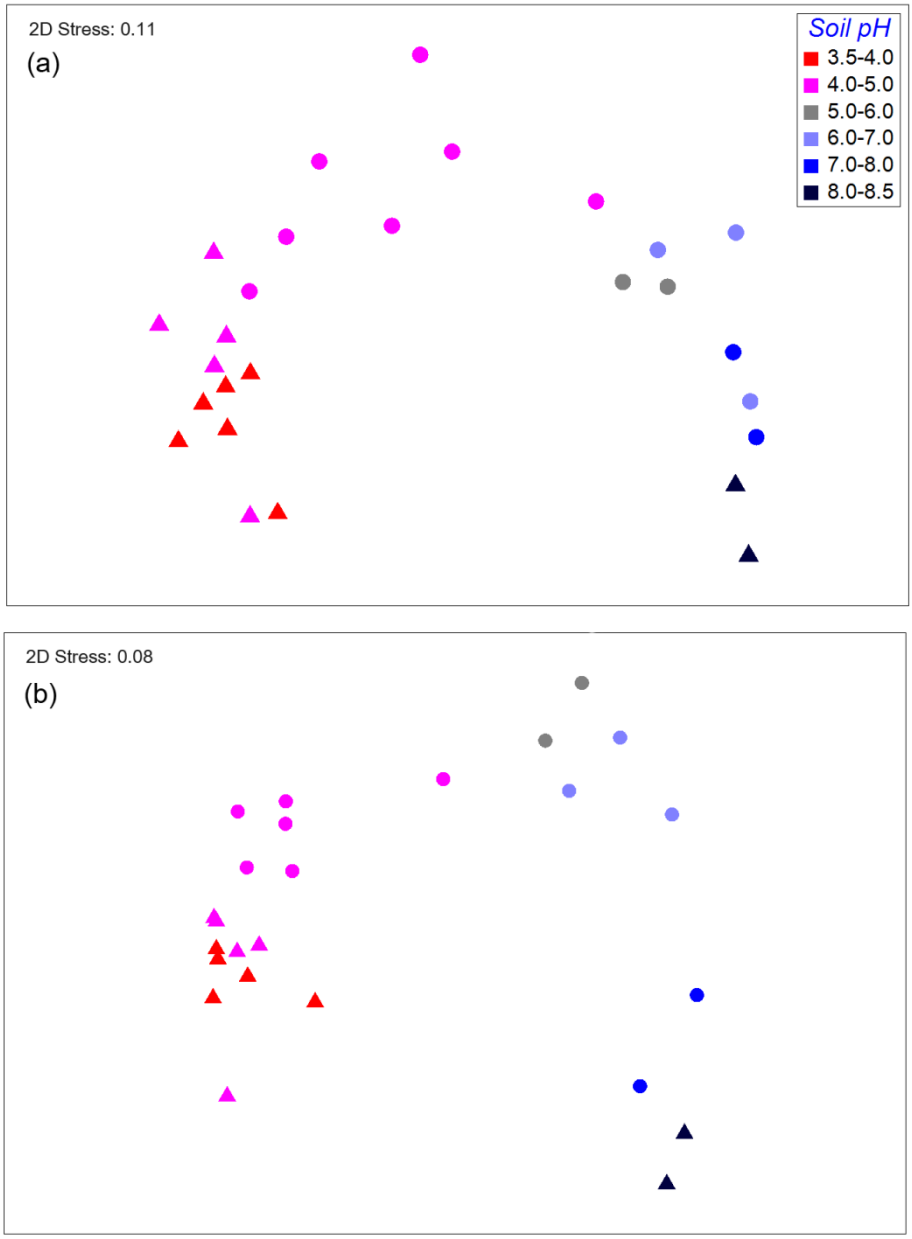

Figure 12. NMDS plot of Bray–Curtis dissimilarities of (a) total archaea and (b) AOA between samples; a red-to-blue panel showing soil pH; triangles and circles representing the forest and non-forest land use types.

2.2.3. Discussion

Community composition, diversity, and soil pH

Soil pH was found to be the best predictor of community composition and of the relative abundance of major subgroups of total Archaea and AOA-analogous to the trend found in the bacterial community (Tripathi *et al.* 2012). In the 16S rRNA gene dataset, soil samples were overwhelmingly dominated by the thaumarchaeotal groups 1.1b and 1.1c (87.3% of all archaeal sequences), which have previously been shown to dominate terrestrial habitats (Auguet *et al.* 2010). The relative abundance of these two groups and of the euryarchaeotal *Thermoplasmata* varied significantly with soil pH. For instance, the relative abundance of group 1.1c and *Thermoplasmata* declined sharply with increasing soil pH (between pH 3.65 and 4.95), and the opposite was true for group 1.1b. These results are in agreement with those of Lehtovirta et al. (2009) in the temperate zone, where they also observed a steep decline in abundance of group 1.1c with increasing pH (between pH 4.5 and 6.0). *Thaumarchaeota* group 1.1c is also commonly the dominant group of Archaea in acidic forest soils (Kemnitz *et al.* 2007; Nicol *et al.* 2007; Nicol *et al.* 2005). Similarly, *Thermoplasmata* are seen as moderately thermophilic or mesophilic microorganisms growing at low pH values (pH optima 0.7 to 3) in acidic thermal springs, solfataras, and volcanic soils (Kletzin 2007). In contrast, the positive relation of group 1.1b with pH is in accordance with the study by Ayton et al. (2010), where they found that group 1.1b was prevalent in mineral alkaline soils of the Ross Sea region. Gubry-Rangin et al. (2011) also found the same relationship

with pH for group 1.1b using a different marker gene (*amoA*). The results here are much clearer than those found by Bates et al. (2011), who surprisingly found less of a strong relationship between different taxonomic groups and soil pH. In the *amoA* gene dataset, the *Nitrososphaera* cluster was detected as the most dominant cluster in both biomes. In several previous *amoA* diversity studies, *Nitrososphaera* has also been shown to dominate the AOA community in soil (Cao *et al.* 2013; Gubry-Rangin *et al.* 2011; Pester *et al.* 2012). Dominant AOA clusters also showed variation with soil pH and these results are in agreement with previous studies (Gubry-Rangin *et al.* 2011; Zhalnina *et al.* 2012).

Archaeal diversity in all of the soil samples was very low in comparison with bacterial diversity in the same samples (Tripathi *et al.* 2012), a pattern which is in agreement with other studies on soil Archaea (Auguet *et al.* 2010; Singh *et al.* 2012a). Total archaeal OTU richness and Shannon index were higher in acidic soils and this could be due to the fact that most of the archaeal classes group 1.1a, group 1.1c and *Thermoplasmata* are well adapted to acidic soils. However, AOA diversity showed a humpback relationship with soil pH, with diversity peaking around at neutral pH. This is consistent with other studies on AOA around the world (Gubry-Rangin *et al.* 2011; Pester *et al.* 2012), but showing the pattern more clearly here, perhaps due to larger numbers of samples having been gathered from the same biome. This humpback relationship is consistent with the fact that most of the AOA identified to date are adapted to neutral or slightly alkaline conditions (De La Torre *et al.* 2008; Könneke *et al.* 2005; Kim *et al.* 2012a; Tourna *et al.* 2011), except for Ca. *Nitrosotalea devanaterra*,

which survives in acidic conditions (i.e., pH 4.0-5.5) (Lehtovirta-Morley *et al.* 2011a).

Effect of land use

Although the relationship between archaeal and AOA community composition and soil pH was very strong, there was also a significant effect of land use on the archaeal and AOA community. These results are in broad agreement with previous studies in tropics, where both land use and soil parameters were found to be key factors in determining the structure of bacterial and archaeal communities (Jesus *et al.* 2009; Taketani & Tsai 2010; Tripathi *et al.* 2012). However, the effect of land use on archaeal and AOA diversity was non-significant after accounting the effect of soil pH. Therefore, the patterns in land use are mainly due to soil pH, and the stronger explanatory power of pH strongly suggests that it is actually the dominant factor. The detection of archaea and AOA might be limited by low primer universality due to non-representative archaeal and AOA databases, so the conclusions of this study must be interpreted with caution.

CHAPTER 3. Spatial Scaling Effects on Bacterial Communities of Tropical Rainforest Soils

3.1. Spatial Scaling Effects on Soil Bacterial Communities in Malaysian Tropical Forests

3.1.1. Introduction

Spatial scaling is known to play an important role in biodiversity patterns of larger organisms. In general, it is common to find greater turnover in species lists between more widely spaced samples (Bell 2001; Huston 1994). For example, for organisms such as trees and vertebrates, the community similarity decreases with increasing spatial distance (Condit *et al.* 2002; Koleff *et al.* 2003; Qian & Ricklefs 2007). In comparison to these macroorganisms, the role of spatial scaling has been less studied in microorganisms and there is no agreement on how their distributions are structured in space (Green & Bohannan 2006). However, with the rapid development of molecular-based techniques to assess the microbial diversity, a new interest in understanding the biogeographical patterns of soil microbial community has emerged (Martiny *et al.* 2006; Nemergut *et al.* 2011).

Soil bacteria constitute a major portion of the biodiversity in soils (Roesch *et al.* 2007) and play a vital role in soil processes (Bardgett *et al.* 2008); thus, understanding their biogeographical patterns and what drives them is crucial for maintaining ecosystems. The classic Baas-Becking hypothesis concerning microbial biogeographical patterns predicted that ‘everything is everywhere, but, the environment selects’ (Baas-Becking 1934). In other words, biogeographic patterns in microbes are thought to result from environmental heterogeneity rather than dispersal limitation. Recent research proposes more complex patterns of microbial biogeography

than the Baas-Becking hypothesis predicts. For example, several studies have indicated some degree of habitat endemicity (Green & Bohannan 2006; Martiny *et al.* 2006) at scales ranging from a few centimeters (Franklin & Mills 2003) to hundreds of kilometers (Cho & Tiedje 2000). Others have attempted to assess the importance of environmental factors versus dispersal limitation of bacterial communities. Some of these studies have found significant environmental effects on bacterial community structure (Angel *et al.* 2010; Fierer & Jackson 2006; Horner-Devine *et al.* 2004), with soil pH being the main responsible for spatial changes at local (Rousk *et al.* 2010), regional (Griffiths *et al.* 2011; Tripathi *et al.* 2012), and continental scales (Lauber *et al.* 2009), whereas others have suggested that dispersal limitation may have a more important role (Chytrý *et al.* 2012; Green *et al.* 2004; Telford *et al.* 2006).

Most studies of microbial biogeography have explored spatial scaling in temperate regions, while tropical rainforests have seldom been investigated, despite being the acme of plant diversity and of other groups of organisms. It is possible that different rules apply to the scaling of microbial life in this biome. For instance, if soil microbial community is partly determined by the presence of certain host organisms, such as trees, and community composition of these vary with distance in ways apparently not related to the environment (Condit *et al.* 2002), then it could be expected that the soil microbial community composition would change in relation to variation on the community of the host organisms. Also, several environmental factors, such as temperature and humidity, which can have a strong influence on soil microbial communities, are less spatially variable over large

expanses of tropical regions than in temperate ones. Therefore, there is a need to study spatial scaling of soil bacterial communities in tropical soils on a more regional scale.

In this study, it was aimed to assess the patterns of bacterial community composition in relation to distance and environmental variables in the tropical rainforest of Southeast Asia. This region is well known for its species-rich tropical rainforest, representing a 'hotspot' of biodiversity (Myers 1988), but it has been poorly studied in relation to its microbial diversity. The tropical soil bacterial communities at local and regional scales were studied by focusing on the following questions: (i) are bacterial communities more similar in samples located closer together than in samples further apart? If so, (ii) are differences in community composition better explained by variation in environmental factors or by spatial distance between samples?

3.1.2. Materials and Methods

Sampling locality and field sampling strategy

Samples were taken from equatorial tropical rainforests across Peninsular Malaysia and Northern Borneo. All sites sampled have an equatorial-wet climate, with precipitation equaling or exceeding potential evapotranspiration in all months of the year. Two distinct peaks of rainfall occur in April–May and October–November in Peninsular Malaysia (McGregor & Nieuwolt 1998), and May– June and October–January in Northern Borneo (Walsh & Newbery 1999). In late February 2010, soil samples were collected at two different sampling scales: local and regional (Figure 13). For the local scale, a 150-m straight line transect was set up at the FRIM Kepong tropical rainforest

reserve. A total of 30 soil samples were taken every 5-m along this transect (Table A1). For the regional scale, a total of 26 samples were collected from lowland and tropical montane forests in sites scattered across Peninsular Malaysia and Northern Borneo (Table A2). All the forest sites sampled were either primary forest (no history of logging) or mature logged forest (records of logging or planting with native species in the last 100 years, but recovery for at least the last 30 years; Table A2). At each sampling site, the top 5 cm of mineral (B horizon) soil (after removing the litter layer) was collected from five selected locations within an area of 1 ha (the four corners of the 1 ha plot, and the central point of the plot), and the five samples were then thoroughly homogenized in the same sterile bag. Samples in all cases were stored at moderate temperatures (20–25℃) for no more than 12 h, until they could be stored in a deep freezer at -20℃.

Environmental variables and soil DNA extraction

Soil pH, total carbon (TC), total nitrogen (TN), available phosphorus, extractable potassium and soil texture were determined for each collected sample at both scales (Tables A1 and A2). Because regional scale sampling sites were widely scattered across Peninsular Malaysia and Northern Borneo, elevation, mean annual temperature (MAT), and mean annual precipitation (MAP) were also included as environmental variables (Table A2). Details on the methods used to determine these characteristics, and the procedure utilized for the soil DNA extraction have been described in previous sections 2.1 and 3.1. The purified and resuspended DNA samples were stored at -20 ℃ and were later used for PCR amplification.

PCR amplification, pyrosequencing, sequence processing, and taxonomic analysis

DNA amplification was performed using primers targeting the V1–V3 regions of the bacterial 16S rRNA gene as described previously by Chun et al. (Chun *et al.* 2010). DNA sequencing was performed by Macrogen (Seoul, South Korea) using 454 GS FLX Titanium sequencing system (Roche), according to the manufacturer's instructions. The sequence data obtained after pyrosequencing were processed using mothur (Schloss *et al.* 2009). Sequences shorter than 200 nt with homopolymers longer than 8 nt, and all reads containing ambiguous base calls or incorrect primer sequences were removed. Next, the sequences were aligned against the EzTaxon-aligned reference sequences (Kim *et al.* 2012b). Sequences were denoised using the 'pre.cluster' command in mothur, which applies a pseudo-single linkage algorithm with the goal of removing sequences that are likely due to pyrosequencing errors (Huse *et al.* 2010). Putative chimeric sequences were detected and removed via the Chimera Uchime algorithm contained within mothur (Edgar *et al.* 2011). Taxonomic classification of each OTU, clustered at ≥97% sequence similarity, was obtained by classifying a representative sequence of each OUT against RDP training dataset number 9 (available at http://www.mothur.org/wiki/RDP_reference_files) using the classify command in mothur at 80% naive Bayesian bootstrap cutoff with 1,000 iterations.

Statistical analyses

To correct for differences in number of reads, which can bias diversity estimates, a subset of 520 sequences was randomly selected from each sample at the local and regional scales. Two samples from the local scale series with less than 520 sequences were removed from the analyses. Additionally, a subset of the most abundant OTUs was also selected at each scale, arbitrarily defined as those containing >100 reads/OTU in the total dataset (i.e., all samples included), following Logares et al. (2012). To determine if the relative abundance of the most abundant bacterial taxa differs between local and regional scales a t-test was used for normal data and Wilcoxon rank-sum test for non-normal data. The Spearman rank correlation was used to test the relationship between the relative abundance of dominant bacterial phyla and soil environmental variables.

For the analyses of community similarity, a Bray–Curtis similarity matrix (1-dissimilarity) was calculated for the whole set of bacterial OTUs and also for only the most abundant OTUs using the vegdist command from the vegan R package (Oksanen *et al.* 2007). To assess the relationship between Bray–Curtis community similarity of the whole community and of the most abundant OTUs with geographical distance (log transformed) and environmental distance (Euclidean) at the local and regional scale, a Mantel test was performed with 999 permutations. Regional scale samples were collected from forests with different logging history (Table A2), thus an analysis of similarity (ANOSIM) was used with 999 permutations to test if bacterial community composition differed in relation to logging history.

A nonmetric multidimensional scaling plot (NMDS) was used based on the Bray–Curtis dissimilarities to visualize the clustering of

bacterial community composition for whole OTUs and the most abundant ones. To examine the relationship of Bray–Curtis community dissimilarity with spatial distance and with environmental distance at both spatial scales, a ranked partial Mantel test was used as implemented in the ecodist R package (Goslee & Urban 2007). The environmental distance matrix was built using a combination of the normalized variables selected by the BEST procedure in PRIMER v6, and then the Euclidean distance was used to calculate the environmental distance between each pair of samples. To further evaluate the relative importance of spatial distance and of each of the environmental variables on bacterial community dissimilarity, a multiple regression on matrices (MRM) approach was used (Legendre *et al.* 1994). To prevent the use of redundant variables in the MRM, correlations were assessed among soil properties using the VARCLUS procedure in the Hmisc R package. All environmental variables were used in the analysis at local scale as no pair of variables was highly correlated (all Spearman's $\rho 2 \geq 0.50$). At the regional scale, MAT was removed as it was highly correlated with clay (Spearman's $\rho 2 \geq 0.57$). Geographic distance was log-transformed and all environmental variables were normalized. Non-significant variables were removed sequentially from the model until a final model with only statistically significant variables was reached. All analyses here involve distance matrices; as pairwise distances are non-independent significance of statistical tests was based on 999 permutations of the samples in the matrices.

A redundancy analysis (RDA)-based variation partitioning analyses (Peres-Neto *et al.* 2006) was used to assess the relative effects

of environmental variables, logging history (only at the regional scale) and spatial distance on species composition of whole community and abundant bacterial OTUs at each scale. For this analysis, Hellinger-transformed OTU abundance data was used as response variable. Environmental variables at the local scale included pH, total C, total N, CN ratio, available P, extractable K, clay, silt, and sand. At the regional scale, elevation and MAP were also included. Logging history was recoded as a dummy variable and included in the analysis at the regional scale. For spatial variables, eigenfunctions were extracted by a principal components of neighbor matrices (PCNM) analysis (Borcard & Legendre 2002; Borcard *et al.* 2004). The PCNM analysis generated ten positive PCNM eigenfunctions from local scale, and five positive PCNM eigenfunctions from regional scale, which were subsequently used in the RDA analysis.

Before the variation partitioning analysis, the importance of environmental variables and the PCNM eigenfunctions in explaining species composition was determined by a separate RDA analysis using Monte Carlo permutation tests (999 unrestricted permutations). A forward selection (Blanchet *et al.* 2008) method was used to remove the non-significant variables from each of the explanatory sets. These calculations were done using vegan, PCNM and ade4 R packages.

To address spatial autocorrelation, a multiscale ordination (MSO) (Wagner 2004) was applied to the results of the RDA (after Hellinger transformation) on species composition matrix of whole and abundant bacterial OTUs at each scale with environmental variables as constraints. MSO is a geostatistical tool, which partitions the variance of the community composition matrix along increasing spatial lags.

Spatial independence was assessed by means of a Mantel test (999 permutations). These analyses were conducted using the vegan R package. Statistical analysis was performed using R version 2.15.1 (RDevelopmentCoreTeam 2008).

Deposited 454 read accession numbers

Sequences were submitted to the sequence read archive (SRA) at NCBI under the accession number SRA049960.

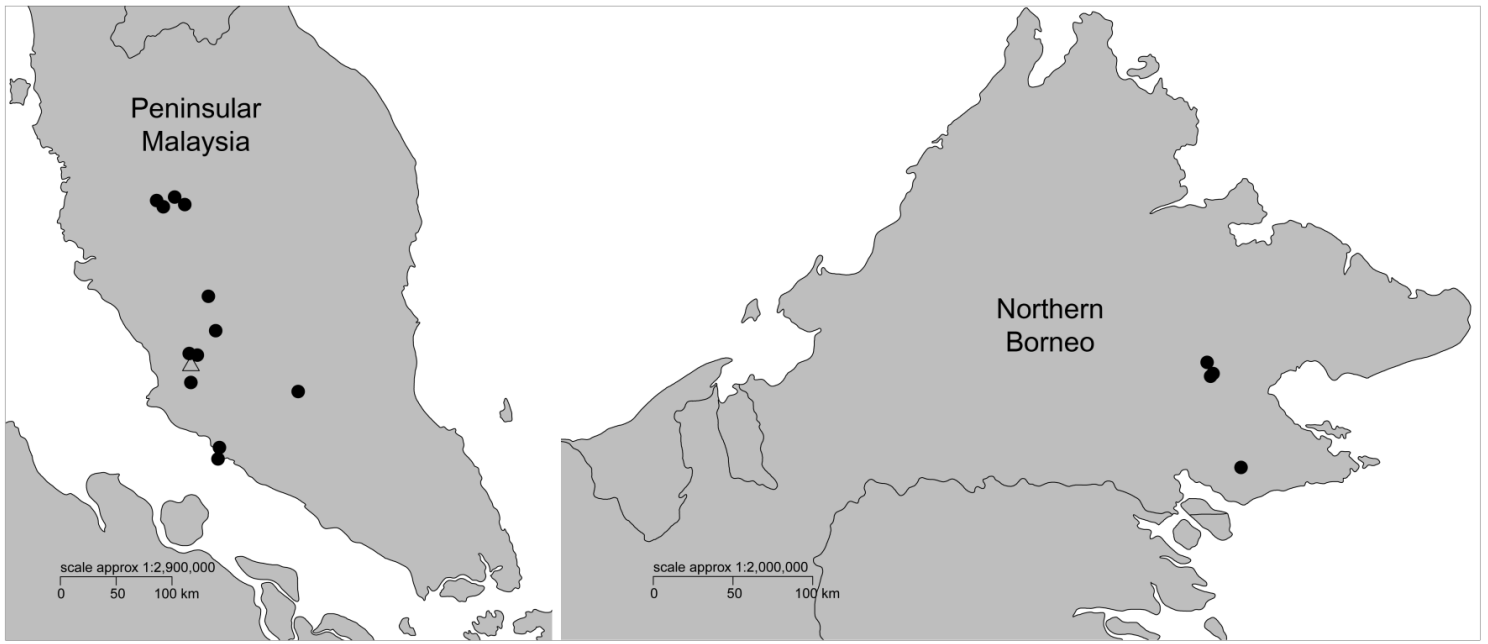

Figure 13. Soil sample locations in Peninsular Malaysia and Northern Borneo. Triangle in Peninsular Malaysia represents the location of the transect sampling point for the local scale. Filled circles across Peninsular Malaysia and Northern Borneo represent regional scale sampling points.

3.1.3. Results

Across all samples, a total of 108,990 quality sequences (average read length=482 bp) were obtained, with coverage ranging from 520 to 9,545 reads per sample. Of the 108,990 good-quality sequences, around 90.3% of sequences could be classified up to phylum level with a total of 11,307 phylotypes (defined at ≥97% sequence similarity level). Abundant OTUs at the local scale (89 OTUs) contributed 36,563 reads (53.7% of total reads at the local scale); at regional scale abundant OTUs (42 OTUs) contributed 14,528 reads (36.8 % of total reads at the regional scale). The most abundant taxa across all soil samples were *Acidobacteria* (35.6%), *Alphaproteobacteria* (23.1%), *Actinobacteria* (6.4%), *Gammaproteobacteria* (5.9%), *Betaproteobacteria* (3.4%), *Bacteroidetes* (2.4%), and *Chloroflexi* (2.3%) (Figure 14). The relative abundance of the most abundant bacterial taxa differed between local and regional scale samples (all $P \leq 0.05$) except for *Acidobacteria*. Soil pH, TC, CN ratio, extractable K, clay and silt percentage were significantly correlated ($P < 0.05$) with the relative abundance of at least one of the most abundant bacterial phyla at the local scale (Table 6); whereas at the regional scale significant correlations ($P < 0.05$) of the relative abundance of at least one of the most abundant phyla were found for soil pH, TC, TN, extractable K, clay, silt, sand, altitude, rainfall, and MAT (Table 6). The relative abundance of *Actinobacteria*, *Alphaproteobacteria*, and *Betaproteobacteria* was significantly correlated with at least one environmental variable at both scales, although not necessarily the same environmental variable (Table 6). *Gammaproteobacteria* only showed significant correlations with

environmental variables at the local scale, while *Bacteroidetes* and *Chloroflexi* showed them only at the regional scale (Table 6). The relative abundance of *Acidobacteria* was not correlated with any environmental variable at both scales (Table 6).

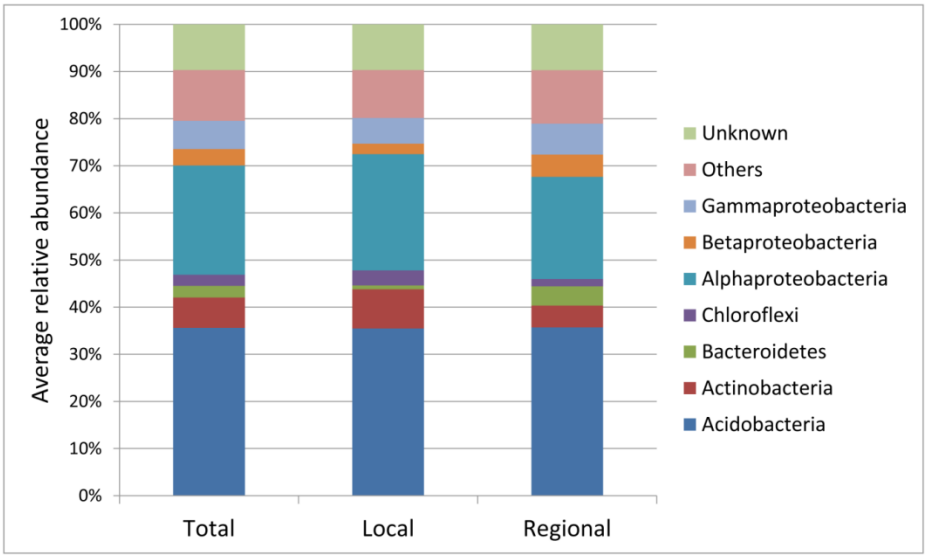

Figure 14. Average relative abundances of dominant bacterial taxa in all samples combined, and in samples collected at local and regional scales.

Table 6. Spearman rank correlations between environmental variables and the relative abundance of dominant bacterial taxa at two spatial scales in Malaysian tropical rainforest.

Environmental variables	*Acidobacteria*	*Actinobacteria*	*Bacteroidetes*	*Chloroflexi*	*Alpha-proteobacteria*	*Beta-proteobacteria*	*Gamma-proteobacteria*
Local scale							
pH	-0.01	-0.48**	0.27	0.11	-0.70**	0.67**	-0.58**
Total carbon (%)	-0.05	0.37	-0.25	0.08	0.40*	-0.40*	0.67**
Total nitrogen (%)	-0.28	-0.09	-0.24	-0.19	-0.14	0.21	0.09
CN ratio	0.07	0.52**	-0.04	0.07	0.48*	-0.50**	0.49**
Available phosphorus	0.07	0.22	0.13	0.16	0.36	-0.46	0.08
Extractable potassium	-0.21	0.09	-0.02	-0.22	0.07	-0.02	0.41*
Clay (%)	0.20	0.24	0.03	0.09	0.32	-0.17	0.40*
Silt (%)	-0.14	0.39*	-0.30	-0.04	0.00	-0.13	0.05
Sand (%)	-0.22	-0.27	-0.05	-0.10	-0.22	0.15	-0.37
Regional scale							
pH	-0.26	-0.36	0.53**	-0.13	-0.66**	0.41*	0.12
Total carbon (%)	0.16	-0.29	0.53**	-0.29	-0.48*	0.22	0.27

Table 6. Continued.

Total nitrogen (%)	0.07	-0.44*	0.53**	-0.47*	-0.28	0.53**	0.02
CN ratio	0.05	0.14	0.07	0.02	-0.33	-0.21	0.23
Available phosphorus	-0.36	0.09	0.06	0.18	-0.03	0.29	-0.24
Extractable potassium	-0.10	-0.48*	0.49*	-0.32	-0.18	0.67**	0.20
Clay (%)	-0.08	0.34	-0.35	0.51**	-0.12	-0.36	-0.17
Silt (%)	0.32	-0.29	0.26	-0.35	-0.36	0.51**	-0.28
Sand (%)	-0.26	-0.04	0.07	-0.12	0.45*	-0.19	0.32
Altitude (m)	-0.13	-0.53**	0.48*	-0.70**	-0.10	0.29	0.06
Rainfall (mm)	-0.23	-0.36	0.23	-0.27	-0.14	0.49*	-0.23
MAT (^{o}C)	-0.05	0.59**	-0.47*	0.69**	0.05	-0.41*	-0.13

*$P < 0.05$ and **$P < 0.01$.

There was a distance–decay curve in bacterial community similarity (i.e., Bray–Curtis) at the local scale (statistics for whole community Figure 15), but no relationship was found at the regional scale (Figure 15). When assessing environmental distance, a strong negative relationship was found between bacterial community similarity and environmental distance at both sampling scales (Local: Mantel R = −0.51, P = 0.0001; Regional: Mantel R = −0.62, P = 0.0001) (Figure 16). The community of the most abundant OTUs showed similar results to those of the whole community at the local and regional scales (Figure 16). There was no significant effect of logging history on bacterial community composition at the regional scale (ANOSIM: R = 0.01, P = 0.37 for whole OTUs, R = −0.01, P = 0.49 for most abundant ones). The NMDS plots of pairwise Bray–Curtis dissimilarities showed that the bacterial community structure among samples at the local and regional scales varied to a similar extent (Figure 17). This was true for the whole bacterial community (Figure 17a), as well as for the most abundant OTUs (Figure 17b).

Controlling for environmental distance, the whole bacterial community dissimilarity was significantly correlated with the geographic distance only at the local scale (ranked partial Mantel test: ρ = 0.13, P = 0.03; Table 7). In contrast, environmental distance was correlated to community dissimilarity at both scales, even when controlling for geographical distance (ranked partial Mantel test: ρ = 0.46, P = 0.001 at the local scale; ρ = 0.55; P = 0.0001 at the regional scale; Table 7). Similar results were observed for the most abundant bacterial OTUs (Table 7). The MRM model explained a large and significant proportion of the variability in bacterial community

structure at the local ($R^2 = 0.51$, $P < 0.0001$) and the regional ($R^2 = 0.66$, $P < 0.0001$) scales (Table 8). Among the environmental variables measured, only soil pH was significantly correlated to whole bacterial community dissimilarity at the local scale, whereas total N and altitude were also significant at the regional scale (Table 8). Similar results were found for the most abundant bacterial OTUs, except that for abundant OTUs community dissimilarity on the local scale was explained by soil pH and total N (Table 8).

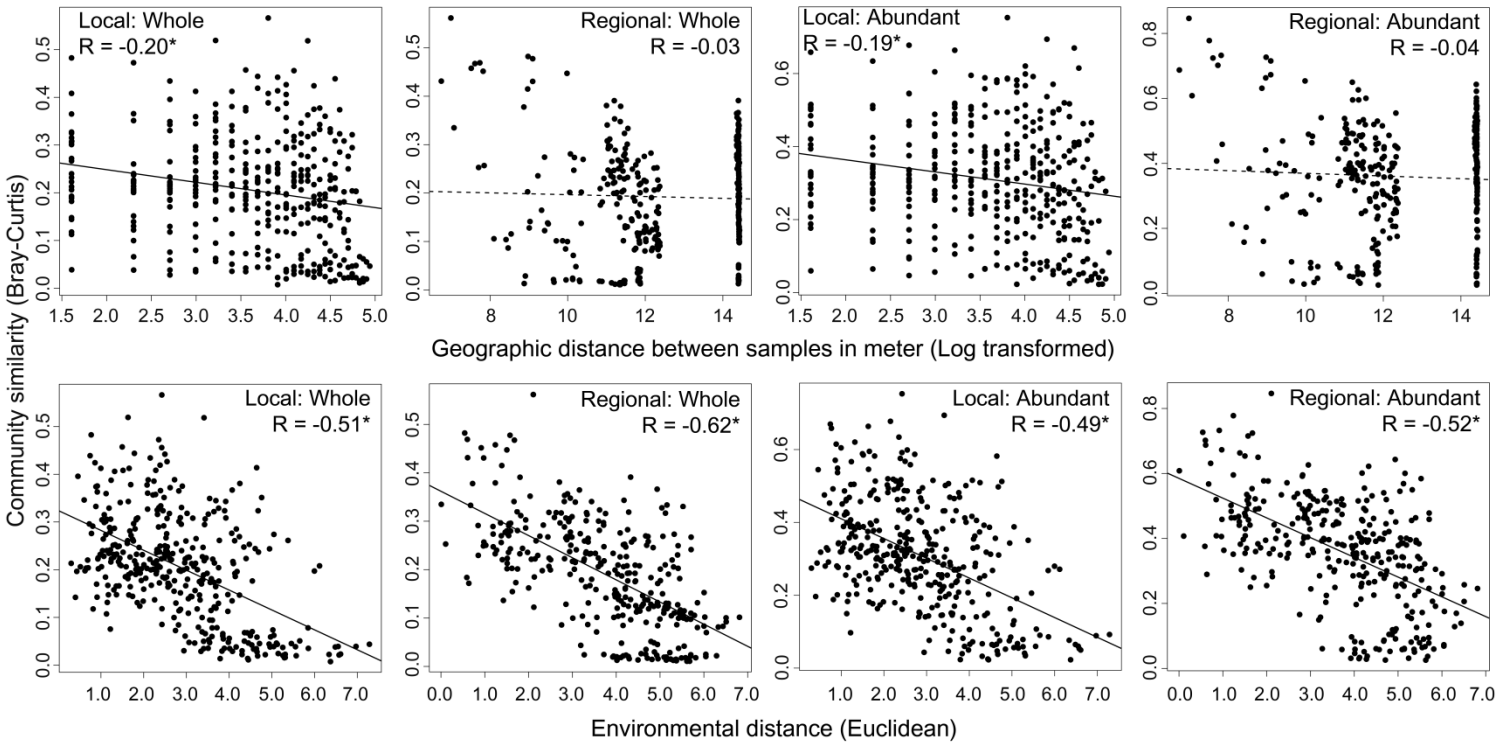

Figure 15. Relationship between Bray-Curtis community similarity and log geographic and Euclidean environmental distance for whole and abundant bacterial community at local and regional scale of sampling.

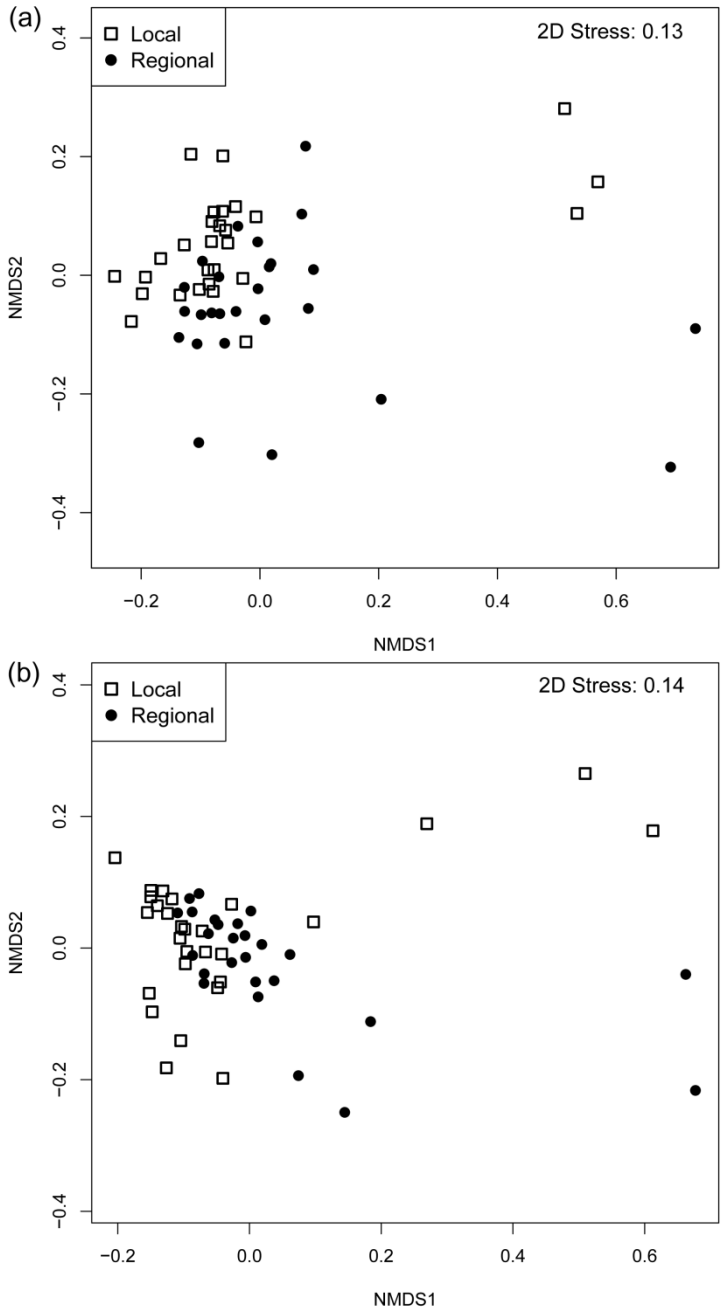

Figure 16. NMDS of Bray-Curtis pairwise dissimilarity of (A) whole bacterial community and (B) abundant bacterial community at the local and regional scales.

Table 7. Ranked partial mantel test results for whole community and abundant OTUs; Spearman ρ between the bacterial community dissimilarity (Bray-Curtis) and either geographic distance or environmental distance.

Correlation between bacterial community similarity and	Controlling for	Whole				Abundant			
		Local		Regional		Local		Regional	
		ρ	*P*	ρ	*P*	ρ	*P*	ρ	*P*
Geographic distance	Environment	0.13	**0.03**	-0.09	0.795	0.12	**0.04**	-0.04	0.613
Environment distance	Geographic distance	0.46	**0.001**	0.55	**0.0001**	0.45	**0.001**	0.48	**0.0001**

Table 8. MRM analysis for whole and abundant bacterial communities at two spatial scales in Malaysian tropical rainforest.

Environmental parameters	Whole		Abundant	
	Local $R^{2\#}$ = 0.51***	Regional $R^{2\#}$ = 0.66***	Local $R^{2\#}$ = 0.44***	Regional $R^{2\#}$ = 0.61***
pH	0.07***	0.07***	0.08***	0.11***
Nitrogen (%)	NS	0.03*	0.02*	0.04*
Altitude (meter)	NA	0.03**	NA	0.03*

*$P \leq 0.05$, **$P \leq 0.001$, and ***$P \leq 0.0001$.

The variation (R^2) of bacterial community dissimilarity (Bray-Curtis) that is explained by the remaining variables and the partial regression coefficients (#) of the final model are reported.

NS = not significant; NA = not applicable.

The RDA and variation partitioning analysis showed that environmental variables significantly explained the largest proportion of variation at both scales (11.67 and 11.95 % for local and regional scales, respectively) for the whole community. Spatial distribution of the samples as given by the PCNM, significantly explained 10.91 and 3.24 % of the variation at the local and regional scale, respectively (Table 9). Accordingly, the shared variation between environment and PCNM was 7.45 % at the local and 0.07 % at the regional scale (Table 9). However, a large proportion of the variation remained unexplained at both scales (Table 9). Similar results were found for abundant bacterial OTUs (Table 9).

The variogram from the MSO (Figure 17) showed that the total inertia (i.e., variance) increased with distance at the local scale, and that there is significant spatial autocorrelation at four distance classes (10, 15, 40, and 140 m). At the regional scale, the variance of data showed no clear pattern with distance (Figure 17), and no significant autocorrelation was found in any distance class. Similar results were found at both scales with abundant bacterial OTUs (Figure 17).

Table 9. Percentage of variation in soil bacterial community explained by environmental variables and principle coordinates neighbor matrices (PCNM) of the whole bacterial community and the community of the most abundant OTUs at two spatial scales.

Variation	Whole				Abundant			
	Local		Regional		Local		Regional	
	Variation explained (%)	*P*	Variation explained (%)	*P*	Variation explained (%)	*P*	Variation explained (%)	*P*
[E + S]	15.13	0.005	14.49	0.005	28.01	0.005	31.13	0.005
[E]	11.67	0.005	11.95	0.005	21.63	0.005	26.54	0.005
[S]	10.91	0.005	3.24	0.005	18.18	0.005	7.55	0.005
[E/S]	4.21	0.005	11.24	0.005	9.83	0.005	23.58	0.005
[S/E]	3.46	0.005	2.54	0.01	6.38	0.005	4.59	0.01
[E ⋂ S]	7.45		0.07		11.79		2.96	
1 - [E + S]	84.86		85.50		71.98		68.86	

[E + S]: Amount of variation explained by environmental variables and PCNM together; [E]: variation explained by environmental variables; [S]: variation explained by PCNM; [E/S]: the pure environmental fraction controlling by spatial factors; [S/E]: the pure PCNM fraction controlling by environmental variables; [E ⋂ S]: the PCNM-environmental covariation, i.e. [E] - [E/S] = [S] - [S/E]; 1 - [E ⋂ S]: the amount of unexplained variation.

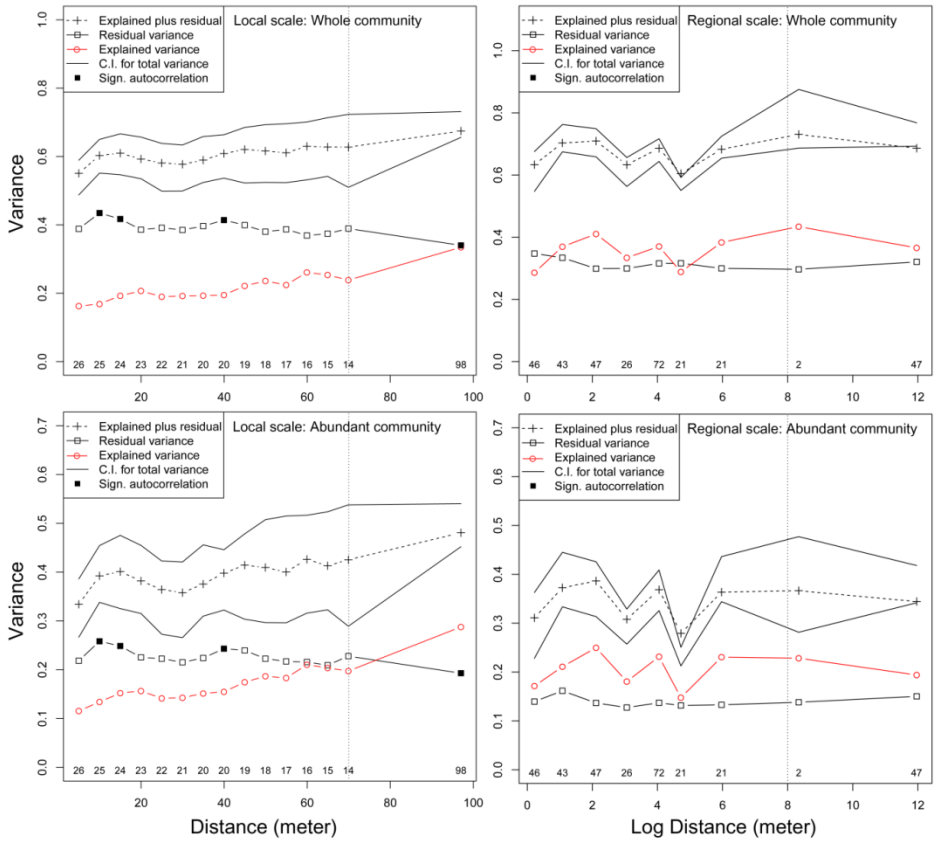

Figure 17. Variogram plot of the multi-scale ordination (MSO) of redundancy analysis (RDA) on species composition matrix for whole and abundant bacterial community at local and regional scale with the environmental variables as constraints. The numbers above the X-axis indicate the number of unique pairs of observations within each distance class.

3.1.4. Discussion

The major goal of this study was to better understand the degree of spatial heterogeneity of the soil bacterial community and its association with distance and environmental conditions in tropical rainforests of Malaysia at local and regional scales. The results show that community similarity decreases with spatial distance at a local scale of 150 m, but not at a larger regional scale. This is true both for the whole bacterial community and for the community of the most abundant OTUs. These results are in agreement with previous studies in other types of ecosystem, where an effect of geographic distance on bacterial community similarity has been found at local scales but not at larger spatial scales (Martiny *et al.* 2011; Pasternak *et al.* 2013). This distance-decay pattern at the local scale could be due to unmeasured spatially auto-correlated environmental variables. Indeed, a large amount of bacterial community variation remains unexplained after the analyses of environmental and spatial interactions. The pattern found here for soil bacteria in the tropical rainforest appears fundamentally different from patterns found for larger organisms, such as plants, where distance effects are strong. Spatial studies have generally shown that in both neotropical rainforests (Condit *et al.* 2002) and North American temperate forests (Gilbert & Lechowicz 2004), plant community similarity decreases with distance, and that most of this decrease cannot be explained in terms of identifiable environmental factors. The distance-decay in community similarity for plants thus appears to leave greater room for an element of dispersal lag and 'neutrality' in broader scale patterns in these larger organisms than in

bacteria. Larger spatial scale sampling over hundreds of kilometers of both tropical and non-tropical communities also clearly shows such turnover and decreasing similarity with distance, in both plants and vertebrates (Koleff *et al.* 2003; Qian & Ricklefs 2007). These results provide no indication that dispersal limitation plays a major role in the biogeography of bacteria in these tropical soils, but it seems that environmental variables have a substantial effect.

Bacterial communities in these samples varied by about the same amount from one another (Figure 16) whether separated by a few kilometers or thousands of kilometers (at the regional scale), and whether separated by a few meters or more than a hundred meters (at the local scale). Environmental variables explained a larger fraction of variation than geographical distance at both scales in the whole-community and the most abundant OTUs. Soil pH, total N and elevation were significant drivers for the bacterial community at the regional scale, as was soil pH alone at the local scale. These results are in broad agreement with studies that report the role of soil pH (Lauber *et al.* 2009; Tripathi *et al.* 2012), total N (Campbell *et al.* 2010; Wessén *et al.* 2010) and elevation (Singh *et al.* 2012b) in explaining variation in bacterial community composition.

Abundance and community composition of most of the dominant bacterial phyla (*Actinobacteria*, *Alphaproteobacteria*, and *Betaproteobacteria*) at both scales were significantly correlated to environmental variables. However, some of the phyla (*Gammaproteobacteria*, *Bacteroidetes*, *Chloroflexi*) showed a response at one scale and not at the other scale. Surprisingly, *Acidobacteria* which was the most abundant bacterial phylum in the collected samples,

did not show a significant correlation to any of the environmental variables at any scale, even though it has been found to be strongly correlated to soil pH (Jones *et al.* 2009; Lauber *et al.* 2009; Tripathi *et al.* 2012). The consistently low pH from the acidic soil of tropical rainforest can explain why *Acidobacteria* are the most abundant phylum (Tripathi *et al.* 2012), and it is possible that the range in pH among the samples was not variable enough to affect *Acidobacterial* relative abundance. The overall phylum-level results indicate that although bacterial community composition was clearly influenced by the environmental variables included, there were no predictable relationships among the variables measured and the identity of the most abundant phyla at both scales.

The proportion of variation explained by environment and/ or distance was higher for the most abundant bacterial OTUs, in comparison to whole bacterial community at both scales (Table 9). This may be because whole community samples include many rare OTUs whose true abundance and occurrence cannot be completely sampled, increasing the proportion of variation that remains unexplained (Pedrós-Alió 2006). Nevertheless, it is plausible that there are in fact deterministic environmental factors that could explain most of this unexplained community variation, but they have just not been discovered in this study. A significant spatial autocorrelation in the community shown by MSO at the local scale suggests that there are factors at certain distance classes that structure the community. These factors could be biological interactions and colonization processes, or other abiotic factors not measured. In contrast, at the regional scale, the residual variance did not show a spatial autocorrelation, suggesting that

species–environment correlations do not depend on geographical distance at this larger scale.

Another possibility is that there is simply a strong stochastic element in soil bacterial communities that reflects local scale population crashes and recolonizations. The best approximation might be a 'lottery model' (Chesson & Warner 1981), with localized unpredictable bursts of resources, where winners and losers are recruited by luck rather than competitive ability or even dispersal ability. A lottery model would explain community composition and variation in terms of very local scale dispersal lags, over maybe a few meters or a few centimeters (or less). By contrast it appears that on the broad spatial scale, dispersal limitation is much reduced or even nonexistent for bacteria, even on large spatial scales covering thousands of kilometers. This general pattern of spatial uniformity of the range in community composition goes against the predictions of the neutral theory of coexistence (Hubbell 2001), which emphasizes the role of lags and limitations of dispersal. It also appears to support the long-held belief in microbial ecology that 'everything is everywhere, but, the environment selects' (Baas-Becking 1934).

Tree species composition in tropical forests has been found to vary with distance (Tuomisto *et al.* 2003), and there is evidence that the root surfaces of tropical forest trees have their own characteristic host-tree-specific associations of bacteria (Oh *et al.* 2012). Thus it is unclear why bacterial communities would not produce a strong distance-decay pattern if the association with the trees were an important factor in their formation. It may be that the community of bacteria in the bulk soil between roots is relatively unaffected by host tree species identity, in

contrast to the community on the rhizoplane itself which Oh et al. (2012) studied. It is also necessary to bear in mind that within that study, only a few tropical tree species were actually sampled. If many more tropical tree species had been included, Oh et al. (2012) might hypothetically have found more strongly overlapping communities occurring again and again amongst the range of tree species present in the forest. If this were the case, a large enough sample of different tree species may be more or less interchangeable in terms of their root bacterial communities, even if the tree species themselves are somewhat different from one locality to another. However, even if overlap in root communities is extensive, it remains possible that it is not complete, and that part of the distance decay pattern which is seen is actually due to the distance decay in the plant community, affecting the bacterial communities associated with roots.

It is not known whether including a fundamentally different geographical region of rainforest (e.g., neotropical or African rainforest) in the comparison would result in greater dissimilarity in bacterial species composition. It may be hypothesized that the longtime separation of these regions (Morley 2000), and opportunities for separate evolution between regions, could result in greater distinctiveness than that seen within Southeast Asia. Martiny et al. (Martiny *et al.* 2011) addressed this question in marshes and found that community similarity of *Betaproteobacteria* is negatively related to distance within marshes and between marshes on the same continent, but no effect of distance was found in marshes from different continents. They suggest that a distance decay curve at a local scale may be possible if geographic distance is correlated with some

unmeasured environmental variables only at this scale.

In conclusion, this study suggests that in tropical rainforest soils at both local and regional scales, environmental variables have a strong effect on the bacterial community composition. Only weak evidence was found of an effect of distance alone, and thus only weak evidence of what might be construed as dispersal lag. Even the distance effect which was found could be a result of environmental factors not accounted for in this study including the distance decay in the plant community that acts as a host for root-associated bacteria. There is a need for experimental studies to test the relevant hypotheses (e.g., the role of dispersal lag) that may apply to soil bacteria in this system and or other studies which compare the empirical spatial diversity and turnover pattern found in this study with other parts of the world. It has been predicted (Koleff *et al.* 2003; Qian & Ricklefs 2007) that diversity turnover will be greater in more diverse lower latitude environments such as tropical forests: it would indeed be interesting to know whether for soil bacteria the amount of between-sample heterogeneity, or other aspects of the spatial pattern, is different in forests from different latitudes.

CHAPTER 4. Biome Effect on Community Composition and Diversity of Soil Archaea and Ammonia Oxidizing Archaea

4.1. Both climate and pH are key in determining soil archaeal community structure and diversity

4.1.1. Introduction

Archaea constitute a small but consistent portion of soil microbial communities around the world (Bates *et al.* 2011; Bengtson *et al.* 2012; Cao *et al.* 2012; Tripathi *et al.* 2013). Both culture-dependent and culture-independent studies from terrestrial and marine environments have revealed that many archaea are capable of ammonia oxidation (Konneke *et al.* 2005; Treusch *et al.* 2005) and thus potentially represent important players in the global nitrogen cycle, in making nitrogen more available to photosyntheizers. Recently, ribosom al gene sequence analysis from uncultivated organisms, and the iso lation of certain key reference organisms, have both expanded our knowledge and led to the proposal of various new candidate phyla within the Archaea. The most prominent addition is the candidate phylum *Thaumarchaeota* (formerly described as mesophilic *Crenarchaeota*) (Brochier-Armanet *et al*. 2008). *Thaumarchaeota* dominates archaeal communities in most soils, and contains all the currently known ammonia-oxidizing archaea (AOA) (Stahl & de la Torre 2012).

Although archaea can be found in all or most soils, the community ecology of terrestrial archaea remains poorly understood. Some global scale surveys, together with other local and regional scale studies, have shown that archaeal commuities are mainly structured by environmental factors, such as salinity (Auguet *et al.* 2010), pH (Bengtson *et al.* 2012; Cao *et al.* 2012; Nicol *et al.* 2008; Tripathi *et al.*

2013), elevation (Singh *et al.* 2012a), climate and vegetation cover (Angel *et al.* 2010), and C/N ratio (Bates *et al.* 2011). Similarly, several studies around the world have shown that environmental factors are the key drivers of AOA community composition and diversity (Gubry-Rangin *et al.* 2011; Pester *et al.* 2012). However, these studies typically analyzed total archaea or the AOA community using 16S rRNA and *amoA* gene, respectively, except for a large geographical scale survey of terrestrial Archaea (Hu *et al.* 2013). Hu et al. (Hu *et al.* 2013) studied the *Thaumarchaeota* and AOA community from a wide range of soils and ecosystem types in China, though largely from one biome, the temperate zone.

This study was intended to explore parallels and differences in the broad-scale biodiversity patterns of total archaea and AOA both within and between biomes. As one of the more wide-ranging archaeal diversity studies performed so far, this work addresses the following questions:

(i) What are the dominant archaeal and AOA taxa in tropical and temperate soils? Are they the same, or distinct?

(ii) Within each biome, what are the main environmental variables structuring patterns in the soil archaeal community? Do key soil archaea and AOA inhabit a particular niche, or set of niches, which are predictable, being based on soil edaphic or site characteristics?

(iii) Are the community patterns and levels of diversity in soil archaea/AOA the same in tropical and temperate environments?

The tropical biome soils were sampled in Malaysia, whereas

for temperate biome samples were collected in Korea and Japan, thereby providing a contrast between the almost aseasonal equatorial tropical zone, and a temperate monsoonal climate with cold winters with frost and snow. Combined taxonomic and phylogenetic based methods were used to explore the total archaea community using 16S rRNA gene, and AOA using *amoA* genes.

4.1.2. Materials and methods

Sample collection

Tropical biome

A total of 24 surface soil samples were collected from primary and selectively logged native tropical forest (11 samples) and non-forest (13 samples) sites including plantation crops, agricultural and pasture fields and pasture lands, across central and southern Malay Peninsula and Northern Borneo. These formed the tropical biome part of the study. Neutral and alkaline forest soils are uncommon in the tropical climate of Malaysia, due to soil weathering and leaching, and because the more fertile higher pH forest soils have almost all been cleared for agriculture or are inaccessible on steep cliffs. Thus to increase sample size of higher pH environments the relatively few neutral-alkaline forest soils were combined with a set of agricultural soil samples from across the pH range.

Temperate biome

For the temperate biome, samples were collected (24 samples from forest sites and one sample from a pasture land) across Korea and

Japan across as broad a pH range (see Table A3 for additional information on each of the collected soils). At each sampling site, the top 5 cm of mineral (A horizon) soil was collected from five selected locations within an area of 1 ha. Soil samples were composited, stored, and shipped at 4℃ for 12 h before they were sieved through a 4-mm mesh to thoroughly homogenize and remove roots and plant detritus from the samples. Soils were thoroughly homogenized by passing them through a 2-mm mesh, and sieved soil were frozen at −20°C for further DNA extraction and chemical analyses.

Soil chemical variables analyses

For Korean samples, soil pH, total carbon, total nitrogen, ammonium and nitrate were measured at National Instrumentation Center for Environmental Management (NICEM, South Korea) based on the standard protocol of the Soil Science Society of America (SSSA).The Japanese and Malaysian soil samples were processed at Field Science Education and Research Center, Kyoto University and Soil Analysis Laboratory of University Putra Malaysia, respectively, using the same SSSA methodology as was used on the South Korean samples.

DNA extraction, PCR and pyrosequencing of 16S rRNA and *amoA* gene

Total DNA was extracted from 0.25g of thoroughly-homogenized soil using a PowerSoil DNA extraction kit (MO BIO Laboratories, Carlsbad, CA, USA) following manufacturer's instructions. For PCR amplification, bar-coded primers were used

targeting the V1 to V3 region of the archaeal 16S rRNA gene, with primers and reaction conditions, as described earlier by Hur et al. (Hur *et al.* 2011). PCR amplification of the archaeal *amoA* gene was performed using the primers C*amoA*-19f /C*amoA*-616r, as described earlier (Pester *et al.* 2012). The amplified products were purified using the QIAquick PCR purification kit (Qiagen, CA, USA). Amplicon pyrosequencing was performed by Macrogen Incorporation (Seoul, Korea) using 454/Roche GS-FLX Titanium Instrument (Roche, NJ, USA).

Processing of pyrosequencing data and taxonomic analysis

The sequence data obtained after pyrosequencing were processed using mothur (Schloss *et al.* 2009). Sequences shorter than 200 nt with homopolymers longer than 8 nt and all reads containing ambiguous base calls or incorrect primer sequences were removed. Thereafter, *amoA* sequences were screened for frame shifts (most likely caused by 454 sequencing errors) using MetaGeneTack (Tang *et al.* 2013). MetaGeneTack identifies the region of the frame shift but not the exact deleted or inserted base, deletions and neighbouring bases were masked by 'N's and insertions were removed and deletions were marked by an N. Next, the archaeal 16S rRNA and *amoA* gene sequences were aligned against the EzTaxon-aligned and archaeal *amoA* aligned databases, respectively. The *amoA* reference sequences were obtained from the ARB database created by Pester *et al*. (2012). Putative chimeric sequences were detected and removed via the Chimera Uchime algorithm contained within mothur (Edgar *et al.* 2011).

All the16S rRNA and *amoA* gene sequences were classified against EzTaxon-extended (Kim *et al.* 2012b) and *amoA* reference databases (Pester *et al.* 2012), respectively, using the naïve Bayesian classifier implemented in mothur (at ≥ 80% bootstrap cutoff with 1000 iterations). All the 16S rRNA and *amoA* sequence data are available under the NCBI SRA accession numbers SRA069952.

Statistical processing and analysis of results

To correct for differences in the number of reads, which can bias diversity estimates, a subset was randomly selected of 520 and 1126 sequences of 16S rRNA and *amoA* genes, respectively, from each sample of the tropical and temperate biomes. Operational taxonomic units (OTUs) of the 16S rRNA and *amoA* genes were defined as sequence groups in which sequences differed by 3 and 15%, respectively. For cultured AOA in which both genes can be compared in the same organism, researchers have found that a 3% 16S rRNA distance equates to a 15% *amoA* gene distance (Pester *et al.* 2012). Phylogenetic diversity (PD) was calculated as Faith's PD (Faith 1992) using a maximum-likelihood (ML) tree inferred from partial 16S rRNA and *amoA* gene sequences by FastTree2 with default settings for phylogenetic analyses (Price *et al.* 2010). To determine if the relative abundance of the most abundant archaeal taxa and diversity indices differ between tropical and temperate biomes a t-test was used for normal data and Wilcoxon rank–sum test for nonnormal data. A permutational multivariate analysis of variance (PERMANOVA) was used on dissimilarity matrices (Anderson 2001) with 999 permutations to test if community composition differs according to biomes.

A redundancy analysis (RDA)-based variation partitioning analyses (Peres-Neto *et al.* 2006) was performed to assess the relative effects of environmental, climate and spatial variables on species composition of both 16S rRNA and *amoA* genes OTUs. For this analysis, Hellinger-transformed OTU abundance data was used as response variable. Environmental variables included were soil pH, total carbon, total nitrogen, CN ratio, nitrate, ammonium and altitude. Climate variables included in the analysis were mean annual temperature (MAT) and mean annual precipitation (MAP). For spatial variables, principal coordinates of neighbor matrices (PCNM) was used to model the spatial structure between samples. Before the variation partitioning analysis, the importance of environmental, climate and spatial variables in explaining species composition was determined by a separate RDA analysis using Monte Carlo permutation tests (999 unrestricted permutations). Then a forward selection (Blanchet *et al.* 2008) method was used to remove the non-significant variables from each of the explanatory sets. These calculations were done using vegan, PCNM and ade4 R packages.

Mantel correlograms were used to evaluate phylogenetic signal in 16S rRNA and *amoA* genes across a range of phylogenetic depths and significance was drawn from 999 randomizations (Diniz-Filho *et al.* 2010; Oden & Sokal 1986). Phylogenetic distances were partitioned into classes (0.02 units) and the correlation coefficient was relating between OTU phylogenetic distances to environmental-optimum distances (Diniz-Filho *et al.* 2010). An environmental-optimum for each OTU was calculated for each environmental variable as in Stegen et al. (Stegen *et al.* 2012). Between-OTU environmental optimum

differences were calculated as Euclidean distances using optima for all the environmental variables.

Phylogenetic turnover in total archaeal and AOA community composition was quantified using unweighted Unifrac (Lozupone & Knight 2005) and the mean nearest taxon distance (betaMNTD) (Fine & Kembel 2011; Stegen *et al.* 2012). BetaMNTD is the mean phylogenetic distance to the closest relative in a paired community for all taxa (Fine & Kembel 2011) and is sensitive to the changes of lineages close to the phylogenetic tips. PERMANOVA was used to test the effect of biomes on the distribution of total archaea and AOA community (Anderson 2001).

For each biome, a standardized effect size (ses.betaMNTD) was computed as the number of standard deviations that observed betaMNTD departed from the mean of null distribution (999 null iterations) based on random shuffling of OTU labels across the tips of the phylogeny (Fine & Kembel 2011; Stegen *et al.* 2012; Wang *et al.* 2013). This randomization holds constant observed species richness, species occupancy and species turnover. Therefore it provides an expected level of betaMNTD given observed species richness, occupancy and turnover. The absolute magnitude of ses.betaMNTD reflects the influence of deterministic processes; the larger the magnitude, the greater the influence of deterministic, niche-based processes.

To assess the relationship between phylogenetic dissimilarity of the total archaea and AOA community with spatial distance and environmental distance (Euclidean) in tropics and temperate biomes, phylobetadiversity was regressed against spatial and environmental

distances using a Gaussian generalized linear model and significance was determined using Mantel tests with 999 permutations. An analysis of covariance was used with 999 permutations to test the hypothesis that the regression slopes do not differ among biomes. To further evaluate the relative importance of spatial distance and of each of the environmental variables on total archaea and AOA community phylogenetic dissimilarity, a multiple regression on matrices (MRM) approach was used (Legendre *et al.* 1994). All statistical analysis, graphs and ordinations were produced using R version 3.0.2 (RDevelopmentCoreTeam 2008).

4.1.3. Results

After the random sampling of 684 sequences from each group, a total of 32,832 high-quality archaeal 16S rRNA gene sequences were obtained. Using the 3% cut-off, the 32,832 archaeal 16S rRNA gene sequences were assigned to 1,070 OTUs. The number of OTUs for each sample ranged from 18 to 127. The representative sequences of these 1,070 OTUs were grouped into 3 different phyla (Figure 18a). Among 1,070 archaeal 16S rRNA gene OTUs, 958 OTUs were affiliated with the phylum *Thaumarchaeota*, which accounted for 89.5% of the obtained OTUs (Figure 18a). The remaining 112 OTUs belonged to the *Euryarchaeota* and *Crenarchaeota* (Figure 18a). For, AOA 1,447 sequences were sampled randomly from each group, which resulted a total of 70,903 high-quality AOA *amoA* gene sequences. Using the 15% cut-off that was recommended in a previous study (Pester *et al.* 2012), the 70,903 AOA *amoA* gene sequences were assigned to 245 OTUs. The number of OTUs for each sample ranged from 7 to 43. According

to the previously reported classification of archaeal *amoA* genes (Pester *et al.* 2012), the representative sequences of the 245 OTUs were grouped into four different clusters (Figure 18b). The *Nitrosopumilus* cluster contained 8 OTUs, and the *Nitrosotalea* cluster contained 58 OTUs (Figure 18b). The remaining 167 and 12 OTUs belonged to the *Nitrososphaera* and *Nitrososphaera sister* cluster (Figure 18ab), respectively. *Thaumarchaeota* was the most abundant archaeal phylum (98.0% of all archaeal sequences) followed by *Euryarchaeota* (1.8% of all archaeal sequences), the only other phylum present across all the samples. The majority of the 16S rRNA archaeal phylotypes (48.7% of all archaeal sequences) were confined to *Thaumarchaeota* group 1.1b followed by group1.1c and group 1.1a (24.7% and 24.6% of all archaeal sequences, respectively) (Figure 19a). Out of these, group 1.1a was significantly more abundant in the temperate soils than the tropical ones (W = 41, $P = 0.03$). *Nitrososphaera* was the most abundant cluster (69.8% of all *amoA* gene sequences). The remaining sequences belonged to the *Nitrosotalea*, *Nitrososphaera* sistercluster and *Nitrosopumilus* cluster (26.5, 3.0, and 0.4% of all *amoA* gene sequences) (Figure 19b). The *Nitrososphaera* cluster was more abundant in tropical soils (W=155, $P < 0.01$), whereas the *Nitrosotalea* cluster dominated in temperate soils (W = 439, P = 0.01).

The composition of the soil archaeal communities based on 16S rRNA and *amoA* gene sequences was significantly influenced by biome (Figure 20; PERMANOVA, $P < 0.001$, 999 permutations). Using the forward selection RDA analysis of 16S rRNA and *amoA* genes, the distribution of total archaeal and AOA communities in different soil

samples and their relationships with environmental, climate and spatial factors were analyzed (Figure 20). The total archaeal and AOA communities were mainly distributed along the soil pH gradient. Latitude was the second most important variable, followed by mean annual precipitation, total carbon and nitrate (nitrate only found significant for total archaeal community). Further variance partitioning analysis revealed that edaphic variables (soil pH, total carbon and NO_3^-) significantly explained 31.8% of the total variation (P = 0.005) of archaeal communities, of which 24.2% of the total variation (P = 0.005) was neither spatially structured nor related to climate (Figure 21a). The unique components of the climate and spatial variables explained less variation - about 3.2 and 1.5% respectively, however the percentage of variation explained was significant (all P values were < 0.05) (Figure 21a). Likewise for the total archaeal community of AOA, edaphic variables (soil pH and total carbon) explained the largest proportion (54.5% of the variation; P = 0.005) of the total variation (Figure 21b). Climate variables explained about 2.8% of the total variation (P = 0.01), however unlike 16S-based community, the unique component spatial variables was not significant in explaining any of the variation observed in community (P = 0.11) (Figure 21b), but the combined effect of climate and spatial variables explained 21.8% of the variation (P = 0.005).

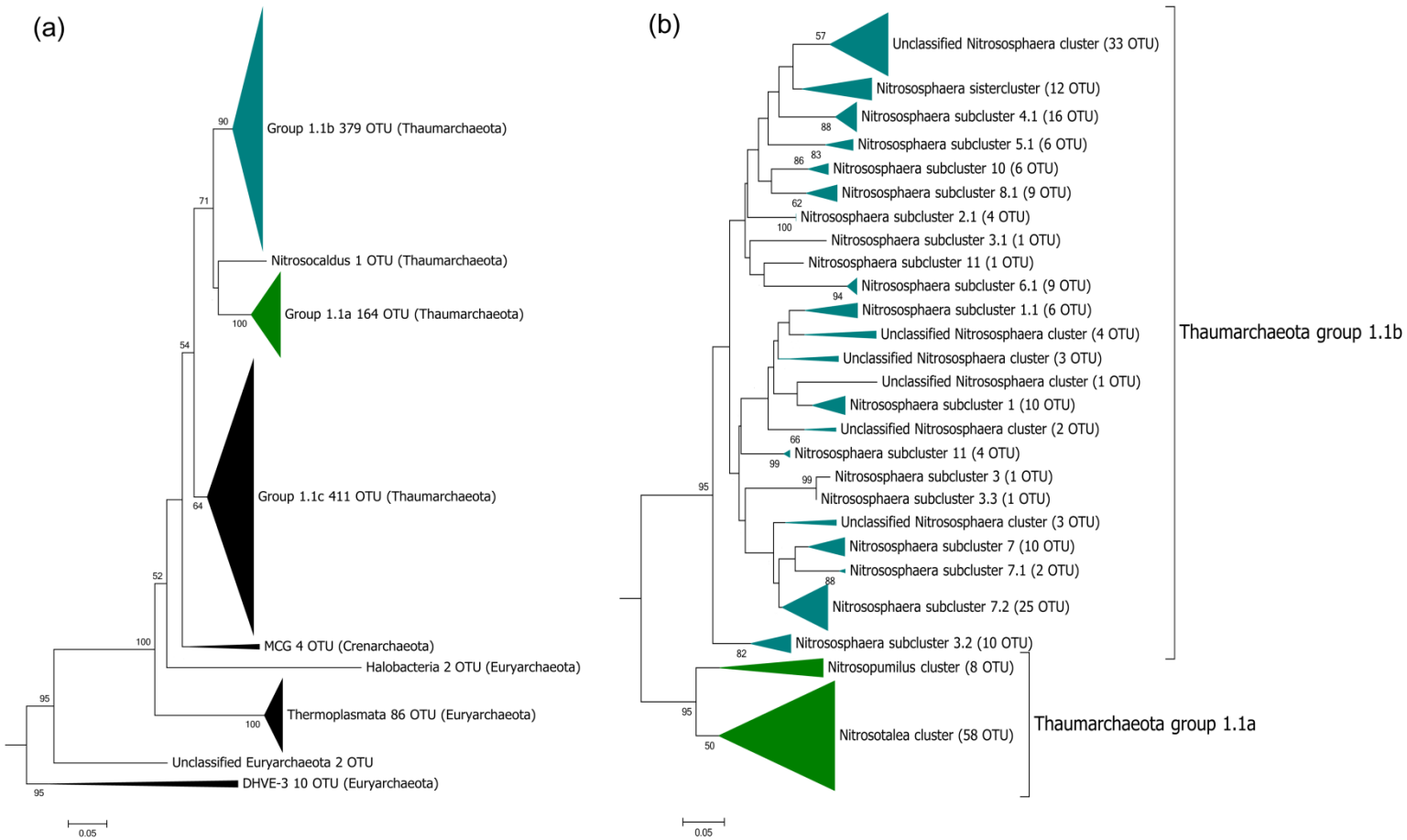
(a)
Group 1.1b 379 OTU (Thaumarchaeota)
Nitrosocaldus 1 OTU (Thaumarchaeota)
Group 1.1a 164 OTU (Thaumarchaeota)
Group 1.1c 411 OTU (Thaumarchaeota)
MCG 4 OTU (Crenarchaeota)
Halobacteria 2 OTU (Euryarchaeota)
Thermoplasmata 86 OTU (Euryarchaeota)
Unclassified Euryarchaeota 2 OTU
DHVE-3 10 OTU (Euryarchaeota)
0.05
(b)
Unclassified Nitrososphaera cluster (33 OTU)
Nitrososphaera sistercluster (12 OTU)
Nitrososphaera subcluster 4.1 (16 OTU)
Nitrososphaera subcluster 5.1 (6 OTU)
Nitrososphaera subcluster 10 (6 OTU)
Nitrososphaera subcluster 8.1 (9 OTU)
Nitrososphaera subcluster 2.1 (4 OTU)
Nitrososphaera subcluster 3.1 (1 OTU)
Nitrososphaera subcluster 11 (1 OTU)
Nitrososphaera subcluster 6.1 (9 OTU)
Nitrososphaera subcluster 1.1 (6 OTU)
Unclassified Nitrososphaera cluster (4 OTU)
Unclassified Nitrososphaera cluster (3 OTU)
Unclassified Nitrososphaera cluster (1 OTU)
Nitrososphaera subcluster 1 (10 OTU)
Unclassified Nitrososphaera cluster (2 OTU)
Nitrososphaera subcluster 11 (4 OTU)
Nitrososphaera subcluster 3 (1 OTU)
Nitrososphaera subcluster 3.3 (1 OTU)
Unclassified Nitrososphaera cluster (3 OTU)
Nitrososphaera subcluster 7 (10 OTU)
Nitrososphaera subcluster 7.1 (2 OTU)
Nitrososphaera subcluster 7.2 (25 OTU)
Nitrososphaera subcluster 3.2 (10 OTU)
Nitrosopumilus cluster (8 OTU)
Nitrosotalea cluster (58 OTU)
Thaumarchaeota group 1.1b
Thaumarchaeota group 1.1a
0.05

Figure 18. Neighbor-joining phylogenetic tree showing the phylogenetic affiliations of the representative sequences of each 16S rRNA gene (a) and *amoA* gene (b) recovered from tropical and temperate soils. The numbers at the nodes are percentages that indicate the levels of bootstrap support from 250 replicates. The scale bar represents 0.05 nucleic acid substitutions per nucleotide position.

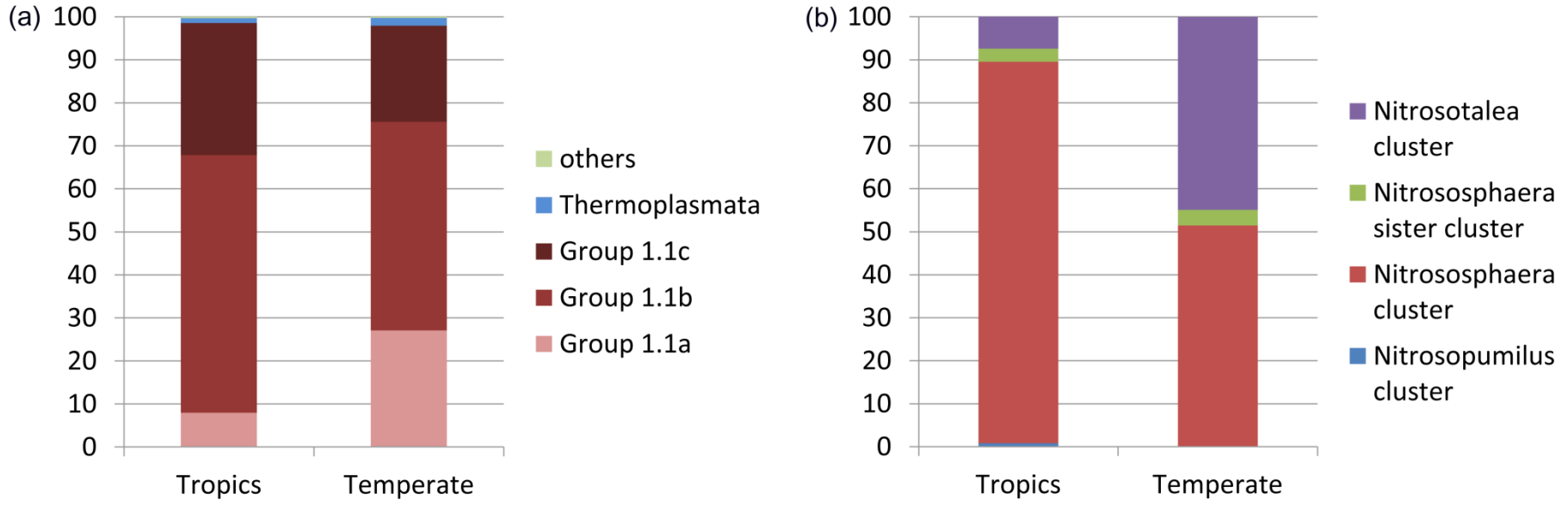

Figure 19. Percentage relative abundances of (a) total archaeal taxa based on 16S rRNA gene sequences and (b) AOA clusters based on *amoA* gene sequences.

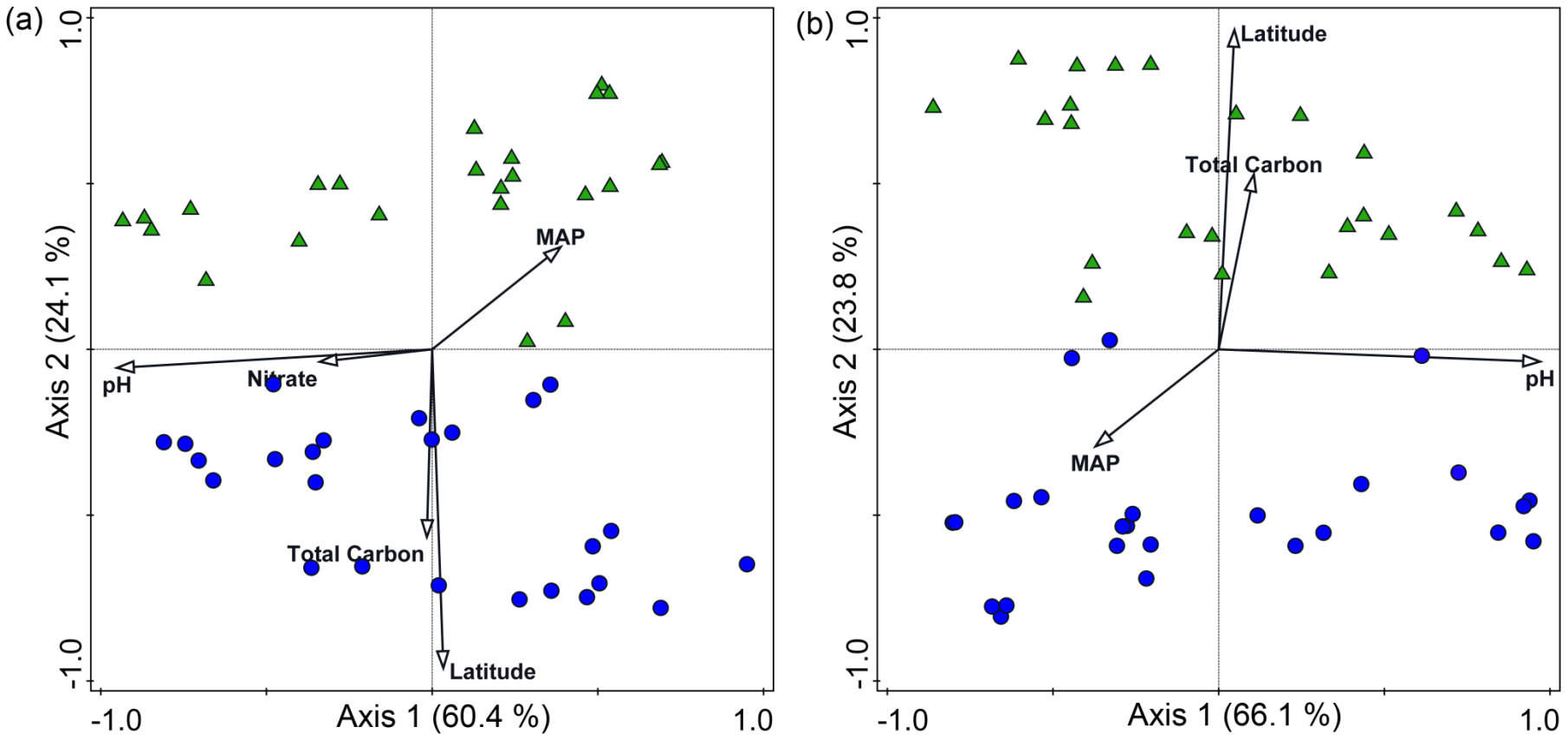

Figure 20. Redundancy analysis plot of (a) total archaeal community structure based on 16S rRNA gene OTUs (97% sequence identity) and (b) AOA community based on *amoA* gene OTUs (85% sequence identity) between tropics and temperate biomes and a vector overlay of the significantly correlated variables. (MAP = mean annual precipitation).

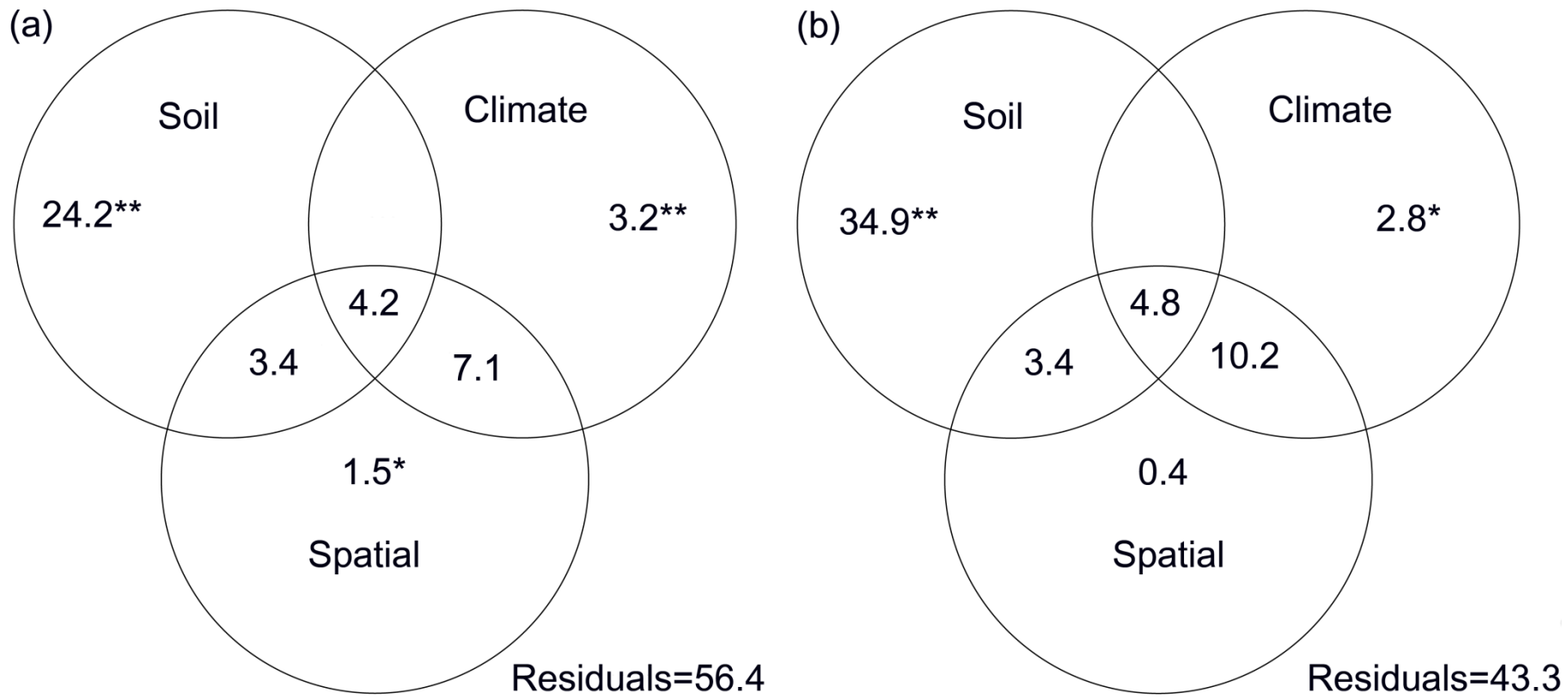

Figure 21. Percentage of variation explained by edaphic, climate and spatial variables (geographical distance) of (a) total archaeal community based on 16S rRNA gene sequences and (b) AOA community based on *amoA* gene sequences. Unique fractions of each explanatory set were evaluated by 999 permutations. Values < 0 are not shown; **P* = 0.01; ***P* = 0.005.

Archaeal 16S rRNA gene OTU richness (i.e., number of OTUs) and Shannon index did not differ between biomes (OTU richness, W = 36, P = 0.12; Shannon index, t = -0.80, P = 0.43) (Figure A3). However, *amoA* gene OTU richness and Shannon index were significantly higher in the temperate biome than in the tropical biome (OTU richness, W = 542, $P < 0.01$; Shannon index, t = 4.7, $P < 0.01$) (Figure A3). Regression analysis results showed that soil pH was significantly correlated to both OTU richness and Shannon index of the total archaeal community, as well as of the AOA (Figure 22). Total archaeal OTU richness and Shannon index negatively correlated to soil pH (Figure 22a and Figure 22c), and AOA OTU richness and Shannon index showed a unimodal relationship with soil pH (Figure 22b and Figure 22d).

Mantel correlograms showed significant correlations across short phylogenetic distances for both 16S rRNA and *amoA* genes (all $P < 0.05$, Figure 23a). However, for 16S rRNA gene there was significant correlations at intermediate phylogenetic distances ($P < 0.05$, Figure 23b). The phylogenetic community composition of 16S rRNA and *amoA* gene, based on both unweighted Unifrac and betaMNTD, was significantly influenced by biome (PERMANOVA, $P < 0.001$, 999 permutations), except for betaMNTD based phylogenetic community composition of 16S rRNA gene (PERMANOVA, P = 0.36).

In both biomes for both 16S rRNA and *amoA* genes, the mean values of ses.betaMNTD were significantly different from expected value of zero (Figure 24; t-test, $P < 0.001$). There was no distance-decay relationship and climate effect observed between pairwise phylobetadiversity and spatial and climate distance for 16S rRNA and

amoA gene in both biomes using Unifrac and betaMNTD (Table 10). However, when assessing environmental distance against phylobetadiversity, a strong positive relationship was found for 16S rRNA and *amoA* gene in both biomes (Table 10). For both Unifrac and betaMNTD, the slope of this relationship varied significantly between the two biomes ($P < 0.01$). Turnover rate in phylogenetic community composition was greater for both genes in the tropical than the temperate biome. The MRM model explained a large and significant proportion of the variability in total archaea and AOA community structure in both tropics and temperate biomes (Table 11). Among the environmental variables measured, only soil pH was found to be strongly correlated to both Unifrac and betaMNTD distances of 16S rRNA and *amoA* genes in both biomes (Table 11).

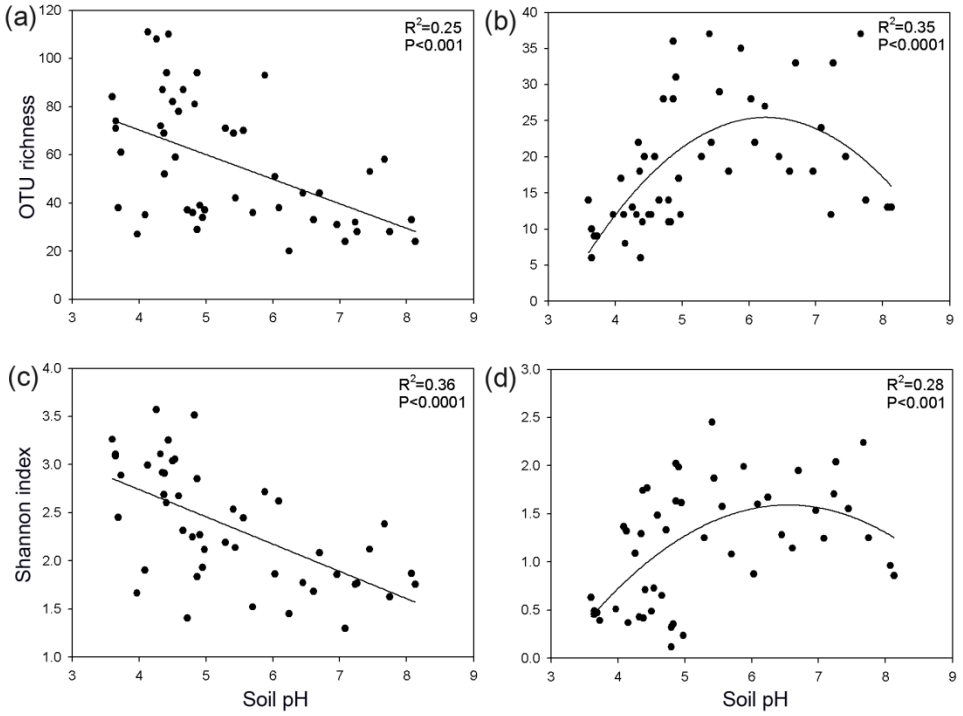

Figure 22. Relationship between soil pH, OTU richness and diversity of total archaea (a and c) and AOA (b and d) across tropical and temperate biome.

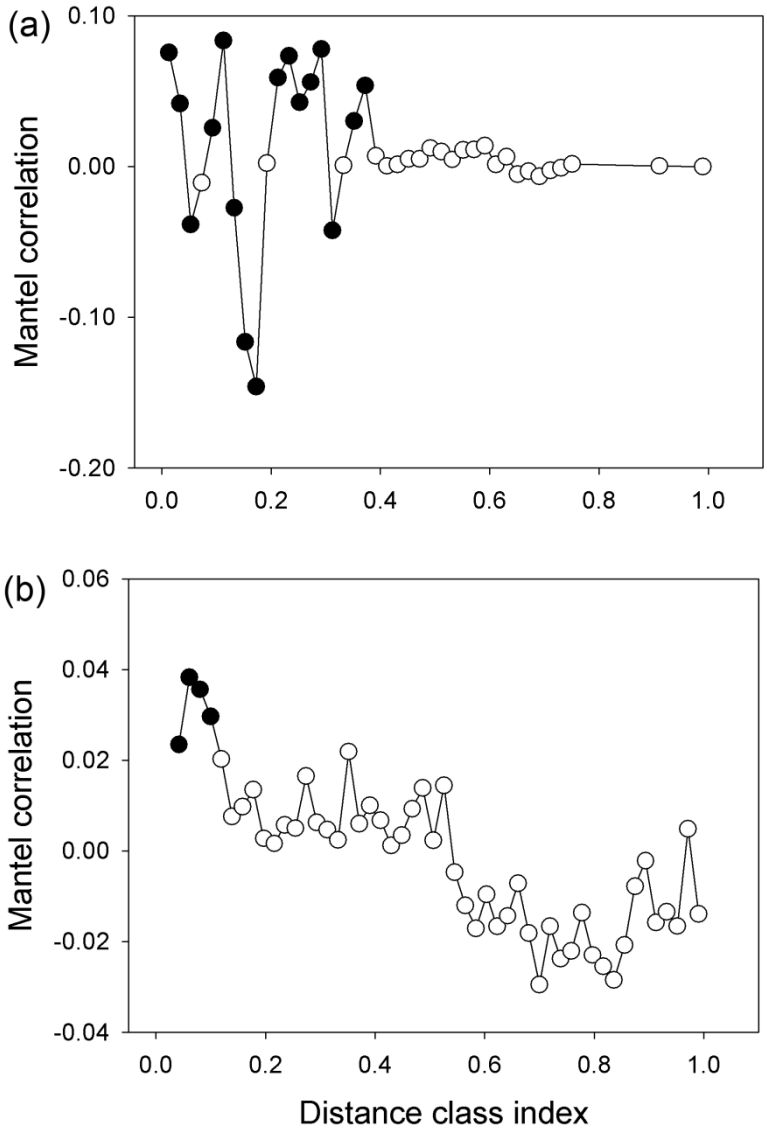

Figure 23. Mantel correlogram between the pairwise matrix of OTU niche distances and phylogenetic distances in (a) 16S rRNA gene and (b) *amoA* gene. Significant correlations ($P < 0.05$, solid circles) indicate phylogenetic signal in species ecological niches.

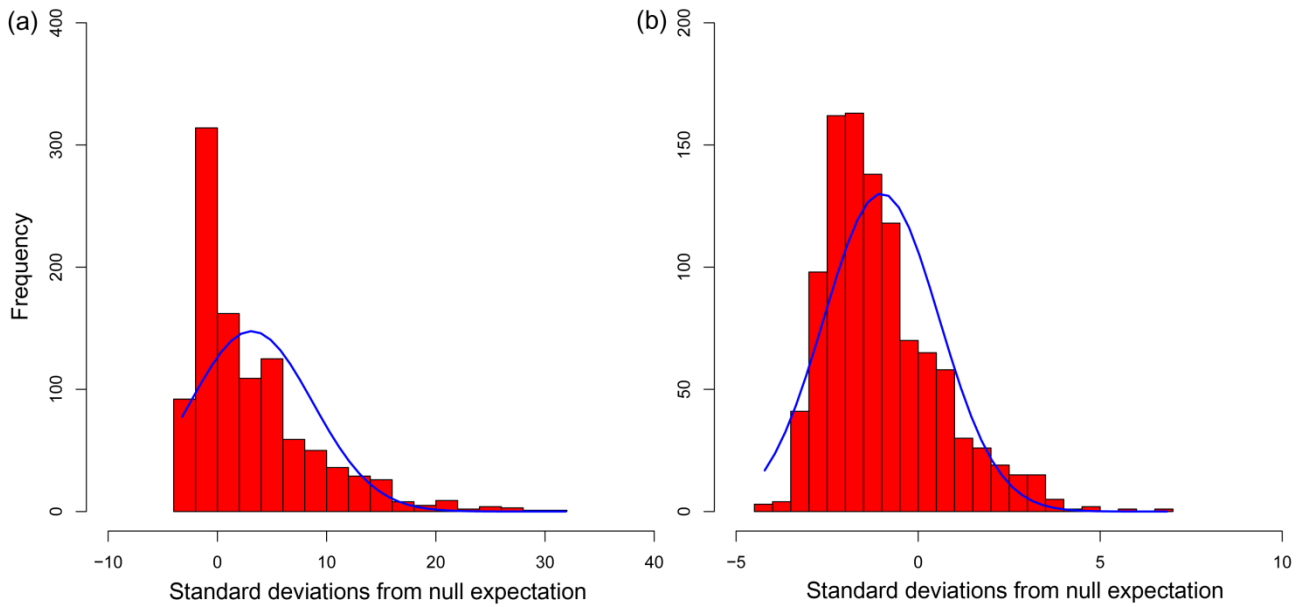

Figure 24. Frequency estimates for distributions of ses.betaMNTD of (a) 16S rRNA and (b) *amoA* gene. Each observation is the number of null model standard deviations the observed value is from the mean of its associated null distribution. Ses.betaMNTD values < -2 indicate less than expected turnover; values > +2 indicate greater than expected turnover. The mean values of ses.betaMNTD of both 16S rRNA and *amoA* genes were significantly different from expected value of zero for random data ($P < 0.001$, t-test).

Table 10. Mantel and partial Mantel tests for the correlation between Unifrac and betaMNTD and the explanatory distances (geographic, climate and environmental distance) using Spearman's rho in tropics and temperate biome.

Effect of	**Tropics**		**Temperate**	
	16S	***amoA***	**16S**	***amoA***
Unifrac				
Geographic	-0.108	0.019	0.071	0.147
Climate	0.028	0.041	0.075	0.107
Environmental	0.585***	0.474***	0.565***	0.288**
betaMNTD				
Geographic	-0.013	0.048	0.134	0.152
Climate	0.075	0.027	0.040	0.061
Environmental	0.429***	0.509***	0.365***	0.237*

***<0.001; **<0.01; *<0.05.

Table 11. MRM analysis of total and AOA communities using unweighted Unifrac and betaMNTD distance in tropics and temperate soils.

Environmental parameters	**Tropics**		**Temperate**	
	16S	***amoA***	**16S**	***amoA***
Unifrac	**R^2=0.620*** **	**R^2=0.584*** **	**R^2=0.524*** **	**R^2=0.577*** **
pH	0.124***	0.076***	0.096***	0.101***
NO_3	0.031*	NS	NS	NS
NH_4	NS	NS	NS	-0.026***
betaMNTD	**R^2=0.724*** **	**R^2=0.422*** **	**R^2=0.490*** **	**R^2=0.458*** **
pH	0.102***	0.056***	0.084***	0.122***
NH_4	NS	NS	NS	-0.037***

***<0.001; **<0.01; *<0.05.

The variation (R^2) is explained by the remaining variables and the partial regression coefficients of the final model are reported.

NS=not significant.

4.1.4. Discussion

The phylogenetic analysis of 16S rRNA gene based tree shows that soil samples are mainly dominated by *Thaumarchaeota* (Figure 18a), and previous molecular surveys based on the archaeal 16S rRNA gene have also revealed great abundance of *Thaumarchaeota* across a multitude of soil and ecosystem types (Auguet *et al.* 2010; Bates *et al.* 2011). Both tropical and temperate soils were dominated by *Thaumarchaeota* group 1.1b. This is consistent with the findings of an archaeal 16S rRNA gene-based survey of Bates et al. (Bates *et al.* 2011), where they showed that only two phylotypes (at 97% sequence identity) belonging to the group 1.1b *Thaumarchaeota* constituted > 70% of the retrieved archaeal sequences in 146 analyzed soils from around the world. In another comprehensive survey of archaeal populations by Auget et al. (Auguet *et al.* 2010), group 1.1b *Thaumarchaeota* has been shown to be dominant in soils globally. Thaumarchaeota group 1.1b relative abundance increased with soil pH, which is consistent with earlier findings (Hu *et al.* 2013; Tripathi *et al.* 2013). This is also consistent with the fact that till to date, the isolated and enriched cultures of group 1.1b (i.e. *Nitrososphaera viennensis* and *Nitrososphaera gargensis*) have been adapted to neutral or slightly alkaline conditions (Hatzenpichler *et al.* 2008; Tourna *et al.* 2011).

It is clear from the previous study by Tripathi et al. (Tripathi *et al.* 2013) of tropical soils, and the present study which adds temperate soils, that *Thaumarchaeota* group 1.1a is more abundant in acidic soils. This is also consistent with the recently described archaeon *Candidatus Nitrosotalea devanaterra* being acidophilic and belonging to group 1.1a (Lehtovirta-Morley *et al.* 2011b). However

they were significantly more abundant, relative to other groups, in temperate soils than in the tropics. This difference in relative abundance pattern suggests that apart from soil pH, there could be some other factors influencing the abundances of group 1.1a *Thaumarchaeota.*

There was cultured members of AOA within Group 1.1a and Group 1.1b *Thaumarchaeota*, that is mirrored in the respective *amoA* phylogeny (Figure 18b). However, cultured members of AOA are also reported in *Thaumarchaeota*-group *Nitrosocaldus*, though no sequence was recoved related to *Nitrosocaldus* group. In the *amoA* dataset, the *Nitrososphaera* cluster was detected as the most dominant cluster in both biomes. In several previous *amoA* diversity studies, *Nitrososphaera* has also been shown to dominate the AOA community in soil (Cao *et al.* 2013; Gubry-Rangin *et al.* 2011; Pester *et al.* 2012). However, the relative abundance of *Nitrosotalea* cluster was significantly higher in temperate soils. Previous studies have referred *Nitrososphaera* and *Nitrosotalea* clusters as alkaliphilic (Zhalnina *et al.* 2012) and acidophilic (Gubry-Rangin *et al.* 2011), respectively. However, their differential relative abundance in similar pH soils in tropical and temperate biomes shows that together with soil pH, biome is also an important factor structuring the broad community composition of total archaea as well as AOA.

Total archaea and AOA community composition varied significantly between biomes and were influenced by soil, spatial and climate factors. Soil pH was the primary environmental factor influencing the total archaea and AOA communities. The total archaea and AOA community composition was strongly structured along the soil pH gradient in both biomes. The most abundant OTUs of both total

archaea and AOA, though strongly structured along the soil pH gradient, also inhibit biome-specific patterns (Figure A4 and Figure A5). For instance in total archaea, OTU3, 5 and 4 which belongs to *Thaumarchaeota* groups 1.1a and 1.1c are highly abundant in temperate acidic soils, whereas, OTU7, 15 and 16 which belongs to *Thaumarchaeota* groups 1.1b and 1.1c are more abundant in tropical acidic soils (Figure A4). Similarly for AOA, OTU1 which affiliates to *Nitrososphaera* cluster are more abundant in tropical acidic soils, whereas, OTU2 which belongs to *Nitrosotalea* cluster are highly abundant in temperate acidic soils (Figure A5). These results are in contrast with previous studies on bacterial communities (Chu *et al.* 2010b; Fierer & Jackson 2006; Lauber *et al.* 2009; Rousk *et al.* 2010; Tripathi *et al.* 2012), which showed that soil bacterial community composition was strongly structured by soil pH and biome definitions are not useful for predicting variability in soil bacterial communities across the globe. The result clearly shows that even if the archaeal community inhabits a similar pH environment, the community composition can be different if those soil pH environments are in different climatic zones.

The total archaea OTU richness and diversity did not differ among biomes, however it is interesting that AOA was more diverse in temperate soils. A possible reason for this "inverse" latitudinal diversity trend might be the greater fluctuation of ammonium concentration throughout the year due to temperature variation in temperate soils, which could select for several different *amoA* variants specialized to function under these various ammonium concentrations. It has been observed that the conversion of soil organic nitrogen occurs more

readily than subsequent nitrification at soil temperatures below 10^{o}C (Campbell & Biederbeck 1972; Cookson *et al.* 2002; Emmer & Tietema 1990), therefore NH_4^+-N often builds up in the soil over winter. When soil temperature goes up after the winter season it is quickly converted to NO_2 and NO_3 by the nitrification process. Besides influencing the abundance of individual taxa and community composition of total archaea and AOA, soil pH also influenced OTU richness and diversity of total archaea and AOA. Total archaeal OTU richness and Shannon index were higher in acidic soils and this could be due to the fact that most of the archaeal classes group 1.1a, group 1.1c and *Thermoplasmata* are well adapted to acidic soils. However, AOA diversity showed a humpback relationship with soil pH and diversity peaked around at neutral pH. This is consistent with other studies on AOA around the world (Gubry-Rangin *et al.* 2011; Pester *et al.* 2012), but shows this pattern more clearly here, perhaps due to larger numbers of samples having been gathered from the same biome. This humpback relationship is consistent with the fact that most of the AOA identified to date are adapted to neutral or slightly alkaline conditions (De La Torre *et al.* 2008; Könneke *et al.* 2005; Kim *et al.* 2012a; Tourna *et al.* 2011), except for Ca. *Nitrosotalea devanaterra*, which survives in acidic conditions (i.e., pH 4.0-5.5) (Lehtovirta-Morley *et al.* 2011a).

Phylogenetic signals in OTU niches are needed to infer ecological processes using phylogenetic information. Significant phylogenetic signals were detected for both 16S rRNA and *amoA* genes, this is consistent with previous studies on bacteria, where significant signals have been detected across short phylogenetic distances. This

pattern in phylogenetic signals strongly suggests that at shorter phylogenetic distances both total archaeal and AOA taxa are phylogenetically clustered, which means that – as with bacteria (Stegen *et al.* 2012; Wang *et al.* 2013) closely related archaeal as well as AOA taxa are ecologically coherent (i.e. tending to occur together in same environment), being phylogenetically conserved in their niches. However, across intermediate taxonomic distances there were both significant negative and positive signals for 16S rRNA genes. This shows that at greater phylogenetic/evolutionary distances, both phylogenetic clustering and dispersion acts together to produce the total archaeal community.

In both the tropical and temperate biomes the ses.betaMNTD distribution mean deviated significantly from zero for both 16S rRNA and *amoA* gene, suggesting a dominant role of deterministic processes in archaeal community composition. Furthermore, 16S rRNA genes in both biomes showed distributions greater than zero, suggesting that across communities there are shifts in environmental variables, which deterministically govern total archaeal community composition. However, *amoA* genes in both tropical and temperate soils had mean ses.betaMNTD values less than zero, suggesting that for AOA communities there was relative consistency in the environmental variables that deterministically affect the AOA community composition. This could be due to the fact that AOA are believed to be a physiologically constrained group, surviving solely on the oxidation of ammonia and at each place the AOA community is composed of relatively similar taxa because of environmental filtering.

In conclusion, through more intensive sampling in two biomes,

it was shown that both total archaea and AOA community composition is strongly driven by both soil pH and biome. The data also indicate that total archaea and AOA have definable niches along the soil pH gradient. *Thaumarchaeota* groups 1.1b, 1.1a and 1.1c appear to be the dominant archaeal lineages in both the temperate and tropical biomes, and the *Nitrososphaera and Nitrosotalea* clusters were the most abundant AOA taxa in both biomes. The diversity of total archaea and AOA showed a different pattern, where total archaeal diversity peaked in acidic soils and AOA diversity peaked in soils having pH close to neutral. For now, it remains an open question as to why more amoA gene variants can coexist at neutral pH, and why more 16S rRNA gene variants occur at low pH. The AOA diversity was greater in the temperate latitudes than the tropics. The phylogenetic community assembly of both total archaea and AOA was found deterministically governed and for both genes, closely related OTUs were are ecologically coherent and tend to co-occur in the same environments. However, at greater phylogenetic distances, both phylogenetic clustering and dispersion acts together to produce the total archaeal community.

Overall, this study is an early step towards understanding the true community ecology of soil archaea. To gain further understanding, there is a need for further more extensive sampling in both tropical and temperate forests, at different times of year in the temperate soil community, ideally using expressed RNA as a measure of cell activity. It is important to bear in mind that all the samples in this study came from just two biomes, within one geographical region (East Asia). Sample sets from other biomes will also enable testing of whether these

same patterns apply within and between other biomes, and geographical regions.

GENERAL CONCLUSIONS

Tropical rainforests of Southeast Asia are well known for its species-rich tropical rainforest, representing a 'hotspot' of biodiversity, but it has been poorly studied in relation to its microbial diversity. Microorganisms not only make up a large proportion of the biological diversity of the rainforest environment but also are a fundamental component of nutrient cycling and productivity. This study investigated the extent of soil microbial diversity in tropical soils of Malaysia and how they respond to land use changes, spatial scaling effect and biome effect, as well as the extent to which ecological processes or other environmental factors contribute to structuring the soil microbial communities in tropical soils.

At first, the effect of land use changes (i.e. forest vs. agriculture) was investigated on soil bacterial, archaeal and AOA communities. Soil pH emerged as the best predictor of abundance, community composition, and diversity of all bacteria, archaea and AOA in tropical soils. While land use as such appears to be mostly just a correlate of pH, and not a predominant determinant of microbial communities, the pH changes associated with conversion to agriculture (e.g. due to soil liming) may be expected to alter microbial communities structure and diversity. Bacteria and AOA were more diverse in soils having pH close to neutral, whereas, archaea were more diverse in acidic tropical soils. *Acidobacteria* and *Proteobacteria* were the most dominant bacterial phyla across all soil samples. *Thaumarchaeota* groups 1.1b and 1.1c appear to be the dominant archaeal lineages in tropical lowland soils, whereas, *Nitrososphaera*

and *Nitrosotalea* clusters were recovered as the most dominant AOA taxa. Comparison of relative abundances of these dominant microbial taxa together with dominant phylotypes strongly supports the concept that individual microbial lineages found in soil are adapted to specific pH ranges.

How the spatial scaling affects the bacterial communities in tropical rainforest soils of Southeast Asia was examined at two spatial scales: a local scale with samples spaced every 5 m over a 150 m transect, and a regional scale with samples 1 to 1,800 km apart. This study suggested that in tropical rainforest soils at both local and regional scales, environmental variables strongly affect the bacterial community composition. Soil pH was the only environmental parameter that significantly explained the variance in bacterial community at the local scale, whereas total nitrogen and elevation were additional important factors at the regional scale. Only a weak evidence of distance alone was found, and thus only weak evidence of what might be construed as dispersal lag. Even the distance effect which was found could be a result of environmental factors not accounted for in this study including the distance decay in the plant community that acts as a host for root-associated bacteria.

On a geographical scale, total soil archaeal and AOA communities in two different climatic zones, -tropical Malaysia and temperate Korea and Japan - were investigated using 454-pyrosequencing targeting partial 16S rRNA and *amoA* genes, respectively. In this study, it has been shown that both total archaea and AOA community composition though strongly structured by soil pH are also influenced by biome. The data also indicate that total archaea and

AOA have definable niches along the soil pH gradient. *Thaumarchaeota* groups 1.1b, 1.1a nd 1.1c appear to be the dominant archaeal lineages in both the temperate and tropical biomes, and the *Nitrososphaera and Nitrosotalea* clusters were the most abundant AOA taxa in both biomes. The diversity of total archaea and AOA showed a different pattern, where total archaeal diversity peaked in acidic soils and AOA diversity peaked in soils having pH close to neutral. The phylogenetic community assembly of both total archaea and AOA was found deterministically governed and for both genes, closely related OTUs were are ecologically coherent and tend to co-occur in the same environments. However, at greater phylogenetic distances, both phylogenetic clustering and dispersion acts together to produce the total archaeal community.

Overall, it appears that patterns of community composition and diversity of microorganisms in tropical soils of Southeast Asia are somewhat different from those observed in macro-organisms in tropics. Soil microbial community structure and diversity were more strongly influenced by edaphic factors rather than habitat types and much soil microbial OTUs appear to have clearly defined environmental niches. However, there are several consistencies with macroorganisms. Total archaeal and AOA communities were distinctive to each climatic zone, which indicates that there are biome- or climate-specific microbial taxa adapted to each environment, which is the typical of biogeography in macroscopic taxa. Together these results provide a baseline ecological framework with which to pursue future research on both tropical soil microbial function, and more explicit biome based assessments of the local ecological drivers of microbial biodiversity.

REFERENCES

Acinas SG, Klepac-Ceraj V, Hunt DE, *et al.* (2004) Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430**, 551-554.

Adams JM (2009) *Species richness: patterns in the diversity of life* Springer in association with Praxis, Chichester.

Aebischer NJ, Robertson PA, Kenward RE (1993) Compositional analysis of habitat use from animal radio-tracking data. *Ecology* **74**, 1313-1325.

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells without Cultivation. *Microbiological Reviews* **59**, 143-169.

Anderson MJ (2001) A new method for non-parametric multivariate analysis of variance. *Austral ecology* **26**, 32-46.

Angel R, Soares MIM, Ungar ED, Gillor O (2010) Biogeography of soil archaea and bacteria along a steep precipitation gradient. *The ISME Journal* **4**, 553-563.

Auguet JC, Barberan A, Casamayor EO (2010) Global ecological patterns in uncultured Archaea. *The ISME Journal* **4**, 182-190.

Auguet JC, Casamayor EO (2008) A hotspot for cold crenarchaeota in the neuston of high mountain lakes. *Environmental Microbiology* **10**, 1080-1086.

Ayton J, Aislabie J, Barker GM, Saul D, Turner S (2010) *Crenarchaeota* affiliated with group 1.1b are prevalent in coastal mineral soils of the Ross Sea region of Antarctica. *Environmental Microbiology* **12**, 689-703.

Béja O, Aravind L, Koonin EV, *et al.* (2000) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902-1906.

Baas-Becking LGM (1934) Geobiologie of inleiding tot de milieukunde (eds. van Stockum WP, Zoon NV), The Hague: Netherlands.

Bardgett RD, Freeman C, Ostle NJ (2008) Microbial contributions to climate change through carbon cycle feedbacks. *The ISME Journal* **2**, 805-814.

Barton AD, Dutkiewicz S, Flierl G, Bragg J, Follows MJ (2010) Patterns of diversity in marine phytoplankton. *Science* **327**, 1509-1511.

Bates ST, Berg-Lyons D, Caporaso JG, *et al.* (2011) Examining the global distribution of dominant archaeal populations in soil. *The ISME Journal* **5**, 908-917.

Bell G (2001) Ecology- neutral macroecology. *Science* **293**, 2413-2418.

Bengtson P, Sterngren AE, Rousk J (2012) Archaeal abundance across a pH gradient in an arable soil and its relationship to bacterial and fungal growth rates. *Applied and Environmental Microbiology* **78**, 5906-5911.

Berry NJ, Phillips OL, Ong RC, Hamer KC (2008) Impacts of selective logging on tree diversity across a rainforest landscape: the importance of spatial scale. *Landscape Ecology* **23**, 915-929.

Bierregaard RO (2001) *Lessons from Amazonia: the ecology and conservation of a fragmented forest* Yale University Press.

Blanchet FG, Legendre P, Borcard D (2008) Forward selection of explanatory variables. *Ecology* **89**, 2623-2632.

Borcard D, Legendre P (2002) All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling* **153**, 51-68.

Borcard D, Legendre P, Avois-Jacquet C, Tuomisto H (2004) Dissecting the spatial structure of ecological data at multiple scales. *Ecology* **85**, 1826-1832.

Borneman J, Triplett EW (1997) Molecular microbial diversity in soils from eastern Amazonia: Evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Applied and Environmental Microbiology* **63**, 2647-2653.

Bossio DA, Girvan MS, Verchot L, *et al.* (2005) Soil microbial community response to land use change in an agricultural landscape of Western Kenya. *Microbial Ecology* **49**, 50-62.

Bradley RD, Hillis DM (1997) Recombinant DNA sequences generated by PCR amplification. *Molecular Biology and Evolution* **14**, 592-593.

Bray RH, Kurtz LT (1945) Determination of total, organic, and available forms of phosphorus in soils. *Soil Science* **59**, 39-45.

Bremner JM, Mulvaney CS (1982) Nitrogen- total. In: *Method of Soil analysis Part-2* (eds. Page AL, Miller RH, Kenny DR), pp. 595-624. American Society of Agronomy, Madison, WI.

Brochier-Armanet C, Boussau B, Gribaldo S, Forterre P (2008) Mesophilic *Crenarchaeota*: proposal for a third archaeal phylum, the *Thaumarchaeota*. *Nature Reviews Microbiology* **6**, 245-252.

Brook BW, Sodhi NS, Ng PKL (2003) Catastrophic extinctions follow deforestation in Singapore. *Nature* **424**, 420-423.

Campbell BJ, Polson SW, Hanson TE, Mack MC, Schuur EA (2010) The effect of nutrient deposition on bacterial communities in Arctic tundra soil. *Environmental Microbiology* **12**, 1842-1854.

Campbell C, Biederbeck V (1972) Influence of fluctuating temperatures and constant soil moistures on nitrogen changes in amended and unamended loam. *Canadian Journal of Soil Science* **52**, 323-336.

Cao H, Auguet J-C, Gu J-D (2013) Global ecological pattern of ammonia-oxidizing archaea. *Plos One* **8**, e52853.

Cao P, Zhang LM, Shen JP, *et al.* (2012) Distribution and diversity of archaeal communities in selected Chinese soils. *FEMS Microbiology Ecology* **80**, 146-158.

Cardinale BJ, Duffy JE, Gonzalez A, *et al.* (2012) Biodiversity loss and its impact on humanity. *Nature* **486**, 59-67.

Cenciani K, Lambais MR, Cerri CC, de Azevedo LCB, Feigl BJ (2009) Bacteria diversity and microbial biomass in forest, pasture and fallow soils in the Southwestern Amazon basin. *Revista Brasileira De Ciencia Do Solo* **33**, 907-916.

Chaban B, Ng SYM, Jarrell KF (2006) Archaeal habitats - from the extreme to the ordinary. *Canadian Journal of Microbiology* **52**, 73-116.

Chesson PL, Warner RR (1981) Environmental variability promotes coexistence in lottery competitive-systems. *American Naturalist* **117**, 923-943.

Cho JC, Tiedje JM (2000) Biogeography and degree of endemicity of fluorescent Pseudomonas strains in soil. *Applied and Environmental Microbiology* **66**, 5448-5456.

Chu H, Fierer N, Lauber CL, *et al.* (2010a) Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environmental Microbiology* **12**, 2998-3006.

Chu HY, Fierer N, Lauber CL, *et al.* (2010b) Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environmental Microbiology* **12**, 2998-3006.

Chun J, Kim KY, Lee JH, Choi Y (2010) The analysis of oral microbial communities of wild-type and toll-like receptor 2-deficient mice using a 454 GS FLX Titanium pyrosequencer. *BMC Microbiology* **10**, 101.

Chytrý M, Lososová Z, Horsák M, *et al.* (2012) Dispersal limitation is stronger in communities of microorganisms than macroorganisms across Central European cities. *Journal of Biogeography* **39**, 1101-1111.

Clarke KR, Gorley RN (2006) PRIMER v6: User Manual/Tutorial. Primer-E Ltd., Plymouth, UK.

Colwell RK, Lees DC (2000) The mid-domain effect: geometric constraints on the geography of species richness. *Trends in Ecology and Evolution* **15**, 70-76.

Condit R, Pitman N, Leigh EG, *et al.* (2002) Beta-diversity in tropical forest trees. *Science* **295**, 666-669.

Cookson W, Cornforth I, Rowarth J (2002) Winter soil temperature (2–15˚C) effects on nitrogen transformations in clover green manure amended or unamended soils; a laboratory and field study. *Soil Biology and Biochemistry* **34**, 1401-1415.

Corlett RT, Primack RB (2011) *Tropical rain forests: an ecological and*

biogeographical comparison John Wiley & Sons.

De La Torre JR, Walker CB, Ingalls AE, Könneke M, Stahl DA (2008) Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environmental Microbiology* **10**, 810-818.

DeLong EF (1998) Everything in moderation: Archaea as 'non-extremophiles'. *Current Opinion in Genetics and Development* **8**, 649-654.

DeLong EF, Preston CM, Mincer T, *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**, 496-503.

Diniz-Filho JAF, Terribile LC, Da Cruz MJR, Vieira LCG (2010) Hidden patterns of phylogenetic non-stationarity overwhelm comparative analyses of niche conservatism and divergence. *Global Ecology and Biogeography* **19**, 916-926.

Dirzo R, Raven PH (2003) Global state of biodiversity and loss. *Annual Review of Environment and Resources* **28**, 137-167.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-2200.

Edwards DP, Hodgson JA, Hamer KC, *et al.* (2010) Wildlife-friendly oil palm plantations fail to protect biodiversity effectively. *Conservation Letters* **3**, 236-242.

Edwards DP, Larsen TH, Docherty TDS, *et al.* (2011) Degraded lands worth protecting: the biological importance of Southeast Asia's repeatedly logged forests. *Proceedings of the Royal Society B-Biological Sciences* **278**, 82-90.

Ehrhardt CJ, Haymon RM, Lamontagne MG, Holden PA (2007) Evidence for hydrothermal Archaea within the basaltic flanks of the East Pacific Rise. *Environmental Microbiology* **9**, 900-912.

Eichorst SA, Kuske CR, Schmidt TM (2011) Influence of plant polymers on the distribution and cultivation of bacteria in the phylum Acidobacteria. *Applied and Environmental Microbiology* **77**, 586-596.

Eilers KG, Debenport S, Anderson S, Fierer N (2012) Digging deeper to find unique microbial communities: The strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biology and Biochemistry* **50**, 58-65.

Emmer I, Tietema A (1990) Temperature-dependent nitrogen transformations in acid oak-beech forest litter in the Netherlands. *Plant and Soil* **122**, 193-196.

Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation* **61**, 1-10.

Ferrer M, Golyshina OV, Chernikova TN, *et al.* (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environmental Microbiology* **7**, 1996-2010.

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 626-631.

Fine PV, Kembel SW (2011) Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities. *Ecography* **34**, 552-565.

Fitzherbert EB, Struebig MJ, Morel A, *et al.* (2008) How will oil palm expansion affect biodiversity? *Trends in Ecology and Evolution*

23, 538-545.

Fox GE, Wisotzkey JD, Jurtshuk P (1992) How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *International Journal of Systematic Bacteriology* **42**, 166-170.

Franklin RB, Mills AL (2003) Multi-scale variation in spatial heterogeneity for microbial community structure in an eastern Virginia agricultural field. *FEMS Microbiology Ecology* **44**, 335-346.

Fulthorpe RR, Roesch LFW, Riva A, Triplett EW (2008) Distantly sampled soils carry few species in common. *The ISME Journal* **2**, 901-910.

Gans J, Wolinsky M, Dunbar J (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**, 1387-1390.

Gibbs HK, Ruesch AS, Achard F, *et al.* (2010) Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16732-16737.

Gibson L, Lee TM, Koh LP, *et al.* (2011) Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature* **478**, 378-381.

Gilbert B, Lechowicz MJ (2004) Neutrality, niches, and dispersal in a temperate forest understory. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7651-7656.

Gilbert JA, Mühling M, Joint I (2008) A rare SAR11 fosmid clone

confirming genetic variability in the ‘Candidatus *Pelagibacter* ubique’genome. *The ISME Journal* **2**, 790-793.

Gill SR, Pop M, DeBoy RT, *et al.* (2006) Metagenomic analysis of the human distal gut microbiome. *Science* **312**, 1355-1359.

Goberna M, Insam H, Klammer S, Pascual JA, Sanchez J (2005) Microbial community structure at different depths in disturbed and undisturbed semiarid Mediterranean forest soils. *Microbial Ecology* **50**, 315-326.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**, 1-19.

Gouvenain RCd, Silander JA (2003) Do tropical storm regimes influence the structure of tropical lowland rain forests? *Biotropica* **35**, 166-180.

Green J, Bohannan BJM (2006) Spatial scaling of microbial biodiversity. *Trends in Ecology and Evolution* **21**, 501-507.

Green JL, Holmes AJ, Westoby M, *et al.* (2004) Spatial scaling of microbial eukaryote diversity. *Nature* **432**, 747-750.

Griffiths RI, Thomson BC, James P, *et al.* (2011) The bacterial biogeography of British soils. *Environmental Microbiology* **13**, 1642-1654.

Grime JP (1987) Dominant and subordinate components of plant communities: implications for succession. In: *Colonization, Succession and Stability* (eds. Gray AJ, Crawley MJ, Edwards PJ), pp. 413-429. Blackwell Scientific Publications, Oxford.

Gubry-Rangin C, Hai B, Quince C, *et al.* (2011) Niche specialization of terrestrial archaeal ammonia oxidizers. *Proceedings of the*

National Academy of Sciences of the United States of America **108**, 21206-21211.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry and Biology* **5**, R245-R249.

Hansen MC, Stehman SV, Potapov PV, *et al.* (2008) Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 9439-9444.

Hartman WH, Richardson CJ, Vilgalys R, Bruland GL (2008) Environmental and anthropogenic controls over bacterial communities in wetland soils. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17842-17847.

Hatzenpichler R, Lebedeva EV, Spieck E, *et al.* (2008) A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2134-2139.

Hoffmann M, Hilton-Taylor C, Angulo A, *et al.* (2010) The impact of conservation on the status of the world's vrtebrates. *Science* **330**, 1503-1509.

Horner-Devine MC, Lage M, Hughes JB, Bohannan BJ (2004) A taxa–area relationship for bacteria. *Nature* **432**, 750-753.

Hu H-W, Zhang L-M, Dai Y, Di H-J, He J-Z (2013) pH-dependent distribution of soil ammonia oxidizers across a large

geographical scale as revealed by high-throughput pyrosequencing. *Journal of Soils and Sediments* **13**, 1439-1449.

Hubbell SP (2001) *The unified neutral theory of biodiversity and biogeography* Princeton University Press, Princeton, N.J. ; Oxford.

Hugenholtz P, Huber T (2003) Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *International Journal of Systematic and Evolutionary Microbiology* **53**, 289-293.

Hur M, Kim Y, Song HR, *et al.* (2011) Effect of gnetically mdified pplars on sil mcrobial cmmunities during the pytoremediation of wste mne tilings. *Applied and Environmental Microbiology* **77**, 7611-7619.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Mcrobiology* **12**, 1889-1898.

Huston MA (1994) *Biological diversity : the coexistence of species on changing landscapes* Cambridge University Press, Cambridge.

Jesus ED, Marsh TL, Tiedje JM, Moreira FMD (2009) Changes in land use alter the structure of bacterial communities in Western Amazon soils. *The ISME Journal* **3**, 1004-1011.

Jones RT, Robeson MS, Lauber CL, *et al.* (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *The ISME Journal* **3**, 442-453.

Joppa LN, Roberts DL, Myers N, Pimm SL (2011) Biodiversity hotspots house most undiscovered plant species. *Proceedings of the National Academy of Sciences of the United States of*

America **108**, 13171-13176.

Könneke M, Bernhard AE, José R, *et al.* (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543-546.

Kemnitz D, Kolb S, Conrad R (2007) High abundance of *Crenarchaeota* in a temperate acidic forest soil. *Fems Microbiology Ecology* **60**, 442-448.

Kim HM, Jung JY, Yergeau E, *et al.* (2014) Bacterial community structure and soil properties of a subarctic tundra soil in Council, Alaska. *FEMS Microbiology Ecology* **89**, 465-475.

Kim JG, Jung MY, Park SJ, *et al.* (2012a) Cultivation of a highly enriched ammonia-oxidizing archaeon of thaumarchaeotal group I.1b from an agricultural soil. *Environmental Microbiology* **14**, 1528-1543.

Kim M, Boldgiv B, Singh D, *et al.* (2013) Structure of soil bacterial communities in relation to environmental variables in a semi-arid region of Mongolia. *Journal of Arid Environments* **89**, 38-44.

Kim OS, Cho YJ, Lee K, *et al.* (2012b) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *International Journal of Systematic and Evolutionary Microbiology* **62**, 716-721.

Kletzin A (2007) General Characteristics and Important Model Organisms. In: *Archaea: Cellular and Molecular Biology* (ed. Cavicchioli R), pp. 14-93. ASM Press, Washington.

Koh LP, Wilcove DS (2008) Is oil palm agriculture really destroying tropical biodiversity? *Conservation Letters* **1**, 60-64.

Koleff P, Lennon JJ, Gaston KJ (2003) Are there latitudinal gradients in species turnover? *Global Ecology and Biogeography* **12**, 483-498.

Konneke M, Bernhard AE, de la Torre JR, *et al.* (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**, 543-546.

Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology* **75**, 5111-5120.

Lauber CL, Strickland MS, Bradford MA, Fierer N (2008) The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biology and Biochemistry* **40**, 2407-2415.

Lavelle P, Pashanasi B (1989) Soil macrofauna and land management in Peruvian Amazonia (Yurimaguas, Loreto). *Pedobiologia* **33**, 283-291.

Legendre P, Lapointe FJ, Casgrain P (1994) Modeling brain evolution from behavior - a permutational regression approach. *Evolution* **48**, 1487-1499.

Lehtovirta-Morley LE, Stoecker K, Vilcinskas A, Prosser JI, Nicol GW (2011a) Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15892-15897.

Lehtovirta-Morley LE, Stoecker K, Vilcinskas A, Prosser JI, Nicol GW (2011b) Cultivation of an obligate acidophilic ammonia oxidizer

from a nitrifying acid soil. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15892-15897.

Lehtovirta LE, Prosser JI, Nicol GW (2009) Soil pH regulates the abundance and diversity of Group 1.1c Crenarchaeota. *Fems Microbiology Ecology* **70**, 367-376.

Lliros M, Casamayor EO, Borrego C (2008) High archaeal richness in the water column of a freshwater sulfurous karstic lake along an interannual study. *FEMS Microbiology Ecology* **66**, 331-342.

Logares R, Lindström ES, Langenheder S, *et al.* (2012) Biogeography of bacterial communities exposed to progressive long-term environmental change. *The ISME Journal*.

Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* **71**, 8228-8235.

Magurran AE (2004) *Measuring biological diversity* Blackwell.

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133-141.

Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC (2011) Drivers of bacterial beta-diversity depend on spatial scale. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 7850-7854.

Martiny JBH, Bohannan BJM, Brown JH, *et al.* (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology* **4**, 102-112.

McGregor GR, Nieuwolt S (1998) *Tropical Climatology: An Introduction to the Climates of the Low Latitudes, 2nd Edition*

John Wiley and Sons, New York.

Morley RJ (2000) *Origin and evolution of tropical rainforests* John Wiley & Sons.

Myers N (1988) Threatened biotas: "hot spots" in tropical forests. *Environmentalist* **8**, 187-208.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858.

Nacke H, Thurmer A, Wollherr A, *et al.* (2011) Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS One* **6**.

Navarrete AA, Taketani RG, Mendes LW, *et al.* (2011) Land-use systems affect archaeal community structure and functional diversity in western Amazon soils. *Revista Brasileira De Ciencia Do Solo* **35**, 1527-1540.

Nemergut DR, Costello EK, Hamady M, *et al.* (2011) Global patterns in the biogeography of bacterial taxa. *Environmental Microbiology* **13**, 135-144.

Nicol GW, Campbell CD, Chapman SJ, Prosser JI (2007) Afforestation of moorland leads to changes in crenarchaeal community structure. *FEMS Microbiology Ecology* **60**, 51-59.

Nicol GW, Leininger S, Schleper C, Prosser JI (2008) The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environmental Microbiology* **10**, 2966-2978.

Nicol GW, Tscherko D, Embley TM, Prosser JI (2005) Primary

succession of soil *Crenarchaeota* across a receding glacier foreland. *Environmental Microbiology* **7**, 337-347.

Nolte V, Pandey RV, Jost S, *et al.* (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Molecular Ecology* **19**, 2908-2915.

Nusslein K, Tiedje JM (1999) Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. *Applied and Environmental Microbiology* **65**, 3622-3626.

Ochsenreiter T, Selezi D, Quaiser A, Bonch-Osmolovskaya L, Schleper C (2003) Diversity and abundance of *Crenarchaeota* in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environmental Microbiology* **5**, 787-797.

Oden NL, Sokal RR (1986) Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Systematic Biology* **35**, 608-617.

Oh YM, Kim M, Lee-Cruz L, *et al.* (2012) Distinctive bacterial communities in the rhizoplane of four tropical tree species. *Microbial Ecology* **64**, 1018-1027.

Oksanen J, Kindt R, Legendre P, *et al.* (2007) The vegan package. *Community ecology package. Disponível em: http://www. R-project. org. Acesso em* **10**, 2008.

Oline DK, Schmidt SK, Grant MC (2006) Biogeography and landscape-scale diversity of the dominant crenarchaeota of soil. *Microbial Ecology* **52**, 480-490.

Parmentier I, Malhi Y, Senterre B, *et al.* (2007) The odd man out? Might climate explain the lower tree α-diversity of African rain forests relative to Amazonian rain forests? *Journal of Ecology*

95, 1058-1071.

Pasternak Z, Al-Ashhab A, Gatica J, *et al.* (2013) Spatial and temporal biogeography of soil microbial communities in arid and semiarid Regions. *Plos One* **8**, e69705.

Pedrós-Alió C (2006) Marine microbial diversity: can it be determined? *Trends in Microbiology* **14**, 257-263.

Peres-Neto PR, Legendre P, Dray S, Borcard D (2006) Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology* **87**, 2614-2625.

Pester M, Rattei T, Flechl S, *et al.* (2012) amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions. *Environmental Microbiology* **14**, 525-539.

Pitman NC, Jørgensen PM (2002) Estimating the size of the world's threatened flora. *Science* **298**, 989-989.

Porazinska DL, Giblin-Davis RM, Powers TO, Thomas WK (2012) Nematode spatial and ecological patterns from tropical and temperate rainforests. *Plos One* **7**.

Price MN, Dehal PS, Arkin AP (2010) FastTree 2-approximately maximum-likelihood trees for large alignments. *Plos One* **5**.

Prosser JI, Bohannan BJ, Curtis TP, *et al.* (2007) The role of ecological theory in microbial ecology. *Nature Reviews Microbiology* **5**, 384-392.

Qian H, Ricklefs RE (2007) A latitudinal gradient in large-scale beta diversity for vascular plants in North America. *Ecology Letters* **10**, 737-744.

RDevelopmentCoreTeam (2008) R: A language and environment for

statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodrigues JL, Pellizari VH, Mueller R, *et al.* (2013) Conversion of the Amazon rainforest to agriculture results in biotic homogenization of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 988-993.

Roesch LF, Fulthorpe RR, Riva A, *et al.* (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal* **1**, 283-290.

Rosen C (2000) *World Resources 2000-2001: People and ecosystems: The fraying web of life* Elsevier.

Rousk J, Baath E, Brookes PC, *et al.* (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME Journal* **4**, 1340-1351.

Schloss PD, Handelsman J (2006) Toward a census of bacteria in soil. *PLoS Computational Biology* **2**, e92.

Schloss PD, Westcott SL, Ryabin T, *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**, 7537-7541.

Sheldon FH, Styring A, Hosner PA (2010) Bird species richness in a Bornean exotic tree plantation: A long-term perspective. *Biological Conservation* **143**, 399-407.

Singh D, Shi L, Adams JM (2013) Bacterial diversity in the mountains of South-West China: Climate dominates over soil parameters. *Journal of Microbiology* **51**, 439-447.

Singh D, Takahashi K, Adams JM (2012a) Elevational patterns in archaeal diversity on Mt. Fuji. *Plos One* **7**.

Singh D, Takahashi K, Kim M, Chun J, Adams JM (2012b) A hump-backed trend in bacterial diversity with elevation on Mount Fuji, Japan. *Microbial Ecology* **63**, 429-437.

Soares-Filho BS, Nepstad DC, Curran LM, *et al.* (2006) Modelling conservation in the Amazon basin. *Nature* **440**, 520-523.

Sodhi NS, Koh LP, Brook BW, Ng PKL (2004) Southeast Asian biodiversity: an impending disaster. *Trends in Ecology and Evolution* **19**, 654-660.

Stahl DA, de la Torre JR (2012) Physiology and diversity of ammonia-oxidizing Archaea. *Annual Review of Microbiology* **66**, 83-101.

Stegen JC, Lin X, Konopka AE, Fredrickson JK (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal* **6**, 1653-1664.

Suenaga H (2012) Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environmental Microbiology* **14**, 13-22.

Taketani RG, Tsai SM (2010) The influence of different land uses on the structure of archaeal communities in Amazonian anthrosols based on 16S rRNA and amoA genes. *Microbial Ecology* **59**, 734-743.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596-1599.

Tang S, Antonov I, Borodovsky M (2013) MetaGeneTack: ab initio detection of frameshifts in metagenomic sequences.

Bioinformatics **29**, 114-116.

Telford RJ, Vandvik V, Birks HJB (2006) Dispersal limitations matter for microbial morphospecies. *Science* **312**, 1015-1015.

Thomas GW (1982) Exchangeable cations. In: *Method of Soil analysis Part-2* (eds. Page AL, Miller RH, Kenny DR). American Society of Agronomy, Madison, WI.

Torsvik V, Ovreas L, Thingstad TF (2002) Prokaryotic diversity - Magnitude, dynamics, and controlling factors. *Science* **296**, 1064-1066.

Tourna M, Stieglmeier M, Spang A, *et al.* (2011) *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 8420-8425.

Treusch AH, Leininger S, Kletzin A, *et al.* (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environmental Microbiology* **7**, 1985-1995.

Tripathi BM, Kim M, Lai-Hoe A, *et al.* (2013) pH dominates variation in tropical soil archaeal diversity and community structure. *FEMS Microbiology Ecology* **68**, 303-311.

Tripathi BM, Kim M, Singh D, *et al.* (2012) Tropical soil bacterial communities in Malaysia: pH dominates in the equatorial tropics too. *Microbial Ecology* **64**, 474-484.

Tuomisto H, Ruokolainen K, Yli-Halla M (2003) Dispersal, environment, and floristic variation of western Amazonian forests. *Science* **299**, 241-244.

Tyson GW, Chapman J, Hugenholtz P, *et al.* (2004) Community

structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37-43.

Unno T, Jang J, Han D, *et al.* (2010) Use of barcoded pyrosequencing and shared OTUs to determine sources of fecal bacteria in watersheds. *Environmental Science and Technology* **44**, 7777-7782.

von Wintzingerode F, Gobel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews* **21**, 213-229.

Wagner HH (2004) Direct multi-scale ordination with canonical correspondence analysis. *Ecology* **85**, 342-351.

Wakelin SA, Macdonald LM, Rogers SL, *et al.* (2008) Habitat selective factors influencing the structural composition and functional capacity of microbial communities in agricultural soils. *Soil Biology and Biochemistry* **40**, 803-813.

Waldrop M, Balser T, Firestone M (2000) Linking microbial community composition to function in a tropical soil. *Soil Biology and Biochemistry* **32**, 1837-1846.

Walsh RPD, Newbery DM (1999) The ecoclimatology of Danum, Sabah, in the context of the world's rainforest regions, with particular reference to dry periods and their impact. *Philosophical Transactions of the Royal Society B-Biological Sciences* **354**, 1869-1883.

Wang J, Shen J, Wu Y, *et al.* (2013) Phylogenetic beta diversity in bacterial assemblages across ecosystems: deterministic versus stochastic processes. *The ISME Journal*.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**, 5261-5267.

Wang Y, Zhang Z, Ruan J, Wang Y, Ali S (1999) Investigation of actinomycete diversity in the tropical rainforests of Singapore. *Journal of Industrial Microbiology and Biotechnology* **23**, 178-187.

Ward NL, Challacombe JF, Janssen PH, *et al.* (2009) Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology* **75**, 2046-2056.

Wardle DA, Bardgett RD, Klironomos JN, *et al.* (2004) Ecological linkages between aboveground and belowground biota. *Science* **304**, 1629-1633.

Wessén E, Hallin S, Philippot L (2010) Differential responses of bacterial and archaeal groups at high taxonomical ranks to soil management. *Soil Biology and Biochemistry* **42**, 1759-1765.

Whitmore TC, Flenley JR, Harris DR (1982) The tropics as the norm in biogeography. *Geographical Journal* **148**, 8-21.

Will C, Thurmer A, Wollherr A, *et al.* (2010) Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. *Applied and Environmental Microbiology* **76**, 6751-6759.

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms - proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of*

the United States of America **87**, 4576-4579.

Yang J, Smith HG, Sherratt TN, Wilkinson DM (2010) Is there a size limit for cosmopolitan distribution in free-living microorganisms? A biogeographical analysis of testate amoebae from polar areas. *Microbial Ecology* **59**, 635-645.

Youssef NH, Ashlock-Savage KN, Elshahed MS (2012) Phylogenetic fiversities and community structure of members of the extremely halophilic Archaea (Order *Halobacteriales*) in multiple saline sediment habitats. *Applied and Environmental Microbiology* **78**, 1332-1344.

Zhalnina K, de Quadros PD, Camargo FA, Triplett EW (2012) Drivers of archaeal ammonia-oxidizing communities in soil. *Frontiers in Microbiology* **3**.

Zhang LM, Wang M, Prosser JI, Zheng YM, He JZ (2009) Altitude ammonia-oxidizing bacteria and archaea in soils of Mount Everest. *FEMS Microbiology Ecology* **70**, 208-217.

APPENDIX

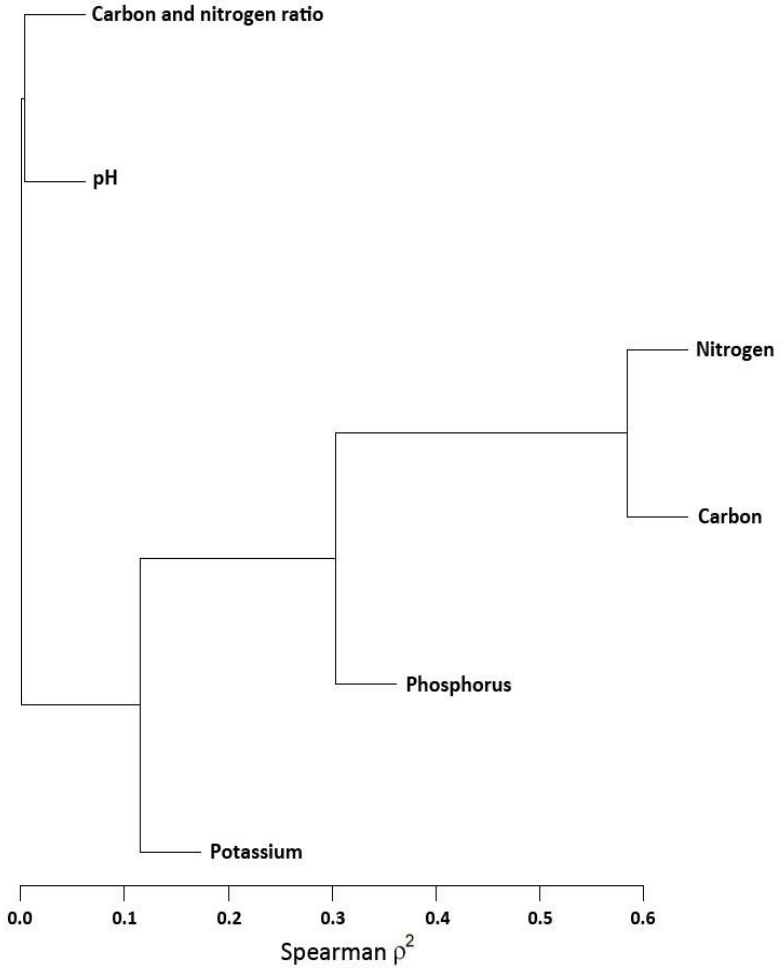

Figure A1. Cluster analysis of the measured environmental variables. The analysis was performed and plotted using varclus in the Hmisc R package.

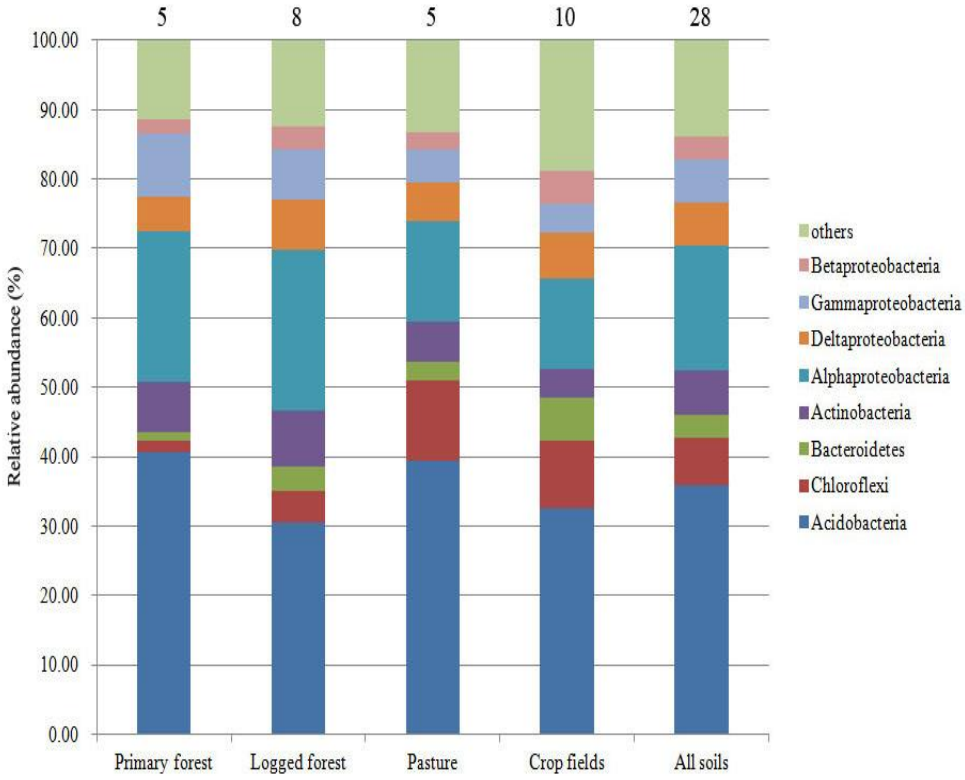

Figure A2. Relative abundances of dominant bacterial taxa in soils with different land-use categories and all soils combined. The numbers above the columns indicate the number of soils included in each category.

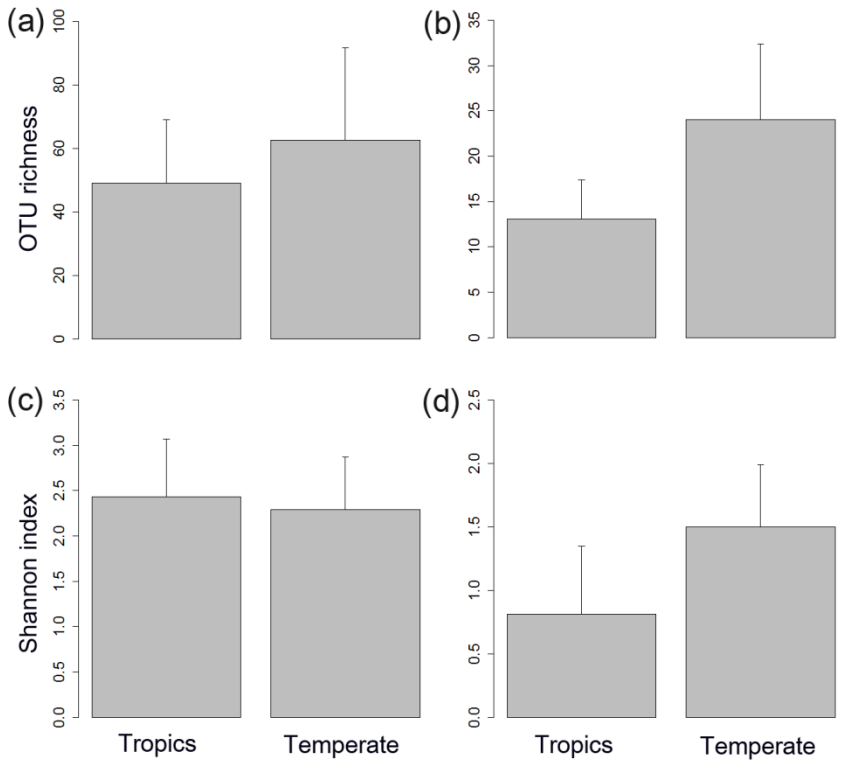

Figure A3. Archaeal OTU richness of (a) total archaea (16S rRNA gene based) and (b) AOA (*amoA* gene based); and Shannon diversity index of (c) total archaea and (d) AOA in tropical and temperate soils.

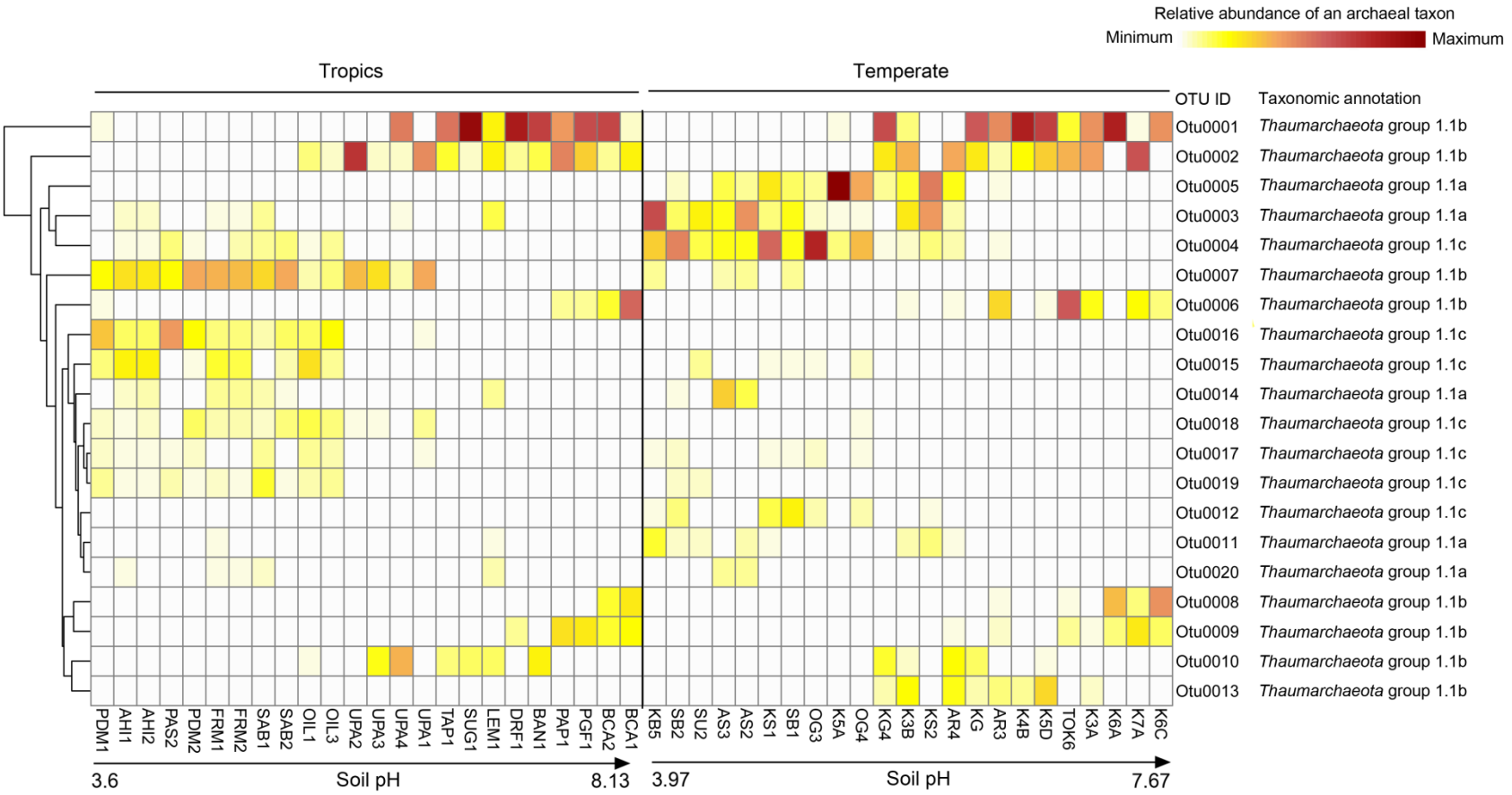

Figure A4. The heat map shows the relative abundances of the 20 most abundant archaeal OTUs in tropical and temperate biome against soil pH gradient.

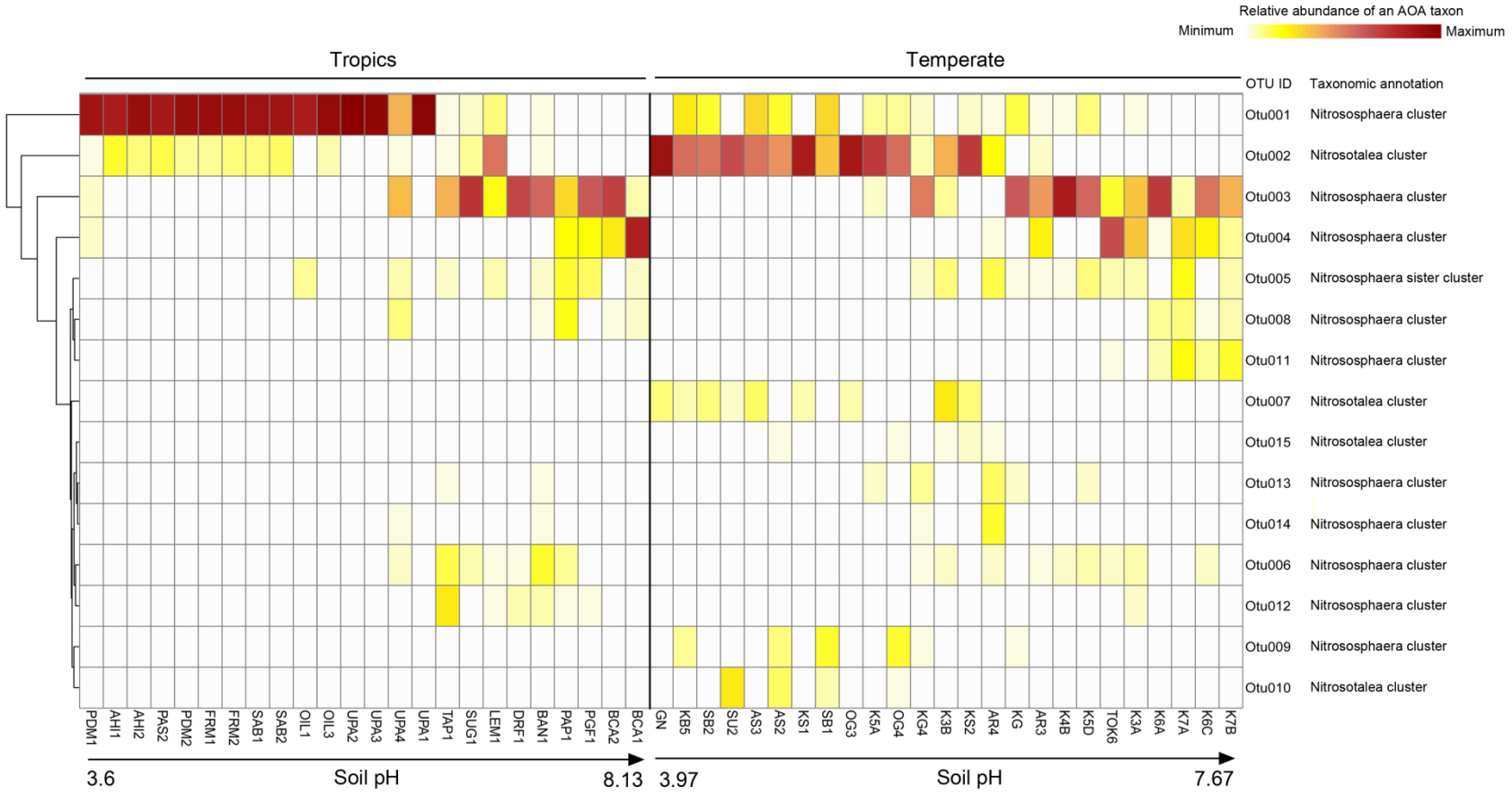

Figure A5. The heat map shows the relative abundances of the 15 most abundant AOA OTUs in tropical and temperate biome against soil pH gradient.

Table A1. Sites sampled at a local scale transect from FRIM forest reserve in Peninsular Malaysia.

Transect points	Latitude	Longitude	pH	TC (%)	TN (%)	CN ratio	Av. P (µg/g)	Ex. K (µg/g)	Clay (%)	Silt (%)	Sand (%)
TR02	N 03 14' 17.58"	E 101 37' 52.44"	6.43	1.46	0.18	8.11	16.70	40.1	17.30	4.30	78.16
TR03	N 03 14' 17.68"	E 101 37' 52.57"	6.45	1.55	0.41	3.78	5.87	63.9	29.05	6.40	64.40
TR04	N 03 14' 17.77"	E 101 37' 52.70"	6.36	1.56	0.33	4.73	4.98	65.0	23.10	5.68	71.04
TR05	N 03 14' 17.87"	E 101 37' 52.83"	3.89	2.40	0.16	15.00	16.50	75.6	32.76	5.22	61.92
TR06	N 03 14' 17.97"	E 101 37' 52.96"	3.97	2.24	0.24	9.33	16.50	69.4	39.86	5.97	54.03
TR07	N 03 14' 18.06"	E 101 37' 53.09"	4.27	3.72	0.33	11.27	4.90	97.3	44.20	6.53	49.17
TR08	N 03 14' 18.16"	E 101 37' 53.22"	4.45	2.88	0.36	8.00	4.83	85.4	34.47	5.86	59.51
TR09	N 03 14' 18.25"	E 101 37' 53.35"	4.32	1.82	0.18	10.11	16.20	60.8	35.65	6.44	57.82
TR10	N 03 14' 18.35"	E 101 37' 53.48"	4.58	1.88	0.20	9.40	15.60	45.3	36.90	6.95	56.05
TR11	N 03 14' 18.45"	E 101 37' 53.61"	4.15	1.88	0.20	9.40	23.30	53.2	37.35	6.12	56.44
TR12	N 03 14' 18.54"	E 101 37' 53.74"	3.76	3.04	0.28	10.86	16.00	54.5	43.48	7.62	48.78
TR13	N 03 14' 18.64"	E 101 37' 53.87"	3.64	3.90	0.28	13.93	16.70	71.7	39.48	6.82	53.59
TR14	N 03 14' 18.74"	E 101 37' 54"	3.49	3.42	0.23	14.87	38.20	68.3	44.22	5.85	49.84
TR15	N 03 14' 18.83"	E 101 37' 54.13"	3.60	2.94	0.27	10.89	14.90	72.2	37.98	7.31	54.59
TR16	N 03 14' 18.93"	E 101 37' 54.26"	4.03	2.56	0.26	9.85	6.94	56.7	42.31	5.14	52.40

Table A1. Continued.

TR17	N 03 14' 19.02"	E 101 37' 54.40"	4.05	2.12	0.27	7.85	14.9	61.6	37.28	6.11	56.50
TR18	N 03 14' 19.12"	E 101 37' 54.53"	3.61	3.19	0.21	15.19	17.4	66.2	33.89	6.97	58.97
TR19	N 03 14' 19.22"	E 101 37' 54.66"	4.04	2.42	0.24	10.08	3.4	56.8	35.43	5.05	59.34
TR20	N 03 14' 19.31"	E 101 37' 54.79"	4.03	2.46	0.22	11.18	16.6	44.8	30.15	7.30	62.40
TR21	N 03 14' 19.41"	E 101 37' 54.92"	3.92	2.64	0.17	15.53	15.4	65.4	20.10	5.68	74.04
TR22	N 03 14' 19.51"	E 101 37' 55.05"	4.02	2.13	0.20	10.65	26.4	69.2	37.87	5.04	56.94
TR24	N 03 14' 19.70"	E 101 37' 55.31"	4.29	1.58	0.18	8.78	24.0	50.1	42.84	4.99	51.03
TR25	N 03 14' 19.79"	E 101 37' 55.44"	3.83	1.76	0.16	11.00	69.2	56.0	46.30	6.23	47.27
TR26	N 03 14' 19.89"	E 101 37' 55.57"	4.05	1.74	0.17	10.24	14.8	54.2	33.47	4.86	61.51
TR27	N 03 14' 19.99"	E 101 37' 55.70"	4.14	1.56	0.29	5.38	16.5	53.6	38.17	5.96	55.78
TR28	N 03 14' 20.08"	E 101 37' 55.83"	4.04	2.70	0.29	9.31	47.8	50.8	38.43	4.05	57.34
TR29	N 03 14' 20.18"	E 101 37' 55.96"	3.50	2.17	0.22	9.86	14.9	56.1	19.30	4.30	80.16
TR30	N 03 14' 20.28"	E 101 37' 56.09"	3.95	1.89	0.22	8.59	15.5	51.0	37.65	4.44	57.82

TC = total carbon, TN = total nitrogen, CN ratio = carbon and nitrogen ratio, Av. P = available phosphorus, Ex. K = extractable potassium.

Table A2. Sites sampled at regional scale across Peninsular Malaysia and Northern Borneo.

Sites*	Lat.	Lon.	pH	TC (%)	TN (%)	CN ratio	Av. P (μg/g)	Ex. K (μg/g)	Alt.[a] (m)	MAP[b] (mm)	MAT[c] (℃)	Clay (%)	Silt (%)	Sand (%)
AHI1	N03 01′ 23″	E101 38′ 12″	3.65	2.21	0.12	18.4	26.3	24.2	54	2702	28	36.9	6.9	56.0
AHI2	N03 01′ 14″	E101 38′ 48″	3.65	2.21	0.12	18.4	26.3	24.2	52	2705	28	37.3	6.1	56.4
BCA1	N03 14′ 36″	E101 41′ 18″	8.13	11.5	0.62	18.6	17.1	117.4	60	2721	27.4	35.4	7.6	57.7
BCA2	N03 15′ 54″	E101 45′ 37″	8.07	5.61	0.25	22.4	34.4	50.5	65	2610	27.2	39.3	6.8	53.8
BTH2	N03 26′ 23″	E101 50′ 11″	5.36	11.9	0.28	42.5	19.2	120.3	1015	2628	25.2	11.0	38.8	50.0
CLF1	N04 27′ 16″	E101 35′ 15″	5.96	8.44	0.64	13.1	17.0	65.4	550	2902	18.1	22.3	44.6	33.0
CRA1	N02 24′ 29″	E101 51′ 18″	4.13	2.88	0.25	11.5	16.5	45.6	45	1940	26.4	19.3	30.8	49.6
CUF1	N04 29′ 13″	E101 21′ 39″	5.54	11.4	0.34	33.5	18.8	127.8	1955	2960	16.9	24.8	25.3	49.7
CUF2	N04 30′ 19″	E101 23′ 50″	4.20	4.90	0.48	10.2	20.9	120.3	1950	2959	16.9	10.0	15.7	74.0
FCF3	N03 43′ 00″	E101 46′ 40″	4.04	12.4	0.25	49.6	18.1	37.1	1489	2618	23.9	19.3	30.8	49.6
FRM1	N03 15′ 23″	E101 37′ 23″	4.15	1.88	0.20	9.4	23.3	53.2	48	2846	27.3	38.9	5.7	55.1
FRM2	N03 13′ 59″	E101 37′ 29″	4.32	1.82	0.18	10.1	16.2	60.8	40	2861	27.4	40.9	5.9	52.6
MAF1	N04 30′ 53″	E101 30′ 21″	4.91	2.06	0.27	7.6	20.2	111.4	1107	2934	17.6	18.1	17.2	64.4
PAS1	N02 57′ 02″	E101 30′ 01″	3.62	4.48	0.43	10.4	14.8	80.1	80	1690	25.3	13.6	12.2	74.1
PAS2	N02 59′ 22″	E102 18′ 34″	3.69	4.36	0.23	18.9	16.0	68.9	83	1759	25.4	23.6	10.0	66.2
PDM1	N02 30′ 00″	E101 52′ 00″	3.60	3.23	0.37	8.7	7.4	97.5	23	1862	26.4	13.9	11.0	73.8

Table A2. Continued.

PDM2	N02 30′ 54″	E101 52′ 46″	3.73	3.23	0.37	8.73	7.4	97.5	23	1854	26.4	14.0	11.4	74.3
RFM1	N04 26′ 12″	E101 24′ 52″	3.53	2.17	0.36	6.03	16.9	61.6	1412	2956	17.2	21.1	20.7	57.9
RFM2	N04 27′ 45″	E101 24′ 00″	3.98	3.94	0.78	5.05	40.1	106.5	1434	2959	17.2	5.9	40.6	53.4
SAB1	N04 26′ 11″	E117 58′ 44″	4.38	5.92	0.52	11.3	13.8	105.5	102	2087	26.6	21.0	20.7	57.9
SAB2	N04 22′ 25″	E117 55′ 41″	4.50	3.51	0.31	11.3	17.2	70.9	65	2091	26.7	22.4	44.5	33.0
DAN1	N04 57′ 53″	E117 49′ 16″	3.80	3.14	0.30	10.4	21.3	77.2	49	2820	26.8	19.2	31.2	49.6
TEM1	N04 56′ 59″	E117 48′ 25″	3.96	2.88	0.24	12.0	18.19	71.8	44	2828	26.8	28.0	38.4	33.6
TEM2	N04 57′ 32″	E117 48′ 14″	4.52	3.20	0.32	10.0	32.89	84.0	45	2828	26.8	28.0	38.4	33.6
DAN2	N04 58′ 13″	E117 48′ 57″	4.32	2.62	0.28	9.36	24.07	95.0	44	2830	26.8	39.2	26.4	34.4
TAK1	N05 01′ 41″	E117 47′ 14″	3.67	3.77	0.29	13.0	24.44	72.3	42	2863	26.8	32.8	27.2	40.0

*Unlogged forest sites: CLF = Cameron lower forest, CRA = Forest Hillslope Cape Rachado, CUF = Cameron uppermost forest, FCF = Fraser close forest, MAF = Mid altitude forest, PAS = Pasoh forest reserve, RFM = Robinson fall montane forest, SAB = Tawau hill forest reserve, TEM = Tembaling, TAK = Takala.

*Logged forest sites: AHI = Ayer Hitam forest reserve, BCA = Batu cave forest reserve, BTH = Bukit Japan tea house, FRM = FRIM forest reserve, , PDM = PD Meranti forest reserve, , DAN = Danum,

[a]Alt. = Altitude, [b]MAP = mean annual precipitation, [c]MAT = mean annual temperature.

Table A3. Soil samples and determined soil parameters for cross biome study of total archaea and AOA.

Sites	Lat.	Lon.	Biome*	pH	TC (%)	TN (%)	CN ratio	NH_4^+ (mgN/ kg soil)	NO_3^- (mgN/ kg soil)	Alt. (m)	MAT (℃)	MAP (mm)
AHI1	N03 01′ 24″	E101 38′ 12″	TRO	3.65	2.21	0.12	18.42	3.56	0.82	54.0	28.0	2702.0
AHI2	N03 01′ 14″	E101 38′ 48″	TRO	3.65	2.21	0.12	18.42	5.81	1.84	52.0	28.0	2705.0
BAN1	N03 10′ 09″	E101 33′ 37″	TRO	6.96	1.28	1.12	9.14	28.37	26.94	32.0	27.4	2772.0
BCA1	N03 14′ 37″	E101 41′ 19″	TRO	8.13	11.58	0.66	18.68	1.05	0.65	60.0	27.4	2721.0
BCA2	N03 15′ 54″	E101 45′ 37″	TRO	8.07	5.61	0.89	22.44	2.32	2.75	65.0	27.2	2610.0
DRF1	N03 58′ 51″	E101 43′ 25″	TRO	6.61	2.92	0.50	9.73	7.13	10.52	548.0	22.4	2781.0
FRM1	N03 15′ 23″	E101 37′ 24″	TRO	4.15	1.88	0.14	9.40	7.27	1.23	48.0	27.3	2846.0
FRM2	N03 13′ 59″	E101 37′ 29″	TRO	4.32	1.82	0.62	10.11	5.63	3.32	40.0	27.4	2861.0
LEM1	N03 11′ 55″	E102 14′ 58″	TRO	6.09	1.52	0.25	7.60	6.61	7.32	52.0	26.1	2142.0
OIL1	N02 52′ 25″	E101 34′ 28″	TRO	4.54	2.65	0.30	12.05	14.80	21.54	6.0	27.6	2612.0
OIL2	N02 43′ 10″	E101 38′ 42″	TRO	4.83	2.44	0.20	11.09	7.85	10.42	22.0	27.1	2343.0
PAP1	N02 47′ 09″	E102 20′ 19″	TRO	7.23	1.84	0.18	16.73	16.50	26.45	63.0	25.9	1930.0
PAS1	N02 59′ 22″	E102 18′ 34″	TRO	3.69	4.36	0.09	18.96	4.15	2.65	83.0	25.4	1759.0
PDM1	N02 30′ 00″	E101 52′ 00″	TRO	3.60	3.23	0.22	8.73	4.33	3.87	23.0	26.4	1862.0
PDM2	N02 30′ 54″	E101 52′ 47″	TRO	3.73	3.23	0.31	8.73	7.81	2.65	23.0	26.4	1854.0

Table A3. Continued.

SUG1	N02 38′ 25″	E101 41′ 38″	TRO	5.70	1.34	0.55	9.57	8.96	15.83	7.0	26.7	2132.0
TAP1	N02 02′ 10″	E102 43′ 10″	TRO	5.44	1.61	0.33	9.47	27.86	18.91	26.0	26.6	2350.0
UPM1	N02 59′ 31″	E101 29′ 25″	TRO	4.98	4.27	1.21	11.24	3.65	1.85	10.0	27.6	2671.0
UPM2	N03 16′ 35″	E102 10′ 11″	TRO	4.80	4.16	0.32	11.89	4.13	2.66	46.0	26.5	2375.0
UPM3	N02 47′ 09″	E102 20′ 33″	TRO	4.80	3.20	0.77	10.32	12.10	3.50	62.0	25.9	1930.0
UPM4	N02 57′ 49″	E101 43′ 54″	TRO	4.95	2.74	0.37	15.22	4.67	4.10	57.0	27.7	2656.0
PGF1	N03 00′ 40″	E101 36′ 16″	TRO	7.75	1.75	0.30	7.61	8.51	4.32	15.0	27.1	2328.0
SAB1	N04 26′ 11″	E117 58′ 44″	TRO	4.38	5.92	0.21	11.38	1.19	0.43	102.0	26.6	2087.0
SAB2	N04 22′ 25″	E117 55′ 41″	TRO	4.50	3.51	0.36	11.31	3.06	4.63	65.0	26.7	2091.0
GNF	N37 24′ 58″	E126 55′ 53″	TEM	3.97	1.00	0.29	11.08	12.50	4.30	163.0	11.9	1381.0
DOF1	N37 23′ 50″	E129 09′ 4″	TEM	6.70	2.72	0.74	12.36	10.80	8.17	141.0	12.4	1371.0
DOF2	N37 23′ 52″	E129 09′ 32″	TEM	4.91	3.24	0.66	10.46	22.40	6.02	138.0	12.4	1371.0
SHF1	N37 20′ 49″	E129 03′ 39″	TEM	6.03	3.66	0.20	12.21	15.10	2.15	197.0	12.4	1374.0
SIF1	N37 14′ 35″	E128 40′ 02″	TEM	4.72	1.51	0.46	7.19	20.20	3.01	545.0	11.7	1377.0
SIF2	N37 14′ 34″	E128 40′ 05″	TEM	6.24	4.28	0.62	11.90	18.90	8.60	557.0	11.7	1377.0
HAF1	N37 10′ 51″	E128 19′ 25″	TEM	7.08	5.06	0.22	9.19	104.10	2.58	280.0	10.8	1415.0
HAN2	N37 10′ 53″	E128 19′ 27″	TEM	7.45	10.03	0.22	30.39	90.70	1.72	267.0	10.8	1415.0
THF1	N37 03′ 47″	E128 18′ 05″	TEM	7.26	22.74	0.11	18.79	119.10	3.44	202.0	10.9	1403.0

Table A3. Continued.

THF2	N37 03′ 49″	E128 18′ 03″	TEM	7.67	11.36	0.23	35.49	194.80	2.58	192.0	10.9	1403.0
DEF1	N35 53′ 49″	E127 46′ 37″	TEM	4.09	7.38	0.37	9.58	37.00	5.59	622.0	12.3	1355.0
NAF1	N35 29′ 47″	E126 55′ 20″	TEM	5.56	4.59	0.37	12.41	26.20	20.60	159.0	13.2	1359.0
NAF2	N35 29′ 46″	E126 55′ 23″	TEM	4.87	3.65	0.23	12.58	22.80	12.90	156.0	13.2	1359.0
ASF1	N43 11′ 32″	E143 18′ 07″	TEM	5.88	14.50	0.52	12.94	18.79	20.58	351.6	5.9	782.0
ASF2	N43 11′ 40″	E143 18′ 07″	TEM	5.41	8.51	0.31	12.95	10.47	9.20	267.1	5.9	782.0
ASH1	N35 12′ 21″	E135 27′ 07″	TEM	4.37	11.76	0.72	13.18	6.84	14.53	677.1	10.0	2495.0
ASH2	N35 12′ 19″	E135 26′ 44″	TEM	4.35	6.22	0.88	12.39	1.90	8.24	679.5	10.0	2495.0
KUF1	N36 19′ 28″	E139 15′ 05″	TEM	4.41	13.27	0.30	17.89	13.18	1.43	978.0	9.0	1807.0
OHF1	N36 20′ 04″	E139 12′ 38″	TEM	5.29	11.92	0.14	17.97	3.77	15.78	815.0	9.7	1713.0
OGF1	N36 33′ 45″	E140 21′ 06″	TEM	4.66	9.13	0.17	19.93	24.71	0.01	661.0	10.7	1910.0
OGF2	N36 33′ 44″	E140 21′ 04″	TEM	4.87	9.34	0.69	15.06	18.34	3.26	626.0	10.7	1910.0
SHI1	N32 13′ 24″	E131 03′ 18″	TEM	4.44	9.35	0.38	12.95	6.11	6.83	1189	13.3	3472.0
SHI2	N32 14′ 07″	E131 06′ 07″	TEM	4.13	15.79	0.35	18.05	8.49	1.45	1100	13.3	3472.0
KIG1	N31 33′ 53″	E130 30′ 37″	TEM	4.26	4.48	0.31	14.98	0.11	0.07	1260	16.3	2569.1
TOK1	N27 24′ 43″	E128 33′ 47″	TEM	6.45	8.76	0.18	12.78	11.19	6.50	155.0	21.6	1896.8

*TRO = Tropical, TEM = Temperate