LINKING MICROBIAL COMMUNITY STRUCTURE TO ECOSYSTEM FUNCTION USING MICROBIOME ASSOCIATION MAPPING AND ARTIFICIAL ECOSYSTEM SELECTION

by

ANDREW H. MORRIS

A DISSERTATION

Presented to the Department of Biology
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2022

DISSERTATION APPROVAL PAGE

Student: Andrew H. Morris

Title: Linking Microbial Community Structure to Ecosystem Function Using Microbiome Association Mapping and Artificial Ecosystem Selection

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

Scott Bridgham	Chair
Brendan Bohannan	Advisor
William Cresko	Core Member
Matt Polizzotto	Core Member
Lucas de Carvalho Ramos Silva	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded March 2022

DISSERTATION ABSTRACT

Andrew H. Morris

Doctor of Philosophy

Department of Biology

March 2022

Title: Linking Microbial Community Structure to Ecosystem Function Using Microbiome Association Mapping and Artificial Ecosystem Selection

Microbiomes mediate a variety of important ecosystem functions. However, it remains unclear what attributes of the microbiome are important for determining the rate of ecosystem functions. Past attempts to elucidate this relationship have either looked too broadly at microbiome diversity or have assumed a priori that we know which taxa are limiting to the rate of function. To overcome this challenge, I borrowed strategies from population genetics including association mapping and artificial selection to robustly identify microbial markers of ecosystem function. I observed high heritability of methane oxidation rate in soil microbiomes demonstrating that variation in the microbial community can generate variation in ecosystem function independent of the environment. In addition, I characterized soil metagenomes along a land-use change gradient with increasing methane emissions. By looking agnostically across all microbial metabolic pathways, I identifed a surprising relationship between the relative abundance of nitrogen fixation genes and the rate of methane emissions. Using this conceptual framework to investigate biodiversity-ecosystem function relationships will deepen our understanding of microbiome function for ecosystem services and human health.

This dissertation includes previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Andrew H. Morris

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA
The Pennsylvania State University, State College, PA, USA
Cornell University, Ithaca, NY, USA

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2022, University of Oregon
Master of Science, Soil Science, 2017, The Pennsylvania State University
Bachelor of Science, Plant Sciences, 2014, Cornell University

AREAS OF SPECIAL INTEREST:

Microbial biodiversity-ecosystem function relationships

PROFESSIONAL EXPERIENCE:

Faculty, Juneau Icefield Research Program, 2018
Teaching Assistant, (Ecology and Evolution, Genetics and Molecular Biology, Cells), Department of Biology, University of Oregon, 2017-2018
Guest Instructor, Impacts of Changing Hydrology on Ecosystem Services in Glacial Systems, The Pennsylvania State University, 2017
Teaching Assistant, Soil Science, The Pennsylvania State University, 2017

GRANTS, AWARDS AND HONORS:

Elma Hendricks Scholarship, University of Oregon, 2018
William R. Sistrom Memorial Scholarship, University of Oregon, 2018
Oregon ARCS Scholar, University of Oregon, 2017

Distinguished Master's Thesis Award, The Pennsylvania State University, 2017

NSF Graduate Research Fellowship Award, The Pennsylvania State University, 2016

Annie's Sustainable Agriculture Scholarship, The Pennsylvania State University, 2016

Scarlet Graduate Fellowship in Watershed Stewardship Award, The Pennsylvania State University, 2015

Katherine Mabis McKenna Fellowship Award, The Pennsylvania State University, 2015

Hatch Multistate Grant, Cornell University, 2013

PUBLICATIONS:

Morris, A. H., Isbell, S. A., Saha, D., & Kaye, J. P. (2021). Mitigating nitrogen pollution with under-sown legume–grass cover crop mixtures in winter cereals. *Journal of Environmental Quality*, 50(2), 324–335.

Meyer, K. M., Morris, A. H., Webster, K., Klein, A. M., Kroeger, M. E., Meredith, L. K., Brændholt, A., Nakamura, F., Venturini, A., Fonseca de Souza, L., Shek, K. L., Danielson, R., van Haren, J., Barbosa de Camargo, P., Tsai, S. M., Dini-Andreote, F., de Mauro, J. M. S., Barlow, J., Berenguer, E., . Bohannan, B. J. M. (2020). Belowground changes to community structure alter methane-cycling dynamics in Amazonia. *Environment International,* 145, 106131.

Meyer, K. M., Hopple, A. M., Klein, A. M., Morris, A. H., Bridgham, S. D., & Bohannan, B. J. M. (2020). Community structure – Ecosystem function relationships in the Congo Basin methane cycle depend on the physiological scale of function. *Molecular Ecology*, 29(10), 1806–1819.

Morris, A., Meyer, K., & Bohannan, B. (2020). Linking microbial communities to ecosystem functions: What we can learn from genotype–phenotype mapping in organisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 375(1798), 20190244.

ACKNOWLEDGEMENTS

Funding for this work was provided in part by the National Science Foundation Graduate Research Fellowship Program (grant no. DGE 1255832) and the ARCS Foundation Florence and Mike Nudelman Scholarship. Data in Chapter 2 were generated with the permission of the Government of Gabon, Centre National de la Recherche Scientifique et Technologique (permit no. AR0035/14/MESRS/CENAREST/CG/CST/CSAR) and the assistance of F. Bivigou, H. Memiaghe, L. Tchignoumba, E. Tobi and I. Akendengue. The University of Oregon, the Gabon–Oregon Transnational Center on Environment and Development, the Smithsonian Conservation Biology Institute and Shell Gabon provided financial and logistical support for Chapter 2.

Thank you to my advisor Brendan Bohannan for always encouraging me to think more broadly. And thank you for allowing me the freedom to follow my creativity.

Thank you to Kyle Meyer for teaching me how to be a microbial ecologist. Your insights into the microbiome of soil and wine have been invaluable to my development as a scientist. Karen Adair was essential to the success of Chapter 3 and always has an answer to every question. She keeps the Bohannan Lab pirate ship afloat. And thank you to my committee – Scott, Bill, Lucas, and Matt – for your constant support and for encouraging me to take my time to explore my data.

To the members of the Bohannan Lab, the Bridgham Lab, and IE^2: Jess, Sophia, Claire, Shola, Max, Kayla, Caitlin, Beth, Becca, Felipe, Jarrett, Anya, Laura, Laurel, and everyone else, you helped make graduate school fun, from tea time to Rennie's to Microbe Brews. I will always remember this community. Sarah:

thank you for sharing a lab, an office, and a home. I am so grateful to have you as my academic twin.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Microbial communities (a.k.a., "microbiomes") mediate important ecosystem functions such as decomposition, greenhouse gas cycling, and carbon storage (Van Der Heijden, Bardgett, & Van Straalen, 2008). Despite the recent influx of genomic data from high-throughput sequencing, microbial ecologists have identified few general rules regarding how variation in microbiome attributes, such as diversity, taxonomic composition, or gene content, contribute to variation in the rate of nutrient cycles (Fuhrman, 2009; Hall et al., 2018). There is growing interest in managing microbiomes for specific ecosystem functions such as crop productivity, carbon sequestration, and greenhouse gas mitigation (Verstraete, 2007). However, achieving these goals requires an understanding of how microbiome variation contributes to variation in ecosystem function.

There is some evidence to suggest that microbiome variation may be important to ecosystem functioning. Many studies have compared ecosystem functions across microbiomes of different origins using common garden and reciprocal transplant experiments (Balser & Firestone, 2005; Cavigelli & Robertson, 2000; Reed & Martiny, 2007; Strickland, Lauber, Fierer, & Bradford, 2009). For example, one study observed differences in the rate and pathway of N_2O production when microbiomes from different soils were incubated under identical conditions, suggesting that microbiome variation contributed to variation in this ecosystem process (Cavigelli & Robertson, 2000). Another study reported that reciprocal transplants of intact soil cores between forest and grassland ecosystems showed functional differences in nitrogen pools and gas fluxes even when subjected to the same climatic conditions (Balser & Firestone, 2005). Similarly, when microbiomes

isolated from plant leaf substrates were re-inoculated onto sterile substrates, microbiomes decomposed their native plant litter faster than communities unfamiliar with that substrate (Strickland et al., 2009). This indicates that different microbiomes exhibit different functional rates under the same conditions. Clearly, different microbiomes are correlated with different rates and controls on ecosystem functions, but it is not yet clear which attributes of microbiomes contribute most strongly to this structure-function relationship.

One hypothesis is that an increase in the abundance of a particular microbial functional group results in an increase in the rate of the process mediated by that functional group (Rocca et al., 2015). The attributes of microbial functional groups (e.g., presence/absence, diversity, membership, etc.) can be inferred from 16S rRNA gene sequences based on taxonomic assignment, from metagenomes based on protein-coding gene annotations, or from the presence, absence, and/or sequence of genes that encode an enzyme involved in the function of interest. For example, the abundance of the methyl coenzyme M reductase (*mcrA*) gene, which encodes a subunit of an enzyme involved in CH_4 production has been shown to be correlated with CH_4 flux in some cases (Freitag & Prosser, 2009; Friedrich, 2005). The percent abundance of methanogens as inferred from the gene content of metagenomes has also been correlated with CH_4 production, for example, in arctic soil (Wagner, Zona, Oechel, & Lipson, 2017). However, measuring functional group abundance via marker genes has not proven very successful at revealing microbial community structure-function relationships since only a third of studies report a significant correlation between the abundance of a marker gene and its corresponding function (Graham et al., 2016; Rocca et al., 2015).

Though it is intuitive to consider that the abundance of a functional group should be related to the rate of the function it mediates, it is possible that other attributes of microbiomes may also be related to a given ecosystem function. For example, there are other functional groups that interact directly or indirectly with CH_4, such as ammonium-oxidizers, which have been reported to oxidize CH_4 (Bédard & Knowles, 1989). This relationship would not be identified by simply measuring the abundance of a marker gene for CH_4 oxidation, such as the gene *pmoA*, which codes for the primary enzyme involved in CH_4 oxidation in most methanotrophs (Kolb, Knief, Stubner, & Conrad, 2003). To fully understand microbiome structure-function relationships, we need to look more agnostically at broader microbiome variation as it relates to ecosystem function.

One approach to looking at broader microbiome composition is to perform a comparative survey of natural variation in an ecosystem function and the associated microbiome variation. This is analogous to surveying variation in an organismal trait and the associated genetic variation. In organismal biology, this is often accomplished via genome-wide association studies (GWAS) where many genetic loci are correlated with a complex trait (Bush & Moore, 2012). Generally, these studies take into account population structure, genetic architecture, and multiple hypothesis-testing in order to identify loci likely linked to the trait of interest (Bush & Moore, 2012; Price et al., 2006; Storey, 2002). An analogous approach has been used for identifying connections between the human microbiome and disease (Gilbert et al., 2016; Qin et al., 2012). However, there are comparatively few applications of this approach to non-host-associated microbiomes and ecosystem function (H.-X. Chang, Haudenshield, Bowen, & Hartman, 2017). Applying this framework to microbiome structure-function relationships while

accounting for ecological covariance structure could help reveal the mapping between microbiome structure and function. I explore this idea in more detail in Chapter 2 of this dissertation. This chapter was previously published with co-authors Brendan J. M. Bohannan and Kyle M. Meyer.

Another common approach for investigating the genetic basis of traits in organismal biology is through artificial selection (Conner, 2003). Microbiomes have been shown to respond to artificial, ecosystem-level selection on a variety of functions (Panke-Buisse, Poole, Goodrich, Ley, & Kao-Kniffin, 2015; Swenson, Arendt, & Wilson, 2000; Swenson, Wilson, & Elias, 2000). However, this approach has not yet been used to identify microbiome attributes (such as, microbial taxa, traits, or genes) likely linked to the trait under selection. This could be a useful approach to enrich for and identify members of the microbiome that are associated with an ecosystem function. I provide an example of the application of this approach in Chapter 3 of this dissertation. This unpublished work was co-authored with Brendan J. M. Bohannan.

Applying these ideas to a major challenge in ecosystem science could reveal new answers to classic questions. For example, a major area of research in ecology wants to address the question: Does the loss of biodiversity resulting from global change lead to a loss in ecosystem function? An important soil ecosystem function is the consumption of atmospheric CH_4 by tropical forest soils (Dirzo & Raven, 2003; Laurance, Sayer, & Cassman, 2014). Deforestation is progressing rapidly throughout the tropics and conversion of rainforest to agriculture flips these soils from a CH_4 sink to a CH_4 source (Meyer, Morris, et al., 2020; Steudler et al., 1996; Verchot, Davidson, Cattânio, & Ackerman, 2000). While there are many soil changes that result from deforestation, it remains unclear what drives this change

in ecosystem function. In Chapter 4, I apply the ideas from this dissertation about how microbiomes mediate ecosystem function to the problem of CH_4 and land-use change. I investigate a land-use change gradient and analyze the functional composition of the soil microbiome to understand what drives changes in CH_4 emissions in tropical soils. This unpublished work was co-authored with Brendan J. M. Bohannan and Kyle M. Meyer.

For my dissertation research, I have chosen to focus on the ecosystem function of soil CH_4 oxidation to deduce microbiome structure-function relationships. The primary reason for this choice is that results of previous research suggest that CH_4 oxidation is more readily explained by microbiome attributes than other components of the CH_4 cycle (Meyer, Morris, et al., 2020). In addition, CH_4 oxidation is one of the most highly phylogenetically-conserved functional traits of microorganisms and so offers the greatest potential for compositional changes to affect functional rate (A. C. Martiny, Treseder, & Pusch, 2013). Because of this trait conservatism, methanotroph ecology is well characterized relative to other microbial functional groups (Ho et al., 2013) and is thus an ideal model with which to study microbial community structure-function relationships. My ultimate goal, however, is to develop and test robust approaches that apply beyond CH_4 oxidation and that could potentially reveal any microbial structure-function relationship.

In addition to the ecological reasons for focusing on CH_4 oxidation, there are broader reasons in the context of global change. Methane is the second most important greenhouse gas after CO_2 and represents 16% of global greenhouse gas emissions (IPCC, 2013). Global wetlands have the greatest uncertainty in CH_4 emissions of any source and CH_4-consuming microbes are the primary biological sink for CH_4 (Kirschke et al., 2013; Saunois et al., 2016). In addition, since

2006, the proportion of microbial CH_4 sources has been increasing relative to other sources (Schaefer et al., 2016). Therefore, not only is microbial community structure potentially an important driver of CH_4 oxidation, but understanding the CH_4 cycling community may have a disproportionate effect on our ability to predict future CH_4 emissions. The ultimate goal of my dissertation research is to contribute to the ability of scientists to predict CH_4 emissions by developing novel approaches to identifying relationships between microbiome variation and variation in CH_4 cycling in soil.

CHAPTER II

LINKING MICROBIAL COMMUNITIES TO ECOSYSTEM FUNCTIONS: WHAT WE CAN LEARN FROM GENOTYPE–PHENOTYPE MAPPING IN ORGANISMS

The main ideas in the manuscript were my own resulting from discussions with co-authors. I assisted with molecular laboratory work, performed the data analysis and drafted the manuscript; Kyle M. Meyer collected field data, performed molecular laboratory work and critically revised the manuscript; Brendan J. M. Bohannan conceived of the study, designed the study and helped draft the manuscript.

From Morris, A., Meyer, K., & Bohannan, B. (2020). Linking microbial communities to ecosystem functions: What we can learn from genotype–phenotype mapping in organisms. Philosophical Transactions of the Royal Society B: Biological Sciences, 375(1798), 20190244.

2.1 Introduction

Ecology is broadly focused on understanding biodiversity and how that biodiversity shapes the ecosystems that humans depend on. Many ecosystem processes essential to all of life are mediated by microorganisms and therefore understanding the relationship between microbial biodiversity and ecosystem function is important (Schimel & Gulledge, 1998; Singh, Bardgett, Smith, & Reay, 2010). Certain ecosystem functions are correlated with microbial diversity, indicating that we should be able to determine what aspects of microbial biodiversity influence ecosystem function. However, attempts to describe that mapping have borne little fruit (Graham et al., 2016; Rocca et al., 2015). We argue that to overcome this challenge we should look to other successful attempts at

mapping biological variation onto higher order processes. In particular, population genetics and the process of genotype-phenotype mapping provide a number of potentially useful insights. For example, genotype-phenotype mapping often makes few assumptions about the nature of the map, i.e. it is 'agnostic.' In addition, population geneticists have developed rigorous methods for reducing potentially confounding relationships such as geographic structuring of populations. Finally, the ultimate goal of genotype-phenotype mapping is to identify the unique contribution of genotype to phenotype separately from other drivers of phenotypic variation such as the environment. Inspired by this, we suggest a reframe of the question, 'is microbial biodiversity related to ecosystem function?' to, 'what is the unique contribution of the microbiome to ecosystem function independent of the environment?'

2.1.1 There is evidence that microbial biodiversity matters for ecosystem function. There is increasing evidence for microbial biodiversity-ecosystem function relationships. For example, there are positive correlations between microbial diversity and ecosystem multifunctionality for a variety of ecosystems and most major lineages of microorganisms (Delgado-Baquerizo et al., 2016, 2017; Jing et al., 2015). Adding microbial diversity or abundance to ecosystem models can in some cases improve model accuracy (Graham et al., 2016). Contrived communities that vary in richness and communities created through sequential dilution or varying filter size to generate differences in diversity can also exhibit differences in ecosystem function (Maron et al., 2018; Philippot et al., 2013; Schnyder, Bodelier, Hartmann, Henneberger, & Niklaus, 2018; Wagg, Bender, Widmer, & van der Heijden, 2014). Finally, reciprocal transplant and common garden experiments that manipulate the connection between

community composition and environment reveal differences in ecosystem function for communities of distinct origins (Cavigelli & Robertson, 2000; Glassman et al., 2018; Strickland et al., 2009). Given these relationships, many investigators have now moved on to the challenge of identifying the aspects of microbial biodiversity (e.g. specific taxa, genes, functional groups, etc.) that influence a given ecosystem function; however, this has proven especially challenging.

2.1.2 The mapping between microbial biodiversity and ecosystem function has been elusive. Most studies that attempt to identify the aspects of microbial biodiversity that influence a given ecosystem function focus on 'functional' gene or transcript abundance. In this case, qPCR or shotgun metagenomic sequencing is used to estimate the abundance of a gene or transcript that is a putative marker for a microbial process (and thus a marker for the functional group that performs that process). For example, the gene *mcrA*, which encodes a subunit of the enzyme that performs the final step in methanogenesis, is commonly used as a marker for methanogenesis and for the methanogen functional group. Other examples include *pmoA* and methanotrophy, *nifH* and nitrification, and *nosZ* and denitrification. It is often hypothesized that the abundance of these markers is predictive of the rate of the associated processes (for example, it is hypothesized that the abundance of *mcrA* is related to the rate of methanogenesis).

Some ecosystem functions in certain ecosystems can be predicted from the abundance or transcriptional activity of genetic markers for those functions. For example, soil methane production and consumption can under some circumstances be predicted from the genetic markers *mcrA* and *pmoA* (Freitag & Prosser, 2009; Freitag, Toet, Ineson, & Prosser, 2010; McCalley et al., 2014). However, for most ecosystem functions, the abundance of a functional gene or transcript is rarely

positively correlated with the rate of the corresponding process. The cases where there is a positive correlation tend to be restricted to agricultural ecosystems and certain functions within the nitrogen cycle(Rocca et al., 2015). In general, including aspects of microbial biodiversity (e.g. functional gene abundance or diversity) improves models of ecosystem function less than one-third of the time and increases variance explained by an average of only eight percentage points over environmental variables (Graham et al., 2016).

2.2 Genotype-phenotype mapping as a source of inspiration

In the approaches described above, microbial ecologists often use microbiome data to infer taxonomic composition, essentially creating species lists from data such as 16S rRNA marker genes or shotgun metagenomes. Interpreting microbiome data in this way has allowed us to use approaches from biodiversity-ecosystem function research (which are often focused on taxonomic or functional groups), but it has generally not been useful for creating more detailed descriptions of the relationship between microbial biodiversity and ecosystem function. But this approach is not the only way we could determine the relationship between a complex set of highly variable data and an aggregate function.

This kind of 'many-to-one' mapping is analogous to the challenge of identifying the genetic basis of complex traits in organismal populations. In such 'genotype–phenotype' mapping studies, a population exhibits variation in a phenotype (e.g. height or disease state) as well as variation in potentially thousands of single-nucleotide polymorphisms (SNPs). To identify the genetic basis for a trait, investigators sample from this population and correlate phenotype with genotype. While some phenotypes (e.g. the propensity for diseases such as Parkinson's) are controlled by a single locus (Kerem et al., 1989; MacDonald et al., 1992), most

traits depend on a large number of genes that control variation in phenotype (Hindorff et al., 2009; Reich & Lander, 2001). In addition, there is often no a priori expectation about which regions of the genome control that trait so we must search for genetic markers throughout the genome. If a marker is significantly correlated with the phenotype of interest, this either indicates it is inside a gene with a direct or indirect effect on phenotype or that it is in linkage disequilibrium (i.e. non-random association between two alleles) with a causal gene.

There are a number of parallels between this challenge faced by organismal biologists and that facing microbial community ecologists. They both involve large numbers of statistical comparisons. Both are attempting to identify causal relationships that are potentially confounded by complex patterns of covariation. There is often no strong expectation about which entities (i.e. which genomic regions or which microbial genes or lineages) are most likely to be causally related to phenotype or function, and thus 'agnostic' approaches are needed. For some ecosystem functions, it is possible that a single taxon could substantially influence its rate. For example, methane flux from permafrost in Sweden may be controlled by a single taxon (McCalley et al., 2014). But for most ecosystem functions, there could be many taxa of small effect that contribute to the rate of ecosystem function. Finally, both ultimately require manipulation (of genes or taxa) to establish causation.

2.2.1 The importance of a taxonomically 'agnostic' approach.

Most microbial biodiversity-ecosystem function research up to this point has used an approach analogous to that used by plant ecologists studying biodiversity-ecosystem function relationships. This approach is to measure or manipulate the diversity of a taxonomic group (e.g. plants) and look for an association with

the function performed by that group (e.g. primary productivity). We can think of plants as a 'functional group,' i.e. a group of taxa united by their ability to perform a particular ecosystem function. For microbes, estimating functional group abundance can be much more challenging. From a small number of cultured isolates, we have a provisional understanding of which microbes might be involved in some ecosystem functions. By sequencing the genomes of these isolates, we have identified genetic markers for certain functions, which we call 'functional genes.' But most microbial taxa remain uncultured and we do not know the function of most microbial taxa detected in environmental samples (Hug et al., 2016; A. C. Martiny, 2019). In addition, there have been recent discoveries of functions in unexpected taxonomic groups, for example methanogenesis by fungi and cyanobacteria, a function previously considered restricted to archaea (Bižić et al., 2020; Lenhart et al., 2012).

As stated earlier, these functional markers are not correlated with their corresponding ecosystem function in most ecosystems and for most processes. In addition, they provide little explanatory power to the models of ecosystem function. Because of this, it might be prudent to look more agnostically at microbial communities to identify taxa, groups of taxa or genes that are important for predicting the rates of ecosystem functions rather than assuming that the genetic markers we have provisionally identified for a given function represent the most likely taxa or genes involved. This agnostic approach is analogous to the approach of many genotype–phenotype mapping studies (e.g. genome-wide association studies), which often look for associations between a phenotype and loci anywhere in the genome.

Beyond finding new physiologies in unexpected lineages, there are other reasons for looking agnostically. In the case of microbial functions, it may be that the organism that performs a function is not the limiting factor for the rate of that function. For example, the rate of soil-to-atmosphere methane flux could be limited by methanogens or methanotrophs or the balance of the two. However, it could also be limited by the bacteria that produce the fermentative byproducts that methanogens use as substrates. Or there could be indirect limitation by organisms that liberate nitrogen or phosphorus into mineral forms. In other words, the influence of microbial communities on the rate of ecosystem function could represent a complex metabolic network much like the regulation of gene expression in organisms that partially determines their phenotype. These broader patterns of biodiversity-ecosystem function relationships would be invisible to any study that solely focuses on the most relevant functional group without considering the possible influence of other taxa.

2.2.2 Controlling for population stratification. It is widely accepted that organisms, including microorganisms, exhibit population stratification due to geographic and environmental separation (J. B. H. Martiny et al., 2006; Wright, 1943). This can lead to spurious associations between phenotypes and genetic markers that are at high frequency in isolated sub-populations. Association studies generally control for population stratification by accounting for shared ancestry among organisms in a population when modelling the connection between genotype and phenotype. Typically, microbial biodiversity-ecosystem function studies do not account for population stratification (i.e. community similarity among ecosystems), although there are some exceptions (Lloyd-Price et al., 2019; Meyer, Hopple, et al., 2020; Qin et al., 2012). Community similarity (the

community analogue of shared ancestry among organisms) is not as tightly linked to geography or environment as is shared ancestry. Therefore, it could be useful to account for these separately in microbial studies, particularly if we are interested in quantifying the effect of microbial communities on ecosystem function independent of these other factors.

Genome-wide association studies correct for stratification using a variety of methods. Generally, they ignore the underlying environmental and spatial distance between samples and instead use shared ancestry as a proxy for local selection and assortative mating. A common approach is to perform a regression of phenotype and shared ancestry (computed as the first one or more principal components of a genotype matrix) and then use the residuals from this model as the values for phenotype in a subsequent regression using the genotypes directly (Price et al., 2006). This principal component correction is designed to test the effect of individual genes after removing the effect of shared ancestry among individuals. Another approach, employed in our example, is variance component modelling (or mixed modelling, hierarchical modelling, etc.), where genotypic similarity is included as a covariate in the model to control for stratification while testing the genotype–phenotype connection(Kang et al., 2010).

If we control for covariates such as community similarity, geographic distance or environmental similarity, it changes the nature of our question. For example, if we test the correlation between the relative abundance of a taxon and the rate of methane flux, we are asking 'is this taxon correlated with methane flux?' If we find a significant result, that may be because variation in the abundance of that organism directly or indirectly contributes to methane flux. However, it might also be that that organism lives only in ecosystems that happen to have a

high rate of methane flux. In this scenario, we are unable to distinguish between these possibilities. However, if we add environmental variables or environmental similarity as a covariate in our model, we can ask, 'Is this taxon uniquely associated with function in a way that it is independent of the environment?' By 'uniquely associated', we mean those taxa associated with the function irrespective of environmental conditions, local community structure or spatial proximity. This slight reframing of the question could be especially rewarding for microbial biodiversity-ecosystem function research, particularly as it relates to incorporating microbial community data into ecosystem models. Finally, it is interesting in its own right to understand whether microbial communities are selected by the underlying environmental conditions to produce a particular rate of ecosystem function or whether community variation has functional consequences independent of the environment.

2.3 An example: high-affinity methane oxidation

To illustrate the ideas outlined above, we reanalysed a subset of previously published data from a paper that has demonstrated one successful approach for applying genotype-phenotype mapping to microbial communities (Meyer, Hopple, et al., 2020). In our reanalysis, we do not intend to challenge the conclusions of that paper, but instead we want to demonstrate how to perform this type of study for microbial ecologists unfamiliar with association studies. A full description of the study design, samples and data generation can be found in that article. Briefly, these data were gathered from intact soil cores taken from diverse ecosystems of the Congo Basin in Gabon, Africa. Cores were incubated in the laboratory under different concentrations of methane to identify the rates of specific methane cycling pathways. In this example, we analyse data from just one of these pathways,

high-affinity methane oxidation (the oxidation of atmospheric concentrations of methane), which we will refer to simply as 'methane oxidation'. In addition, we only include amplicon sequences from the DNA-inferred community and not the RNA-inferred community, both of which are presented in the original paper (Meyer, Hopple, et al., 2020). The data we analysed include methane oxidation rates, amplicon sequence variants (ASVs) generated using the 'DADA2' pipeline and inferred from unique 16S rRNA gene sequences (Callahan, McMurdie, & Holmes, 2017), pmoA abundance estimates (via qPCR), and four environmental covariates (soil moisture, bulk density, carbon and nitrogen).

Analyses were conducted in the 'R' statistical environment using the 'phyloseq' package (McMurdie & Holmes, 2013; R Core Team, 2018). The relative abundances of ASVs were corrected using the variance stabilizing transformation from 'DESeq2' (Love, Huber, & Anders, 2014; McMurdie & Holmes, 2014). We first tested the correlation between ecosystem function and typical measures of microbial community structure: functional gene abundance and community richness, which were estimated using the 'breakaway' package (A. Willis & Bunge, 2015). We then tested covariation between community structure (estimated as Bray–Curtis distance using 'vegan'), environmental variation (Euclidean distance) and geographic distance (Euclidean distance) using Mantel tests (Bray & Curtis, 1957; Oksanen et al., 2019). Finally, we identified taxa that were significantly associated with function independent of the environment by fitting variance component models using 'varComp' to test the relationship between relative abundance of each ASV and methane oxidation rate (Kang et al., 2010; Qu, Guennel, & Marshall, 2013). To illustrate how including different covariates (environmental, geographic and community) can result in different conclusions about which taxa are associated

with function, we fitted this model with and without random effects variance components for environmental similarity, geographic site ID and Bray-Curtis similarity. Significant taxa were determined by controlling the false discovery rate at q-value < 0.05 (Storey, 2002). Figures were created using 'ggplot2' (Wickham, 2016). All raw data and scripts required to recreate this analysis are available in the electronic supplementary material.

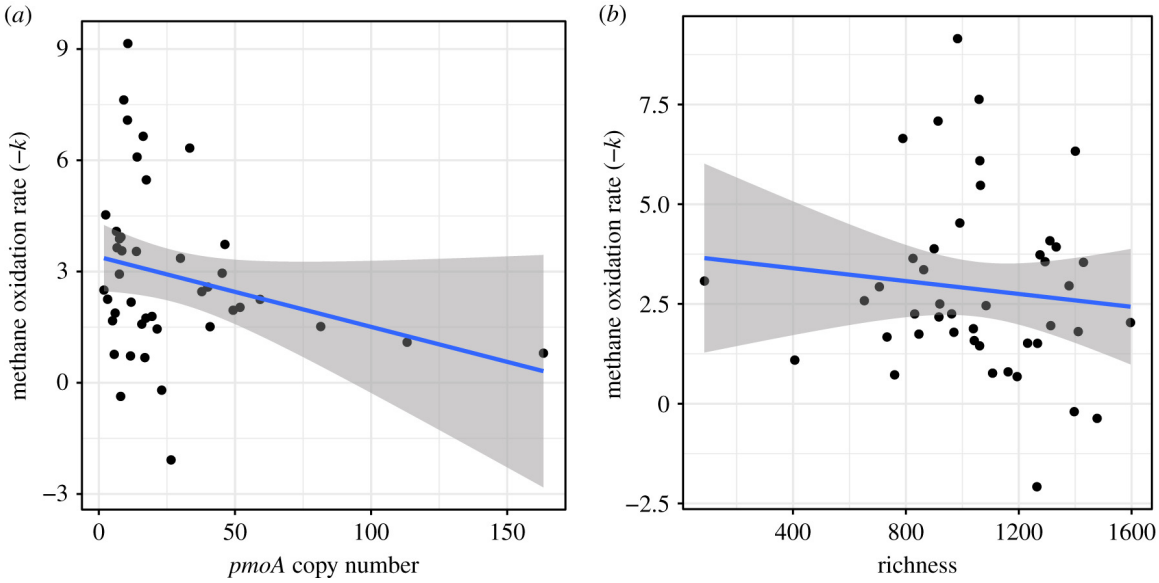

Figure 1. Methane oxidation rate is not correlated with functional gene abundance or ASV richness. Correlations between methane oxidation rate and (a) abundance of the functional gene pmoA (n = 42), and (b) ASV richness (n = 44). Lines represent the ordinary least squares regression lines with standard errors.

2.3.1 Results and Discussion. Microbial biodiversity-ecosystem function studies typically test functional group abundance or community richness as it relates to ecosystem function. In our case, methane oxidation rate was not significantly correlated with *pmoA* gene abundance or 16S rRNA gene-based taxonomic richness (table 1 and Figure 1). To demonstrate that the covariance structure of the data might alter our conclusions about which taxa regulate ecosystem function, we tested collinearity between each pair of distance matrices

for community, environment and geography. We found a moderate and significant correlation between community composition and environmental variation, geography and community composition, and geography and environmental variation (table 2 and Figure 2). To visualize this population stratification, principal coordinate plots show that beta diversity of samples separated by site ID and by ecosystem type (wetland or upland; Figure 2), which indicates substantial spatial and environmental structuring of microbial populations. This suggests that the presence or abundance of certain taxa will be elevated in specific ecosystems. In this case, high-affinity methane oxidation is typically greater in upland ecosystems than in wetland ecosystems and so any taxa differentially abundant in uplands will tend to be correlated with methane oxidation regardless of their involvement in that process. It is necessary to control for this stratification to rigorously identify associations between taxa and function.

Table 1. Estimates for the linear relationship between methane oxidation rate and two measures of microbial community structure: *pmoA* functional gene abundance and ASV richness.

term	estimate	s.e.	t-statistic	p-value
pmoA copy number	0.019	0.011	1.705	0.096
richness	0.001	0.001	0.694	0.491

Table 2. Mantel tests for each pair of dissimilarity matrices. Community distance matrix was based on Bray-Curtis distance while both environment and geography distance matrices were based on Euclidean distance. p-values were determined by permutation test with 999 permutations.

terms	Mantel statistics (r)	95% upper quantile of permutations	p-value
community $\sim$ geography	0.353	0.056	0.001
community $\sim$ environment	0.474	0.109	0.001
geography $\sim$ environment	0.241	0.055	0.001

To demonstrate this approach, we tested the effect of the relative abundance of each ASV on methane oxidation rate while controlling for different sets of covariates including environment, geography and community. After controlling the false discovery rate, 460 unique ASVs were identified as significantly correlated with function when no covariates were included in the model. We found the different numbers of taxa significantly associated with methane oxidation depending on which covariates were included in the model (table 3). Each of these sets of taxa represents different versions of the biodiversity-ecosystem function mapping question. For example, by attempting to control for environmental variation statistically, we can identify taxa whose traits may contribute to variation in function that is independent of environmental conditions. Similarly, by controlling for geographic distance among samples, we can reduce the likelihood that the taxa we identify are only related to function because of an association with unmeasured environmental variation that is spatially structured or because of differences in dispersal history among sites. In the model that controlled for all three covariates (community, environment and geography), only six ASVs were significantly correlated with methane oxidation rate (Figure 3). These taxa could be useful indicators of methane oxidation rate across space and different ecosystems. Researchers could elaborate on these findings using targeted cultivation and manipulative experiments to further understand their contribution to methane oxidation.

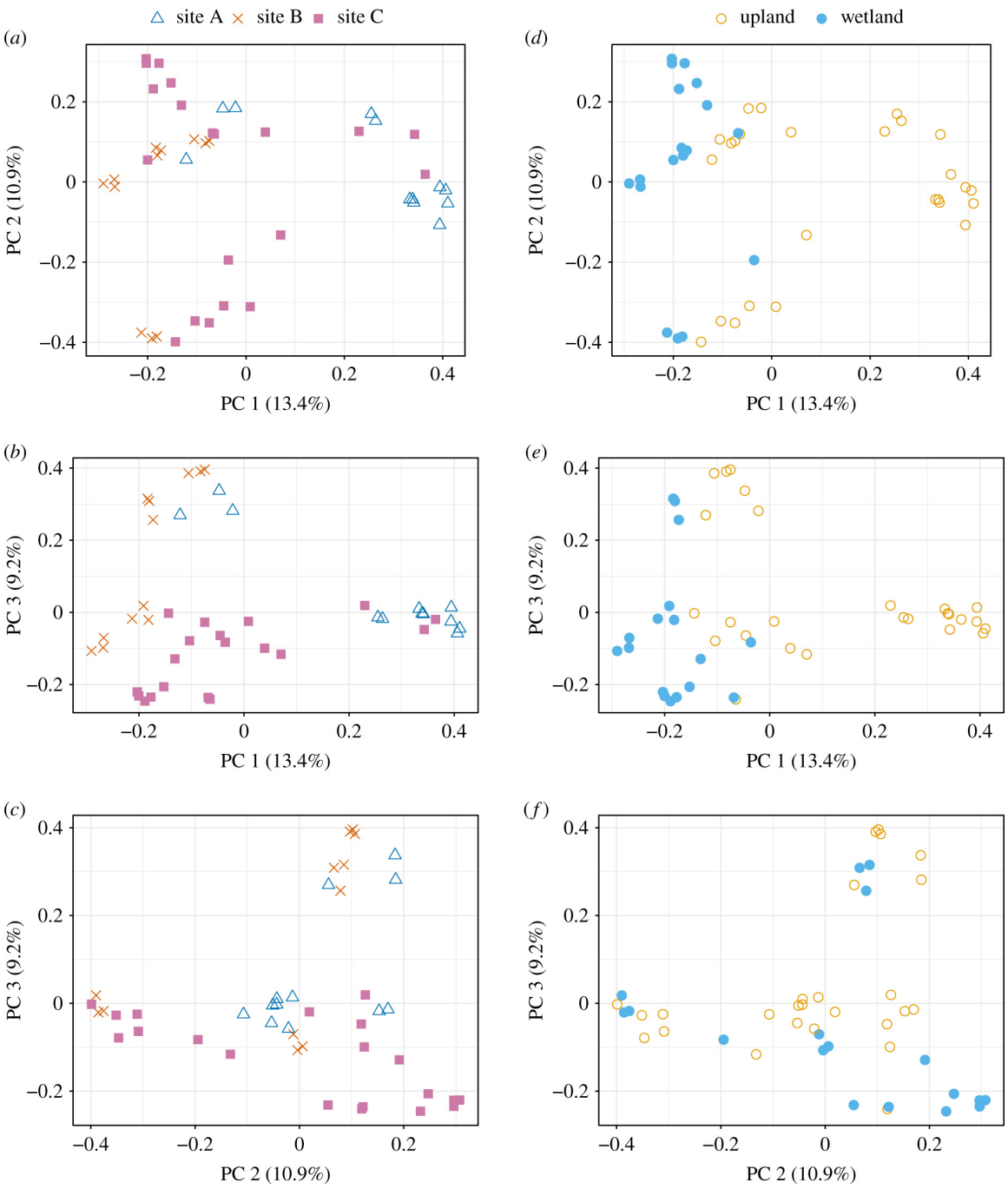

Figure 2. Microbial community composition is spatially and environmentally structured. Principal coordinate plots of Bray–Curtis distance representing the first three axes of community composition. In (a–c) points are identified by site ID, and in (d–f) points are identified by wetland or upland ecosystem. All four environmental covariates separate strongly by wetland/upland. Axis length is proportional to variance explained as indicated in parentheses. PC, principal coordinate.

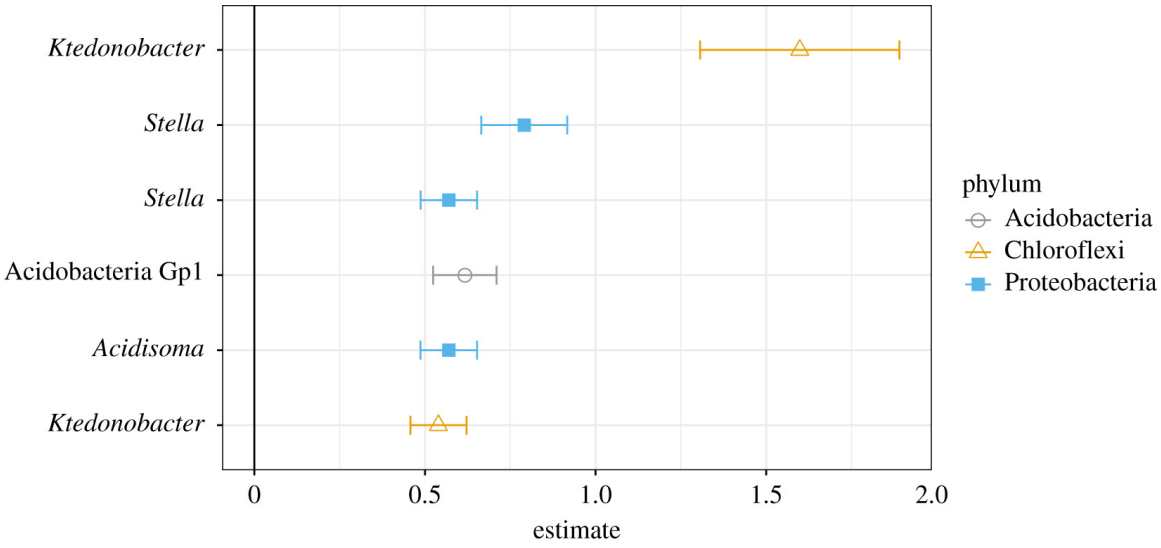

Figure 3. Taxa associated with methane oxidation rates after controlling for geographic location, environmental similarity and community composition. Points are estimates for the linear relationship between the relative abundance of a single ASV and methane oxidation rate with standard errors from variance component models including similarity matrices as covariates for community and environment and site ID for geographic location. Amplicon sequence variants are labelled at the finest resolution available: genus for all except the Group 1 Acidobacteria. Points are identified by phylum. Significant taxa were determined by controlling the false discovery rate at q-value <0.05.

Table 3. Number of significant taxa after including each set of covariates in a variance component model. 'removed' and 'added' columns are relative to the no-covariate model. Significance was determined by controlling the false discovery rate at q-values <0.05.

term(s)	removed	added	significant
none	0	0	460
geo	338	21	143
com	460	0	0
env	281	1	180
geo + com	458	0	2
geo + env	377	13	96
com + env	447	0	13
geo + com + env	454	0	6

Notably, these six taxa fall into three genera and one class with cultured representatives that are not known to consume methane (Belova, Pankratov, Detkova, Kaparullina, & Dedysh, 2009; Y.-j. Chang et al., 2011; Domeignoz-Horta, DeAngelis, & Pold, 2019; Fritz, Strömpl, & Abraham, 2004). These taxa could be related to ecosystem function in multiple ways. The most interesting possibility is that each of these taxa is statistically related because it is causally connected to the function. This could be direct—for example, an organism that consumes methane—or indirect—for example, an organism that regulates substrates necessary for methane oxidizers. Alternatively, a significant association could occur for non-causal reasons. For example, any organism that tends to be in high abundance where methane oxidation rates are high would be correlated with methane oxidation, even if it has no causal relationship. This could be because such an organism is favoured under the same conditions that favour methane oxidation. Such covariation can drive associations that are not causal, but the effects of covariation would have been reduced by controlling for covariates in our tests.

2.4 Caveats and future directions

Once taxa have been identified with an association test (such as the one we outline above), there are multiple ways they could be used for future study. One approach common in genetics, particularly for markers of genetic disorders, is to generate a polygenic score based on the summed effect of many genes on a phenotype of interest, such as the probability of developing a disorder. A similar aggregate bioindicator could be generated for ecosystems that would summarize the probability of the rate or occurrence of a particular ecosystem function. This would be accomplished by measuring the abundance of the taxa identified in an association study and determining their association with the rate of an ecosystem

function. Alternatively, the identified taxa could be incorporated into a structural equation model in an attempt to better understand the individual effects and interactions among taxa as they contribute to the rate of ecosystem function (Grace, Anderson, Olff, & Scheiner, 2010). This might give an indication of the relative importance of different taxa as compared with other factors, such as environmental variables, and also identify underlying latent variables that explain variation in ecosystem function.

Ultimately, the relationships identified in any comparative mapping study must be verified. For genotype-phenotype studies in organisms, there are multiple ways that this verification is accomplished. In some cases, organisms can be artificially selected for a particular phenotype (e.g. through experimental evolution in an environment that favours the phenotype of interest) and the genetic changes that occur in response to selection can be compared with those identified via mapping studies. An analogous approach for microbial biodiversity-ecosystem function studies would be to apply artificial ecosystem selection (*sensu* Swenson, Wilson, and Elias (2000)) on a given function and compare the taxa (or genes) that change in response to selection with those identified via a comparative approach (such as the one illustrated in our example).

The most common way that loci identified in a genotype-phenotype mapping study are verified is through manipulative genetics. The identified loci can be knocked out or over-expressed and the effect on phenotype compared with that predicted from mapping studies. In the case of microbial biodiversity studies, it may be possible to inhibit a particular functional group through the use of specific antimicrobial or chemical inhibitors or using phages that exhibit high host-specificity (Koskella & Meaden, 2013; Maxson & Mitchell, 2016),

but this is not generally possible. In some cases, we may be able to isolate a microorganism of interest in pure culture and add it back to an ecosystem, transiently increasing its abundance (roughly analogous to 'overexpressing' a gene). A greater focus on culture-based approaches could increase the success of these kinds of microbial enrichments. Finally, synthetic communities (contrived assemblages of microorganisms) may be the most powerful way to test hypotheses about microbial biodiversity-ecosystem function relationships, but currently these approaches are limited by the small number of taxa that can be routinely cultured from most environments (but see Schnyder et al. (2018) and Berg and Koskella (2018)).

There are a number of limitations to the biodiversity-ecosystem function mapping approach we describe, some in common with organismal mapping studies and others unique. For example, simple linear models such as the variance component model used in this study are typical for genetics studies, but may not be the best way to identify correlations for microbiomes because of the unique challenges of microbial data. Marker gene and metagenome sequences are inherently compositional, reads are often absent from most samples (i.e. they are zero-inflated), and differences in sequencing depth make it difficult to compare relative abundances across samples, challenges that are not faced by population geneticists. We have addressed these challenges using a variance stabilizing transformation, but other models that test differential abundance and differential variance which can control for differences in sequencing depth and are robust to zero-inflation might be more appropriate (e.g., Martin, Witten, and Willis (2020)). Clustering reads at higher taxonomic levels could circumvent zero-inflation by providing more continuous variation in taxon abundances across ecosystems.

However, this approach introduces biases based on the completeness of taxonomic databases, the accuracy of 16S-based taxonomic assignment, and the removal of reads that lack a taxonomic assignment (although, newer approaches to taxonomic classification might help (Shah, Meisel, & Pop, 2019)). Alternatively, decreasing the threshold of sequence similarity to cluster reads without taxonomy could be analogous to aggregating at higher taxonomic levels, but it is uncertain whether these larger aggregates of taxa have any trait conservatism related to function. Here, we chose to test ASVs at the level of the individual read so as not to bias our results in these ways. Finally, we have applied this approach to ASVs inferred from 16S rRNA gene sequences, but any unit of microbiome data such as metagenomic reads or metatranscriptomic mRNA reads could be tested in an association study.

Experimentally, future studies could improve on our example by sampling a more homogeneous set of ecosystems. Our survey includes an especially broad assortment of ecosystems, including grasslands, plantations, forests, peatlands and mineral soil wetlands among others. These ecosystems represent a range of moisture conditions that could regulate the abundance and activity of methane oxidizers and access to methane and oxygen, which methane oxidizers rely on. While this captured substantial variation in methane oxidation rates, sampling from such diverse ecosystems could result in spurious associations between taxa and function. For example, taxa differentially abundant in upland ecosystems that are unrelated to methane oxidation might appear correlated simply as a result of their presence in those ecosystems with high oxidation rates. Future studies could try restricting their search to a more homogeneous population of ecosystems specific to the question at hand.

2.5 Conclusion

Microbial biodiversity-ecosystem function research has demonstrated positive correlations between diversity and ecosystem function. However, the abundances of microbial functional groups (as currently defined) are often poor predictors of ecosystem function and commonly do not add substantial explanatory power to ecosystem models. Therefore, a new perspective on how to determine the relationship between microbial communities and ecosystem functions is sorely needed. Organismal biologists have over a hundred years of experience identifying relationships between complex sets of highly variable data (genotypes or genome sequences) and aggregate functions (organismal phenotypes). We assert that combining the approaches of traditional biodiversity-ecosystem function research with ideas from genotype-phenotype mapping could provide this new perspective. This integration could not only make underutilized approaches such as covariate modelling and artificial selection more available to microbial ecologists, but also provide instructive examples of how best to conceive of microbial biodiversity-ecosystem function questions. If this integration is successful, it is possible that in the not-so-distant future our field will be able to robustly identify taxa, genes, or even molecules that will allow us to accurately predict the response of ecosystems to environmental change. Doing so will not only generate novel hypotheses about how complex microbial communities drive ecosystem function, but also help scientists predict and manage changes to ecosystem functions resulting from human activities.

2.6 Bridge

The ideas in this chapter represent the conceptual framework that will guide the rest of my dissertation. This framework was designed to address the problem of

understanding how biodiversity loss due to global change will contribute to changes in ecosystem function. This framework draws on an analogy between population genetics and biodiversity-ecosystem function research. Based on this analogy, I will borrow two approaches from population genetics: an experimental approach using artificial selection and a comparative approach using a natural gradient of variation in CH_4 emissions. The following Chapter 3 will apply the artificial selection approach and the later Chapter 4 will apply the comparative approach.

CHAPTER III

ARTIFICIAL ECOSYSTEM SELECTION TO DEDUCE THE MAPPING BETWEEN MICROBIOME STRUCTURE AND ECOSYSTEM FUNCTION

For the unpublished work in this chapter, I conceived of the study, designed the study, carried out the experiment, performed molecular laboratory work, performed the data analysis, and drafted the manuscript. Brendan J. M. Bohannan was the principal investigator, designed the study, and critically revised the manuscript.

3.1 Introduction

Microbiomes mediate a variety of important functions in ecosystems, and there is great interest in understanding how attributes of microbiomes may influence variation in ecosystem functions (Crowther et al., 2019). Biodiversity-ecosystem function relationships have been described for a variety of macroorganismal communities. For example, plant species richness correlates with productivity and marine community diversity correlates with a variety of ecosystem functions (Cardinale et al., 2006; Gamfeldt et al., 2015; Hooper et al., 2012). While there is evidence that microbiomes are important for determing ecosystem functions, describing this relationship has been elusive (Graham et al., 2016; Rocca et al., 2015).

One reason for this difficulty is that it is challenging to directly manipulate microbiome composition and diversity they way we can for macroorganismal communities. Instead, we must use genetics through marker gene or metagenomic sequencing to characterize the microbiome. However, studies that measure the taxonomic or functional diversity of a microbiome through these approaches rarely find a correlation between diversity and function (Graham et al., 2016). Another

approach is to quantify cells or genes using flow cytometry or quantitative PCR. Studies such as this still rarely find a correlation between the abundance of a functional group and the rate of the corresponding process (Rocca et al., 2015). As a result, it remains unclear how and when microbiome variation contributes to variation in the rate of ecosystem functions.

Another likely reason that microbiome-ecosystem function relationships have been difficult to document is that the most common approaches for looking for a relationship requires some prior knowledge of the likely causal links between microbiomes and ecosystem functions. For example, many studies have looked for relationships with the abundance or diversity of a "marker gene" for a particular function. Marker genes are microbial genes that code for an enzyme or enzyme-subunit known to be involved in a particular function. But this approach assumes that one can identify the "right" marker gene *a priori* which, given how little is understood about microbial diversity, seems highly unlikely. As I argue in Chapter 2, what is needed are approaches to microbiome-ecosystem function mapping that do not require this degree of prior knowledge.

Yet another likely reason that microbiome-ecosystem function relationships have been difficult to document is that microbiomes can be related to ecosystem functions in two related but distinct ways that historically have been difficult to separate. One way is that microbiomes may simply be conduits through which the environment drives function. That is to say that the environmental conditions completely regulate the attributes of the microbiome and a shift in the environment causes a shift in microbiomes attributes, such as the abundance of a microbial functional group. This shift results in a change in the rate of an ecosystem function. In this case, microbiome attributes would be statistically associated

with variation in ecosystem function, but also strongly covary with environmental conditions. Therefore, there would be no association between microbiomes and ecosystem functions independent of the environmental conditions. Under this scenario, knowing anything about the microbiome would not necessarily provide you with information regarding the causes of the change in ecosystem function or improve one's ability to predict changes in ecosystem function. Furthermore, attempting to alter ecosystem function by altering the microbiome without changing the underlying environmental conditions would likely be fruitless.

Another possibility is that the microbiome itself alters the rate of ecosystem function independent of the environment. In this case, one cannot predict the rate of ecosystem function without understanding changes in the microbiome. In addition, managing the microbiome through inoculation or selective antibiotics would potentially be an effective approach for enhancing or mitigating certain changes to ecosystem functions. These two scenarios are not mutually exclusive, nor are they likely unique to microbes, but methods to determine the relative important of these two scenarios have not until recently been employed.

The problem of connecting microbiome attributes to the rate of an ecosystem function is analogous to the problem of connecting genomic variation to phenotypic traits in organisms, and solutions to this problem may be found by exploring this analogy (Morris, Meyer, & Bohannan, 2020). For example, one approach to the problem of genotype-phenotype mapping is to perform artificial selection on a trait of interest and identify genes that respond to selection. Such genes are likely to be causally linked to the trait under selection. An analogous approach for exploring microbiome-ecosystem function relationships is gaining acceptance among microbiologists (Panke-Buisse et al., 2015; Swenson, Wilson,

& Elias, 2000). Microbiomes have been shown to respond to selection at the ecosystem-trait level (Swenson, Wilson, & Elias, 2000), and some ecosystem-scale microbiome traits have been shown to be transferable through microbiome inoculations (Panke-Buisse et al., 2015). Applying artificial selection to whole microbiomes could be a novel approach for quantifying how much of the variation in the rate of an ecosystem function could be attributed independently to microbiome variation (Goodnight, 2000). In addition, this approach could be used to identify microbiome attributes, such as genes or taxa, associated with the rate of an ecosystem function, although, to my knowledge this has not previously been accomplished (Arias-Sánchez, Vessman, & Mitri, 2019; Morris et al., 2020). Such an approach has the potential to expand our ability to model microbial ecosystem-function relationships more accurately and allow us to manage ecosystems for particular outcomes (Fuhrman, 2009; Schimel & Gulledge, 1998). Here I use this approach to estimate how much of the variation in ecosystem function (the flux of CH_4 gas between the soil and atmosphere) can be attributed to variation in the microbiome and to identify microbial markers associated with this ecosystem function.

Soil CH_4 oxidation is a suitable function for this study because there is evidence that it may vary with microbiome attributes. For example, variation in CH_4 emissions in arctic permafrost is correlated with the transcriptional activity of certain methanogens (Freitag & Prosser, 2009; Freitag et al., 2010). In addition, soil microbiome phylogenetic variation is a strong predictor of CH_4 oxidation in forests and pastures of the Brazilian Amazon (Meyer, Morris, et al., 2020). Finally, methanogenesis and methanotrophy are two of the most deeply conserved microbial physiologies and are represented in a narrow range of taxa and so the taxonomic

composition of the microbiome is more likely to be associated with the rate of CH_4 flux than other broader or more shallowly conserved functions (A. C. Martiny et al., 2013; Schimel & Gulledge, 1998). This suggests that microbiome variation might drive variation in CH_4 oxidation in soils.

For the current project, I applied artificial ecosystem selection on soil microbiomes by selecting on the ecosystem-scale CH_4 oxidation rate in order to estimate the additive genetic variance of soil CH_4 oxidation (Swenson, Wilson, & Elias, 2000). To identify which microorganisms could be markers of CH_4 oxidation rate, I compared the composition of the artificially selected microbiomes to a control set of microbiomes without selection. Artificial ecosystem selection has a similar effect to enrichment culturing by amplifying the population of interest (Swenson, Wilson, & Elias, 2000). This will reduce the diversity of the soil microbiome and allow for greater power in detecting significant markers of CH_4 oxidation rate. I then evaluate whether these markers meet our underlying assumptions about which taxa limit the rate of ecosystem function.

3.2 Materials and Methods

3.2.1 Experimental design. I performed an artificial ecosystem selection experiment (*sensu* Swenson, Wilson, and Elias (2000) by passaging replicate soil microbiomes. The trait I selected on was CH_4 oxidation rate. My experiment had two selection lines with twelve jars each for a total of twenty-four jars per passage. One line was a positive selection line where the two or three jars with the highest CH_4 oxidation rate were chosen to inoculate the next set of positive jars. The other line was a neutral selection line where an equal number of jars as the positive line were chosen at random to inoculate the next set of neutral jars. The number of jars chosen was based on the distribution of fluxes

among the positive jars, i.e., I chose the top three jars unless only two jars had considerably greater CH_4 oxidation rates based on visual inspection of histograms. The experiment was carried out over five passages until a significant divergence in functional rates was observed between the two selection treatments.

The initial soil microbiome was sampled from the top 10 cm of an upland mineral soil under a deciduous forest ecosystem near the University of Oregon campus in Eugene, OR, USA. Incubations were performed in 500 mL mason jars with rubber septa installed in the lids. Each jar was sterilized with 70% ethanol to which was added 45 g of autoclaved potting mix, 5 g of living soil, and 3.5 mL of sterile deionized water to bring them to 60% of field capacity. The jars were then homogenized, capped, and injected with 4.3 mL of 99% CH_4 to bring the headspace concentration to approximately 1000 ppm CH_4. To create the two treatment groups, twenty-four jars were created in an identical manner and then randomly assigned to either the positive or neutral selection line. Jars were flushed and respiked twice per week to maintain aerobic conditions and elevated CH_4 concentrations and were incubated at ambient temperature for approximately three weeks. Methane oxidation rates were determined at the end of the incubation period. For the positive treatment, the three jars with the greatest CH_4 oxidation rate were chosen to inoculate the next generation. For the neutral treatment, three jars were randomly selected to inoculate the next generation. For each treatment, these three jars were homogenized and this homogenized soil was used as the 5 g inoculum for the next set of jars. The next set of jars were created in an identical manner to the first generation with fresh autoclaved potting mix and the same moisture and CH_4 content.

3.2.2 Methane oxidation rate. Methane oxidation rates were determined after flushing and spiking jars to approximately 1000 ppm CH_4. Headspace samples of 1 mL were collected from each jar immediately after spiking and then at time points 3, 6, 24, and 48 hours for a 5-point curve. Samples were immediately injected into a SRI model 8610C gas chromatograph equipped with a flame ionization detector (Torrance, CA, USA) to determine the headspace CH_4 concentration. Fluxes were calculated from a first-order exponential decay function as decay constant k with units day^{-1}. Oxidation rates are presented as the additive inverse of k (i.e., $-k$) so that a more positive value represents a greater oxidation rate. The jars selected for the positive treatment in passage 2 had the lowest CH_4 oxidation rate of the twelve jars due to a calculation error in the CH_4 oxidation rate. All other passages correctly used the top three jars.

3.2.3 Soil DNA extraction and sequencing. A subsample of soil from the starting inoculum and from every jar in passages 2 and 5 was collected and stored at -80°C. Soil DNA was extracted from 0.25 g soil. Negative controls were extracted from autoclaved potting mix and DNase-free water. Extractions were performed using the DNeasy PowerSoil kit (QIAGEN, Germany) and quantified using Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, USA). In order to estimate the diversity and relative abundance of the bacterial and archaeal taxa in my soil ecosystems, I sequenced the V4 region of the 16S rRNA gene using the 515F - 806R primer combination (Caporaso et al., 2011). PCR mixtures were: 10 μl NEBNext Q5 Hot Start HiFi PCR master mix, 9.2 μl primer mixture (1.09 μM concentration), and 0.8 μl of DNA template. Reaction conditions were: 98°C for 30 s (initialization); 35 cycles of 98°C for 10 s (denaturation), 61°C for 20 s (annealing), and 72°C for 20 s (extension); and 72°C for 2 m (final extension).

Amplicons were purified twice using 0.8x ratio Mag-Bind RxnPure Plus isolation beads (Omega Bio-Tek, USA). Sequencing libraries were prepared using a dual-indexing approach (Fadrosh et al., 2014; Kozich, Westcott, Baxter, Highlander, & Schloss, 2013). Amplicon concentrations were quantified using Qubit (Thermo Fisher Scientific Technologies, USA) and multiplexed at equimolar concentration. Sequencing was performed at the University of Oregon Genomics Core Facility on the Illumina NovaSeq 6000 with paired-end 150 bp reads.

3.2.4 Bioinformatics. Bioinformatics processing was performed in 'R' (R Core Team, 2018). Demultiplexed sequencing reads were denoised using 'DADA2' to generate a table of amplicon sequence variants (ASVs) (Callahan et al., 2016). Taxonomic assignment for the 16S reads was performed using the RDP naive Bayesian classifier (Wang, Garrity, Tiedje, & Cole, 2007). The presence of contaminants was evaluated using both the prevalence and frequency methods from 'DECONTAM' by comparing samples to negative controls (Davis, Proctor, Holmes, Relman, & Callahan, 2018). I did not identify any obvious contamination using these methods.

3.2.5 Statistical Analysis. Statistical analyses were performed in 'R' (R Core Team, 2018). To test whether there was a significant change in CH_4 oxidation rate (k) as a response to selection, I tested a difference in slopes between the positive and neutral selection lines. The CH_4 oxidation rates were strongly right skewed with most values close to zero and few large, positive values. This resulted in residuals that did not meet the assumptions of constant variance and normal distribution. Therefore, CH_4 oxidation rates were log10 transformed to better meet the assumptions of a linear model and to make figures easier to interpret. First, I tested the effect of treatment by fitting two nested models with and without

treatment using 'lm'. I compared these models using the likelihood ratio test with the 'anova' function. I then fit the full model to estimate the slope of the positive line, which represents the change in CH_4 oxidation rate per passage as a response to selection.

To estimate the proportion of variance in ecosystem function due to variation in the microbiome, I estimated narrow-sense heritability (h^2) as the regression of mid-offspring on mid-parent (Falconer & MacKay, 1996). The mid-parent was the mean for all three selected jars and the mid-offspring was the mean for all jars produced by those parents. First, I tested if there was an effect of treatment on the heritability estimate. I compared these models using the likelihood ratio test with the 'anova' function. I then fit the full model to estimate the heritability of the Positive and Neutral treatments.

Richness estimates and tests were performed using 'breakaway' (A. D. Willis & Martin, 2020). I tested a difference in richness between Passages 2 and 5 with both treatments combined. Then I subset the samples from Passage 5 and compared richness between the Positive and Neutral treatment. Next, I compared beta-diversity as both Aitchison dissimilarity, which is the Euclidean distance after center-log ratio transformation, and the binary Jaccard distance for presence or absence (Aitchison, 1982). I tested a difference in centroid and dispersion of beta diversity between the Positive and Neutral treatment and Passage 2 and 5 with permutational analysis of variance (PERMANOVA) using the 'adonis2' function from 'vegan' (McArdle & Anderson, 2001; Oksanen et al., 2019).

To identify taxa that responded to selection on CH_4 oxidation rate, I tested differential abundance using a beta-binomial model (Martin et al., 2020). I first grouped ASVs at the family level keeping ASVs without a family assignment. I

then subset the samples in Passage 5 and removed all families with a prevalence of less than 10%. Taxa present in $< 10\%$ of samples would only be present in one or two jars and therefore it would be unreasonable to test differential abundance on these taxa. I compared families in passage 5 between the Positive and Neutral treatment to identify taxa that were enriched or depleted as a response to selection. I used the likelihood ratio test to estimate p-values with a significance threshold of false-discovery rate < 0.05 (Martin et al., 2020). Estimates are presented as the expected relative abundance in the positive treatment relative to the neutral treatment.

3.3 Results

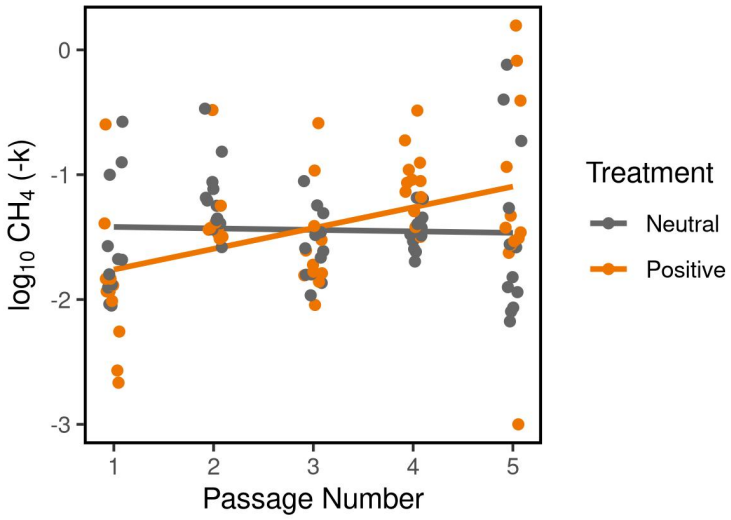

Figure 4. Response to selection on soil CH_4 oxidation rate fit by ordinary least squares regression. The y-axis is the log10 of the additive inverse of the first-order exponential decay constant k (i.e., $-k$) with units day^{-1}. The x-axis is the passage number. The orange line and points are the Positive treatment and the gray line and points are the Neutral treatment.

I observed a response to selection on whole-ecosystem soil CH_4 oxidation rate (Figure 4). In addition, the response to selection varied with treatment (LRT

of nested models with and without treatment: df = 2, ss = 1.86, p = 0.02). At the start of the experiment, the Positive treatment had a mean CH_4 oxidation rate that was 24% lower than the Neutral treatment (difference of y-intercepts = -0.34, SE = 0.16, t = -2.14, p = 0.03). There was no change in CH_4 oxidation rate of the Neutral treatment over the five passages (slope = -0.01, SE = 0.05, t = -0.26, p = 0.80). By contrast, the Positive treatment had a 51% increase in CH_4 oxidation rate per passage (slope = 0.18, SE = 0.06, t = 2.76, p = 0.01).

I estimated heritability as the regression of mid-offspring on mid-parent (Figure 5). Offspring CH_4 oxidation rates were correlated with parental CH_4 oxidation rates in both the Positive treatment (slope = 1.08, SE = 0.24, t = 4.43, p = 0.01) and the Neutral treatment (slope = -0.58, SE = 0.20, t = -2.89, p = 0.04). Notably, the sign of the effect was reversed with a positive heritability for the Positive treatment and a negative heritability for the Neutral treatment.

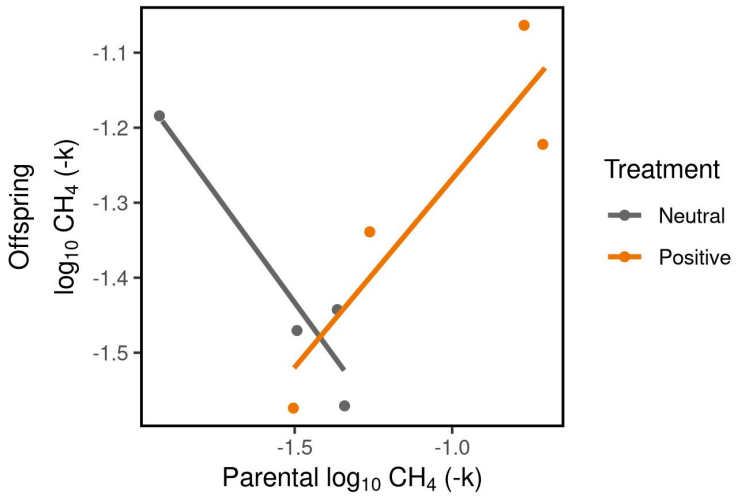

Figure 5. Ordinary least squares regression of mid-offspring CH_4 flux on mid-parent CH_4 flux ($-k$ with units day^{-1}). Mid-parent is the mean of the jars selected to inoculate the next passage. Mid-offspring is the mean of all twelve jars produced in one passage. The orange lines and points are the Positive treatment and the gray lines and points are the Neutral treatment.

Sequencing of 16S rRNA amplicons revealed that there were 9717 unique amplicon sequence variants (ASVs) across all 46 jars sampled. Richness of ASVs decreased among all of the jars between Passage 2 and 5 (Richness difference between Passage 5 and Passage 2 = -2450.8, SE = 285.09, $p < 0.001$). However, there was no difference in richness between the Positive and the Neutral treatment in Passage 5 (Difference = 6.1, SE = 36.04, p = 0.866). In addition, there was no correlation between richness and CH_4 oxidation rate across all passages and treatments. (CH_4 = -18.3, SE = 55.41, p = 0.742).

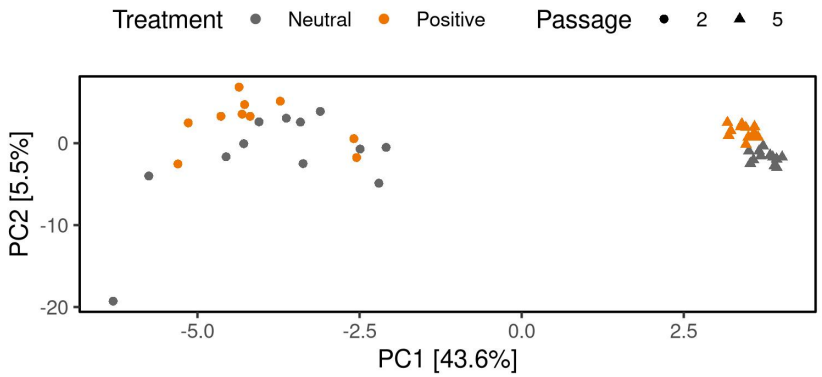

Figure 6. Principal component analysis (PCA) plot of beta diversity for all jars. Dissimilarities are based on Aitchison dissimilarity, which is Euclidean distance after center-log ratio transformation. Colors represent treatment and shapes represent passage number. The ratio of the axes is proportional to the variance explained by each principal component in order to accurately represent the distance between the points.

Taxonomic dissimilarity of the soil microbiome varied strongly by passage and weakly by treatment (Figures 6). Permutational analysis of variance (PERMANOVA) demonstrated an effect of passage, treatment, and their interaction on Aitchison dissimilarity (Table 4). Passage explained 41.9% of the variation in Aitchison dissimilarity and 47.4% in Jaccard dissimilarity. Treatment explained 4.7% for Aitchison and 3.4% for Jaccard. Variation in microbiome

Table 4. PERMANOVA on Aitchison and Jaccard dissimilarities.

Term	df	SS	R2	F	p
Aitchison					
Passage	1	192687.59	0.42	34.96	0.001
Treatment	1	21430.52	0.05	3.89	0.006
Pass:Treat	1	13888.56	0.03	2.52	0.038
Residual	42	231507.31	0.50	NA	NA
Total	45	459513.98	1.00	NA	NA
Jaccard					
Passage	1	3.68	0.47	42.51	0.001
Treatment	1	0.26	0.03	3.04	0.019
Pass:Treat	1	0.19	0.02	2.22	0.072
Residual	42	3.64	0.47	NA	NA
Total	45	7.77	1.00	NA	NA

composition decreased as a result of passaging in jars ($F = 80.5$, $p < 0.001$), but did not differ between selection treatments ($F = 0.40$, $p = 0.54$).

To investigate which taxa responded to selection on soil CH_4 oxidation rate, I first looked at which taxa were unique to the Positive and Neutral treatments aggregated at the family level. Several taxa were present in the Positive treatment and absent in the Neutral treatment. Most of these families had low prevalance (present in 3 or fewer samples) and low abundance (median < 5 reads). There were two families unique to the Positive treatment with relatively high prevalence. This included an ASV that was a member of the Bacteroidia Class with no lower taxonomic designation. This ASV had a prevalence of 10/12 and a median abundance of 2 reads. The other prevalent family was a member of the Silvanigrellaceae, a newly described family placed in its own order. This family was present in all 12 samples and had a median abundance of 8 reads. Silvanigrella is the only cultivated member of Silvanigrellaceae and was isolated from a temperate fresh water lake (Hahn et al., 2017). Of the families unique to

the Neutral treatment, only one had a prevalence greater than 2/12. This family, Armatimonadaceae, was present in half of the Neutral samples (prevalence = 6/12). The type strain for Armatimonadaceae was isolated from the rhizosphere of *Phragmites australis* (Tamaki et al., 2011).

The remaining families were shared between the Positive and Neutral treatment, but some differed in their relative abundance. To identify taxa that responded to selection on soil CH_4 oxidation rate, I tested the differential relative abundance of families in the Positive jars relative to the Neutral jars within Passage 5 using a beta-binomial model. I identified 37 families that were enriched or depleted in the Positive treatmeant relative to the Neutral treatment (Figure 7).

Overall, none of the families enriched in the Positive selection treatment are known methanotrophs. Several taxa identified had a higher taxonomic designation that contains methanotrophs, for example, the Gammaproteobacteria class had a large effect size. This class is one of the groups containing methanotrophic families such as Methylococcaceae. However, gammaproteobacteria is among the most diverse groups in the Prokaryotes, so this is not strong evidence for a selection response by methanotrophs (Garrity, Bell, & Lilburn, 2005). The next family was Puniceicoccaceae, which is a member of the phylum Verrucomicrobia and contains marine microbes. The Verrucomicrobia is a diverse group that contain known methanotrophs as well as ammonia-oxidizing bacteria (Freitag & Prosser, 2003). An ASV from the order Armatimonadales was also enriched in the Positive treatment. Interestingly, this order also contains the family Armatimonadaceae, which was found to be unique to the Neutral treatment. Cytophagaceae was also enriched in the Positive treatment and contains a number of mainly aerobic heterotrophs that can digest a variety of macromolecules (McBride, Liu, Lu, Zhu, & Zhang, 2014).

Notably, the family Gemmatimonadaceae has a single bacterial bin, Candidatus '*Methylotropicum kingii*', that contains methanotrophy genes, but this family was depleted in the Positive selection treatment (Bay et al., 2021).

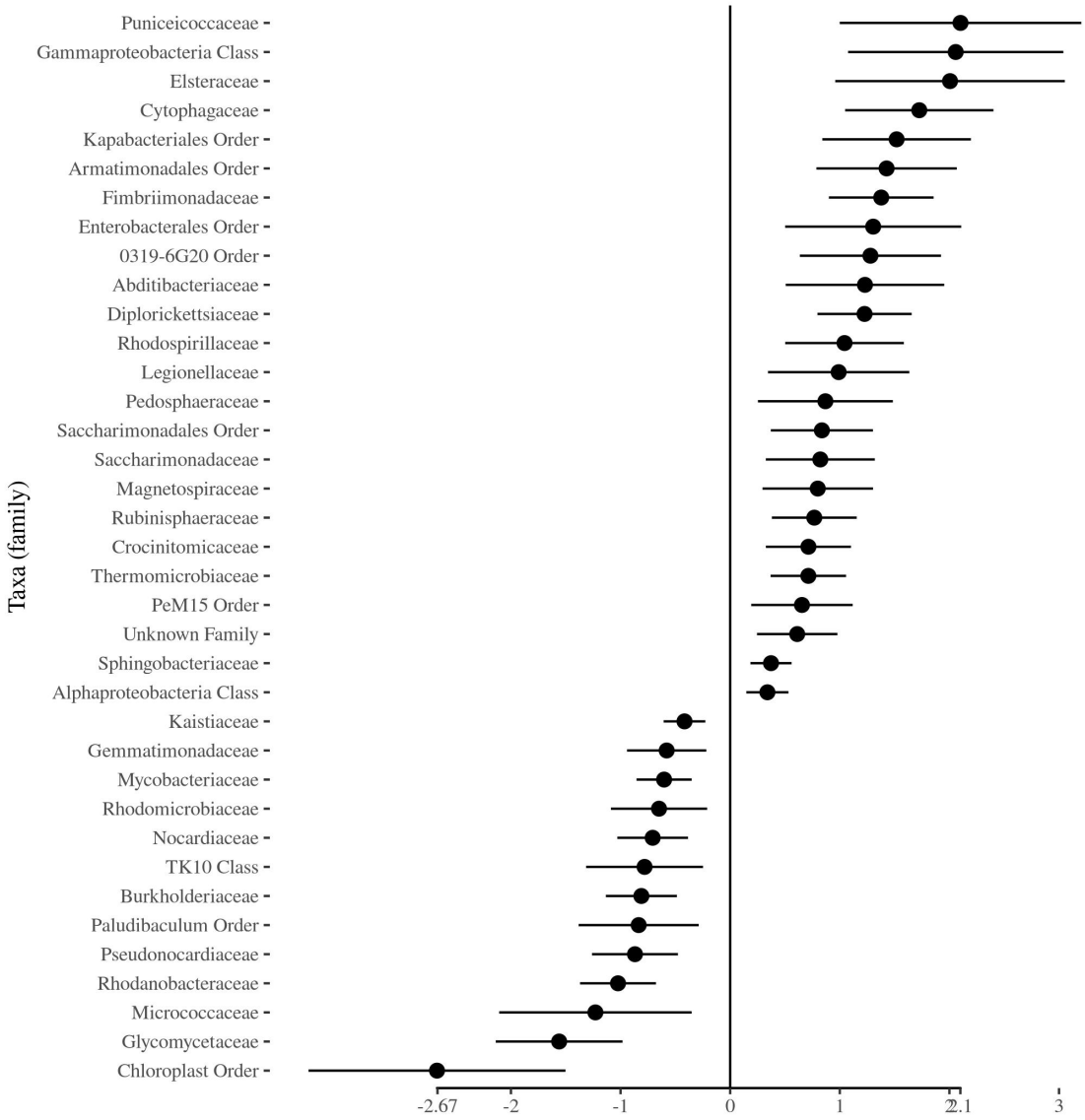

Figure 7. Differentially abundant family-level taxa based on a beta-binomial model. Values on the x-axis are estimates and standard errors of the expected difference in the logit-transformed relative abundance between the two treatments. Positive values are enriched in the Positive treatment and negative values are enriched in the Neutral treatment. Taxa presented here are significant at FDR < 0.5.

3.4 Discussion

I observed a response to selection on whole-ecosystem soil CH_4 oxidation rate. This type of response has been observed for other functions, such as chloroaniline degradation in water (Swenson, Wilson, & Elias, 2000), but this is the first instance, to my knowledge, of selection being performed directly on soil biogeochemical cycling. This experimental design attempted to create multiple ecosystems that were very similar with respect to the type and quantity of soil, soil moisture content, and the headspace concentration of CH_4. I further controlled for environmental variation across replicates by having twelve replicates per treatment, using controlled laboratory conditions, and by having a random selection treatment as a control. Therefore, the observed response to selection on CH_4 oxidation rate is likely due to changes in the microbiome.

To investigate the relationship between microbiome structure and ecosystem function, I wanted to determine how much of the variation in soil CH_4 oxidation was attributable to variation in the microbiome. This ecological question is analogous to the problem of estimating heritability in quantitative genetics (Morris et al., 2020). Variation in an organismal trait is determined by the sum of genetic variation, environmental variation, and the interaction between the two (Falconer & MacKay, 1996). One goal of quantitative genetics is to determine the proportion of phenotypic variation attributable to genetic variation. This is commonly estimated as the narrow-sense heritability defined as the additive genetic variance. This goal can be achieved through artificial selection experiments with selection on the character of interest (Falconer & MacKay, 1996).

The heritability I observed in the Positive treatment was large ($h^2 = 1.08$) relative to other studies of biological heritability (Visscher, Hill, & Wray,

2008). However, given the small sample size and wide confidence intervals (95% CI: 0.40, 1.75) it is likely that a larger study could more precisely estimate the true heritability. I intentionally eliminated much of the environmental variation that would be present in natural ecosystems, but it is unclear how much jar-to-jar variation there is in the composition of the soil substrate. Given these data, though, I conclude that variation within the soil microbiome has a very strong effect on soil CH_4 oxidation rates independent of any environmental variation.

The negative heritability observed in the Neutral treatment could indicate a negative genetic correlation between CH_4 oxidation rate and traits associated with persistence in a jar (Conner, 2003). The conditions in the jar favor organisms that can grow in potting mix, reproduce within a three week incubation, and survive passaging between jars. These traits might be negatively correlated with a community's ability to oxidize CH_4 indicating a potential tradeoff (Conner, 2003). This tradeoff would result in a negative heritability of CH_4 oxidation rate in the Neutral treatment, which did not experience selection on CH_4 oxidation rate.

Since there was a response to selection, I wanted to identify which taxa responded to selection. That is, the taxa that were enriched in the Positive treatment relative to the Neutral treatment. Studies of microbiome structure-function relationships typically quantify marker genes for the final enzyme in a pathway to try to predict the rate of flux through that pathway. However, of the 37 taxa identified by the differential abundance test, none were members of a family known to oxidize CH_4. This suggests that in my system the taxa that perform the function CH_4 oxidation do not limit the rate of flux through the pathway. Traits that we assume are important for regulating function (such as the diversity or relative abundance of methanotrophs estimated from phylogenetic marker

genes) may not be important for determining variation in function. Perhaps once a sufficient number of methanotrophs are present, the process itself is limited by other metabolic processes in the ecosystem, such as nitrogen cycling mediated by non-methanotrophic organisms.

I found that 24 taxa were enriched in the Positive treatment, which might indicate that the soil CH_4 oxidation rate in these jars is controlled by multiple taxa. This conclusion is analogous to the conclusion in quantitative genetics that most traits at the organismal level are the result of the interaction between many independent genes. Indeed, very few traits or genetic diseases are the result of a single gene or mutation (Kerem et al., 1989; MacDonald et al., 1992). This results in the use of polygenic risk scores for predicting phenotype or disease risk in humans (Hindorff et al., 2009; Reich & Lander, 2001). An analogous conclusion can be drawn for predicting the rate of ecosystem function from the presence or relative abundance of microbial taxa. The rate of an ecosystem function at the whole-ecosystem level is the result of interactions among a variety of disparate taxa with different traits. To better understand microbiome structure-function relationships, my results suggest that we should move away from single marker genes and instead investigate the role of multiple functional groups in determining the function of ecosystems whether it is CH_4 flux from soil or host-microbiome health.

Taxonomic richness dropped precipitously between Passage 2 and Passage 5. There are several explanations for this drop in diversity. The initial community was a diverse microbiome sampled from a natural soil. With the initial inoculation and at each subsequent passaging, the ecosystem was subsampled to 10% (5 g of living soil was combined with 45 g of sterilized substrate). This subsampling likely explains a large part of the drop in diversity. In addition, beyond the imposed

selection regime based on CH_4 oxidation rate, these soil microbiomes were under a variety of selection pressures such as persistence in a jar, survival during the transfer from one jar to the next, and the ability to colonize a new jar in a couple of weeks, to name a few. These selection pressures would have further filtered the microbiome for taxa that could persist in this laboratory environment.

In addition to the drop in richness, there was also a decrease in beta diversity among the jars as a result of passaging. This could be due to the homogenization step between each passage. Once the jars were selected to generate the next set of jars, the soil from those jars was homogenized and this homogenate was used to inoculate the next set of jars. The goal of this step was to "shuffle" membership in the microbiomes among the jars in order to test different combinations and relative abundances of taxa in terms of their effect on CH_4 oxidation rate. This likely also made the communities more similar in their membership by increasingly the likelihood that each taxon was represented in each jar. Another explanation for this biotic homogenization is the aforementioned selection for persistence in a jar. These selection pressures would have favored the subset of taxa suited to these laboratory conditions regardless of whether they were in the Positive or Neutral selection line. This can be seen by the relative similarity of the Positive and Neutral jars in Passage 5. Despite the Positive treatment undergoing selection on CH_4 oxidation rate, both the Positive and Neutral jars were experiencing selection on a variety of other traits that made the overall soil microbiomes appear similar.

One caveat that is important to keep in mind is that 16S rRNA genes are not markers of functional characteristics of microorganisms. They are simply phylogenetic markers useful for determining the relative position of taxa in a

phylogenetic tree. In addition, assigning functions to taxa based solely on their taxonomy is a fraught exercise because many microbial functions are not very deeply conserved, are spread across disparate lineages of the tree of life, or are easily transferred between distantly related taxa. However, methanotrophy and methanogenesis are two of the most deeply conserved microbial traits (A. C. Martiny et al., 2013). This results in part from the fact that those functions require multiple genes to perform and are therefore not easy to evolve independently or to transfer between lineages through horizontal gene transfer. However, in order to be confident that the response I observed was not simply due to a relative increase in methanotrophs, I would need to apply a functional approach to characterizing the microbiome, for example, by sequencing whole metagenomes from my samples.

Future research in this area should use methods that are able to detect functions and metabolic pathways within the community that respond to selection. For example, genome-resolved metagenomics could identify pathways that are enriched in the Positive selection treatment. This would further advance our understanding of the exact traits that are enriched in high-flux ecosystems and therefore could be markers of ecosystem function. In addition, future studies could build on experimental results such as mine to sample a gradient of ecosystem CH_4 fluxes to determine whether the taxa identified in our selection experiment could be useful indicators of CH_4 flux in the field. These studies would allow us to build a picture of the multiple traits that drive variation in ecosystem function at the whole-ecosystem level. I will address some of this challenges in the final chapter of this dissertation.

3.5 Conclusion

I performed an artificial ecosystem selection experiment on whole-ecosystem CH_4 oxidation rate to determine whether variation in microbiome composition can contribute to variation in ecosystem function. I observed a significant response to selection on CH_4 oxidation rate - the first example of an experiment performing selection on soil biogeochemical cycling. In addition, I observed a strong heritability of CH_4 flux between passages, suggesting that variation in microbiome composition could be a major source of variation in CH_4 flux in ecosystems. Surprisingly, the taxonomic groups that experienced selection on CH_4 oxidation rate were not enriched in methanotrophs. This suggests that understanding a single functional group is insufficient for predicting the effect of microbiome composition on the rate of ecosystem function. Instead, we may need to investigate alternative functional groups or multiple interacting taxa to understand the role of microbiomes in ecosystem function. While it has often been assumed that microorganisms play a major role in determining variation in ecosystem function, this has not previously been shown empirically. This experiment demonstrates that variation in microbiome composition can contribute to considerable variation in ecosystem function.

3.6 Bridge

The overarching questions for this dissertation is whether variation in the microbiome is necessary to predict variation in ecosystem function. In this chapter, I addressed this question by experimentally demonstrating that variation in the soil microbiome can generate variation in an important ecosystem function–CH_4 oxidation. In addition, to begin to unravel what aspects of microbiome variation are important for generating variation in ecosystem function, I identified taxonomic

groups from 16S rRNA gene variants that were enriched in the ecosystems that underwent selection on CH_4 oxidation rate. One limitation to this approach is the contrived nature of the experiment in artificial potting mix in a laboratory environment under elevated CH_4 concentrations. In addition, using 16S markers did not allow me to identify functional groups within my microbiomes. To address these limitations, in the next chapter I will use shotgun metagenomics of soil microbiomes along an ecological gradient of CH_4 oxidation rates to identify which functional groups are altered by global change and which are association with high rates of CH_4 emissions. The objective of this chapter will be to understand whether biodiversity loss due to global change will lead to a loss of essential ecosystem functions.

CHAPTER IV

IDENTIFYING THE METAGENOMIC DRIVERS OF METHANE EMISSIONS FROM PASTURES OF THE BRAZILIAN AMAZON

For the unpublished work in this chapter, field samples were collected by Kyle M. Meyer and Brendan J. M. Bohannan. I was the primary contributor to the experimental design, laboratory work, and data analysis and did all the writing. Brendan J. M. Bohannan was the principal investigator for this work and helped conceptualize the experiment and revise the writing.

4.1 Introduction

One objective for biological research is to understand how global change will alter biodiversity patterns (Foley et al., 2005; Newbold et al., 2015). These changes can, in turn, affect the provisioning of ecosystem functions, such as gas exchange between the soil and atmosphere (Hooper et al., 2005; Schimel & Gulledge, 1998). A major global change affecting tropical rainforests is the conversion of forests to agricultural pastures (Dirzo & Raven, 2003; Laurance et al., 2014). This has been widely noted to shift ecosystems from net CH_4 sinks to net CH_4 sources (Meyer, Morris, et al., 2020; Steudler et al., 1996; Verchot et al., 2000). In addition, one of the strongest predictors of CH_4 flux in these systems is microbiome composition (Meyer, Morris, et al., 2020). Based on these observations, I want to understand how changes in belowground biodiversity patterns alter ecosystem-scale CH_4 fluxes.

A shift in the rate of an ecosystem function as a response to global change could be driven by changes to abiotic conditions, biotic communities, or both. For example, such a response could follow predictable patterns of reaction kinetics based on temperature, redox conditions, and the availability of substrates. If changes in these factors lead to changes in the microbiome as well as changes in

ecosystem function, it would be unnecessary to measure microbiome composition in order to accurately predict the response of an ecosystem function to environmental change. One could accurately predict the rate of future ecosystem functions from soil abiotic characteristics alone. However, if global change introduces or eliminates certain functional groups or if the relative abundances and distributions of species are altered directly by global change, then future functional rates might be unpredictable from basic reaction parameters (Schimel & Gulledge, 1998). In that case, understanding which microbial functional groups are shifting in response to environmental change will be essential for predicting future ecosystem functions.

One system in which microbiome composition may be an important driver of the rate of ecosystem function is in soil CH_4 emissions from the Brazilian Amazon. Past research in the Brazilian Amazon has shown that forest-to-pasture conversion switches an ecosystem from a net CH_4 sink in forest to a net CH_4 source in pasture (Meyer, Morris, et al., 2020; Steudler et al., 1996). Deep metagenomic sequencing of forests and pastures in the western Amazon revealed that forest-to-pasture conversion resulted in shifts in the CH_4 cycling community. Specifically, the proportion of methanogens increased relative to methanotrophs (Meyer et al., 2017). In addition, laboratory incubations using stable-isotope probing revealed an increase in the abundance and activity of methanogens as a result of deforestation (Kroeger et al., 2018). While it seems that the shift from CH_4 sink to CH_4 source could easily be explained by these changes in the CH_4-cycling community, these studies lacked CH_4 flux measurements and so it is not certain that high-flux sites can be explained only by the relative abundance of methanogens.

One study directly compared 16S rRNA gene variants with CH_4 flux measurements co-located in time and space (Meyer, Morris, et al., 2020).

This study found that microbiome composition was the strongest predictor of variation in CH_4 flux. In this study, CH_4 flux was correlated with the richness of methanogens based on phylogenetic markers (Meyer, Morris, et al., 2020). While methanogens and methanotrophs were associated with CH_4 flux in this study, most associated taxa were not known to produce or consume CH_4 (Meyer, Morris, et al., 2020). In addition, my project in Chapter 3 revealed that selection on soil CH_4 oxidation rate was associated with shifts in the 16S-inferred microbiome composition without any changes in the relative abundance of CH_4 oxidizing taxa, suggesting that CH_4 flux may not necessarily be controlled only by CH_4 cycling taxa. These studies are limited in the inferences that can be made about microbiome function since they used sequence similarity in regions of the 16S rRNA genes to identify microbial taxa; this approach only provides limited phylogenetic information about the taxa present and does not provide direct information about functional potential.

To address the shortcomings of these previous studies, I combined in-situ CH_4 flux measurements in two regions of the Brazilian Amazon co-located in time and space with soil metagenomic sequencing to identify aspects of the soil microbiome that are associated with increases in CH_4 flux. This approach is powerful, because metagenomic sequencing allows the direct identification of specific metabolic pathways without the need to infer function from phylogenetic affiliation. In addition, I compared the relative abundance of microbial pathways directly with rates of CH_4 flux to identify functional groups that increase as a proportion of the soil microbiome in high flux ecosystems. By using a comparative approach across a chronosequence of land use change in the field, as opposed to a laboratory study, the results of this study will be maximally applicable to a real-

world system. However, there are limitations to this approach; by not relying on a manipulative experiment, evaluating mechanisms underlying the association was difficult. Despite this tradeoff, the work I describe here will be essential to guide future studies using controlled field experiments or laboratory incubations to evaluate hypotheses generated from this study.

The core goal of this study was to determine whether it is necessary to measure variation in microbiome composition in order to predict future CH_4 emissions. In addition I wanted to identify specific aspects of the microbiome that are informative for predicting variation in CH_4 emissions. By modeling the relationship between microbiome composition and CH_4 flux while controlling for underlying environmental covariates and the geographic distance between sites, I was able to identify microbiome functional pathways that were associated with CH_4 flux independent of the measured environmental variables. These results will help guide future efforts to model how variation in the soil microbiome can be used to predict CH_4 emissions under future global change.

4.2 Materials and Methods

This study was conducted along land-use change gradients in two regions of the Brazilian Amazon: Fazenda Nova Vida (FNV) in the state of Rondônia and Santarem in the state of Pará. The full site information can be found in Meyer, Morris, et al. (2020). Briefly, soil samples were collected from primary forest, cattle pasture, and secondary forest. In FNV, samples included 10 forest, 10 pasture, and 9 secondary forest samples. In santarem, samples included 10 forest samples, 10 pasture samples, and 13 secondary forest samples. At each site, samples were collected along a 200 m transect at 50 m intervals. At each sampling location, soil CH_4 flux was measured in real time using a Fourier-transform infrared spectrometer

(Gasmet, DX 4015, Vantaa, Finland). Immediately following gas sampling, air temperature, soil temperature, and soil moisture data were collected. In addition, five soil cores were collected with a sterilized corer (5 cm diameter and 10 cm depth). These cores were homogenized, subsampled, and stored at 4°C for chemical analysis or frozen for DNA extraction. Soils were analyzed for 19 attributes which were used as the environmental covariates. These included pH, organic matter, P, S, K, Ca, Mg, Al, H + Al, sum of exchangeable bases, cation exchange capacity, base saturation, Al saturation, Cu, Mn, Zn, and N. Total DNA was extracted from 0.25 g soil using the DNeasy PowerSoil kit (Qiagen Inc., Valencia, CA, USA).

4.2.1 Metagenomic sequencing, assembly, and annotation.

Sequencing was perfomed on the Illumina NovaSeq S4 platform with 151 bp paired-end reads. Quality filtering was performed on raw reads using the 'rqcfilter2.sh' script from 'bbtools' version 38.88 (Bushnell, 2021; Bushnell, Rood, & Singer, 2017). Sequencing and quality filtering were performed by the Joint Genome Institute (JGI). Filtered reads were downloaded from the JGI Genome Portal and were normalized using 'bbnorm.sh' from 'bbtools' version 38.90 (Bushnell, 2021). Assembly was performed on normalized reads using 'megahit' version 1.2.9 with settings '–min-contig-len 1000 –k-min 27 –k-max 127 –k-step 10 –kmin-1pass' (Li, Liu, Luo, Sadakane, & Lam, 2015). Filtered (not normalized) reads were mapped to the assembly using 'bowtie2' (Langmead & Salzberg, 2012). Contigs and coverages were imported to 'anvio-7' (Eren et al., 2021). Open reading frames were identified using 'prodigal' (Hyatt et al., 2010). Contig coverages were quantified as coverage per nucleotide position using 'anvi-profile' in Anvio with a minimum contig length of 2500 bp/nts. Functional genes and pathways were annotated using KEGG

KoFams (Aramaki et al., 2020). Finally, functional gene abundances were estimated as mean coverage across the contig.

4.2.2 Microbiome Analysis. Microbiome functional gene annotations and mean coverage data were exported from 'Anvi'o' and imported to 'R' (R Core Team, 2018). Functional genes were aggregated at the KEGG Module level by summing the mean coverage of genes from the same Module. The resulting 297 pathways were used in subsequent analyses. Microbiome analyses in 'R' were performed using the 'phyloseq' and 'microViz' packages (Barnett, Arts, & Penders, 2021; McMurdie & Holmes, 2013). For the distance-based analyses and ordinations, microbiome functional dissimilarities were calculated using the Aitchison dissimilarity, which is the Euclidean distance after center-log ratio transformation (Aitchison, 1982). This dissimilarity preserves the ratios between functional pathway abundances while controlling for the compositional nature of the data. Ordinations were created using principal component analysis (PCA) on Aitchison dissimilarity.

4.2.3 Statistical Analysis. Statistical analyses were perfomed in the 'R' programming environment version 4.1.2 (R Core Team, 2018). Plots were created using 'ggplot2' (Wickham, 2016). Difference in mean CH_4 flux by Region and Land Type were tested with a Kruskal-Wallis test using the 'kruskal.test' function (Hollander & Wolfe, 1973). Multiple comparisons were perfomed using Dunn's test with the Benjamini-Hochberg p-value adjustment using the 'dunnTest' function from the 'FSA' package (Dunn, 1964). Multivariate differences in soil physicochemical characteristics by Region and Land Type were tested with PERMANOVA using the 'adonis2' function from the 'vegan' package using the Euclidean distance on all soil abiotic data (McArdle & Anderson, 2001;

Oksanen et al., 2019). Multivariate differences in soil microbiome functional composition by Region and Land Type were similarly tested with PERMANOVA using the Aitchison dissimilarity (Aitchison, 1982; McArdle & Anderson, 2001). Multivariate correlation between soil physicochemical data and microbiome functional composition were tested using a Mantel test with the 'mantel.rtest' function from 'ade4' and the p-value was determined with 999 permutations (Mantel, 1967).

To identify microbiome functional markers for CH_4, I used a regression approach based on the ideas described in Chapter 2 of this dissertation (Morris et al., 2020). Methane fluxes were strongly right-skewed and so were $\log_{10}$ transformed after adding a constant to make all values positive. Next, I used a two step approach to remove multi-collinear variables. First, soil abiotic variables were evaluated for multi-collinearity based on Pearson's correlation coefficients. Any variables with a correlation coefficient > 0.70 were removed. The variables I selected for inclusion in subsequent analyses were pH, organic matter, N, P, K, S, Fe, and a principal component axis correlated with micronutrients (Cu, Mn, and Zn). Multiple linear regression was used to model CH_4 as a function of all selected abiotic variables and the geographic distance between sites. Models were evaluated for normality of the residuals and homoscedasticity. Variables were again evaluated for collinearity based on variance-inflation factors (VIF) using the 'vif' function from the 'car' package (Fox & Weisberg, 2018). Variables with VIF greater than 3 were removed from subsequent models. This analysis was repeated individually for each Region since the response of CH_4 to forest-to-pasture conversion differed by Region. For the individual regional models, latitude and longitude were converted

to distance in meters and were rotated using a PCA to pull out the main axis of the transect and the first PC of this analysis was included as geographic distance.

Using the previous multiple regression models, the model was refit for each KEGG pathway with coverage as the last variable in the model. In this way, each pathway was tested in terms of its marginal effect on CH_4 flux after controlling for the distance between sites and the underlying environmental variation. The significance threshold was adjusted for multiple comparisons using the Bonferroni correction (i.e., α/n). This process was repeated without the covariates to test each KEGG pathway on its own.

4.3 Results

Past research showed that CH_4 fluxes differed across land use types in the Amazon (Meyer, Morris, et al., 2020). In these data, CH_4 flux varied by Region and Land Type (Kruskal-Wallis: $X = 33.98, df = 5, p < 0.001$; Figure 8). However, there was a region-by-land-type interaction where forest-to-pasture conversion increased CH_4 flux in FNV (Dunn's: FNV Forest:FNV Pasture $Z = -3.06$, adj. $p < 0.001$), but not in Santarem (Dunn's: Santarem Forest:Santarem Pasture $Z = -1.70, p = 0.22$; Figure 8).

Soil physicochemical characteristics differed by Region and Land Type (Table 5; Figure 9A). However, only 10% of the variation in soil characteristics could be explained by Land Type compared to 28% across Regions. Looking at broad patterns of microbiome composition, both Land Type and Region and their interaction separated strongly by the functional composition of the soil microbiome based on Aitchison dissimilarity of KEGG pathways (Table 6; Figure 9B). Lastly, there was a weak multivariate correlation between soil physicochemical properties and soil microbiome functional composition (Mantel: $R^2 = 0.18, p = 0.003$).

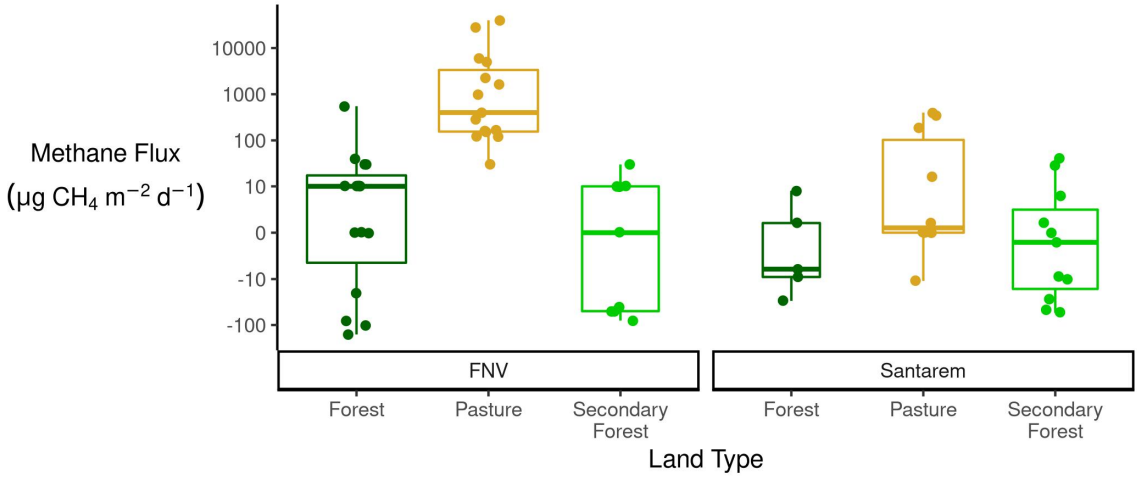

Figure 8. Soil-to-atmosphere CH_4 flux. The y-axis has been transformed with the psuedo-log base 10 for ease of interpretation. Positive values are net CH_4 production and negative values are net CH_4 consumption. Separate plots for each region (FNV and Santarem). Land-use is indicated by color (forest, pasture, or secondary forest). Recreated from Meyer, Morris, et al. (2020).

Table 5. PERMANOVA on Euclidean distance of soil characteristics.

Term	df	SS	R2	F	p
Region	1	336.88	0.28	29.16	0.001
Land_type	2	116.36	0.10	5.04	0.001
Region:Land_type	2	81.10	0.07	3.51	0.002
Residual	59	681.66	0.56	NA	NA
Total	64	1216.00	1.00	NA	NA

Next, I tested the relationship between microbial functional pathway coverage and CH_4 flux. This analysis had the goal of identifying microbiome markers for CH_4 flux in ecosystems and to generate hypotheses about what might be limiting CH_4 flux in these regions. Since the response of CH_4 to land-use change varied by region with increased CH_4 in FNV Pastures, but no increase in Santarem Pastures, I decided to test the relationship between microbial functional pathways and CH_4 flux separately for each region.

Table 6. PERMANOVA on Aitchison dissimilarity of microbiome functional composition.

Term	df	SS	R2	F	p
Region	1	10408.15	0.24	59.20	0.001
Land_type	2	13193.21	0.31	37.52	0.001
Region:Land_type	2	8651.78	0.20	24.60	0.001
Residual	59	10373.69	0.24	NA	NA
Total	64	42626.84	1.00	NA	NA

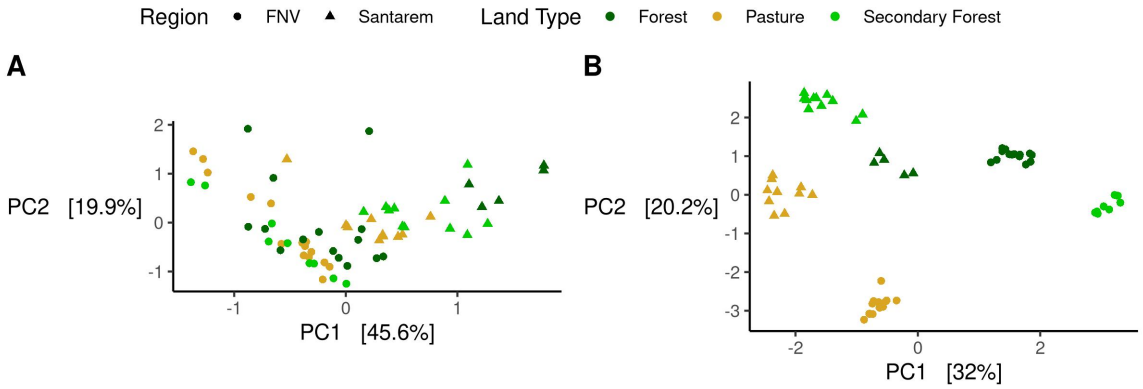

Figure 9. Principal component analysis (PCA) plot of environmental dissimilarity and microbiome function dissimilarity. **A**, PCA plot of environmental dissimilarity using euclidean distance with all soil metadata. **B**, PCA plot of microbiome functional dissimilarity using Aitchison dissimilarity of all KEGG pathways. Aitchison is euclidean distance after center-log ratio transformation. **A, B**, Land-use (forest, pasture, or secondary forest) is indicated by the color of the points and Region (FNV or Santarem) by shape. The ratio of the axes is proportional to the variance explained by each principal component in order to accurately represent the distance between the points.

For the FNV region, the full abiotic model of CH_4 included the predictors geographic distance, organic matter, nitrogen, phosphorus, potassium, iron, and sulfur after removing multicollinear variables based on variance-inflation factors. This model explained 22.5% of the variation in CH_4 flux (adj. $R^2 = 0.225$, F = 2.58, df = 7, p = 0.03). After adding KEGG pathway coverage to the model, two KEGG pathways were identified by this analysis. The first was the abscisic

acid biosynthesis pathway, which is a plant metabolic pathway (adj. $R^2 = 0.53$, estimate = 0.08, se = 0.017, t = 4.6, adj. p = 0.019). The second was the nitrogen fixation pathway (adj. $R^2 = 0.527$, estimate = 0.018, se = 0.004, t = 4.6, adj. p = 0.022). When each KEGG pathway model was fit without abiotic covariates and with KEGG pathway coverage as the only predictor, 21 pathways were significantly correlated with CH_4 flux. All but 1 were positively correlated with CH_4. Of these 21 pathways, none were directly related to CH_4 cycling. They included nitrogen fixation, sulfate-sulfur assimilation, two antimicrobial resistance genes, and a variety of metabolic genes related to carbohydrate biosynthesis and degradation.

For the Santarem region, the full abiotic model of CH_4 included the predictors geographic distance, organic matter, nitrogen, potassium, iron, and micronutrients (copper, manganese, and zinc) after removing multicollinear variables based on variance-inflation factors. This model was not significant (adj. $R^2 = 0.126$, F = 1.60, df = 6, p = 0.20). This model was then refit with coverage for each KEGG pathway as the ultimate parameter in the model. These models exhibited a uniform p-value distribution and no significant pathways were returned after Bonferroni correction. The same result was observed for models fit with only individual KEGG pathways - uniform p-value distribution and no significant pathways after Bonferroni correction.

4.4 Discussion

In this study, I wanted to understand whether it is important to measure variation in the microbiome to predict changes in ecosystem function. In addition, I wanted to identify which aspects of the microbiome drive variation in CH_4 flux independent of geographic and environmental factors. To address these questions, I analyzed CH_4 fluxes across a land-use change gradient in the Brazilian Amazon.

Forest-to-pasture conversion increased CH_4 flux in these ecosystems, as has been demonstrated before (Meyer, Morris, et al., 2020; Steudler et al., 1996; Verchot et al., 2000). The response of CH_4 to deforestation varied by region with increased CH_4 flux in FNV but not in Santarem. This suggests that the soil, climate, and/or biota differ between these regions resulting in unique responses to land-use change. Indeed, the overall soil characteristics separated much more strongly by region than by land-use type. Therefore, I chose to test the relationship between CH_4, soil characteristics, and microbiome composition separately in these two regions.

The full abiotic model of CH_4 in the Santarem region was not significant. The pasture systems in Santarem and FNV are managed differently. Santarem was converted relatively recently and is not intensively managed whereas FNV was converted much longer ago and has a more intensive rotation system. The lack of model fit may indicate that there was not enough variation in CH_4 flux in Santarem to model rates based on the variables measured. Sampling this region more intensively might reveal a significant affect, but that cannot be determined with these data.

Microbiome functional composition based on coverage of KEGG Pathways separated strongly by region, land type, and their interaction. This indicates that the regions have different microbiome compositions and displayed unique microbiome trajectories as a response to land-use change. For Santarem, I did not find any KEGG pathways that were associated with CH_4 flux with or without controlling for abiotic covariates. This may be due to relatively low variation in CH_4 at these sites. This does not necessarily mean that the microbiome is unimportant to CH_4 flux, but that in order to detect significant features at this small effect size would require greater sampling replication. In FNV, 21 pathways

were associated with CH_4 before controlling for environmental covariates, though none were related to CH_4 cycling. After controlling for abiotic covariates, nitrogen fixation was the only microbial pathway that was correlated with CH_4 emissions.

Others have also identified nitrogen fixation as an important response to land-use change. It has been reported that the number of gene copies of *nifH*, the marker gene for the nitrogen fixation pathway, increased ten-fold in pastures relative to forests and secondary forests at the FNV sites (Mirza, Potisap, Nüsslein, Bohannan, & Rodrigues, 2014). In addition, the *nifH*-inferred nitrogen fixing community composition was strongly correlated with soil nitrogen concentration and the soil C:N ratio. In that study, pastures exhibited a decrease in soil nitrogen concentration and an increase in the soil C:N ratio relative to forests, possibly driven by increased N demand by pasture plants. These conditions would favor nitrogen fixation and indicate a general state of nitrogen limitation in pastures (Mirza et al., 2014).

There are several hypotheses regarding how nitrogen fixation might be related to CH_4 cycling. One possibility is that nitrogen-fixing methanogens are increasing in relative abundance in high flux sites. Nitrogen fixation is one way for methanogens to overcome N-limitation. However, methanogenic diazotrophs are mostly prevalent in wetlands and aquatic ecosystems and so are unlikely to be important in upland pastures (Bae, Morrison, Chanton, & Ogram, 2018; Bodelier & Steenbergh, 2014). Another possibility is that nitrogen fixation provides substrates for methanogenesis. One of the primary substrates for hydrogenotrophic methanogenesis is H_2 ($CO_2 + 4\,H_2 \longrightarrow CH_4 + 2\,H_2O$), which is a byproduct of nitrogen fixation ($N_2 + 8\,H + \longrightarrow 2\,NH_3 + H_2$) (Min & Sherman, 2010). A third possibility is that nitrogen fixation could be a sign of decreased NO_3 concentrations

in pastures. Denitrification competes with methanogenesis for H_2, but NO_3 is more thermodynamically favorable than substrates for methanogenesis, such as CO_2 and acetate. Therefore, in the presence of NO_3, methanogenesis is inhibited (Klüber & Conrad, 1998). Rather than being directly related to CH_4 production through substrate production, nitrogen fixation genes could simply be a sign of low NO_3 sites where methanogenesis is more oxidatively favorable. This would not require an increase in the relative abundance of methanogens, only an increase in the thermodynamic favorability of methanogenesis.

To address these hypotheses, one could ask whether NO_3 is more limiting or nitrification is less favored in pastures. This seems to be the case in the western Amazon where the FNV sites are located. Conversion from forest to pasture decreases soil NO_3 concentrations and net nitrification rates (Neill et al., 1997). In addition, it has been reported that microbial genes involved in nitrification are more strongly associated with primary and secondary forests than pastures (Paula et al., 2014). These results would indicate that pastures are less favorable for nitrifiers and/or that soil NO_3 concentrations would be decreased in these environments, which would promote nitrogen fixation and potentially make methanogenesis more thermodynamically favorable and/or less substrate limited.

Why were there no CH_4-cycling genes associated with CH_4 flux rates when past research has demonstrated shifts in the CH_4 cycling community as a response to forest-to-pasture conversion (Kroeger et al., 2018; Meyer et al., 2017; Meyer, Morris, et al., 2020)? First, CH_4 cycling taxa are not the only ones identified by past studies; many nitrogen-cycling taxa were identified as associated with CH_4 flux in a previous 16S rRNA survey at these sites (Meyer, Morris, et al., 2020). In addition, it is not mutually exclusive that the ratio of methanogens to

methanotrophs could increase in pastures without a correlation with CH_4 flux rates. One possibility is that methanogens increase in abundance or methanotrophs decrease in abundance in pastures due to more favorable conditions, but CH_4 flux is not limited by the abundance of these functional groups. Instead, it could be limited by reactants, redox conditions, or other nutrients, which might be alleviated by or indicated by other functional groups, such as nitrogen fixers. Another possibility is that the relative abundance of CH_4-cycling organisms is strongly correlated with soil abiotic variables. Therefore, methanogens and methanotrophs track the changing abiotic conditions, but these changes do not inform our understanding of variation in CH_4 beyond what we could glean from measuring simple abiotic soil characteristics. However, the previous 16S study controlled for environmental and geographic covariates before testing the association between taxa and CH_4 flux so this should have been accounted for (Meyer, Morris, et al., 2020). In addition, I did not detect any significant CH_4 cycling pathways with the no-covariate model. It may be that since metagenomic sequencing samples the microbiome at a relatively shallow level compared to 16S that this metagenomic survey missed some relatively low abundance CH_4-cycling taxa.

One limitation of this study design and any metagenomic or marker gene survey is that sequencing data are fundamentally compositional (Calle, 2019). Therefore, an increase in relative abundance of one feature does not guarantee an increase in the absolute abundance of that feature. Future research could use results from compositional studies such as this to guide measurements of the absolute abundance of specific functional genes, such as with the *nifH* gene in Mirza et al. (2014). Alternatively, future analyses could employ compositional approaches to the analysis of metagenomic datasets to draw more robust

conclusions from sequencing data (Calle, 2019; Fernandes et al., 2014). In addition, while this study is useful for generating hypotheses about what might drive variation in CH_4 at the ecosystem-scale, evaluating those hypotheses requires controlled studies. For example, future research could target the sequencing or quantification of nitrogen fixing organisms in these field sites or perform laboratory manipulations through the addition of pure cultures or the targeted removal of nitrogen fixers with antibiotics. Alternatively, soil abiotic conditions, such as NO_3 concentrations, could be experimentally manipulated to see what effect they have on the CH_4-cycling community independent of other factors in these soils.

4.5 Conclusion

Increased CH_4 emissions from soils is a major consequence of land-use change in tropical rainforests. Deforestation leads to a switch from CH_4 consumption to CH_4 production in certain soil ecosystems. Therefore, a major research goal in the face of global change is to identify what attributes of these ecosystems lead to this shift in ecosystem function. Past research has been inconclusive, likely due to the use of phylogenetic marker genes with limited functional inference, because metagenomic surveys were not co-located in time and space with in situ CH_4 flux measurements, and because abiotic covariates were not consistently considered. In this study, I found that soil abiotic variables were not strongly informative of CH_4 flux rates. In addition, CH_4-cycling organisms were not correlated with CH_4 flux, possibly due to collinearity with abiotic variables or because CH_4 fluxes are not limited by the abundance of these organisms. Nitrogen fixation was the pathway most associated with CH_4 flux. This could be because nitrogen fixation directly provides substrates for methanogenesis or because it is an indicator of environmental conditions favorable for methanogenesis, such as low

NO_3 concentrations. Microbiome biodiversity-ecosystem function studies typically assume that the gene most limiting to the rate of an ecosystem function is the gene coding for the final enzyme in a pathway, such as the enzyme encoded by the *mcrA* gene in methanogenesis. This assumption may not be valid for natural ecosystems in which a process may be limited by the products of other pathways or by unfavorable redox conditions. Therefore, future studies should take a more agnostic approach to testing genes associated with the rate of ecosystem function.

CHAPTER V

CONCLUSION

Microbiomes mediate important ecosystem functions and yet it remains unclear if understanding variation in the microbiome is important for predicting the rate of ecosystem function. While it is generally true that the rate of some ecosystem functions is correlated with microbiome composition, describing a more detailed mapping between microbial biodiversity and ecosystem function remains elusive. To address this challenge, I proposed a framework for understanding the microbiome-ecosystem function mapping as analogous to the mapping between genomic variation and phenotype in organisms. Drawing on this analogy, I designed both an experimental approach and a comparative approach for evaluating the relationship between the soil microbiome and the rate of CH_4 emissions from soil.

In the experimental approach, I performed artificial ecosystem selection at the whole-community level by selecting for microbiomes with high rates of CH_4 oxidation. I found a strong response to selection on ecosystem-level microbiome traits and strong heritability of CH_4 oxidation by the soil microbiome. This indicates that understanding variation in microbiome composition is potentially important for predicting future variation in the rate of ecosystem CH_4 flux. I also identified many taxa that were enriched as a response to selection. To apply this to a real-world example, I used a land-use change study in the Brazilian Amazon to test which functional pathways were associated with elevated CH_4 emissions after forest-to-pasture conversion. The only microbial function associated with high CH_4 flux sites was nitrogen fixation. In both of these studies, the strongest microbiome predictors of CH_4 flux were not directly related to CH_4 cycling.

The typical assumption for microbiome ecosystem function relationships is that the abundance of a functional group is the primary limiting factor for the function performed by that group. My results show that this is not necessarily the case for CH_4 cycling in soil. Instead, the rate of CH_4 emissions could be limited by other functional groups that are not directly related to CH_4, such as nitrogen fixers. Alternatively, CH_4 fluxes could be regulated by multiple taxa, much like quantitative traits in organisms that are controlled by many genetic loci. Using this framework, mcirobial ecologists can begin to more precisely describe the mechanisms behind variation in the rate of microbial ecosystem functions. By better understanding what drives variation in ecosystem function, we will be better equipped to predict future ecosystem functions under global change and to manage ecosystems for particular outcomes.

REFERENCES CITED

Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(2), 139–160. doi: 10.1111/j.2517-6161.1982.tb01195.x

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020, April). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, *36*(7), 2251–2252. doi: 10.1093/bioinformatics/btz859

Arias-Sánchez, F. I., Vessman, B., & Mitri, S. (2019, August). Artificially selecting microbial communities: If we can breed dogs, why not microbiomes? *PLOS Biology*, *17*(8), e3000356. doi: 10.1371/journal.pbio.3000356

Bae, H.-S., Morrison, E., Chanton, J. P., & Ogram, A. (2018, April). Methanogens Are Major Contributors to Nitrogen Fixation in Soils of the Florida Everglades. *Applied and Environmental Microbiology*, *84*(7), e02222-17. doi: 10.1128/AEM.02222-17

Balser, T. C., & Firestone, M. K. (2005, April). Linking microbial community composition and soil processes in a California annual grassland and mixed-conifer forest. *Biogeochemistry*, *73*(2), 395–415. doi: 10.1007/s10533-004-0372-y

Barnett, D. J., Arts, I. C., & Penders, J. (2021, July). microViz: An R package for microbiome data visualization and statistics. *Journal of Open Source Software*, *6*(63), 3201. doi: 10.21105/joss.03201

Bay, S. K., Dong, X., Bradley, J. A., Leung, P. M., Grinter, R., Jirapanjawat, T., ... Greening, C. (2021, February). Trace gas oxidizers are widespread and active members of soil microbial communities. *Nature Microbiology*, *6*(2), 246–256. doi: 10.1038/s41564-020-00811-w

Bédard, C., & Knowles, R. (1989, March). Physiology, biochemistry, and specific inhibitors of CH4, NH4+, and CO oxidation by methanotrophs and nitrifiers. *Microbiological Reviews*, *53*(1), 68–84. doi: 10.1128/mr.53.1.68-84.1989

Belova, S. E., Pankratov, T. A., Detkova, E. N., Kaparullina, E. N., & Dedysh, S. N. . (2009). Acidisoma tundrae gen. nov., sp. nov. and Acidisoma sibiricum sp. nov., two acidophilic, psychrotolerant members of the Alphaproteobacteria from acidic northern wetlands. *International Journal of Systematic and Evolutionary Microbiology*, *59*(9), 2283–2290. doi: 10.1099/ijs.0.009209-0

Berg, M., & Koskella, B. (2018, August). Nutrient- and Dose-Dependent Microbiome-Mediated Protection against a Plant Pathogen. *Current Biology*, *28*(15), 2487-2492.e3. doi: 10.1016/j.cub.2018.05.085

Bižić, M., Klintzsch, T., Ionescu, D., Hindiyeh, M. Y., Günthel, M., Muro-Pastor, A. M., ... Grossart, H.-P. (2020, January). Aquatic and terrestrial cyanobacteria produce methane. *Science Advances*, *6*(3), eaax5343. doi: 10.1126/sciadv.aax5343

Bodelier, P. L., & Steenbergh, A. K. (2014, November). Interactions between methane and the nitrogen cycle in light of climate change. *Current Opinion in Environmental Sustainability*, *9–10*, 26–36. doi: 10.1016/j.cosust.2014.07.004

Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, *27*(4), 325–349. doi: 10.2307/1942268

Bush, W. S., & Moore, J. H. (2012, December). Chapter 11: Genome-Wide Association Studies. *PLOS Computational Biology*, *8*(12), e1002822. doi: 10.1371/journal.pcbi.1002822

Bushnell, B. (2021). *BBMap.*

Bushnell, B., Rood, J., & Singer, E. (2017, October). BBMerge – Accurate paired shotgun read merging via overlap. *PLOS ONE*, *12*(10), e0185056. doi: 10.1371/journal.pone.0185056

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017, December). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*(12), 2639–2643. doi: 10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016, July). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. doi: 10.1038/nmeth.3869

Calle, M. L. (2019, March). Statistical Analysis of Metagenomics Data. *Genomics & Informatics*, *17*(1), e6. doi: 10.5808/GI.2019.17.1.e6

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., ... Knight, R. (2011, March). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, *108*(Supplement 1), 4516–4522. doi: 10.1073/pnas.1000080107

Cardinale, B. J., Srivastava, D. S., Emmett Duffy, J., Wright, J. P., Downing, A. L., Sankaran, M., & Jouseau, C. (2006, October). Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature*, *443*(7114), 989–992. doi: 10.1038/nature05202

Cavigelli, M. A., & Robertson, G. P. (2000). The Functional Significance of Denitrifier Community Composition in a Terrestrial Ecosystem. *Ecology*, *81*(5), 1402–1414. doi: 10.1890/0012-9658(2000)081[1402:TFSODC]2.0.CO;2

Chang, H.-X., Haudenshield, J. S., Bowen, C. R., & Hartman, G. L. (2017). Metagenome-Wide Association Study and Machine Learning Prediction of Bulk Soil Microbiome and Crop Productivity. *Frontiers in Microbiology*, *8*, 519. doi: 10.3389/fmicb.2017.00519

Chang, Y.-j., Land, M., Hauser, L., Chertkov, O., Glavina Del Rio, T., Nolan, M., ... Lapidus, A. (2011, July). Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium Ktedonobacter racemifer type strain (SOSP1-21T). *Standards in Genomic Sciences*, *5*(1), 97–111. doi: 10.4056/sigs.2114901

Conner, J. K. (2003). Artificial Selection: A Powerful Tool for Ecologists. *Ecology*, *84*(7), 1650–1660.

Crowther, T. W., van den Hoogen, J., Wan, J., Mayes, M. A., Keiser, A. D., Mo, L., ... Maynard, D. S. (2019, August). The global soil community and its influence on biogeochemistry. *Science*, *365*(6455). doi: 10.1126/science.aav0550

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018, December). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*(1), 226. doi: 10.1186/s40168-018-0605-2

Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D., ... Singh, B. K. (2016, January). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nature Communications*, *7*(1), 10541. doi: 10.1038/ncomms10541

Delgado-Baquerizo, M., Powell, J. R., Hamonts, K., Reith, F., Mele, P., Brown, M. V., ... Bissett, A. (2017). Circular linkages between soil biodiversity, fertility and plant productivity are limited to topsoil at the continental scale. *New Phytologist*, *215*(3), 1186–1196. doi: 10.1111/nph.14634

Dirzo, R., & Raven, P. H. (2003). Global State of Biodiversity and Loss. *Annual Review of Environment and Resources*, *28*(1), 137–167. doi: 10.1146/annurev.energy.28.050302.105532

Domeignoz-Horta, L. A., DeAngelis, K. M., & Pold, G. (2019). Draft Genome Sequence of Acidobacteria Group 1 Acidipila sp. Strain EB88, Isolated from Forest Soil. *Microbiology Resource Announcements*, *8*(1), e01464-18. doi: 10.1128/MRA.01464-18

Dunn, O. (1964). Multiple comparisons using rank sums. *Technometrics*(6), 241–252.

Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., ... Willis, A. D. (2021, January). Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology*, *6*(1), 3–6. doi: 10.1038/s41564-020-00834-3

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014, February). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, *2*(1), 6. doi: 10.1186/2049-2618-2-6

Falconer, D. S. D. S., & MacKay, T. F. C. (1996). *Introduction to quantitative genetics*. Harlow : Prentice Hall.

Fernandes, A. D., Reid, J. N., Macklaim, J. M., McMurrough, T. A., Edgell, D. R., & Gloor, G. B. (2014, May). Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome*, *2*(1), 15. doi: 10.1186/2049-2618-2-15

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., ... Snyder, P. K. (2005, July). Global Consequences of Land Use. *Science*. doi: 10.1126/science.1111772

Fox, J., & Weisberg, S. (2018). *An R Companion to Applied Regression* (Third Edition ed.). Sage.

Freitag, T. E., & Prosser, J. I. (2003, March). Community Structure of Ammonia-Oxidizing Bacteria within Anoxic Marine Sediments. *Applied and Environmental Microbiology*.

Freitag, T. E., & Prosser, J. I. (2009, November). Correlation of Methane Production and Functional Gene Transcriptional Activity in a Peat Soil. *Applied and Environmental Microbiology*, *75*(21), 6679–6687. doi: 10.1128/AEM.01021-09

Freitag, T. E., Toet, S., Ineson, P., & Prosser, J. I. (2010, July). Links between methane flux and transcriptional activities of methanogens and methane oxidizers in a blanket peat bog. *FEMS Microbiology Ecology*, *73*(1), 157–165. doi: 10.1111/j.1574-6941.2010.00871.x

Friedrich, M. W. (2005, January). Methyl-Coenzyme M Reductase Genes: Unique Functional Markers for Methanogenic and Anaerobic Methane-Oxidizing Archaea. In *Methods in Enzymology* (Vol. 397, pp. 428–442). Academic Press. doi: 10.1016/S0076-6879(05)97026-2

Fritz, I., Strömpl, C., & Abraham, W.-R. . (2004). Phylogenetic relationships of the genera Stella, Labrys and Angulomicrobium within the 'Alphaproteobacteria' and description of Angulomicrobium amanitiforme sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *54*(3), 651–657. doi: 10.1099/ijs.0.02746-0

Fuhrman, J. A. (2009, May). Microbial community structure and its functional implications. *Nature*, *459*(7244), 193–199. doi: 10.1038/nature08058

Gamfeldt, L., Lefcheck, J. S., Byrnes, J. E. K., Cardinale, B. J., Duffy, J. E., & Griffin, J. N. (2015). Marine biodiversity and ecosystem functioning: What's known and what's next? *Oikos*, *124*(3), 252–265. doi: 10.1111/oik.01549

Garrity, G. M., Bell, J. A., & Lilburn, T. G. (2005). Class III. Gammaproteobacteria class. nov., p. 1. In D. J. Brenner, N. R. Krieg, J. T. Staley, & G. M. Garrity (Eds.), *Bergey's manual of systematic bacteriology* (Second ed., Vol. 2). New York, NY: Springer.

Gilbert, J. A., Quinn, R. A., Debelius, J., Xu, Z. Z., Morton, J., Garg, N., ... Knight, R. (2016, July). Microbiome-wide association studies link dynamic microbial consortia to disease. *Nature*, *535*(7610), 94–103. doi: 10.1038/nature18850

Glassman, S. I., Weihe, C., Li, J., Albright, M. B. N., Looby, C. I., Martiny, A. C., ... Martiny, J. B. H. (2018, November). Decomposition responses to climate depend on microbial community composition. *Proceedings of the National Academy of Sciences*, *115*(47), 11994–11999. doi: 10.1073/pnas.1811269115

Goodnight, C. J. (2000, August). Heritability at the ecosystem level. *Proceedings of the National Academy of Sciences*, *97*(17), 9365–9366. doi: 10.1073/pnas.97.17.9365

Grace, J. B., Anderson, T. M., Olff, H., & Scheiner, S. M. (2010). On the specification of structural equation models for ecological systems. *Ecological Monographs*, *80*(1), 67–87. doi: 10.1890/09-0464.1

Graham, E. B., Knelman, J. E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., ... Nemergut, D. R. (2016). Microbes as Engines of Ecosystem Function: When Does Community Structure Enhance Predictions of Ecosystem Processes? *Frontiers in Microbiology*, *0*. doi: 10.3389/fmicb.2016.00214

Hahn, M. W., Schmidt, J., Koll, U., Rohde, M., Verbarg, S., Pitt, A., ... Lang, E. . (2017). Silvanigrella aquatica gen. nov., sp. nov., isolated from a freshwater lake, description of Silvanigrellaceae fam. nov. and Silvanigrellales ord. nov., reclassification of the order Bdellovibrionales in the class Oligoflexia, reclassification of the families Bacteriovoracaceae and Halobacteriovoraceae in the new order Bacteriovoracales ord. nov., and reclassification of the family Pseudobacteriovoracaceae in the order Oligoflexales. *International Journal of Systematic and Evolutionary Microbiology*, *67*(8), 2555–2568. doi: 10.1099/ijsem.0.001965

Hall, E. K., Bernhardt, E. S., Bier, R. L., Bradford, M. A., Boot, C. M., Cotner, J. B., ... Wallenstein, M. D. (2018, September). Understanding how microbiomes influence the systems they inhabit. *Nature Microbiology*, *3*(9), 977–982. doi: 10.1038/s41564-018-0201-z

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009, June). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, *106*(23), 9362–9367. doi: 10.1073/pnas.0903103106

Ho, A., Kerckhof, F.-M., Luke, C., Reim, A., Krause, S., Boon, N., & Bodelier, P. L. E. (2013). Conceptualizing functional traits and ecological characteristics of methane-oxidizing bacteria as life strategies. *Environmental Microbiology Reports*, *5*(3), 335–345. doi: 10.1111/j.1758-2229.2012.00370.x

Hollander, M., & Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. New York: John Wiley & Sons.

Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E. K., Hungate, B. A., Matulich, K. L., ... O'Connor, M. I. (2012, June). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*, *486*(7401), 105–108. doi: 10.1038/nature11118

Hooper, D. U., Chapin, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., ... Wardle, D. A. (2005). Effects of Biodiversity on Ecosystem Functioning: A Consensus of Current Knowledge. *Ecological Monographs*, *75*(1), 3–35. doi: 10.1890/04-0922

Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., ... Banfield, J. F. (2016, April). A new view of the tree of life. *Nature Microbiology*, *1*(5), 1–6. doi: 10.1038/nmicrobiol.2016.48

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010, March). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*(1), 119. doi: 10.1186/1471-2105-11-119

IPCC. (2013). Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *[Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA*, 1535.

Jing, X., Sanders, N. J., Shi, Y., Chu, H., Classen, A. T., Zhao, K., ... He, J.-S. (2015, September). The links between ecosystem multifunctionality and above- and belowground biodiversity are mediated by climate. *Nature Communications*, *6*(1), 8159. doi: 10.1038/ncomms9159

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S.-y., Freimer, N. B., ... Eskin, E. (2010, April). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, *42*(4), 348–354. doi: 10.1038/ng.548

Kerem, B., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., ... Tsui, L. C. (1989, September). Identification of the cystic fibrosis gene: Genetic analysis. *Science*, *245*(4922), 1073–1080. doi: 10.1126/science.2570460

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., ... Zeng, G. (2013, October). Three decades of global methane sources and sinks. *Nature Geoscience*, *6*(10), 813–823. doi: 10.1038/ngeo1955

Klüber, H. D., & Conrad, R. (1998, March). Effects of nitrate, nitrite, NO and N2O on methanogenesis and other redox processes in anoxic rice field soil. *FEMS Microbiology Ecology*, *25*(3), 301–318. doi: 10.1111/j.1574-6941.1998.tb00482.x

Kolb, S., Knief, C., Stubner, S., & Conrad, R. (2003, May). Quantitative Detection of Methanotrophs in Soil by Novel pmoA-Targeted Real-Time PCR Assays. *Applied and Environmental Microbiology*, *69*(5), 2423–2429. doi: 10.1128/AEM.69.5.2423-2429.2003

Koskella, B., & Meaden, S. (2013, March). Understanding Bacteriophage Specificity in Natural Microbial Communities. *Viruses*, *5*(3), 806–823. doi: 10.3390/v5030806

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013, September). Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*.

Kroeger, M. E., Delmont, T. O., Eren, A. M., Meyer, K. M., Guo, J., Khan, K., ... Nüsslein, K. (2018). New Biological Insights Into How Deforestation in Amazonia Affects Soil Microbial Communities Using Metagenomics and Metagenome-Assembled Genomes. *Frontiers in Microbiology*, *9*, 1635. doi: 10.3389/fmicb.2018.01635

Langmead, B., & Salzberg, S. L. (2012, April). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. doi: 10.1038/nmeth.1923

Laurance, W. F., Sayer, J., & Cassman, K. G. (2014, February). Agricultural expansion and its impacts on tropical nature. *Trends in Ecology & Evolution*, *29*(2), 107–116. doi: 10.1016/j.tree.2013.12.001

Lenhart, K., Bunge, M., Ratering, S., Neu, T. R., Schüttmann, I., Greule, M., ... Keppler, F. (2012, September). Evidence for methane production by saprotrophic fungi. *Nature Communications*, *3*(1), 1046. doi: 10.1038/ncomms2049

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015, May). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. doi: 10.1093/bioinformatics/btv033

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., ... Huttenhower, C. (2019, May). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*, *569*(7758), 655–662. doi: 10.1038/s41586-019-1237-9

Love, M. I., Huber, W., & Anders, S. (2014, December). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. doi: 10.1186/s13059-014-0550-8

MacDonald, M. E., Novelletto, A., Lin, C., Tagle, D., Barnes, G., Bates, G., ... Gusella, J. F. (1992, May). The Huntington's disease candidate region exhibits many different haplotypes. *Nature Genetics*, *1*(2), 99–103. doi: 10.1038/ng0592-99

Mantel, N. (1967, February). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2), 209–220.

Maron, P.-A., Sarr, A., Kaisermann, A., Lévêque, J., Mathieu, O., Guigue, J., ... Ranjard, L. (2018, February). High Microbial Diversity Promotes Soil Ecosystem Functioning. *Applied and Environmental Microbiology*, *84*(9), e02738-17. doi: 10.1128/AEM.02738-17

Martin, B. D., Witten, D., & Willis, A. D. (2020, March). Modeling microbial abundances and dysbiosis with beta-binomial regression. *The Annals of Applied Statistics*, *14*(1), 94–115. doi: 10.1214/19-AOAS1283

Martiny, A. C. (2019, August). High proportions of bacteria are culturable across major biomes. *The ISME Journal*, *13*(8), 2125–2128. doi: 10.1038/s41396-019-0410-3

Martiny, A. C., Treseder, K., & Pusch, G. (2013, April). Phylogenetic conservatism of functional traits in microorganisms. *The ISME Journal*, *7*(4), 830–838. doi: 10.1038/ismej.2012.160

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., ... Staley, J. T. (2006, February). Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology*, *4*(2), 102–112. doi: 10.1038/nrmicro1341

Maxson, T., & Mitchell, D. A. (2016, June). Targeted treatment for bacterial infections: Prospects for pathogen-specific antibiotics coupled with rapid diagnostics. *Tetrahedron*, *72*(25), 3609–3624. doi: 10.1016/j.tet.2015.09.069

McArdle, B. H., & Anderson, M. J. (2001). Fitting Multivariate Models to Community Data: A Comment on Distance-Based Redundancy Analysis. *Ecology*, *82*(1), 290–297. doi: 10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2

McBride, M. J., Liu, W., Lu, X., Zhu, Y., & Zhang, W. (2014). The Family Cytophagaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (pp. 577–593). Berlin, Heidelberg: Springer. doi: 10.1007/978-3-642-38954-2_382

McCalley, C. K., Woodcroft, B. J., Hodgkins, S. B., Wehr, R. A., Kim, E.-H., Mondav, R., ... Saleska, S. R. (2014, October). Methane dynamics regulated by microbial community response to permafrost thaw. *Nature*, *514*(7523), 478–481. doi: 10.1038/nature13798

McMurdie, P. J., & Holmes, S. (2013, April). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, *8*(4), e61217. doi: 10.1371/journal.pone.0061217

McMurdie, P. J., & Holmes, S. (2014, April). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology*, *10*(4), e1003531. doi: 10.1371/journal.pcbi.1003531

Meyer, K. M., Hopple, A. M., Klein, A. M., Morris, A. H., Bridgham, S. D., & Bohannan, B. J. M. (2020). Community structure – Ecosystem function relationships in the Congo Basin methane cycle depend on the physiological scale of function. *Molecular Ecology*, *29*(10), 1806–1819. doi: 10.1111/mec.15442

Meyer, K. M., Klein, A. M., Rodrigues, J. L. M., Nüsslein, K., Tringe, S. G., Mirza, B. S., ... Bohannan, B. J. M. (2017). Conversion of Amazon rainforest to agriculture alters community traits of methane-cycling organisms. *Molecular Ecology*, *26*(6), 1547–1556. doi: 10.1111/mec.14011

Meyer, K. M., Morris, A. H., Webster, K., Klein, A. M., Kroeger, M. E., Meredith, L. K., ... Bohannan, B. J. M. (2020, December). Belowground changes to community structure alter methane-cycling dynamics in Amazonia. *Environment International*, *145*, 106131. doi: 10.1016/j.envint.2020.106131

Min, H., & Sherman, L. A. (2010, July). Hydrogen Production by the Unicellular, Diazotrophic Cyanobacterium Cyanothece sp. Strain ATCC 51142 under Conditions of Continuous Light. *Applied and Environmental Microbiology*. doi: 10.1128/AEM.00146-10

Mirza, B. S., Potisap, C., Nüsslein, K., Bohannan, B. J. M., & Rodrigues, J. L. M. (2014, January). Response of Free-Living Nitrogen-Fixing Microorganisms to Land Use Change in the Amazon Rainforest. *Applied and Environmental Microbiology*. doi: 10.1128/AEM.02362-13

Morris, A., Meyer, K., & Bohannan, B. (2020, May). Linking microbial communities to ecosystem functions: What we can learn from genotype–phenotype mapping in organisms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1798), 20190244. doi: 10.1098/rstb.2019.0244

Neill, C., Piccolo, M. C., Cerri, C. C., Steudler, P. A., Melillo, J. M., & Brito, M. (1997, April). Net nitrogen mineralization and net nitrification rates in soils following deforestation for pasture across the southwestern Brazilian Amazon Basin landscape. *Oecologia*, *110*(2), 243–252. doi: 10.1007/s004420050157

Newbold, T., Hudson, L. N., Hill, S. L. L., Contu, S., Lysenko, I., Senior, R. A., ... Purvis, A. (2015, April). Global effects of land use on local terrestrial biodiversity. *Nature*, *520*(7545), 45–50. doi: 10.1038/nature14324

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2019). *Vegan: Community ecology package.*

Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., & Kao-Kniffin, J. (2015, March). Selection on soil microbiomes reveals reproducible impacts on plant function. *The ISME journal*, *9*(4), 980–989. doi: 10.1038/ismej.2014.196

Paula, F. S., Rodrigues, J. L. M., Zhou, J., Wu, L., Mueller, R. C., Mirza, B. S., ... Pellizari, V. H. (2014, June). Land use change alters functional gene diversity, composition and abundance in Amazon forest soil microbial communities. *Molecular Ecology*, *23*(12), 2988–2999. doi: 10.1111/mec.12786

Philippot, L., Spor, A., Hénault, C., Bru, D., Bizouard, F., Jones, C. M., ... Maron, P.-A. (2013, August). Loss in microbial diversity affects nitrogen cycling in soil. *The ISME Journal*, *7*(8), 1609–1619. doi: 10.1038/ismej.2013.34

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006, August). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. doi: 10.1038/ng1847

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., ... Wang, J. (2012, October). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*, *490*(7418), 55–60. doi: 10.1038/nature11450

Qu, L., Guennel, T., & Marshall, S. L. (2013). Linear Score Tests for Variance Components in Linear Mixed Models and Applications to Genetic Association Studies. *Biometrics*, *69*(4), 883–892. doi: 10.1111/biom.12095

R Core Team. (2018). *R: A language and environment for statistical computing* [Manual]. Vienna, Austria.

Reed, H. E., & Martiny, J. B. H. (2007). Testing the functional significance of microbial composition in natural communities. *FEMS Microbiology Ecology*, *62*(2), 161–170. doi: 10.1111/j.1574-6941.2007.00386.x

Reich, D. E., & Lander, E. S. (2001, September). On the allelic spectrum of human disease. *Trends in Genetics*, *17*(9), 502–510. doi: 10.1016/S0168-9525(01)02410-6

Rocca, J. D., Hall, E. K., Lennon, J. T., Evans, S. E., Waldrop, M. P., Cotner, J. B., ... Wallenstein, M. D. (2015, August). Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *The ISME Journal*, *9*(8), 1693–1699. doi: 10.1038/ismej.2014.252

Saunois, M., Bousquet, P., Poulter, B., Peregon, A., Ciais, P., Canadell, J. G., ... Zhu, Q. (2016, December). The global methane budget 2000–2012. *Earth System Science Data*, *8*(2), 697–751. doi: 10.5194/essd-8-697-2016

Schaefer, H., Fletcher, S. E. M., Veidt, C., Lassey, K. R., Brailsford, G. W., Bromley, T. M., ... White, J. W. C. (2016, April). A 21st-century shift from fossil-fuel to biogenic methane emissions indicated by 13CH4. *Science*, *352*(6281), 80–84. doi: 10.1126/science.aad2705

Schimel, J. P., & Gulledge, J. (1998). Microbial community structure and global trace gases. *Global Change Biology*, *4*(7), 745–758. doi: 10.1046/j.1365-2486.1998.00195.x

Schnyder, E., Bodelier, P. L. E., Hartmann, M., Henneberger, R., & Niklaus, P. A. (2018). Positive diversity-functioning relationships in model communities of methanotrophic bacteria. *Ecology*, *99*(3), 714–723. doi: 10.1002/ecy.2138

Shah, N., Meisel, J. S., & Pop, M. (2019). Embracing Ambiguity in the Taxonomic Classification of Microbiome Sequencing Data. *Frontiers in Genetics*, *10*, 1022. doi: 10.3389/fgene.2019.01022

Singh, B. K., Bardgett, R. D., Smith, P., & Reay, D. S. (2010, November). Microorganisms and climate change: Terrestrial feedbacks and mitigation options. *Nature Reviews Microbiology*, *8*(11), 779–790. doi: 10.1038/nrmicro2439

Steudler, P. A., Melillo, J. M., Feigl, B. J., Ill, C. N., Piccolo, M. C., & Cerri, C. C. (1996). *Consequence of forest-to-pasture conversion on CH 4 fluxes in the Brazilian Amazon Basin.*

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 479–498. doi: 10.1111/1467-9868.00346

Strickland, M. S., Lauber, C., Fierer, N., & Bradford, M. A. (2009). Testing the functional significance of microbial community composition. *Ecology*, *90*(2), 441–451. doi: 10.1890/08-0296.1

Swenson, W., Arendt, J., & Wilson, D. S. (2000, October). Artificial selection of microbial ecosystems for 3-chloroaniline biodegradation. *Environmental Microbiology*, *2*(5), 564–571. doi: 10.1046/j.1462-2920.2000.00140.x

Swenson, W., Wilson, D. S., & Elias, R. (2000, August). Artificial ecosystem selection. *Proceedings of the National Academy of Sciences*, *97*(16), 9110–9114. doi: 10.1073/pnas.150237597

Tamaki, H., Tanaka, Y., Matsuzawa, H., Muramatsu, M., Meng, X.-Y., Hanada, S., ... Kamagata, Y. . (2011). Armatimonas rosea gen. nov., sp. nov., of a novel bacterial phylum, Armatimonadetes phyl. nov., formally called the candidate phylum OP10. *International Journal of Systematic and Evolutionary Microbiology*, *61*(6), 1442–1447. doi: 10.1099/ijs.0.025643-0

Van Der Heijden, M. G. A., Bardgett, R. D., & Van Straalen, N. M. (2008). The unseen majority: Soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, *11*(3), 296–310. doi: 10.1111/j.1461-0248.2007.01139.x

Verchot, L. V., Davidson, E. A., Cattânio, J. H., & Ackerman, I. L. (2000, January). Land-Use Change and Biogeochemical Controls of Methane Fluxes in Soils of Eastern Amazonia. *Ecosystems*, *3*(1), 41–56. doi: 10.1007/s100210000009

Verstraete, W. (2007, May). Microbial ecology and environmental biotechnology. *The ISME Journal*, *1*(1), 4–8. doi: 10.1038/ismej.2007.7

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008, April). Heritability in the genomics era — concepts and misconceptions. *Nature Reviews Genetics*, *9*(4), 255–266. doi: 10.1038/nrg2322

Wagg, C., Bender, S. F., Widmer, F., & van der Heijden, M. G. A. (2014, April). Soil biodiversity and soil community composition determine ecosystem multifunctionality. *Proceedings of the National Academy of Sciences*, *111*(14), 5266–5270. doi: 10.1073/pnas.1320054111

Wagner, R., Zona, D., Oechel, W., & Lipson, D. (2017). Microbial community structure and soil pH correspond to methane production in Arctic Alaska soils. *Environmental Microbiology*, *19*(8), 3398–3410. doi: 10.1111/1462-2920.13854

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007, August). Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. doi: 10.1128/AEM.00062-07

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.

Willis, A., & Bunge, J. (2015). Estimating diversity via frequency ratios. *Biometrics*, *71*(4), 1042–1049. doi: 10.1111/biom.12332

Willis, A. D., & Martin, B. D. (2020, May). Estimating diversity in networked ecological communities. *Biostatistics*(kxaa015). doi: 10.1093/biostatistics/kxaa015

Wright, S. (1943, March). Isolation by Distance. *Genetics*, *28*(2), 114–138.