ABSTRACT

Title of Dissertation: **AUGMENTING SEQUENCING TECHNOLOGY FOR BETTER INFERENCE IN SOIL MICROBIOME ANALYSIS**

Dietrich Jonathan Epp Schmidt, Doctor of Philosophy, 2023

Dissertation directed by: Professor Stephanie Yarwood,
Environmental Science and Technology Department

The advent of DNA sequencing revolutionized the field of microbiome research. Many organisms, by virtue of their codependence and/or growth rate, are either impossible or extremely challenging to get into pure culture. Sequencing allows important taxonomic and phylogenetic information to be obtained independent of culturing. Development of the sequencing technology itself has allowed for high throughput workflow that has allowed low cost and extensive sampling of microbiomes across environments. The co-development of reference datasets for taxonomy and functional assignments, along with open-source bioinformatics pipelines has further empowered scientists to explore microbiomes in many environments. However, there are limitations to sequence data that have constrained the ecological inferences in microbiome research. One such limitation, the compositional nature of sequence data, has impeded our ability to make accurate inferences about the environmental drivers of

taxon abundance and covariance across conditions. In this dissertation I explore the use of quantitative PCR in combination with sequencing techniques to generate “Quantitative Sequencing” data (QSeq) that mitigates the limitations of compositionality on inferences relating to taxon abundance and covariance across environmental gradients.

In chapter 1, I reviewed key characteristics of the soil environment and sequencing as a mechanism for sampling. In chapter 2, I leveraged modeling, synthesis, and literature review methods to establish the questions and data characteristics that demand QSeq methodology. I show that even small amounts of variation in total abundance make determining the effects of environment (biotic and abiotic factors) on any given taxon unreliable without QSeq. In Chapter 3, I extend the logic of quantitative sequencing to improve metagenome prediction from PICRUSt2. Using data synthesis methods, accounting for 16S gene abundance consistently improved the accuracy of predicted functional genes. This was confirmed by high correlations between predicted and measured gene abundance (QPCR). There was however a large variation in prediction accuracy, likely due in part to database biases and in part to decoupling of bacterial function from taxonomy. In Chapter 4, I applied QSeq in the context of an experimental, long-term farming system that has large gradients in total abundance with depth, and I used QSeq to identify taxa that changed in abundance due to different farming system management and soil depth. Finally in Chapter 5, I used QSeq to identify putative N-fixing taxa that responded to glyphosate in four experimental farming systems. I show that the abundance of these taxa were decoupled from other effects of glyphosate on N-fixation in soybean across farming systems.

AUGMENTING SEQUENCING TECHNOLOGY FOR BETTER INFERENCE IN SOIL MICROBIOME ANALYSIS

by

Dietrich Jonathan Epp Schmidt

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Stephanie Yarwood, Chair
Dr. Michel Cavigelli
Dr. Jude Maul
Professor Mihai Pop
Professor. Kate Tully

Dedication

For my parents Heidi and Tim,

for always encouraging my interests in the ecosystems around me.

Acknowledgements

This work would not have been possible without the tangible and emotional support of many people. First, I must thank my advisor, Dr Stephanie Yarwood. If you hadn’t taken the considerable risk agreeing to be my advisor without funding available at the beginning of my MS program and then fought for years to make sure I was supported, I would not be here.

To Dr. Jude Maul, Dr. Michel Cavigelli, Dr. Mihai Pop, and Dr. Kate Tully: thank you for your patient and thoughtful guidance, I could not have asked for a better committee.

To all the members of the Yarwood lab, in roughly chronological order, Glade Bogar, Christine Maietta, Holly Roussey, Tammy Walksy, Ashley and Gina Robey, Martina Gonzalez Mateu, Lindsay Wood, Eni Baballari, Tori Monsaint-Queeney, Zach Berry, Zack Bernstein, Amy Kuritzky, Jessie King, Matthew Silverman, Josh Gaimaro, and Nora Hamovit: thank you for fostering a fun and engaging work environment that was always welcoming; thank you for sharing your expertise and insight, and getting me unstuck when I hit a wall; thank you for the many words of encouragement; and thank you for pushing me to grow. To the many additional graduate students and friends, thank you for being there with me through the late nights, early mornings, successes, and failures. Your friendships made graduate school easy.

To my family: I can’t easily articulate the myriad of ways you have prepared me and supported me through this adventure. Thank you! To Kory, thank you for interrupting EVERY single zoom call. To my wife Audrey: “something witty ;)”

Finally, this work was supported in part by the following:

- NRT-INFEWS: UMD Global STEWARDS (STEM Training at the Nexus of Energy, WAter Reuseand FooD Systems) that was awarded to the University of Maryland School of Public Health by the National Science Foundation National Research Traineeship Program, Grant number 1828910.
- USDA-ARS Long-Term Agricultural Research (LTAR) network
- USDA NIFA Organic Transitions Grant (USDA-NIFA-ICGP-003448).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

amoA: ammonia mono-oxygenase subunit A

DNA: Deoxyribonucleic acid

RNA: Ribonucleic acid

mcrA: methyl coenzyme M reductase alpha subunit

nifH: Nickle-Molybdenum nitrogenase subunit component H

nirK: dissimilatory nitrite reductase containing copper

nirS: dissimilatory nitrite reductase containing cytochrome cd_1

nosZ: nitrous oxide reductase

pmoA: particulate methane mono-oxygenase subunit A

QPCR: Quantitative Polymerase Chain Reaction

QSeq: Quantitative Sequencing

RNAseq: RNA Sequencing

Chapter 1: Sampling Theory and Soil Microbial Communities

Abstract

High throughput sequencing technology has been revolutionary. In molecular ecology, it has provided a culture-free technique to identify simultaneously many taxa that may not be easily grown in pure culture. But sequence data is not without drawbacks: primarily, sequence data is compositional, meaning that it does not represent the true abundance of taxa. Rather, it only captures the abundance of taxa relative to one another. This means that special inferential techniques are required to identify individual taxa that respond to a treatment, or that are correlated with one another across treatments. The inferential techniques used in microbiome research have been heavily influenced by the techniques developed for gene expression data. But there's a problem: the true abundance of transcripts is not important to the interpretation of gene expression data; gene expression is compositional in nature. This is not the case for the interpretation of microbiome data, especially in soils; subsets of the microbiome behave independently, and the true abundance of individual taxa is material to determining the underlying rules and processes governing community assembly across environments or experimental conditions.

This chapter provides context for the dissertation by first establishing that important and ecologically relevant gradients in total soil microbial abundance exist at virtually all scales. Then, I provide a brief overview of a diverse array of intrinsic factors, including niche differentiation and physical isolation, which cause groups within soil microbial communities to have independent (i.e. non-compositional) constraints to survival and growth. I then discuss compositional analysis, and its limitations in the context of the known spatial patterns and diversity of niche strategies in soil microbial communities. I conclude this chapter with the research objectives of this dissertation, and a brief description of how I addressed each objective.

1.1 Spatial organization of soil communities

A single cubic millimeter of soil may contain as many as 3.0×10^6 bacterial cells (Watt et al., 2006). Across the entire soil volume these cells are concentrated in less than 1% of total soil volume (Young et al., 2008), indicating that microbial population structure is patchy at a micro-scale. Within any given soil, microbial biomass exhibits spatial organization at the microscale around soil microsites (Kuzyakov and Blagodatskaya, 2015). As a generality, microbial biomass is highest at the soil surface and decreases exponentially with depth (Eilers et al., 2012; Hao et al., 2021; Leeuwen et al., 2017) and is also concentrated around hotspots such as detritus and roots. The availability of labile carbon is the primary limitation on soil microbial activity and thus population sizes (Blagodatsky et al., 1998; Schimel and

Weintraub, 2003); but variation in the availability of other important substrates, including nitrogen, may modulate the formation and maintenance of high microbial activity in specific microsites through time (Hodge et al., 2000). Important biogeographic cycles, such as carbon and nitrogen cycling, are mediated in high microbial density microhabitats, making the distribution and density of specific types of soil microsites at the landscape scale relevant for modeling global biogeochemical cycles (Blagodatsky and Smith, 2012).

At a global scale, microbial biomass in soil is correlated to edaphic factors including soil organic carbon, total nitrogen, C:N ratio, bulk density, soil pH, and soil texture (He et al., 2020); where mid-latitude and tropical soils have the lowest fungal and bacterial biomass, while arctic soils have the highest fungal and bacterial biomass. In general, soil pH, soil moisture, and soil temperature regulate fungal and bacterial biomass by interacting with their physiology to constrain activity and growth (de Vries et al., 2012). At large spatial scales, these factors are often correlated. For example, at a global scale soil pH tends to be alkaline when annual precipitation is below total potential evapotranspiration, and tends to be acidic when precipitation exceeds this threshold (Slessarev et al., 2016). Similarly, soil carbon also exhibits a global pattern in abundance that is correlated to soil moisture and temperature (Wieder et al., 2013). Potential total evapotranspiration exhibits strong latitudinal gradients, with low temperature strongly associated with lower potential evapotranspiration (Singer et al., 2021). This indicates that at a global scale the wetness and warmth of the climate interact to determine both soil pH and soil carbon. Many anthropogenic factors may also drive local variation in soil moisture, soil carbon, and soil pH (Pouyat et al., 2015, 2010, 2007; Bockheim et al., 2014; Jenny, 1941). Thus, there are substantial gradients of microbial biomass and activity occurring in soils that range in scale from the microsite scale the global scale.

1.2 Decoupling of microbial abundance in soils

While there are large gradients in total microbial abundance across spatial scales, not all taxa in any community respond to the same environmental drivers. There are many trophic strategies that may act to decouple the main drivers of microbial growth across subsets of the soil microbial community. For example, some soil prokaryotes are autotrophic, relying on photosynthesis or mineral redox reactions to capture energy and build biomass. These strategies are widespread; for example, photosynthetic biofilms are active in soils; they form rapidly after tillage in agricultural soils (Peng and Bruns, 2019a, 2019b) and are extremely important for the formation of soil crusts that stabilize soils in highly arid environments (West, 1990). Photosynthesis developed early in bacterial evolution, resulting in a high diversity of photo-responsive metabolism among prokaryotes (Blankenship, 2010, 2002; Cardona, 2019), distributed across numerous phyla (Blankenship, 2010). While photosynthesis is commonly thought of as an autotrophic energy-generating strategy, among bacteria it may also be used to augment heterotrophic metabolism (Selosse et al., 2017). Thus, while the abundance of most soil microbes is dependent on the availability of organic matter, a diverse set of soil microbes are partially or wholly dependent on the availability of sunlight, a resource that exists only within the top centimeter of the soil profile (Tester and Morris, 1987). This implies that there is a subset of the soil microbial community whose abundance should behave independently of the majority of the microbial community with respect to the availability of e.g. soil organic matter.

Autotrophs capable of primary production also exist in other layers of the soil profile. A variety of combinations of elements can be used as an electron donor-acceptor pairs to generate energy for carbon fixation. For example, nitrifying bacteria and archaea can use oxygen to oxidize ammonia for energy (Lehtovirta-Morley, 2018). Methanotrophs may use oxygen to oxidize methane for energy (Op den Camp et al., 2009), but may also use other electron acceptors such as nitrite (Ettwig et al., 2010). Sulfur bacteria such as *Thiobacillus* reduce sulfate to hydrogen sulfide by oxidizing hydrogen (Fischer et al., 1983); and another sulfur reducing bacterium, *Geobacter sulfurreducens*, has recently been shown capable of autotrophic growth using an iron reduction pathway (Zhang et al., 2020). Autotrophic methanogens generate energy by oxidizing hydrogen to reduce carbon dioxide to methane (Ladapo and Whitman, 1990). These alternative metabolic pathways allow microbial taxa to act as primary producers, maintaining biomass without relying on detritus from plants or sunlight. In effect, these pathways decouple the abundance of a subset of the microbial community from the factors driving the abundance of the rest of the microbial community.

These metabolic activities are dependent on specific environmental conditions that may exist as strong spatial and temporal gradients, and may depend in part on the heterotrophic metabolism of other microbes. For example, some chemolithoautotrophic metabolism require specific environmental conditions to be favorable. In soils, hydrogen oxidation-coupled methanogenesis (Hornibrook et al., 1997) and sulfur reduction require strongly reducing conditions (Pester et al., 2012). These conditions are sufficiently common that anoxic methane and ammonia oxidation pathways have been maintained; methane may be oxidized under anoxic conditions by consortia of methane oxidizers and sulfate reducing bacteria, and ammonia oxidation can be paired with nitrate reduction in a process known as anammox (Zhu et al., 2010). While the development of anoxic conditions in soils is driven by the activity of heterotrophic microbes to deplete oxygen, the environmental boundaries that drive soils towards anaerobic conditions (namely available labile carbon for heterotrophic metabolism, saturation of soil with water, and limited diffusion of oxygen) exhibit predictable temporal and spatial patterns, both within the profiles of individual soils and as a function of the local topography, parent material and weather (Weil and Brady, 2017). These environmental boundary conditions define when and where lithoautotrophic strategies are viable in most soil systems, and constrain the spatial and temporal extent to which lithoautotrophic metabolism contributes to microbial biomass. Similarly, light generally does not penetrate more than 5 mm into soils (Tester and Morris, 1987), limiting photosynthesis to the surface of the soils. And, since soil phototrophs depend on moisture availability, photosynthesis is often constrained to short pulses (Lange et al., 1998). Taken together, there is a subset of soil prokaryotes that behave as primary producers – fixing carbon – and whose abundance is not determined directly by the flux of plant-fixed carbon.

For microbes that do depend on the flux of plant-fixed carbon, the amount of microbial biomass obtainable per unit of organic input is determined by the microbial carbon use efficiency (Gill and Finzi, 2016; Soares and Rousk, 2019). The availability of other soil nutrients, especially nitrogen (N), has a large impact on how efficiently microbes may convert carbon substrate to biomass (Feng et al., 2022). When N is less available, microbes use extracellular enzymes to degrade carbon and release nitrogen in a process that has been called “nutrient mining” (Hodge et al., 2001; Tisserant et al., 2013). Soil microbial carbon use efficiency likely interacts with the mineral matrix to determine the amount of soil carbon that is stored long-term (Cotrufo et al., 2013). Among heterotrophs, microbes employ differing growth strategies that result in different community-level carbon use efficiency. Microbes that specialize in fast growth are known as copiotrophs; whereas bacteria specialized in slow growth and nutrient mining are known as oligotrophs (Fierer et al., 2007). Plant associated fungi – especially arbuscular and ectomycorrhizal fungi – are known to alter the carbon use efficiency of soil microbial metabolism (Anthony et al., 2020); in general, ectomycorrhizal fungi slow carbon metabolism and drive higher nutrient mining activities

(Averill and Hawkes, 2016) by competing for nitrogen and suppressing the growth of other fungal decomposers (Fernandez and Kennedy, 2016). Thus, different taxa in the same environment exhibit differences in growth rate and growth efficiency; and differences in growth rate and growth efficiency contribute to variation in total abundance among taxa along environmental gradients. The soil microbial community is well known to contain members whose abundances are governed by different mechanisms and exhibit different response rates to environmental gradients (e.g. heterotroph vs autotroph abundance; or phototroph vs chemoautotroph abundance). This feature of microbial communities becomes important when using sequence data to make inferences about environmental controls on microbial communities.

The examples above highlight trophic strategies that could cause the abundance of taxa sampled from the same environment to behave independently of one another. Even among microbes that share niche characteristics, our current standard sampling practice likely result in a community profile with most taxa never interacting directly. Single-celled soil microbes typically depend on water for as a medium for motility (Yang and van Elsas, 2018); when soil pores are saturated, microbes can disperse from one pore to another. But as soils dry, fewer pores are connected by sufficient water for microbial dispersal. Since microbial taxa occupy less than 1% of the total soil volume (Young et al., 2008), most microbial taxa are unlikely to interact directly with the majority of taxa in the sample (Guseva et al., 2022).

The large size of the soil sample in a typical soil sampling protocol compared to the area inhabited by individual microbes is akin to taking a single sample to analyze the entire coastal transition zone from terrestrial to pelagic; in many soils, a 10 cm core will transect multiple distinct environments with a wide variety of physical and chemical properties that select for distinct communities (Baveye, 2021; Weil and Brady, 2017). Thus, current soil sampling practices generate a snapshot of the microbial community for a given site (Baveye, 2021). This implies that the growth and abundance of most taxa are likely to be determined independently of one another at this scale.

1.3 Compositional data

The factors causing microbes to respond to their environment at different rates are numerous and it is beyond the scope of this dissertation to review all potential mechanisms and contexts. But it is sufficient to understand that some of these factors are almost certainly at play in any given soil sample. This point is critical to the discussion of the compositionality of sequence data and the subsequent ecological inferences in soil research.

Compositional data is data for which the sum of the features (e.g. taxa) in each sample is arbitrary and non-informative (Aitchison, 1982; Gloor and Reid, 2016). In compositional datasets, the abundance of features is not measured, rather their proportion as a function of all other features is measured (Gloor et al., 2016). Therefore, if one feature changes in abundance, the relative abundance of all other features change, even if their abundance does not (Weiss et al., 2017). For example, considering the proportion of taxa from a soil community that are photosynthetic: if the abundance of non-photosynthetic taxa increases, then the proportion of all taxa that are photosynthetic decreases even if the actual abundance of photosynthetic taxa does not change. Thus, compositional datasets are inherently relational; they describe the abundance of features *relative* to the rest of the dataset. When microbial taxa are measured as proportions, it becomes impossible to determine if an increase in taxon proportion is due to an increase in abundance of that taxon, or a decrease in abundance of other taxa. Sampling by sequencing is widely considered to generate compositional data (Bronner et al., 2009; Gloor et al., 2016). Using compositional data to understand community assembly mechanisms implicitly presumes that the total abundance of taxa is invariant (see Chapter 2), and that community assembly mechanisms are zero sum (all taxa compete equally for habitat). For example, one assumption might be that microbes must compete for a fixed amount of substrate, and that growth of one taxon negatively impacts the growth of all other taxa. In the previous sections, I have endeavored to highlight clear evidence that contradict these assumptions at scales that range from micron to global. There is substantial variation in microbial abundance at all spatial scales; and at all spatial scales, subsets of the microbial community experience different constraints to growth. Therefore, patterns in community composition are unlikely to reliably represent the actual abundance of individual taxa across environmental gradients.

1.4 Sequencing, and sampling effort

Sequencing has become a ubiquitous method for sampling microbiomes (Bates et al., 2011; Caporaso et al., 2012, 2011, 2010; Fierer and Jackson, 2006). Environmental microbiome, microbial ecology, and microbial bioindicator analyses widely use metabarcoding to answer fundamental ecological questions (e.g. Auguet et al., 2009; Bahram et al., 2018; Fierer and Jackson, 2006; Pawlowski et al., 2016, 2014). Metabarcoding approaches are popular in part because open-access bioinformatics pipelines are well established for this data type, facilitating sequence data processing (Bokulich et al., 2018; Caporaso et al., 2010; López-García et al., 2018; Schloss, 2020); meaningful inferences can be made from only a few thousand sequences per sample (Lauber et al., 2009); and, several hundred samples may be multiplexed into a single sequencing run, making the cost of sequencing per sample relatively low (Caporaso et al., 2012). Thus, sequencing is now the de facto primary tool for sampling soil microbial communities.

There are several sampling concepts that come into play for sampling microbial communities using sequencing: sequencing depth, sampling depth, and sampling effort. First, sequencing depth is the number of sequences obtained per sample. In contrast sampling depth is the proportion of all sequence variants captured by sequencing (i.e. proportion of total diversity captured). Sampling depth is useful conceptually, but in practice the total diversity of sequence variants is unknown for any given soil sample. These concepts may be distinguished from a third concept, sampling effort, which is the integration of space-time of habitat sampled with the critical assumption that no differences exist in the probability of detecting a taxon if it exists in the sample area. Only studies that control sampling effort conceptually confer information about taxon density in the environment to the taxon counts (i.e. observations). Studies of biodiversity often seek to control sample-wise sampling effort because variation in sampling effort across sites confounds diversity estimation and confounds taxon abundance estimation (Oono, 2017; Watson, 2017). One way that wildlife ecologists ensure that they accurately estimate species abundance is to explicitly control the spatial and temporal dimensions of the sampling (Iijima, 2020). For example, the Pollard Walk defines a uniform transect length, width, height, and duration for sampling butterflies (van Swaay et al., 2008). The resulting count data reflects specific spatial and temporal dimension that allow for the population density of the organism(s) to be estimated and compared.

Sampling microbiome communities using sequencing differs from sampling plants and animals in some fundamentally important ways. As with plants and animals, the initial sampling is usually explicitly constrained by spatial and temporal factors (e.g. same volume of substrate on a specific date). Subsequently DNA is extracted from the sample and prepared for sequencing. In preparing the sample for sequencing, only a very tiny fraction of total DNA is generally used as an input to the DNA sequencer (Bronner et al., 2009; Illumina, 2021). The number of sequences that a sequencer may generate per sample is defined by a stochastic process that has little to do with the total abundance of taxa in the environment (Bronner et al., 2009). Because the number of sequences drawn per sample is both random and not related to the abundance of organisms, sequence data does not have an unbiased detection probability for taxa across samples. As a result, any efforts to constrain sampling effort in microbiome research based on the initial sample collection (e.g. 1.0 gram of soil) does not impart the same information about taxon density that spatially and temporally explicit sampling methods convey for plants and animals. This is a well-known feature of sequence data, and has led microbiome researchers generally to treat sequence data as compositional (Gloor et al., 2016). Since compositional data cannot be used to infer the abundance of features, the challenge therefore is developing a standard method to control sampling effort such that the count data accurately represents the density of each taxon in the environment.

In an attempt to address the lack of information about taxon density, sequence datasets have commonly been subsampled in a process called “rarefaction” such that each sample analyzed would have the same number of sequences (McMurdie and Holmes, 2014). This method may mitigate some of the confounding effects of variable sequencing depth on differential abundance analysis (Weiss et al., 2017). However, controlling sequencing depth is not conceptually equivalent to controlling sampling effort. When observation is constrained explicitly to a specific set of spatial and temporal dimensions, observations (counts) represent an index of taxon abundance per unit of time-volume of habitat. Properly collected data can therefore be explicitly interpreted as representing the density of a taxon (or taxa) in the environment. For sequence data to represent the density of taxa in the environment, additional information about taxon density must be accessed from somewhere. Rarefying to an even sampling depth ignores that in soils there are large gradients of microbial abundance that exist at all spatial scales, meaning different samples are very likely to have different densities of taxa, and should therefore have different total number of counts per sample (see sections 1.1-1.3 and Chapters 2 and 4). For rarefaction to generate sequence counts that can similarly be interpreted as representing the density of taxa, the number of sequences chosen (rarefaction depth) for each sample needs to be proportional to the total abundance of taxa in that sample (Epp Schmidt et al., 2019). This requires the abundance to be measured independently of the sequencing data, and is not typically how rarefaction is implemented (McMurdie and Holmes, 2014). Alternatively, information on taxon density may be imparted by scaling sequence counts (relative abundance) based on the measured total abundance of taxa in the sample, a method called quantitative sequencing (QSeq) (Jian et al., 2020). This method is preferable because it maximizes the amount of information available for each sample (McMurdie and Holmes, 2014).

Statistical inference in microbiome research has been heavily influenced by the standard practices developed for statistical inference in gene expression. For example, tools such as DESeq2 (Love et al., 2014) and limma-voom(Ritchie et al., 2015) that are commonly used for differential abundance analysis were designed for differential expression. The term “differential abundance” is itself a direct analogy to “differential expression” (McMurdie and Holmes, 2014):

> *“Like differentially expressed genes, a species/OTU is considered differentially abundant if its mean proportion is significantly different between two or more sample classes in the experimental design.”*
> *(McMurdie and Holmes, 2014)*

The use of the term “abundance” is misleading. These differential abundance analyses are well-documented to have no conceptual basis for providing information about the abundance of taxa (Aitchison, 1982; Gloor et al., 2017, 2016; Gloor and Reid, 2016; Tsilimigras and Fodor, 2016). This becomes a problem when researchers are interested in understanding how specific taxa respond to an environmental gradient or experimental treatment; current standard practice in environmental microbiome research is to use differential (relative) abundance analysis to determine which taxa change in *relative* abundance (Gloor et al., 2016), and from this draw inferences on the ecological processes driving the microbial community. It is an error to infer that a taxon is positively associated with an environmental gradient because its *relative* abundance significantly increases in association with that environment (Aitchison, 1982; Gloor et al., 2017). Recent empirical and modeling work in gut microbiomes shows clearly that taxa may become more dominant in a community without increasing in abundance (Barlow et al., 2020; Lloréns-Rico et al., 2021). Yet this tacit inference has often been made.

Benchmarking of published software is ad-hoc; benchmarking conditions vary across studies (Bokulich et al., 2020, 2016; Meyer et al., 2021) and software releases (Kurtz et al., 2015; Watts et al., 2019; Weiss et al., 2016). The compositionality of sequence data (non-informative sample-wise sum of features) is a presumed constraint of sequence data analysis in microbiome studies (Aitchison, 1982; Gloor et al., 2017, 2016; Gloor and Reid, 2016; Tsilimigras and Fodor, 2016); yet when these analytic tools are released, they often are not even benchmarked on realistic compositional data (Weiss et al., 2017). All of this makes it difficult to discern which methods are actually the most appropriate. As if this were not enough, a central premise of this dissertation is that compositional analysis itself leads to errors of inference about the mechanisms driving community assembly patterns. In part, this is because assumptions of compositional analysis are often violated by microbiome data (see Chapter 2); but, it is also because without knowing the actual abundance of taxa in each sample researchers are unable to unambiguously identify which taxa are changing in abundance in their experiments.

"*It is a cliché to generate data under a parametric model that incorporates statistical features that a bioinformatic method was built to handle, and to then show how competing models fail to perform in the presence of those features, i.e. the algorithms under investigation are not independent*"

(Bokulich et al., 2020)

1.5 Objectives

The objectives of this dissertation are:

1) Identify the conditions and questions for which quantitative sequencing improves ecological inference.
2) Determine if the QSeq transformation can be applied to improve metagenomics prediction software such as PICRUSt2.
3) Use QSeq methods to understand community dynamics in Maryland agricultural systems that would not be accessible without accurate taxon abundance estimation:
    a. How does farming practice structure the vertical distribution of taxa in the soil profile?
    b. How does glyphosate impact nitrogen fixation, and the community of free-living nitrogen fixing taxa in Maryland agricultural systems?

Chapter 2 addresses objective 1 using literature review, data synthesis and modeling approaches to determine the conditions under which quantitative abundance estimates (QSeq) results in more accurate representations of ecological patterns. Chapter 3 addresses objective 2 using a data synthesis approach to evaluate whether the predicted abundance of certain functional genes can be improved using logic borrowed from QSeq (scaling predicted metagenomes by total abundance). I hypothesized that scaling predictions by the absolute microbial abundance would improve predicted gene abundance. I also hypothesized that there would be differences among genes; that certain genes could be more reliably predicted than others based on their evolutionary past and representation in the reference databases. Chapter 4 addresses objective 3a using an experimental approach to measure the abundance of microbial taxa sampled from different soil depths in three Long-Term Agricultural Research farming systems in collaboration with the USDA Beltsville Agricultural Research Center. I hypothesized that differences in tillage and inputs would lead to distinct distribution (or in the case of tillage, redistribution) of taxa across the depth profile. And finally, Chapter 5 addresses objective 3b using a data synthesis approach to examine how multiple compartments of the N fixation ecosystem function (including community dynamics) react to the application of glyphosate in 4 distinct farming systems. My hypothesis was that the N-fixing microbial community would have a consistent response to glyphosate treatment, and that each of the measured compartment would respond in coordination. For example, if glyphosate impacts N fixation rate, then the abundance of nifH in the soil should change, and the abundance of individual N-fixing taxa should be affected.

1.6 Literature Cited

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

Anthony, M.A., Crowther, T.W., Maynard, D., van den Hoogan, J., Averill, C., 2020. Distinct Assembly Processes and Microbial Communities Constrain Soil Organic Carbon Formation. One Earth 2, 349–360. https://doi.org/10.1016/j.oneear.2020.03.006

Auguet, J.-C., Barberan, A., Casamayor, E.O., 2009. Global ecological patterns in uncultured Archaea. ISME J 4, 182–190. https://doi.org/10.1038/ismej.2009.109

Averill, C., Hawkes, C.V., 2016. Ectomycorrhizal fungi slow soil carbon cycling. Ecol Lett 19, 937–947. https://doi.org/10.1111/ele.12631

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., Huerta-Cepas, J., Medema, M.H., Maltz, M.R., Mundra, S., Olsson, P.A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., Tedersoo, L., Bork, P., 2018. Structure and function of the global topsoil microbiome. Nature 560, 233–237. https://doi.org/10.1038/s41586-018-0386-6

Barlow, J.T., Bogatyrev, S.R., Ismagilov, R.F., 2020. A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. Nature Communications 11, 2590. https://doi.org/10.1038/s41467-020-16224-6

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., Fierer, N., 2011. Examining the global distribution of dominant archaeal populations in soil. ISME J 5, 908–917. https://doi.org/10.1038/ismej.2010.171

Baveye, P.C., 2021. Bypass and hyperbole in soil research: Worrisome practices critically reviewed through examples. European Journal of Soil Science 72, 1–20. https://doi.org/10.1111/ejss.12941

Blagodatsky, S., Smith, P., 2012. Soil physics meets soil biology: Towards better mechanistic prediction of greenhouse gas emissions from soil.

Soil Biology and Biochemistry 47, 78–92. https://doi.org/10.1016/j.soilbio.2011.12.015

Blagodatsky, S.A., Yevdokimov, I.V., Larionova, A.A., Richter, J., 1998. Microbial growth in soil and nitrogen turnover: Model calibration with laboratory data. Soil Biology and Biochemistry 30, 1757–1764. https://doi.org/10.1016/S0038-0717(98)00029-7

Blankenship, R.E., 2010. Early Evolution of Photosynthesis1. Plant Physiol 154, 434–438. https://doi.org/10.1104/pp.110.161687

Blankenship, R.E., 2002. Molecular Mechanisms of Photosynthesis. John Wiley & Sons.

Bockheim, J.G., Gennadiyev, A.N., Hartemink, A.E., Brevik, E.C., 2014. Soil-forming factors and Soil Taxonomy. Geoderma 226–227, 231–237. https://doi.org/10.1016/j.geoderma.2014.02.016

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., Gregory Caporaso, J., 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6, 90. https://doi.org/10.1186/s40168-018-0470-z

Bokulich, N.A., Rideout, J.R., Mercurio, W.G., Wolfe, B., Maurice, C.F., Dutton, R.J., Turnbaugh, P.J., Knight, R., Caporaso, J.G., 2016. mockrobiota: a public resource for microbiome bioinformatics benchmarking (No. e2065v1). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.2065v1

Bokulich, N.A., Ziemski, M., Robeson, M.S., Kaehler, B.D., 2020. Measuring the microbiome: Best practices for developing and benchmarking microbiomics methods. Computational and Structural Biotechnology Journal 18, 4048–4062. https://doi.org/10.1016/j.csbj.2020.11.049

Bronner, I.F., Quail, M.A., Turner, D.J., Swerdlow, H., 2009. Improved Protocols for Illumina Sequencing. Curr Protoc Hum Genet 0 18, 10.1002/0471142905.hg1802s62. https://doi.org/10.1002/0471142905.hg1802s62

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS 108, 4516–4522. https://doi.org/10.1073/pnas.1000080107

Cardona, T., 2019. Thinking twice about the evolution of photosynthesis. Open Biol 9, 180246. https://doi.org/10.1098/rsob.180246

Cotrufo, M.F., Wallenstein, M.D., Boot, C.M., Denef, K., Paul, E., 2013. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? Global Change Biology 19, 988–995. https://doi.org/10.1111/gcb.12113

de Vries, F.T., Manning, P., Tallowin, J.R.B., Mortimer, S.R., Pilgrim, E.S., Harrison, K.A., Hobbs, P.J., Quirk, H., Shipley, B., Cornelissen, J.H.C., Kattge, J., Bardgett, R.D., 2012. Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities. Ecol Lett 15, 1230–1239. https://doi.org/10.1111/j.1461-0248.2012.01844.x

Eilers, K.G., Debenport, S., Anderson, S., Fierer, N., 2012. Digging deeper to find unique microbial communities: The strong effect of depth on the structure of bacterial and archaeal communities in soil. Soil Biology and Biochemistry 50, 58–65. https://doi.org/10.1016/j.soilbio.2012.03.011

Epp Schmidt, D.J., Kotze, D.J., Hornung, E., Setälä, H., Yesilonis, I., Szlavecz, K., Dombos, M., Pouyat, R., Cilliers, S., Tóth, Z., Yarwood, S., 2019. Metagenomics reveals bacterial and archaeal adaptation to urban land-use: N catabolism, methanogenesis, and nutrient acquisition. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02330

Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M.M., Schreiber, F., Dutilh, B.E., Zedelius, J., de Beer, D., Gloerich, J., Wessels, H.J.C.T., van Alen, T., Luesken, F., Wu, M.L., van de Pas-Schoonen, K.T., Op den Camp, H.J.M., Janssen-Megens, E.M., Francoijs, K.-J., Stunnenberg, H., Weissenbach, J., Jetten, M.S.M., Strous, M., 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. Nature 464, 543–548. https://doi.org/10.1038/nature08883

Feng, X., Qin, S., Zhang, D., Chen, P., Hu, J., Wang, G., Liu, Y., Wei, B., Li, Q., Yang, Y., Chen, L., 2022. Nitrogen input enhances microbial carbon use efficiency by altering plant–microbe–mineral interactions. Global Change Biology 28, 4845–4860. https://doi.org/10.1111/gcb.16229

Fernandez, C.W., Kennedy, P.G., 2016. Revisiting the ‘Gadgil effect’: do interguild fungal interactions control carbon cycling in forest soils? New Phytologist 209, 1382–1394. https://doi.org/10.1111/nph.13648

Fierer, N., Bradford, M.A., Jackson, R.B., 2007. Toward an ecological classification of soil bacteria. Ecology 88, 1354–1364. https://doi.org/10.1890/05-1839

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. PNAS 103, 626–631. https://doi.org/10.1073/pnas.0507535103

Fischer, F., Zillig, W., Stetter, K.O., Schreiber, G., 1983. Chemolithoautotrophic metabolism of anaerobic extremely thermophilic archaebacteria. Nature 301, 511–513. https://doi.org/10.1038/301511a0

Gill, A.L., Finzi, A.C., 2016. Belowground carbon flux links biogeochemical cycles and resource-use efficiency at the global scale. Ecology Letters 19, 1419–1428. https://doi.org/10.1111/ele.12690

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 0. https://doi.org/10.3389/fmicb.2017.02224

Gloor, G.B., Reid, G., 2016. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can. J. Microbiol. 62, 692–703. https://doi.org/10.1139/cjm-2015-0821

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., Egozcue, J.J., 2016. It’s all relative: analyzing microbiome data as compositions. Ann Epidemiol 26, 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003

Guseva, K., Darcy, S., Simon, E., Alteio, L.V., Montesinos-Navarro, A., Kaiser, C., 2022. From diversity to complexity: Microbial networks in soils. Soil Biology and Biochemistry 169, 108604. https://doi.org/10.1016/j.soilbio.2022.108604

Hao, J., Chai, Y.N., Lopes, L.D., Ordóñez, R.A., Wright, E.E., Archontoulis, S., Schachtman, D.P., 2021. The Effects of Soil Depth on the Structure of Microbial Communities in Agricultural Soils in Iowa (United States). Applied and Environmental Microbiology 87, e02673-20. https://doi.org/10.1128/AEM.02673-20

He, L., Mazza Rodrigues, J.L., Soudzilovskaia, N.A., Barceló, M., Olsson, P.A., Song, C., Tedersoo, L., Yuan, Fenghui, Yuan, Fengming, Lipson, D.A., Xu, X., 2020. Global biogeography of fungal and bacterial biomass carbon in topsoil. Soil Biology and Biochemistry 151, 108024. https://doi.org/10.1016/j.soilbio.2020.108024

Hodge, A., Campbell, C., Fitter, A., 2001. An arbuscular mycorrhizal fungus accelerates decomposition and acquires nitrogen directly from organic material | Nature.

Hodge, A., Robinson, D., Fitter, A., 2000. Are microorganisms more effective than plants at competing for nitrogen? Trends in Plant Science 5, 304–308. https://doi.org/10.1016/S1360-1385(00)01656-3

Hornibrook, E.R.C., Longstaffe, F.J., Fyfe, W.S., 1997. Spatial distribution of microbial methane production pathways in temperate zone wetland soils: Stable carbon and hydrogen isotope evidence. Geochimica et Cosmochimica Acta 61, 745–753. https://doi.org/10.1016/S0016-7037(96)00368-7

Iijima, H., 2020. A Review of Wildlife Abundance Estimation Models: Comparison of Models for Correct Application. jmam 45, 177–188. https://doi.org/10.3106/ms2019-0082

Illumina, 2021. MiniSeq System Denature and Dilute Libraries Guide (1000000002697).

Jenny, H., 1941. Factors of Soil Formation, a System of Quantitative Pedology. Agronomy Journal 33, 857–858. https://doi.org/10.2134/agronj1941.00021962003300090016x

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., Korpela, K., 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLOS ONE 15, e0227285. https://doi.org/10.1371/journal.pone.0227285

Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bonneau, R.A., 2015. Sparse and Compositionally Robust Inference of Microbial Ecological Networks. PLOS Computational Biology 11, e1004226. https://doi.org/10.1371/journal.pcbi.1004226

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: concept & review. Soil Biology and Biochemistry 83, 184–199. https://doi.org/10.1016/j.soilbio.2015.01.025

Ladapo, J., Whitman, W.B., 1990. Method for isolation of auxotrophs in the methanogenic archaebacteria: role of the acetyl-CoA pathway of autotrophic CO2 fixation in Methanococcus maripaludis. Proceedings of the National Academy of Sciences 87, 5598–5602. https://doi.org/10.1073/pnas.87.15.5598

Lange, O.L., Belnap, J., Reichenberger, H., 1998. Photosynthesis of the cyanobacterial soil-crust lichen Collema tenax from arid lands in southern Utah, USA: role of water content on light and temperature responses of CO2 exchange. Functional Ecology 12, 195–202. https://doi.org/10.1046/j.1365-2435.1998.00192.x

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. Appl. Environ. Microbiol. 75, 5111–5120. https://doi.org/10.1128/AEM.00335-09

Leeuwen, J.P. van, Djukic, I., Bloem, J., Lehtinen, T., Hemerik, L., Ruiter, P.C. de, Lair, G.J., 2017. Effects of land use on soil microbial biomass, activity and community structure at different soil depths in the Danube floodplain. European Journal of Soil Biology 79, 14–20. https://doi.org/10.1016/j.ejsobi.2017.02.001

Lehtovirta-Morley, L.E., 2018. Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. FEMS Microbiology Letters 365, fny058. https://doi.org/10.1093/femsle/fny058

Lloréns-Rico, V., Vieira-Silva, S., Gonçalves, P.J., Falony, G., Raes, J., 2021. Benchmarking microbiome transformations favors experimental quantitative approaches to address compositionality and sampling depth biases. Nat Commun 12, 3562. https://doi.org/10.1038/s41467-021-23821-6

López-García, A., Pineda-Quiroga, C., Atxaerandio, R., Pérez, A., Hernández, I., García-Rodríguez, A., González-Recio, O., 2018. Comparison of Mothur and QIIME for the Analysis of Rumen

Microbiota Composition Based on 16S rRNA Amplicon Sequences. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.03010

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

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Meyer, F., Lesker, T.-R., Koslicki, D., Fritz, A., Gurevich, A., Darling, A.E., Sczyrba, A., Bremges, A., McHardy, A.C., 2021. Tutorial: assessing metagenomics software with the CAMI benchmarking toolkit. Nat Protoc 16, 1785–1801. https://doi.org/10.1038/s41596-020-00480-3

Oono, R., 2017. A confidence interval analysis of sampling effort, sequencing depth, and taxonomic resolution of fungal community ecology in the era of high-throughput sequencing. PLOS ONE 12, e0189796. https://doi.org/10.1371/journal.pone.0189796

Op den Camp, H.J.M., Islam, T., Stott, M.B., Harhangi, H.R., Hynes, A., Schouten, S., Jetten, M.S.M., Birkeland, N.-K., Pol, A., Dunfield, P.F., 2009. Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia. Environmental Microbiology Reports 1, 293–306. https://doi.org/10.1111/j.1758-2229.2009.00022.x

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., Wilding, T.A., 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Molecular Ecology Resources 14, 1129–1140. https://doi.org/10.1111/1755-0998.12261

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., Esling, P., 2016. Protist metabarcoding and environmental biomonitoring: Time for change. European Journal of Protistology, Current trends in protistology – results from the VII ECOP - ISOP Joint Meeting 2015 55, 12–25. https://doi.org/10.1016/j.ejop.2016.02.003

Peng, X., Bruns, M.A., 2019a. Development of a nitrogen-fixing cyanobacterial consortium for surface stabilization of agricultural soils. Journal of Applied Phycology 31, 1047–1056. https://doi.org/10.1007/s10811-018-1597-9

Peng, X., Bruns, M.A., 2019b. Cyanobacterial Soil Surface Consortia Mediate N Cycle Processes in Agroecosystems. Frontiers in Environmental Science 6, 156. https://doi.org/10.3389/fenvs.2018.00156

Pester, M., Knorr, K.-H., Friedrich, M., Wagner, M., Loy, A., 2012. Sulfate-reducing microorganisms in wetlands – fameless actors in carbon cycling and climate change. Frontiers in Microbiology 3.

Pouyat, R., Yesilonis, I., Dombos, M., Szlavecz, K., Setala, H., Cilliers, S., Hornung, E., Kotze, J., Yarwood, S., 2015. A global comparison of surface soil across five cities: a test of the urban ecosystem convergence hypothesis. Soil Science.

Pouyat, R.V., Szlavecz, K., Yesilonis, I.D., Groffman, P.M., Schwarz, K., 2010. Chemical, Physical, and Biological Characteristics of Urban Soils, in: Urban Ecosystem Ecology, Agronomy Monographs. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, pp. 119–152.

Pouyat, R.V., Yesilonis, I.D., Russell-Anelli, J., Neerchal, N.K., 2007. Soil Chemical and Physical Properties That Differentiate Urban Land-Use and Cover Types. Soil Science Society of America Journal 71, 1010. https://doi.org/10.2136/sssaj2006.0164

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47–e47. https://doi.org/10.1093/nar/gkv007

Schimel, J.P., Weintraub, M.N., 2003. The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. Soil Biology and Biochemistry 35, 549–563. https://doi.org/10.1016/S0038-0717(03)00015-4

Schloss, P.D., 2020. Reintroducing mothur: 10 Years Later. Appl Environ Microbiol 86. https://doi.org/10.1128/AEM.02343-19

Selosse, M.-A., Charpin, M., Not, F., 2017. Mixotrophy everywhere on land and in water: the grand écart hypothesis. Ecology Letters 20, 246–263. https://doi.org/10.1111/ele.12714

Singer, M.B., Asfaw, D.T., Rosolem, R., Cuthbert, M.O., Miralles, D.G., MacLeod, D., Quichimbo, E.A., Michaelides, K., 2021. Hourly potential evapotranspiration at 0.1° resolution for the global land surface from 1981-present. Sci Data 8, 224. https://doi.org/10.1038/s41597-021-01003-9

Slessarev, E.W., Lin, Y., Bingham, N.L., Johnson, J.E., Dai, Y., Schimel, J.P., Chadwick, O.A., 2016. Water balance creates a threshold in soil pH at the global scale. Nature 540, 567–569. https://doi.org/10.1038/nature20139

Soares, M., Rousk, J., 2019. Microbial growth and carbon use efficiency in soil: Links to fungal-bacterial dominance, SOC-quality and stoichiometry. Soil Biology and Biochemistry 131, 195–205. https://doi.org/10.1016/j.soilbio.2019.01.010

Tester, M., Morris, C., 1987. The penetration of light through soil. Plant, Cell & Environment 10, 281–286. https://doi.org/10.1111/j.1365-3040.1987.tb01607.x

Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N., Frei dit Frey, N., Gianinazzi-Pearson, V., Gilbert, L.B., Handa, Y., Herr, J.R., Hijri, M., Koul, R., Kawaguchi, M., Krajinski, F., Lammers, P.J., Masclaux, F.G., Murat, C., Morin, E., Ndikumana, S., Pagni, M., Petitpierre, D., Requena, N., Rosikiewicz, P., Riley, R., Saito, K., San Clemente, H., Shapiro, H., van Tuinen, D., Bécard, G., Bonfante, P., Paszkowski, U., Shachar-Hill, Y.Y., Tuskan, G.A., Young, J.P.W., Sanders, I.R., Henrissat, B., Rensing, S.A., Grigoriev, I.V., Corradi, N., Roux, C., Martin, F., 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proceedings of the National Academy of Sciences 110, 20117–20122. https://doi.org/10.1073/pnas.1313452110

Tsilimigras, M.C.B., Fodor, A.A., 2016. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Annals of Epidemiology, The Microbiome and Epidemiology 26, 330–335. https://doi.org/10.1016/j.annepidem.2016.03.002

van Swaay, C.A.M., Nowicki, P., Settele, J., van Strien, A.J., 2008. Butterfly monitoring in Europe: methods, applications and perspectives. Biodivers Conserv 17, 3455–3469. https://doi.org/10.1007/s10531-008-9491-4

Watson, D.M., 2017. Sampling effort determination in bird surveys: do current norms meet best-practice recommendations? Wildl. Res. 44, 183–193. https://doi.org/10.1071/WR16226

Watt, M., Hugenholtz, P., White, R., Vinall, K., 2006. Numbers and locations of native bacteria on field-grown wheat roots quantified by fluorescence in situ hybridization (FISH). Environmental Microbiology 8, 871–884. https://doi.org/10.1111/j.1462-2920.2005.00973.x

Watts, S.C., Ritchie, S.C., Inouye, M., Holt, K.E., 2019. FastSpar: rapid and scalable correlation estimation for compositional data. Bioinformatics 35, 1064–1066. https://doi.org/10.1093/bioinformatics/bty734

Weil, R., Brady, N., 2017. The Nature and Properties of Soils. 15th edition.

Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., Birmingham, A., Cram, J.A., Fuhrman, J.A., Raes, J., Sun, F., Zhou, J., Knight, R., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J 10, 1669–1681. https://doi.org/10.1038/ismej.2015.235

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E.R., Knight, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5, 27. https://doi.org/10.1186/s40168-017-0237-y

West, N.E., 1990. Structure and Function of Microphytic Soil Crusts in Wildland Ecosystems of Arid to Semi-arid Regions, in: Begon, M., Fitter, A.H., Macfadyen, A. (Eds.), Advances in Ecological Research. Academic Press, pp. 179–223. https://doi.org/10.1016/S0065-2504(08)60055-0

Wieder, W.R., Bonan, G.B., Allison, S.D., 2013. Global soil carbon projections are improved by modelling microbial processes. Nature Clim Change 3, 909–912. https://doi.org/10.1038/nclimate1951

Yang, P., van Elsas, J.D., 2018. Mechanisms and ecological implications of the movement of bacteria in soil. Applied Soil Ecology 129, 112–120. https://doi.org/10.1016/j.apsoil.2018.04.014

Young, I.M., Crawford, J.W., Nunan, N., Otten, W., Spiers, A., 2008. Chapter 4 Microbial Distribution in Soils: Physics and Scaling, in: Advances in Agronomy. Academic Press, pp. 81–121. https://doi.org/10.1016/S0065-2113(08)00604-4

Zhang, T., Shi, X.-C., Ding, R., Xu, K., Tremblay, P.-L., 2020. The hidden chemolithoautotrophic metabolism of Geobacter sulfurreducens uncovered by adaptation to formate. ISME J 14, 2078–2089. https://doi.org/10.1038/s41396-020-0673-8

Zhu, G., Jetten, M.S.M., Kuschk, P., Ettwig, K.F., Yin, C., 2010. Potential roles of anaerobic ammonium and methane oxidation in the nitrogen cycle of wetland ecosystems. Appl Microbiol Biotechnol 86, 1043–1055. https://doi.org/10.1007/s00253-010-2451-4

Submitted to BMC Microbiome

Chapter 2: Total abundance correction in microbiome sequence data: when and why it provides better insight

Dietrich Epp Schmidt[1] (corresponding), Jude Maul[2], and Stephanie Yarwood[1]

[1]University of Maryland College Park

[2]United States Department of Agriculture Beltsville Agricultural Research Service

Keywords: Microbiome, Sequencing, Inference, Quantitative Sequencing, Total Abundance Scaling, Total Abundance

Abstract

High-throughput, multiplexed-amplicon sequencing has become a core tool for understanding microbiomes. As researchers have widely adopted sequencing many open-source analysis pipelines have been developed to compare microbiomes based on the relative abundance of assigned sequence variants (ASVs). These data analysis pipelines can distinguish one community from another but lack reliable information about either the true abundance of taxa or the covariance of taxa along ecological gradients. The abundance of taxa provides critical information about ecological processes that structure communities; for example, changes in community composition that are driven by increasing abundance of few taxa, and those driven by decreasing abundance of many taxa are fundamentally different processes. The current standard practices of using compositional analysis in microbiome

research is unable to distinguish among these processes without collecting additional information. Recently, sequencing has been combined with Q-PCR (among other sources of total quantitation), to generate “Quantitative Sequencing” (QSeq) data that more accurately estimate the true abundance of taxa. As far as we know, no systematic evaluation has been done to determine under what conditions and for what inferences QSeq analysis offers an advantage over compositional analysis. In this paper, we use a combination of published data sets, synthesis, and empirical modeling to offer guidance for when QSeq should and should not be implemented.

QSeq is primarily useful for taxon-specific analysis. As little as 5% variation in total abundance among experimental groups resulted in more accurate inference by QSeq than compositional methods. Compositional methods for differential abundance and correlation unreliably detected patterns in abundance and covariance when there was greater than 20% variation in total abundance among experimental groups.

QSeq transformation and analysis should be the default for answering taxon-specific questions of amplicon sequence data. Publicly available bioinformatics pipelines should incorporate support for QSeq transformation and analysis.

2.1 Introduction

Sequencing has become central for understanding environmental microbiomes (Fierer and Jackson, 2006; Caporaso et al., 2010, 2011, 2012;

Bates et al., 2011). Environmental microbial -omics, microbial ecology, and microbial bioindicator analyses overwhelmingly use amplicon sequencing in a metabarcoding approach to answer fundamental ecological questions and accomplish basic environmental monitoring tasks (e.g. Fierer and Jackson, 2006; Auguet et al., 2009; Pawlowski et al., 2014, 2016; Bahram et al., 2018). Metabarcoding approaches are popular in part because open-access bioinformatics pipelines are well established, facilitating sequence data processing (Caporaso et al., 2010; Bokulich et al., 2018; López-García et al., 2018; Schloss, 2020). Meaningful inferences can be made from only a few thousand sequences per sample (Lauber et al., 2009), and several hundred samples may be multiplexed into a single sequencing run, making the cost of sequencing per sample relatively low (Caporaso et al., 2012). Yet, analysis of sequence data alone is inherently constrained by compositionality (Gloor et al., 2017).

A key step that differentiates compositional analysis from non-compositional approaches is the log-ratio transformation (Gloor et al., 2016). The core assumption of compositional analysis is that count abundance in sequence data is confounded by sequencing depth, and therefore cannot convey direct information about the abundance of taxa (Gloor et al., 2016). Instead, the log-ratio transformation (e.g. Aitchison, 1982; Fernandes et al., 2013, 2014) defines the abundance of each taxon in an explicitly relational way (Gloor et al., 2016). For example, rather than saying a taxon has a certain cell density per unit area, compositional analyses might define abundance based on the proportion each taxon represents among all taxa that were sampled (Figure 2.1; Aitchison, 1982; Love et al., 2014). Log-ratio transformation mitigates some of the confounding effects of sequencing depth, but has the disadvantage that it gives only relative information; assuming no response of other taxa, an increase in abundance of one taxon implies decreasing relative abundance of other taxa. Therefore, in most cases relative abundance (i.e. compositional data) cannot be used to infer changes in the true abundance of taxa in the environment (Aitchison, 1982; Gloor et al., 2016, 2017; Tsilimigras and Fodor, 2016; Alteio et al., 2021). Neither extracted DNA concentrations, nor sequencing counts give reliable information about total abundance of taxa. If researchers are interested in how taxon abundance changes across the environment, they must independently measure total abundance.

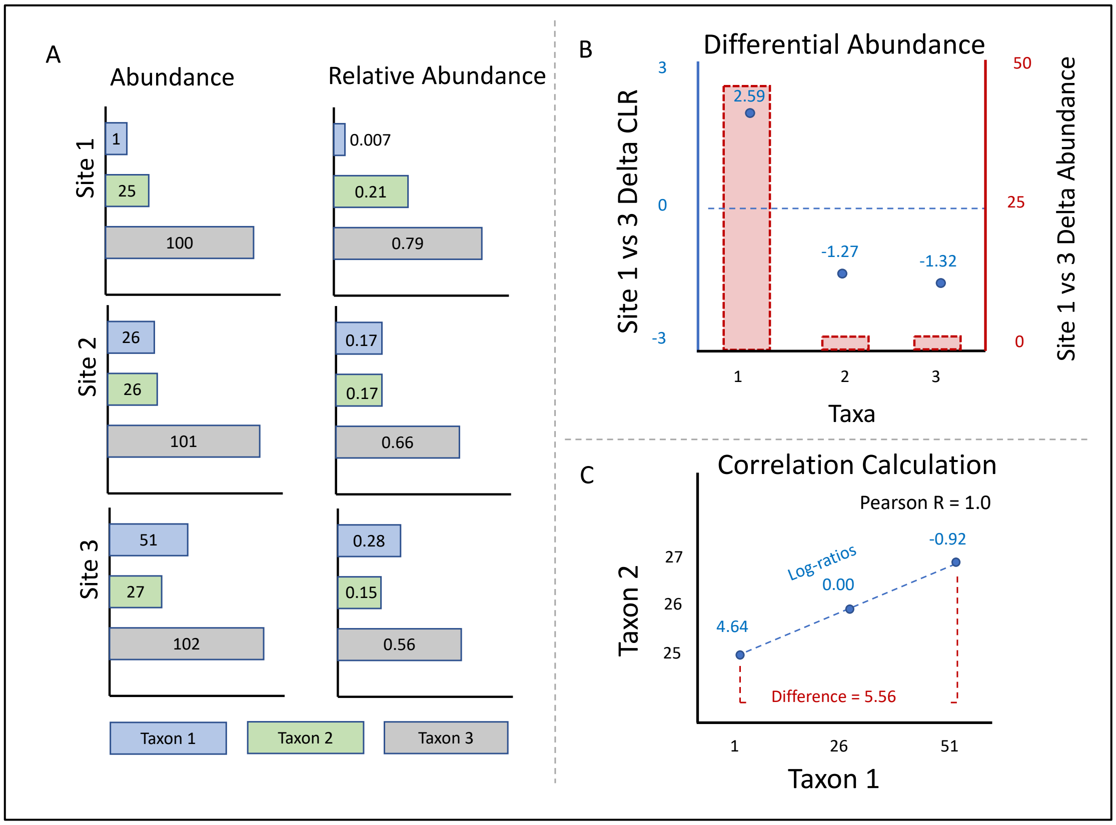

Figure 2.1: Theoretical example of three taxa in three sites. Panel A illustrates how relative abundance differs from true abundance; comparing site 1 to site 3, all taxa increase in abundance, but only taxon 1 increases in relative abundance. Panel B illustrates how correlation calculations are made on compositional data and how abundance estimates differ between true abundance and centered log-ratio methods. Blue axis and points represent the change in centered log-ratio value. The red axis and bars represent the quantitative difference in abundance. While all taxa increase in abundance, only one taxon increases in relative abundance. Panel C illustrates the math underlying how correlations are inferred. Correlations are inferred based on the variance of log-ratio values for pairs of taxa across all sites. The highest variance in log-ratio values is inferred as negative correlations and the lowest variance values are inferred as positive correlations. Panel C shows Taxon 1 and 2 from Panel A, which are a perfectly correlated pair of taxa with large, non - zero variance; this is a proof of concept that perfectly positively correlated taxa can produce high variance log-ratios in compositional analyses when the slope of their response curves is not equal.

The central assumption of QSeq is that sequence data can provide direct information about the abundance of taxa if it is appropriately normalized using an independent measure of the total abundance of taxa in each sample. Thus, both sequence data and an independent quantitative measure of total abundance in each sample are needed to implement QSeq. Some studies have used QPCR to measure total community abundance (Epp Schmidt et al., 2017, 2022; Jian et al., 2020); other methods include flow cytometry (Vandeputte et al., 2017), internal standards (i.e. spike-in: Smets et al., 2016; Stämmler et al., 2016; Tourlousse et al., 2017), and Digital PCR (Barlow et al., 2020). The QSeq transformation is done by first transforming sequence count data into relative abundance, then scaling the counts to the total abundance of each sample (Jian et al., 2020). An alternative approach is to use degenerate tags during sequencing to measure the quantity of genes (Hoshino et al., 2021), though this method is not evaluated here.

Calculating the abundance of specific taxa has a large impact on data interpretation. For example, compositional sequence data were used to infer that Crohn’s disease increased the abundance of some bacteria and decreased the abundance of others (Gevers et al., 2014). In fact, data from Gevers et al. only showed that the relative abundance increased or decreased. It is possible that all taxa in these samples were lower abundance in Crohn’s disease. Knowing the abundance of taxa gives insight into the community assembly process; if the whole community decreased abundance it indicates that host regulation is an important determinant, whereas if a couple taxa consistently increased in abundance, it indicates that certain taxa may have played a causative role in the disease. Although this example was from the human microbiome, it is applicable to all environments and types of sequence data. A global analysis of soil microbiomes suggested that patterns in bacterial and fungal taxonomic and functional diversity were driven by direct competition. This argument was supported by the bacterial:fungal biomass ratio indicating that bacterial:fungal ratios differed in part by latitude, mean annual precipitation, mean annual temperature, and soil C/N ratio (Bahram et al., 2018). Yet the bacterial biomass (shown in their Extended Data 4a&b) did not decrease when fungal biomass increased, rather bacterial and fungal biomass were positively correlated across latitudes. In contrast to the to the main finding, mechanisms other than competition and niche differentiation between bacteria and fungi appear to influence functional and taxonomic diversity. The absolute abundances of taxa (or taxonomic groups), rather than their relative abundance, are critical to accurately inferring ecological interactions.

Quantitative sequencing can improve the quality of inferences about assembly processes (Epp Schmidt et al., 2017; Barlow et al., 2020; Jian et al., 2020). Although some of the benefits of QSeq are documented, it may be unclear when this analysis is appropriate. To clarify when and for what purposes QSeq provides better ecological inferences, we surveyed the literature, conducted a simulation study, and compared the performance of QSeq on three previously published datasets. Our simulation was designed to determine which analysis method produced the most accurate ecological inferences.

There has been disagreement over how to properly interpret compositional data while benchmarking differential abundance tests, because researchers have different objectives (McMurdie and Holmes, 2014; Weiss et al., 2017). Our approach is distinct from previous methods because we do not benchmark our findings against a compositional framework. Our benchmarking reference is the absolute abundance and true correlation of the taxa, not their relative abundances. Our simulation explicitly modeled the confounding effect of sequencing by including a subsampling routine. We supplemented the model with three published datasets with known environmental gradients to compare compositional sequence analysis, rarefaction, and composition analysis with QSeq normalization. Sequencing results were analyzed using PERMANOVA where samples had categorical environmental variables. We tested the ability of these different analysis approaches to distinguish the microbial community according to their respective experimental factors to determine how characteristics of the dataset might influence the quality of ecological inferences.

2.2 Materials and Methods

The data and scripts to recreate the simulation and synthesis analyses are available in the github repository: github.com/djeppschmidt/QSeq_model. The R package QSeq can be installed following the instructions at github.com/djeppschmidt/QSeq. The QSeq package transforms raw count data into relative abundance; then scales the sequence counts using independent user-provided total abundance data. The QSeq OTU table is then output for the user in a phyloseq (McMurdie and Holmes, 2013) compatible format. The package also provides auxiliary wrapper functions for importing sequence data from QIIME2 that rely on the qiime2R package (https://github.com/jbisanz/qiime2R#readme). Users should note that QSeq will not improve the performance of differential abundance tests that assume compositionality; rather it prepares the data for abundance or correlation analyses (e.g., regression against environmental variables, or correlation among taxa).

Simulation

The simulation is based on two primary datasets (Figure S2.1): 1) the Reference community with defined taxa abundances including any gradient in total abundance, and 2) the Test community subsampled from the reference dataset. The test community was the input in each of the statistical inference methods, and the inferred patterns were compared to the actual patterns in the Reference community. The subsampling routine was designed to model the effect of sequencing on data quality. The simulation was run 100 times for each unique set of parameters.

The Reference community consisted of 50 taxa across 10 samples. Reference community data was generated in two steps: 1) define the background variation in taxon abundance, and 2) impose a gradient of total abundance on the dataset. Since microbial communities exhibit skewed taxon abundances, mean abundance of each taxon across samples was defined using a lognormal distribution with a mean value of 50 and a standard deviation of 2.2. The abundance of each taxon in each sample was determined by drawing random numbers from a normal distribution, where each taxon's given mean abundance was defined by the lognormal distribution and the standard deviation was 10% of the mean for that taxon. A small amount of sparsity was imposed on the dataset by randomly assigning 5 observations a value of 0. The gradient in total abundance was defined by selecting 10 values from a normal distribution with a mean value of 10 and a standard deviation that was determined by the intended strength of the environmental gradient. The simulation was run with the standard deviation set to 0, 1, 2, 3, or 4, resulting in standard deviations in total abundance that was approximately 0, 10, 20, 30, or 40 percent of the mean total abundance of the sample. Total abundance values were rounded to the nearest integer and sorted by increasing value before being applied to the dataset. Each sample was multiplied by its respective gradient correction value, generating a dataset with the same probability distribution of the taxon relationships within the sample – but a range of total abundances across samples. This created a community dataset where biotic interactions did not change, but the environment increasingly impacted covariance and abundance patterns. Samples 1-5 were defined as group 1 for the differential abundance analysis; samples 6-10 were defined as group 2. Because the total abundance data was ordered from smallest to largest, the difference in group means will be proportional to the standard deviation in total abundance across sites. Note that our gradient simulation at value 0 did not exclude variation that was due to the random nature of the underlying taxon abundance distribution (Figure

2.2B). This dataset was considered the Reference dataset for benchmarking the differential abundance and taxon correlation inference methods.

To generate the Test community, the effect of sequencing needed to be simulated. This was done by randomizing the number of sequences per sample using a normal distribution. The sequencing simulation mean was 800 and the standard deviation was set to 30. This resulted in total counts of each sample that differed from the reference dataset, but the relative abundances of the taxa were representative of their relative abundance. This emulated the inference environment of actual sequence data, where the count abundance of any given taxon in any sample is determined primarily by its relative abundance and the total sequences captured in that sample.

QSeq methods approximate the total sample abundance using an independent measure. There may be error associated with the estimation of the total abundance of taxa in each sample. To determine the sensitivity of QSeq to this type of measurement error, we imposed a range of error values from 0 to 50% (Figure S2.2); for comparing differential abundance and correlation methods, we imposed a 5% accuracy penalty on the QSeq estimate of total abundance because this was near the error value at which QSeq performs worse than compositional analyses on datasets with no difference in total abundance for correlation analyses. To compare compositional and QSeq methods for differential abundance analysis, we applied ANCOM-BC (Lin and Peddada, 2020) and ALDEx2 (Fernandes et al., 2014) using default parameters on the subsampled data alongside a generalized linear model using the QSeq normalized data. We chose these two packages to represent compositional approaches because unlike many others, they are designed for microbiome analysis. ANCOM-BC infers absolute abundance of the population, while ALDEx2 provides the greatest user flexibility in defining the terms for the log-ratio normalization step. To assess the accuracy of correlation analysis, we applied FastSpar (Kurtz et al., 2015; Watts et al., 2019) and applied Pearson's correlation to the QSeq normalized data (Figure 2.2C).

For benchmarking, we used a generalized linear model and Pearson's correlation on the reference dataset to define the true responses of taxa. Note that as the standard deviation in total abundance increased, an increasing portion of the community differed in abundance between the two groups. On average, out of 50 taxa, 2.5 (or 0.05 percent) were expected to be differentially abundant at an alpha of 0.05 when there was no overlying gradient. We used a Pearson correlation on the reference dataset to define true correlations among taxa. Note that just like in the differential abundance analysis, as the variance in total abundance increased, correlations among the taxa become increasingly positively biased (Figure 2.2C). As the overlying gradient increased, correlations among taxa became less sparse, which is an explicit violation of the assumptions underlying the compositional correlation analysis.

Data Synthesis

Data from three previously published research articles were analyzed using QSeq (Table 2.1). These studies included a global study of urban soils (GLUSEEN; Epp Schmidt et al., 2017); a study involving the Farming Systems Project (FSP) – a USDA research initiative based in Maryland (Epp Schmidt et al., 2022); and a study of tidal fresh water wetland systems in Maryland (TFW; Prasse et al., 2015). Each study included 16S rRNA gene amplicon sequences and QPCR using the same F515 and R806 primers targeting the V3-V4 regions of 16S rRNA gene (Caporaso et al., 2012). The datasets differed in sequencing platforms (Table 2.1). GLUSEEN had a high degree of sparsity. FSP had a steep environmental gradient caused by soil depth. The TFW had a lower sequencing depth. DADA2 (Callahan et al., 2016, 2) was used for sequence error inference and to process the 16S sequence data from each study into Amplicon Sequence Variants (ASVs). Chimeras were removed. Community data was either transformed to relative abundance, transformed to QSeq, or were transformed using the centered log-ratio (CLR) transformed for compositional analysis. We applied two types of CLR as implemented in Vegan (v2.6-2): Aitchison distance with a pseudo-count of 1 per sample, with a CLR transformation that used the geometric mean of all taxa as the reference, and the “robust Aitchison” distance that does not require a pseudo-count and uses only taxa in each sample to calculate the geometric mean (Martino et al., 2019). We used our implementation of QSeq for the QSeq transformation. For each study, the variance explained for two main experimental factors was calculated using the adonis2 function in vegan v2.6-2 (Oksanen et al., 2016). These factors were: City and Land-use for GLUSEEN; Farming System and soil Depth for FSP; and, Plant and Site for TFW. We compared Jaccard, Bray-Curtis, and Aitchison distances.

Table 2.1: Studies used to compare beta diversity analysis, and unique characteristics of the dataset

Environment	Paper	Platform	Characteristic	N Samples (mean sequences per sample)
GLUSEEN	(Epp Schmidt et al., 2017)	Illumina	High sparsity	100 (139,896)
TFW	(Prasse et al., 2015)	Roche 454	Low Sequencing Depth	57 (9,737)
FSP	(Epp Schmidt et al., 2022)	Illumina	Strong Total Abundance Gradient	192 (81,030)

2.3 Results

Benchmarking

At zero variation in total abundance, ANCOM-BC had a higher true detection rate than ALDEx2, but did not differ from QSeq (Table 2.2). Increasing the standard deviation in total abundance slightly decreased the true detection rate of both compositional methods at inferring taxon abundance (Figure 2.2). Of the three methods, ALDEx2 had the lowest true and false detection rates. ALDEx2 offered flexibility in the user defined log-ratio however, so these results were not representative of all use-cases (see discussion). QSeq outperformed both methods when the standard deviation of total abundance exceeds 5% of the mean abundance across samples (Figure 2.2A; Table 2.2). The pattern was repeated in correlation inference. The compositional method FastSpar (modified sparCC) inferred correlations most accurately when there was almost no variation in absolute abundances but had decreased accuracy of inference as the standard deviation of the absolute abundances increased. Like the differential abundance analysis, Pearson correlation using QSeq transformed counts, outperformed compositional methods when the standard deviation in absolute abundance exceeded 5% of the mean absolute abundance and improved as the standard deviation of absolute abundances increased (Figure 2.2). Rarefaction was the worst at inferring taxon correlations regardless of variations in absolute abundance or standard deviations.

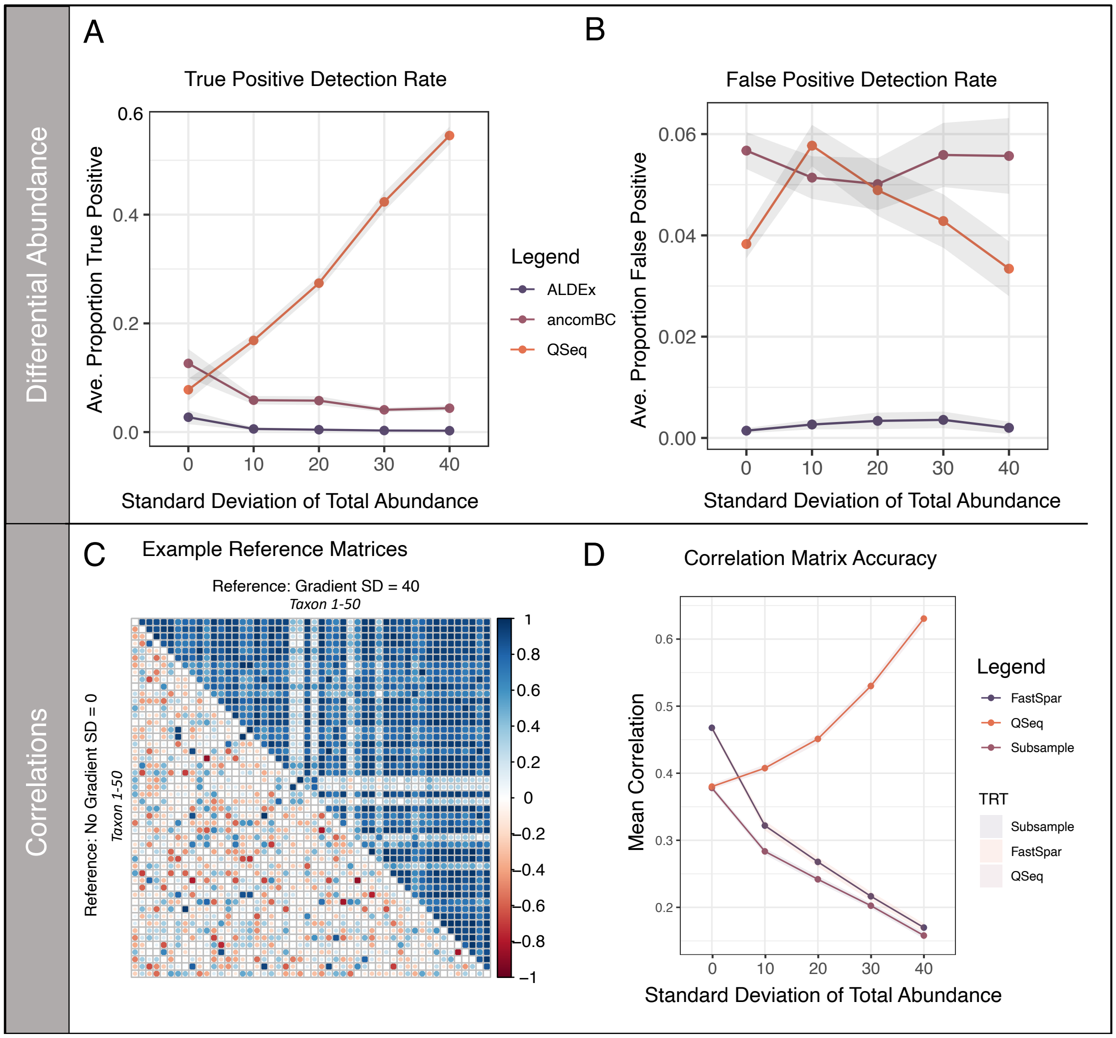

Figure 2.2: Simulated data demonstrating the effect of variation in total abundance on differential abundance and correlation analyses. A) true positive detection rate of three methods. B) false positive detection rate of three methods. For differential abundance tests, we compared ALDEx2 and ANCOM-BC, which represent two differing implementations of compositional differential abundance analysis. Each was implemented with the default settings. C) Correlation matrix demonstrating the difference in correlation patterns among taxa comparing datasets with nominal variation in total abundance (lower diagonal) versus 40 percent variation imposed on the total gradient (upper diagonal). In the latter case, almost all taxa are highly positively correlated. D) Correlation coefficient representing the accuracy of the correlation matrix under each of three construction methods. FastSpar is a python implementation of SparCC (compositional method) that corrects an error in the p-value estimation routine and speeds up calculations. The QSeq and Subsample each use a Pearson correlation under different normalization of the sample – Qseq adjusts the sample counts to represent the differences in total abundance (see methods for details). Except for panel C, the data in this figure represents the average value from 100 runs of the simulation.

Table 2.2: Tukey HSD comparisons of average true positive detection for taxa across methods.

TUKEY GROUP		Total Abundance Standard Deviation				
Group 1	Group 2	SD = 0	SD = 10	SD = 20	SD = 30	SD = 40
ANCOM-BC	ALDEx2	**0.01**	**0.001**	**0.001**	**0.001**	**0.001**
QSeq	ALDEx2	0.11	**0.001**	**0.001**	**0.001**	**0.001**
QSeq	ANCOM-BC	0.63	**0.001**	**0.001**	**0.001**	**0.001**

Table 2.3: Tukey HSD comparisons of correlation inferences among methods.

TUKEY GROUP		Total Abundance Standard Deviation				
Group 1	Group 2	SD = 0	SD = 10	SD = 20	SD = 30	SD = 40
FastSpar	Subsample	**0.001**	**0.001**	**0.012**	0.23	0.43
QSeq	Subsample	0.934	**0.001**	**0.001**	**0.001**	**0.001**
QSeq	FastSpar	**0.001**	**0.001**	**0.001**	**0.001**	**0.001**

QSeq: beta diversity

In the FSP dataset, only Aitchison distance detected a difference in community composition due to farming system (Table 2.4). All methods detected a strong gradient in community composition due to location of sample in the soil profile, but Bray-Curtis had the highest R squared. In the GLUSEEN dataset, all methods detected a strong effect of both the city and land-use category. In each of these cases, Bray-Curtis on relative abundance data provided the highest R squared (Table 2.4). This pattern was also replicated in the TFW study, except that all methods found a significant effect of site, but none of the methods detected a significant effect of plant species on the community composition. Variance in the absolute abundances were different among the three datasets, as determined by a Levene’s test of homogeneity of variance (Figure 2.3, $P = 0.001$). The standard deviation in absolute abundance in GLUSEEN was 75% of the mean; in TFW it was 98% of the mean; and in FSP it was 128% of the mean.

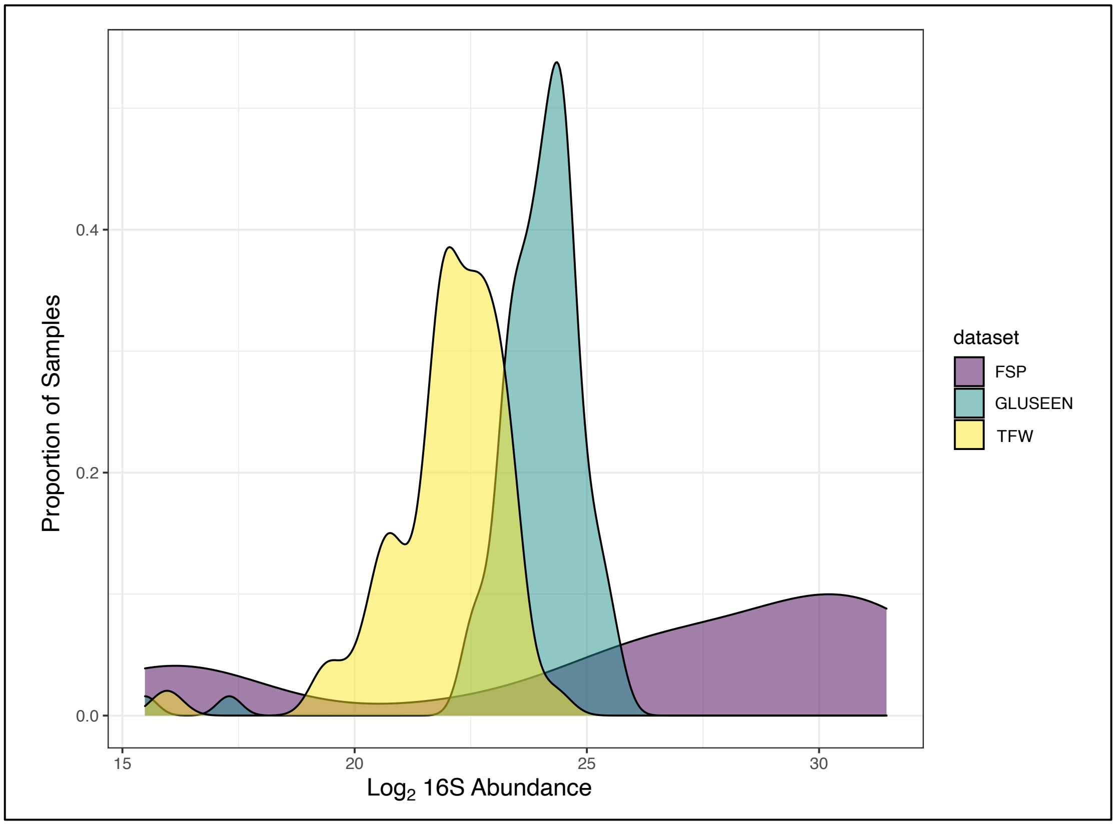

Figure 2.3: Distribution of $\log_2$ transformed total 16S gene abundance in each dataset. FSP has a much larger variance in total abundance compared to GLUSEEN or TFW (Levene’s test: $P < 0.001$). The standard deviation for total abundance in GLUSEEN samples was 75% of the mean total abundance; 98% total abundance in TFW; and 128% of the mean total abundance in FSP.

Table 2.4: explanatory power of different beta diversity methods.

		Relative Abundance		Compositional		QSeq	
		Jaccard	Bray-Curtis	Robust Aitchison	Aitchison	Jaccard	Bray-Curtis
FSP	Farming System	0.036 (0.122)	0.034 (0.229)	0.038 (0.075)	**0.046 (0.039)**	0.030 (0.293)	0.027 (0.519)
	Depth	**0.347 (0.001)**	**0.455 (0.001)**	**0.299 (0.001)**	**0.416 (0.001)**	**0.319 (0.001)**	**0.423 (0.001)**
GLUSEEN	City	**0.164 (0.001)**	**0.237 (0.001)**	**0.139 (0.001)**	**0.164 (0.001)**	**0.147 (0.001)**	**0.212 (0.001)**
	Land-use	**0.083 (0.001)**	**0.111 (0.001)**	**0.057 (0.001)**	**0.068 (0.001)**	**0.067 (0.001)**	**0.090 (0.001)**
MAES	SITE	**0.098 (0.001)**	**0.122 (0.001)**	**0.095 (0.001)**	**0.099 (0.001)**	**0.092 (0.001)**	**0.110 (0.001)**
	PLANT	0.053 (0.752)	0.052 (0.76)	0.053 (0.748)	0.053 (0.693)	0.053 (0.689)	0.052 (0.686)

2.4 Discussion

Compositional and quantitative analysis of empirical data may produce results that appear contradictory. For example, (Barlow et al., 2020) compared compositional and quantitative differential abundance approaches and found that the same taxa were found to respond differently depending on the method. Specifically, they found that a ketogenic diet resulted in nearly all taxa in the gut of mice decreasing abundance, but that the genus *Akkermansia*, in particular, did not change in abundance. *Akkermansia* did change relative to all other taxa (i.e. true positive in a compositional analysis), but did not change in absolute abundance (i.e. true negative for QSeq analysis). This result is broadly supported by conceptual work on compositionality in data (Gloor and Reid, 2016; Gloor et al., 2016, 2017;

Tsilimigras and Fodor, 2016). Our modeled data further support these findings and suggest that QSeq is most useful in determining how individual taxa respond to their environment or other members of the community, especially when there is variation in the absolute abundance of the community (Figures 2.2 & 2.4). Our simulation data agreed with the findings that QSeq can substantially improve the accuracy of abundance estimates for individual taxa in the dataset (Barlow et al., 2020; Jian et al., 2020).

Results of the simulated data provide boundaries for when QSeq should be expected to outperform compositional methods for inferring changes in abundance and correlation of taxa across environmental gradients. When there was a large variation (SD = 40% of the mean) in absolute abundance, QSeq outperformed compositional methods even when the abundance estimation had 50 % error (Figure S2.2). Taxon relationships were better predicted using QSeq when the total abundance measurement error was $<$ 5 % and the standard deviation of sample abundance was > 5% of the mean absolute abundance of taxa across samples (Figure 2.2). Although measurement error can alter the threshold, our models suggested that a standard deviation of 5% of the mean total abundance across samples could be considered a conservative threshold for determining if absolute abundance is “invariant” with respect to the compositional network analyses. This was a surprisingly low value. Across three datasets we analyzed, the lowest relative standard deviation in total abundance was 75% of the mean total abundance (GLUSEEN) and the highest was 128% of the mean total abundance (FSP; Figure 2.3). These results suggest that in many environmental contexts the dataset is unlikely to fit the assumptions of compositional analyses. As a standard practice, researchers should quantify both the variation in absolute gene abundance across their dataset and estimate the error in these measurements to determine whether or not to apply QSeq.

Compositional analyses primarily use a transformation analogous to the centered log-ratio (CLR; Aitchison, 1982). CLR is calculated by dividing the abundance of each taxon in each sample by the geometric mean of that sample. This normalizes the counts within each sample to the average count abundance within that sample, thereby standardizing taxon counts within each sample to the mean abundance of the respective sample. CLR transformations that use the sample-wise geometric mean as reference result in compositional analyses that are unable to account for differences in the mean abundance among samples (Aitchison, 1982; Gloor and Reid, 2016; Gloor et al., 2016, 2017; Tsilimigras and Fodor, 2016). Therefore, the poor performance of compositional inference methods is an expected result when the underlying data have large differences in sample-wise mean taxon abundance values. There is an exception to this generality. There are a variety of methods that may be used to calculate CLR using different subsets of the taxa as reference in the log-ratio (Fernandes et al., 2013, 2014). If a feature within the dataset truly does not differ in abundance across samples, it may be used as the reference (Quinn et al., 2019) and would result in CLR values that accurately reflect the log-fold change of the true abundance of taxa. This situation is unlikely to occur naturally; however, spike-in techniques (Smets et al., 2016; Stämmler et al., 2016; Tourlousse et al., 2017) could generate an appropriate control to create this condition.

Meanwhile compositional correlation analysis is also unlikely to accurately represent patterns in the community because the assumptions of the method are easily violated. Compositional correlation analysis depends on the observation that the ratio of two features in a compositional dataset accurately represents the ratio of their absolute abundance (Aitchison, 1982). This observation is used as the basis to infer correlation between taxa under the assumptions that 1) the ratio of two strongly positively correlated taxa will not change (i.e. will be invariant across samples); 2) conversely, the ratios of two strongly negatively correlated taxa will be highly variable across samples; and 3) that relatively few taxa exhibit strong correlations (Kurtz et al., 2015; Watts et al., 2019). When combined, these assumptions allow the distribution of log-ratio values of taxa across samples to be used to identify the outliers (high and low variance) as the strong positive and negative correlations (Kurtz et al., 2015). However, the assumptions of this method are easily violated under normal environmental conditions. For example, taxa can be strongly positively correlated, and yet have high variance in log-ratios. To illustrate this, consider three sites (Figure 1) that have three taxa each. The abundance of taxon 1 across sites is respectively, 1, 26, and 51. The abundance of taxon 2 across sites is respectively 25, 26, and 27. The abundance of taxon 3 is respectively 100, 101, and 102. When correlated to one another, all three taxa have a Pearson correlation coefficient of 1 across all three sites. But they have different variances in log-ratio; taxa 2 and 3 have low variance (<0.001), and taxon 1 has high variance with each of the other taxa (>4.25). A compositional correlation inference framework would determine that taxon 1 is negatively correlated with taxon 2 and 3; and that taxa 2 and 3 are positively correlated. These inferences are flawed, however. If most taxa covary in abundance due to a strong environmental gradient, then variation in log-ratio does not reflect correlations amongst taxa, rather it reflects the relative response rate of each compared to other taxa. It is common for steep environmental gradients to define ecotones and therefore exhibit strong correlations among large portions of the community across the transition (Smith et al., 2018), and it is common for there to be steep environmental gradients that determine the abundance of microbial taxa (Figure 3; Barlow et al., 2020; Epp Schmidt et al., 2022). Since most environmental studies examine temporal and spatial variations that are expected to delineate differences in the microbial community, researchers should be cognizant of the high likelihood for these patterns in their data.

2.5 Conclusions

Change in the density of taxa, rather than their composition, is critically important for unraveling the underlying ecological drivers of microbial community assembly. Erroneous inferences are easily made if the absolute abundances of taxa are not accurately estimated. We show that when absolute community abundance changed by as little as 5%, QSeq more accurately predicted taxon abundance and covariances among taxa. There is a high likelihood that most, if not all, environmental microbiome datasets exceed this total abundance variability threshold. Thus, differential abundance analysis, and correlation analysis should by default be done using the QSeq data transformation. QSeq offers no advantage to the estimation of alpha diversity. And whether QSeq improves the ecological inference of beta diversity likely depends heavily on the distance index used and the question being investigated. We urge that developers for microbiome analysis pipelines offer support for QSeq analysis as part of their standard pipeline.

2.6 Acknowledgements

Dietrich Epp Schmidt was supported by NRT-INFEWS: UMD Global STEWARDS (STEM Training at the Nexus of Energy, WAter Reuseand FooD Systems) that was awarded to the University of Maryland School of Public Health by the National Science Foundation National Research Traineeship Program, Grant number 1828910.

We thank Dr. Mihai Pop for feedback that substantially improved the manuscript.

2.7 Literature Cited

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

Alteio, L. V., Séneca, J., Canarini, A., Angel, R., Jansa, J., Guseva, K., et al. (2021). A critical perspective on interpreting amplicon sequencing data in soil ecological research. *Soil Biology and Biochemistry* 160, 108357. doi: 10.1016/j.soilbio.2021.108357.

Auguet, J.-C., Barberan, A., and Casamayor, E. O. (2009). Global ecological patterns in uncultured Archaea. *ISME J* 4, 182–190. doi: 10.1038/ismej.2009.109.

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., et al. (2018). Structure and function of the global topsoil microbiome. *Nature* 560, 233–237. doi: 10.1038/s41586-018-0386-6.

Baiser, B., Olden, J. D., Record, S., Lockwood, J. L., and McKinney, M. L. (2012). Pattern and process of biotic homogenization in the New Pangaea. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20121651. doi: 10.1098/rspb.2012.1651.

Barlow, J. T., Bogatyrev, S. R., and Ismagilov, R. F. (2020). A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. *Nature Communications* 11, 2590. doi: 10.1038/s41467-020-16224-6.

Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knight, R., and Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME J* 5, 908–917. doi: 10.1038/ismej.2010.171.

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6, 90. doi: 10.1186/s40168-018-0470-z.

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

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

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6, 1621–1624. doi: 10.1038/ismej.2012.8.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS* 108, 4516–4522. doi: 10.1073/pnas.1000080107.

Epp Schmidt, D., Dlott, G., Cavigelli, M., Yarwood, S., and Maul, J. E. (2022). Soil microbiomes in three farming systems more affected by depth than farming system. *Applied Soil Ecology* 173, 104396. doi: 10.1016/j.apsoil.2022.104396.

Epp Schmidt, D. J., Pouyat, R., Szlavecz, K., Setälä, H., Kotze, D. J., Yesilonis, I., et al. (2017). Urbanization erodes ectomycorrhizal fungal diversity and may cause microbial communities to converge. *Nature Ecology & Evolution* 1, 0123. doi: 10.1038/s41559-017-0123.

Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G., and Gloor, G. B. (2013). ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLOS ONE* 8, e67019. doi: 10.1371/journal.pone.0067019.

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

Fierer, N., and Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *PNAS* 103, 626–631. doi: 10.1073/pnas.0507535103.

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The Treatment-Naive Microbiome in New-Onset Crohn’s Disease. *Cell Host & Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* 0. doi: 10.3389/fmicb.2017.02224.

Gloor, G. B., and Reid, G. (2016). Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can. J. Microbiol.* 62, 692–703. doi: 10.1139/cjm-2015-0821.

Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V., and Egozcue, J. J. (2016). It's all relative: analyzing microbiome data as compositions. *Ann Epidemiol* 26, 322–329. doi: 10.1016/j.annepidem.2016.03.003.

Hoshino, T., Nakao, R., Doi, H., and Minamoto, T. (2021). Simultaneous absolute quantification and sequencing of fish environmental DNA in a mesocosm by quantitative sequencing technique. *Sci Rep* 11, 4372. doi: 10.1038/s41598-021-83318-6.

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., and Korpela, K. (2020). Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLOS ONE* 15, e0227285. doi: 10.1371/journal.pone.0227285.

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. (2015). Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLOS Computational Biology* 11, e1004226. doi: 10.1371/journal.pcbi.1004226.

Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl. Environ. Microbiol.* 75, 5111–5120. doi: 10.1128/AEM.00335-09.

Lin, H., and Peddada, S. D. (2020). Analysis of compositions of microbiomes with bias correction. *Nat Commun* 11, 3514. doi: 10.1038/s41467-020-17041-7.

López-García, A., Pineda-Quiroga, C., Atxaerandio, R., Pérez, A., Hernández, I., García-Rodríguez, A., et al. (2018). Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.03010.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 550. doi: 10.1186/s13059-014-0550-8.

Martino, C., Morton, J. T., Marotz, C. A., Thompson, L. R., Tripathi, A., Knight, R., et al. (2019). A Novel Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems* 4, e00016-19. doi: 10.1128/mSystems.00016-19.

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

McMurdie, P. J., and Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol* 10, e1003531. doi: 10.1371/journal.pcbi.1003531.

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2016). vegan: Community Ecology Package version 2.4-0. Available at: http://CRAN.R-project.org/package=vegan.

Ovaskainen, O., and Abrego, N. (2020). *Joint Species Distribution Modelling: With Applications in R (*. Cambridge University Press.

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., and Wilding, T. A. (2014). Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources* 14, 1129–1140. doi: 10.1111/1755-0998.12261.

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., and Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology* 55, 12–25. doi: 10.1016/j.ejop.2016.02.003.

Prasse, C. E., Baldwin, A. H., and Yarwood, S. A. (2015). Site History and Edaphic Features Override the Influence of Plant Species on Microbial Communities in Restored Tidal Freshwater Wetlands. *Applied and Environmental Microbiology* 81, 3482–3491. doi: 10.1128/AEM.00038-15.

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., and Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *Gigascience* 8, giz107. doi: 10.1093/gigascience/giz107.

Schloss, P. D. (2020). Reintroducing mothur: 10 Years Later. *Appl Environ Microbiol* 86. doi: 10.1128/AEM.02343-19.

Shade, A., and Stopnisek, N. (2019). Abundance-occupancy distributions to prioritize plant core microbiome membership. *Current Opinion in Microbiology* 49, 50–58. doi: 10.1016/j.mib.2019.09.008.

Smets, W., Leff, J. W., Bradford, M. A., McCulley, R. L., Lebeer, S., and Fierer, N. (2016). A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biology and Biochemistry* 96, 145–151. doi: 10.1016/j.soilbio.2016.02.003.

Smith, J. R., Letten, A. D., Ke, P.-J., Anderson, C. B., Hendershot, J. N., Dhami, M. K., et al. (2018). A global test of ecoregions. *Nat Ecol Evol* 2, 1889–1896. doi: 10.1038/s41559-018-0709-x.

Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P. J., et al. (2016). Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome* 4, 28. doi: 10.1186/s40168-016-0175-0.

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., et al. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution* 11, 442–447. doi: 10.1111/2041-210X.13345.

Tourlousse, D. M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., and Sekiguchi, Y. (2017). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Research* 45, e23. doi: 10.1093/nar/gkw984.

Tsilimigras, M. C. B., and Fodor, A. A. (2016). Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of Epidemiology* 26, 330–335. doi: 10.1016/j.annepidem.2016.03.002.

Vandeputte, D., Kathagen, G., D’hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 551, 507–511. doi: 10.1038/nature24460.

Watts, S. C., Ritchie, S. C., Inouye, M., and Holt, K. E. (2019). FastSpar: rapid and scalable correlation estimation for compositional data. *Bioinformatics* 35, 1064–1066. doi: 10.1093/bioinformatics/bty734.

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27. doi: 10.1186/s40168-017-0237-y.

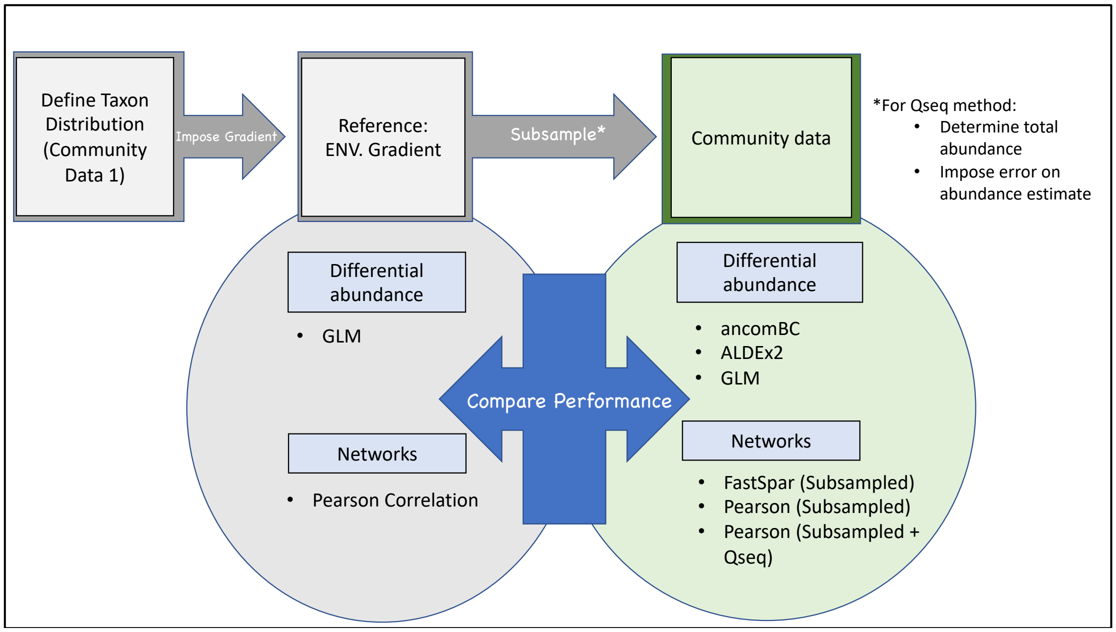

Figure S2.1: diagram of the process for the simulation from Figure 1.

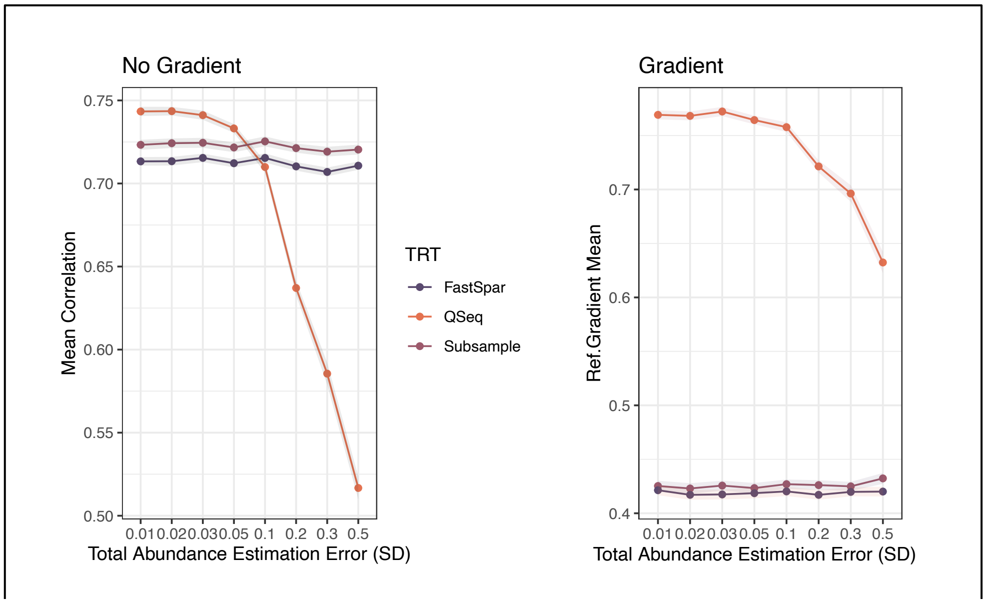

Figure S2.2: Sensitivity of QSeq to total abundance estimation error on a dataset with no skewness with and without an environmental gradient.

Table S2.1: Table describing the justification of quadrants for Figure 2.4.

Abundance Category	**Community Category**	**Analysis**	**Rationale**
Important	Taxon	Abundance due to environment, including other taxa (linear and logistic regressions; correlation & covariance)	Compositional analyses are designed to eliminate gradients in total abundance – thus cannot detect correlations among taxa that are due to environmental gradients, especially those that influence the overall abundance of taxa in the community. Our modeling data shows clearly that small amounts of variation in total abundance (greater than 5%) substantially reduces the accuracy of both compositional and Pearson correlation analyses without QSeq normalization; but that QSeq normalization improves the accuracy of correlation analyses, especially when there is greater than 10% variation in total abundance among samples.
	Community	Correlation or Covariance Network	This is a community-wide application of single correlation or regression analysis among taxa
		Total microbial abundance	This is a direct measure of total microbial abundance
		Abundance-Occupancy Models	Require accurate estimation of taxon abundance; compositional methods measure proportion, not abundance
		Joint Species Distribution Models	Community-wide application of single taxon regression models; require accurate estimation of taxon abundance.
	Taxon	Occupancy	Occupancy is the measure of taxon prevalence (or detection of taxa) – accurate estimation of taxon

Not Important			abundance is not important for this analysis. However, researchers should include in the model an estimation of the effect of sequencing depth on detection probability to account for sequencing effort.
	Community	Alpha Diversity (Richness, Evenness)	These are community-scale metrics of the rank abundance curve, and are not sensitive to total abundance. Richness estimates are strongly influenced by sequencing depth (which should be controlled); measrues of evenness reflect the relative detection probabilities of the taxa. Neither require accurate estimation of true taxon abundance in the environment.
		Co-Occurrence Networks	Unlike correlation and covariance, occurrence reflects occupancy and is not dependent on the abundance of each taxon. Similar to occupancy analyses, the co-occurrence network should account for sequencing effort in each sample, and estimate the likelihood of non-detection due to undersampling.
		Differential Proportions / Relative Abundance	This is in the community category because compositions or proportions of one taxon are determined in part by all taxa. And it is in the not-abundant category because measuring accurately the abundance of the taxon is not important to determining it's proportion.
Either Important,	Taxon	Indicator Taxa	Indicator species can be determined on the basis of occupancy (prevalence within a given environment); but in some applications, especially in environmental quality monitoring, the

or Not Important			abundance of a specific taxon is itself an important indicator. Thus, whether abundance is important depends on the objective of the indicator.
	Community	Beta Diversity	This depends on the research objectives and assumptions. Some similarity indices (such as Jaccard or Bray-Curtis) reflect zeros and total abundance, which is desirable if the community differs in occurrence and abundance across the measured conditions (and if the researcher is interested in these aspects). Other similarity indices (such as Aitchison distance) more accurately reflect relative changes in abundance, but cannot accept zeros and do not reflect total abundance. If the question is about differences in relative proportion, then abundance may not be important to account for.

Chapter 3: Quantitative adjustment of metagenome prediction for improved accuracy

Dietrich Epp Schmidt[1], Jude E Maul[2], Christine Maietta[3], Lindsay Wood[4], Mohamad Bahram[5], Koki Maeta[6], Stephanie A Yarwood[1].

[1]University of Maryland College Park
[2]United States Department of Agriculture Beltsville Agricultural Research Service
[3]The Pew Charitable Trust
[4]Cold Regions Research and Engineering Laboratory, New Hampshire
[5]University of Tartu, Estonia
[6] Crop, Livestock & Environment Division, JIRCAS, Tsukuba, Ibaraki, Japan

Abstract

High throughput sequencing coupled with metabarcoding has widely been adopted as method to monitor environmental microbiomes. Most studies target 16S rRNA genes, but microbial taxonomy, especially in the case of bacteria, is sometimes an unreliable predictor of function. Many software programs have been developed to infer the presence of functional genes based on sequence data from environmental samples, however guidelines for how to interpret these data are poorly defined. Typically, the accuracy of the software is benchmarked against metagenome data, all of which is expressed as relative abundance. The actual quantity of genes in the environment are rarely used to substantiate the accuracy of the predicted metagenomes. To address this issue, we compared the accuracy of gene quantity estimated by PICRUSt2 to gene copy number measured by QPCR. Data from multiple studies of different environmental systems and at different spatial scales were synthesized. Quantification of 16S rRNA genes and sequence data were combined using QSeq to generate estimate of true quantity. We hypothesized

that this method when applied to PICRUSt2 data would improve the accuracy of gene abundance estimates. We further hypothesized that horizontal gene transfer would result in certain genes being less reliably estimated than others. We found that total abundance correction improved PICRUSt2 predications, but there was no difference in prediction quality based on gene identity across studies. Rather, unique characteristics of each study lead to specific genes being more or less accurately predicted, suggesting that the primary barrier to consistently accurate gene abundance prediction is database quality. We suggest that gene prediction software cannot be the sole evidence used to infer absolute abundance of many functional genes.

3.1 Introduction

Microbes are ubiquitous and essential members of all ecosystems on earth (Cavicchioli et al., 2019). These taxa mediate many important biogeochemical cycles (Bardgett and van der Putten, 2014); diseases in human health and the health of crops, and livestock (Suman et al., 2022) and affect the primary productivity of ecosystems. Thus, microbiome function is essential to many ecosystem services. Recent advances in sequencing technology have revolutionized microbiome research, allowing communities to be rapidly sampled at large spatial scales at a reasonable cost (Fierer and Jackson, 2006). Amplicon sequencing has become a staple component of environmental monitoring and inquiry activities (Apothéloz-Perret-Gentil et al., 2017; Cordier et al., 2019, 2018), particularly for bacteria that can be characterized using 16S rRNA genes. 16S rRNA gene studies are often limited in the functional information provided by the sequencing (Fierer et al., 2021). Because whole genome sequencing is considerably more expensive, and because other quantitative methods like QPCR require more laboratory time per unit of information (Porter and Hajibabaei, 2018) a suite of database and annotation tools have been developed to infer metagenomic content from amplicon data (Djemiel et al., 2022).

There are several strategies to predict genomic content of a 16S rRNA gene dataset. The first is exemplified by PICRUSt2 (Douglas et al., 2020), which uses an ancestral state model to predict the genomic content of bacteria and archaea based on where they are placed in the phylogenetic tree. In contrast, Tax4Fun2 predicts genomic content of taxa by averaging the genomic content of closely related genomes in the reference database (Wemheuer et al., 2020, p. 2, 2018, p. 2). A third approach, FAPROTAX predicts genome content at the species level based on a manually curated list of genes that exist in all sequenced strains within the species (Sansupa et al., 2021). All of these methods work by matching taxa from the sample to a reference database of genome content; they differ in how they construct the reference database, and the confidence thresholds for predicting genomic content (i.e. precision of sample to reference match, confidence level in presence of gene, etc). The accuracy of metagenome inference software is typically benchmarked by correlating the predicted metagenome content against measured metagenome content as determined by shotgun metagenomic data (Aßhauer et al., 2015; Douglas et al., 2020; Iwai et al., 2016; Jun et al., 2015; Narayan et al., 2020; Wemheuer et al., 2020, 2018). This approach has the advantage that the detection of multiple genes and pathways can be simultaneously tested, but suffers from several significant disadvantages.

The first weakness of using shotgun metagenomic data as the reference is that they are based on relative abundance; all sequence data is inherently compositional because the number of sequences per sample is random in multiplexed runs (Fernandes et al., 2014; Gloor et al., 2017, 2016). This limitation is virtually never addressed in the context of metagenome prediction however (Djemiel et al., 2022), even though it is an important aspect of all ‘omics data (Quinn et al., 2019). The compositional nature of sequence data implies that neither 16S rRNA gene amplicon nor shotgun metagenomic datasets give direct information about the total abundance of the gene in the environment. Differences in compositional datasets are best interpreted as differences in the *relative response rate* of their components to experimental conditions; they are unreliable as an indicator of true abundance if the overall abundance of features (in this case, genes) changes across experimental conditions (see Chapter 2). Predicted metagenomes are also compositional, making inferences about the effect of environmental gradients on the absolute abundance of genes impossible based solely on sequencing data and metagenome prediction.

The second weakness compounds the first. Within compositional metagenomics data there is a high degree of structural correlation in gene abundance that is due to the actual correlation among certain gene families (Douglas et al., 2020). Genes coding for universal cell functions like DNA maintenance and replication will always be more abundant than genes for non-universal metabolic functions like nitrogen fixation or methanogenesis. This results in a systemic positive correlation bias that can be seen in benchmarking efforts. For example (Douglas et al., 2020) used a “null” control reference genome that was defined as the mean gene family abundance across all reference genomes. This null model indicates how much variation is explained by the relative prevalence of gene families across the reference genomes before considering the treatment effects. The mean correlation between reference metagenomes and this null prediction model ranged from 0.6 to 0.8, indicating that much of the correlation between predicted and measured metagenomes found in the literature is due simply to the relative dominance of certain gene families across all genomes. Efforts to use a differential abundance framework, which is more sensitive to changes in the relative abundance of rarer features (Gloor et al., 2016), to benchmark the metagenome prediction tools generally find that predicted metagenomes poorly predict which genes are differentially abundant (Douglas et al., 2020; Narayan et al., 2020). This indicates that in general predicted metagenomes also poorly reflect the relative abundance of rarer metabolic functions.

Recent work in amplicon sequence datasets has challenged the notion that sequence datasets must be analyzed in a compositional framework (Barlow et al., 2020; Epp Schmidt et al., 2022; Jian et al., 2018). By combining sequence data and QPCR, researchers may generate quantitative estimates of microbial taxa. Theoretically, these datasets could also generate quantitative predictions of functional gene abundances. In this paper, we test whether predictions using 16S rRNA gene Quantitative Sequencing are accurate when compared to Quantitative PCR of target functional genes. We combined data from several previously published datasets that include 16S rRNA gene sequencing and QPCR, and QPCR of at least one functional gene. We focused on the prediction of commonly studied environmental genes that contribute to C and N cycling: nitrite reductases (nirS, nirK), nitrous oxide reductase (nosZ), particulate methane mono-oxygenase subunit A / ammonia mono-oxygenase subunit A (pmoA-amoA), nitrogenase H (nifH), and methyl coenzyme M reductase alpha subunit (mcrA). We had two hypotheses: 1) Using the quantitative sequencing method will substantially improve the accuracy of predicted gene quantity from the predicted metagenomes and 2) genes will differ in how accurately their quantities can be predicted based on how closely gene phylogeny matches 16S rRNA gene phylogeny.

3.2 Materials and Methods

To determine if the abundance of functional genes can be accurately predicted using the 16S rRNA gene markers as a reference, we conducted a meta-analysis using six studies (Table 3.1) that included six different functional genes (amoA, mcrA, nifH, nirS, nirk, and nosZ). The KO and EC values for each of the genes are summarized in Table 3.2. The specific primer pairs used for quantifying each gene in each study are presented in Table 3.3. Each study needed to contain QPCR values for 16S rRNA and at least one functional gene. Although there are numerous studies that collected these data along with 16S rRNA sequence, most present only summary data for QPCR. Since this study needed the sample-wise values of QPCR, as well as the sequence data, we reached out directly to the authors. This resulted in a relatively small list of representative studies that could be included.

Table 3.1: Description of studies included in meta-analysis.

Dataset	Environment	Gene Targets	Citation
Maryland Wetland	Wetland soil – local to Maryland, USA	nirS, nirK, nosZ, mcrA, pmoA	Prasse et al., (2015)
Global Wetland	Wetland soil – Global analysis	amoA, nirS, nirK, nosZI, nosZII, nifH	Bahram et al., (2022)
Industrial Manure	Industrial – Manure maturation experiment	amoA, nirS, nirK, nosZ, mcrA	Nguyen et al., (2022)
Urban	Urban soils – global gradient in land use	amoA (archaea), amoA (bacteria)	Epp Schmidt et al., (2019)
Maryland Agriculture	Agricultural soils – local to Maryland, USA	nifH	Epp Schmidt et al., (in review)
Maryland Riparian	Urbanized riparian soils – local to Maryland, USA	nirS, nirK	Wood et al., (in review)

Study specific data preparation

The Maryland Wetland dataset was originally sequenced on a Roche 454 sequencer; we accessed the sequences from NCBI (PRJNA276421), then used QIIME2 (Bokulich et al., 2018; Caporaso et al., 2010) to demultiplex the samples, removed barcodes, and ran DADA2 amplicon sequence variant (ASV) inference. This dataset measured amoA and pmoA separately; however for many samples amoA was below detection limit. Therefore, we discarded the amoA measures and only kept pmoA for the correlation analysis.

The Global Wetland sequences were downloaded from NCBI (PRJNA718418). The sample metadata was supplied by the authors (Bahram et al., 2022). DADA2 was used for chimera filtering, sequence quality control, and sequence variant inference. Taxonomy was assigned from SILVA (release no. 132); and the dataset was filtered to only include bacterial and archaeal sequences. In this study, there were multiple sequencing runs for each QPCR sample that was taken; therefore, we averaged sequencing runs to produce one community profile per QPCR sample. Additionally, this dataset measured nosZ I and nosZ II abundance separately; we summed the abundance across both primers to estimate total nosZ abundance.

Sequence and meta data for Industrial Manure, Maryland Agricultural, Urban, and Maryland Riparian studies were provided directly by the authors. In each of these studies, DADA2 was used for chimera filtering, sequence quality control, and sequence variant inference. Taxonomy was assigned against SILVA (release 132) database. The upland urban study included bacterial and archaeal amoA gene quantities for QPCR – these were summed to capture total measured amoA.

Table 3.2: KO and EC values for each gene

Gene	KO Number	EC Number
nirS	15864	1.7.2.1 & 1.7.99.1
nirK	00368	1.7.2.1
nosZ	00376	1.7.2.4
pmoA	10944	1.14.18.3
amoA	10944	1.14.99.39
mcrA	00399	2.8.4.1
nifH	02588	--

Table 3.3: Genes and primers in each study. When different, citation for forward (F) and reverse (R) primers are indicated. For amoA, primers that target bacteria (AOB) and archaea (AOA) are indicated. Phylogenetically constraints are also indicated for nosZ; eg clade 1 (nosZ1) and clade 2 (nosZ2).

Dataset	nirS	nirK	nosZ	pmoA-amoA	mcrA	nifH
Maryland Wetland	(Kandeler et al., 2006)	(Henry et al., 2004)	nosZ1: (Henry et al., 2006)	pmoA: (Im et al., 2011)	(Pereyra et al., 2010)	--
Global Wetland	F:(Wei et al., 2015); R: (Kandeler et al., 2006)	(Henry et al., 2004)	nosZ1: (Henry et al., 2006) nosZ2: (Jones et al., 2013)	AOB: (Rotthauwe et al., 1997) AOA: (Tourna et al., 2008)	--	Ueda et al 1995
Industrial Manure	(Throbäck et al., 2004)	(Throbäck et al., 2004)	(Throbäck et al., 2004)	AOB: (Rotthauwe et al., 1997) AOA: (Meinhardt et al., 2015)	(Angel et al., 2011)	--
Urban	--	--	--	AOB: (Rotthauwe et al., 1997) AOA: (Mincer et al., 2007)	--	--
Maryland Agriculture	--	--	--	--	--	(Rösch et al., 2002)
Maryland Riparian	(Kandeler et al., 2006)	(Henry et al., 2004)	--	--	--	--

Predicted functional gene profiles were constructed using PICRUSt2 (v2.4.1) and Tax4Fun2 (v1.1.5). Tax4Fun2 is no longer available in a public repository; so, we provide a tar file in the supplemental to allow for the analysis in this paper to be replicated. Since Tax4Fun2 is no longer being supported, the purpose for including Tax4Fun2 in this analysis is not to compare its accuracy to PICRUSt2. Rather, we include the Tax4Fun2 output in this analysis because the specific parameterization of Tax4Fun2 allows us to explore how reference dataset quality and match precision interact in producing functional gene predictions. In Tax4Fun2, there are two reference datasets, where reference genomes are clustered at the 99% and 100% 16S rRNA gene sequence match. And then the user may additionally determine the percent match of sample 16S rRNA gene sequences must have to the reference. We ran Tax4Fun2 using both reference datasets, and a range of matches from 96% to 100%. PICRUSt2 does not allow for this exploration.

Both packages produce estimates of gene abundance based on the count data supplied; therefore, we generated quantitative estimates by first transforming these estimated abundances to relative abundance. Then, similar to QSeq methods published elsewhere (Epp Schmidt et al., 2022; Jian et al., 2018), we scaled the abundance of functional genes using the total number of 16S rRNA genes to produce a predicted abundance value for each gene. In principle this should produce a predicted gene quantity. PICRUSt2 provides metagenome predictions in terms of KEGG Orthology (KO), and Enzyme Nomenclature (EC) numbers. EC numbers are treated as attributes of KO numbers (Kanehisa et al., 2016a, 2016b). There may be multiple EC numbers for each KO, but there also may be multiple KO numbers that are represented by an EC number. Therefore, we examined the PICRUSt2 predictions using both the KO and EC output.

Statistics

Pearson correlation coefficient were calculated between measured and predicted gene abundance in each dataset using each software output (both raw output and transformed abundance). Using only the PICRUSt2 results, we determined for each pair of normalization methods (quantitative or relative) whether quantitative normalization increased or decreased the correlation. The Chi-squared test was used to determine whether the QSeq consistently increased the correlation between predicted and measured gene abundance across datasets. We also used a linear model to determine if QSeq and gene identity affected the correlation between measured and predicted gene quantity. For this linear model, we used the study of origin as a blocking factor to control for differences among studies. We also calculated the Pearson Correlation coefficient for predicted and measured gene abundance across all datasets for each gene.

3.3 Results

Correcting total abundance consistently improved the correlation between measured and predicted gene abundance, regardless of whether KO or EC annotations were used (Figure 3.1; Chi-Squared test: $P < 0.001$; each ANOVA $P < 0.012$). In contrast, the identity of the gene did not affect the correlation between measured and predicted gene quantity, regardless of whether KO or EC values were used (Figure 3.1; each ANOVA, $P > 0.3$). For each of the focal genes (nirS, nirK, nosZ, and amoA), at least one dataset produced Pearson correlation coefficient greater than 0.7 comparing the measured quantity to a quantitative prediction (Table 3.5). Predicted abundance of nirK was strongly correlated to measured gene abundance across datasets (Figure 3.2; KO Pearson R =0.82; EC Pearson R = 0.83). Conversely, the predicted abundance of nirS was not strongly correlated to measured nirS abundance across datasets (Figure 3.3; Pearson KO R = 0.32, EC R = 0.42). Predicted abundance of pmoA-amoA genes were weakly correlated across datasets (Figure 3.4; KO and EC Pearson R = 0.50). And the predicted abundance of nosZ was weakly correlated to measured abundance (Figures 3.4 & 3.5; KO and EC Pearson R = 0.54). Whether EC or KO provided more reliable predictions depended on the dataset and the gene. For example in Maryland wetlands, EC provided a higher correlation than KO for pmoA and nirK, but not nirS (Table 3.4 & Table 3.5). There was no EC number for nifH, so nifH could not be predicted using EC nomenclature.

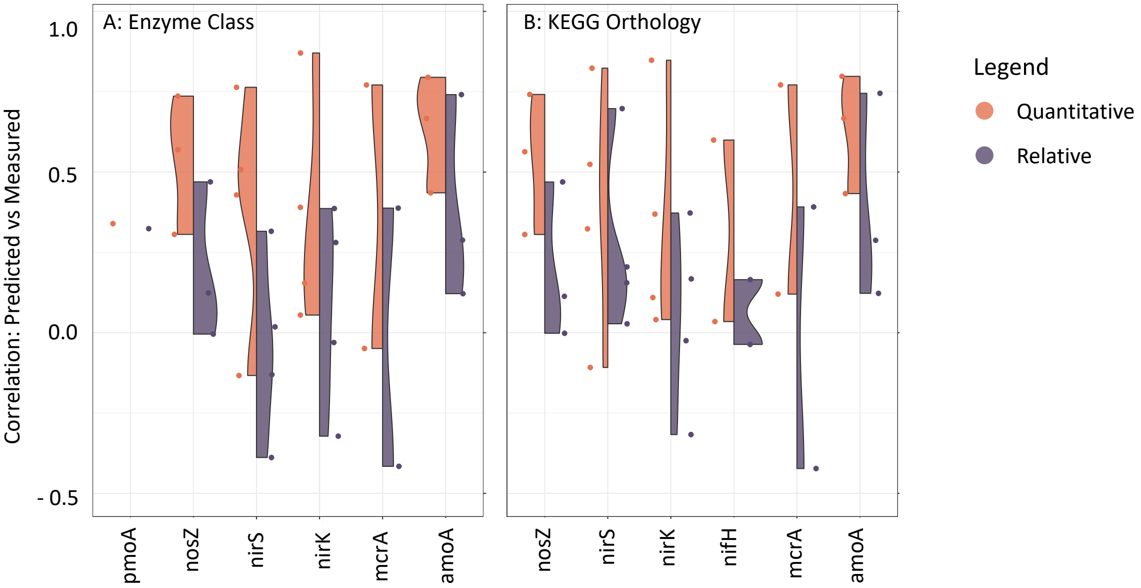

Figure 3.1: Cloud and rain plot depicting the correlation of predicted gene quantity to measured gene quantity across datasets and genes using A) Enzyme Class (EC) or B) Kegg Orthology (KO) values predicted from PICRUSt2. PICRUSt2 outputs were transformed to relative abundance (Relative) or were transformed to relative abundance then scaled sample-wise by 16S QPCR value (Quantitative). Quantitative predictions were consistently more accurate than Relative predictions (Chi-Squared test: P < 0.001; each ANOVA P < 0.012), but there were no differences among genes (each ANOVA, P > 0.3)

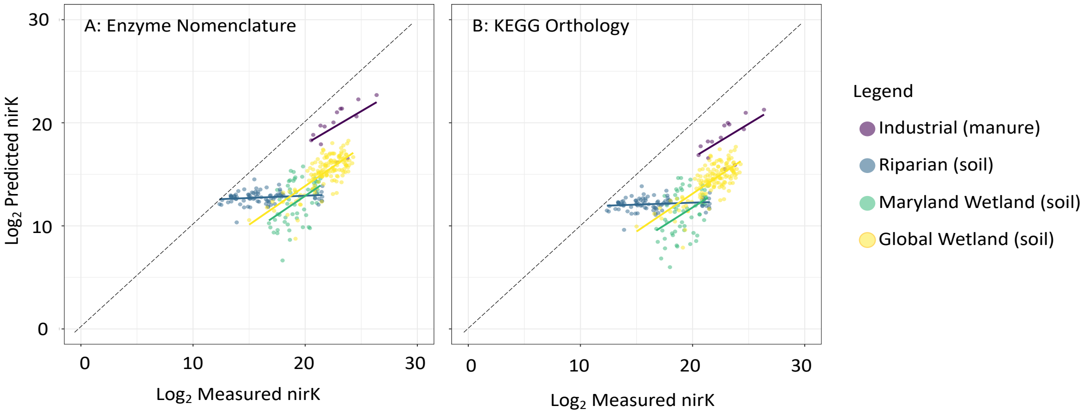

Figure 3.2: Scatterplot of nirK correlation between predicted gene quantity to measured gene quantity across datasets using A) Enzyme Class (EC) or B) Kegg Orthology (KO) values predicted from PICRUSt2. PICRUSt2 predictions were transformed to relative abundance, and scaled using 16S QPCR to estimate total abundance of each gene. Dashed line represents perfect prediction. Points falling above the dashed line represent overestimated gene abundance, whereas points falling below the dashed line indicate underestimation of gene abundance.

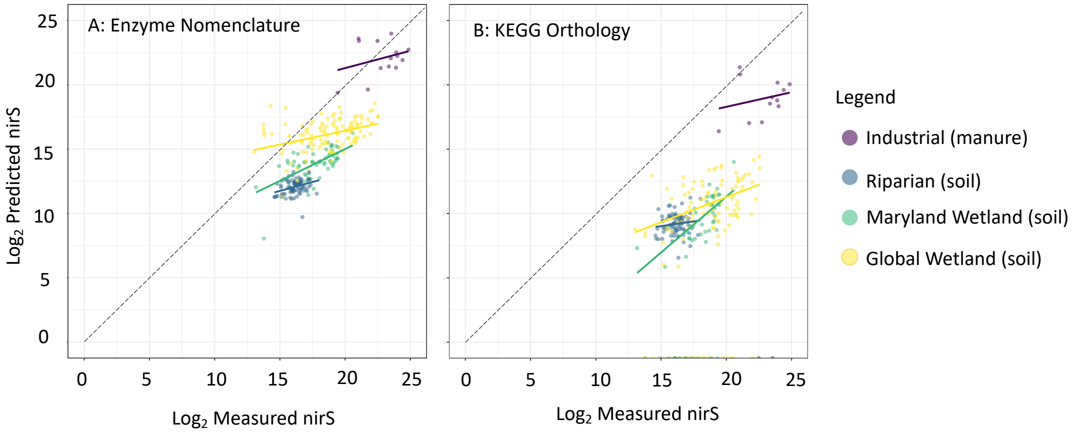

Figure 3.3: Scatterplot of nirS correlation between predicted gene quantity to measured gene quantity across datasets using A) Enzyme Class (EC) or B) Kegg Orthology (KO) values predicted from PICRUSt2. PICRUSt2 predictions were transformed to relative abundance, and scaled using 16S QPCR to estimate total abundance of each gene. Dashed line represents perfect prediction. Points falling above the dashed line represent overestimated gene abundance, whereas points falling below the dashed line indicate underestimation of gene abundance. Points falling on the x-axis indicate samples where zero gene counts were predicted from the 16S amplicon data.

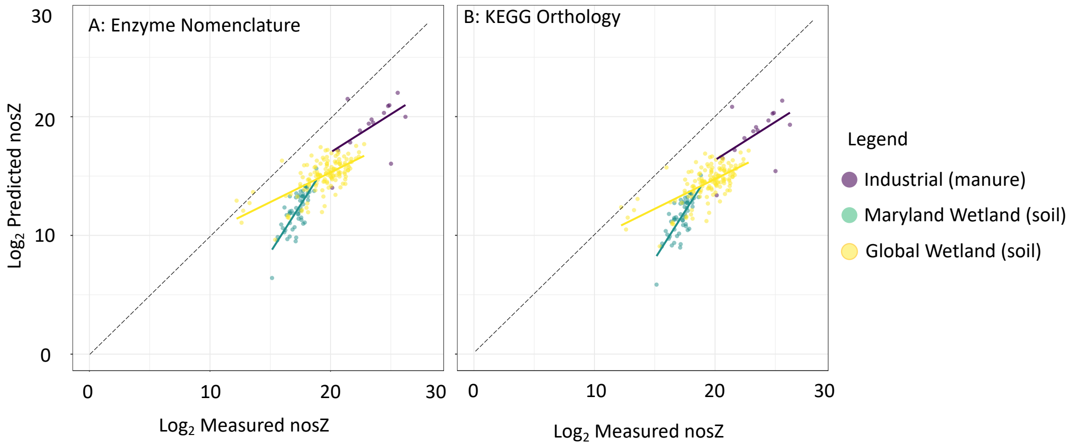

Figure 3.4: Scatterplot of nosZ correlation between predicted gene quantity to measured gene quantity across datasets using A) Enzyme Class (EC) or B) Kegg Orthology (KO) values predicted from PICRUSt2. PICRUSt2 predictions were transformed to relative abundance, and scaled using 16S QPCR to estimate total abundance of each gene. Dashed line represents perfect prediction. Points falling above the dashed line represent overestimated gene abundance, whereas points falling below the dashed line indicate underestimation of gene abundance.

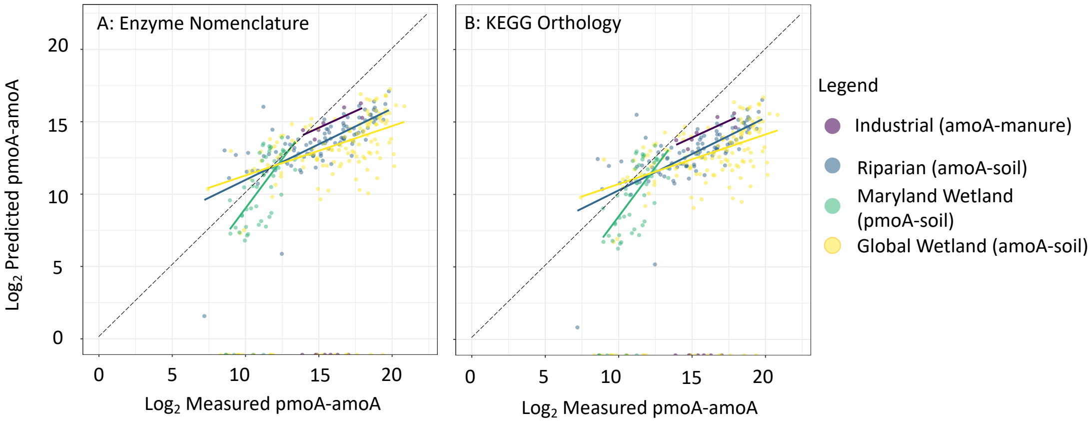

Figure 3.5: Scatterplot of pmoA-amoA correlation between predicted gene quantity to measured gene quantity across datasets using A) Enzyme Class (EC) or B) Kegg Orthology (KO) values predicted from PICRUSt2. PICRUSt2 predictions were transformed to relative abundance, and scaled using 16S QPCR to estimate total abundance of each gene. Dashed line represents perfect prediction. Points falling above the dashed line represent overestimated gene abundance, whereas points falling below the dashed line indicate underestimation of gene abundance. Points falling on the x-axis indicate samples where zero gene counts were predicted from the 16S amplicon data

The environments with the highest correlations differed between genes. For example, nirS had a high correlation in Maryland wetland soil (Pearson’s R = 0.82), but performed poorly in more upland and industrial (manure) datasets (Table 3.4). In contrast, nirK had a strong correlation in industrial manure samples (Pearson’s R = 0.84) and performed poorly in all other environments. The strongest correlation for nosZ was in Maryland wetland soils (Table 3.4; Pearson’s R = 0.74), and for amoA the urban soils and industrial manure settings were all highly correlated (Pearson’s R > 0.6). Although some of the studies used the same primers for each gene, most studies used distinct selection of primers to quantify gene abundance (Table 3.3).

Table 3.4: Correlation coefficients for PICRUSt2 / KO before and after correcting for taxon abundance.

Correction	DataSet	Scope	pmoA-amoA	nirK	nirS	nosZ	mcrA	nifH
Native Estimate	Maryland Wetland	local	0.3196	0.1676	0.6975	0.4692	0.3916	---
	Urban	global	0.7451	---	---	---	---	---
	Industrial manure	local	0.1226	0.3726	0.2049	0.1133	-0.423	---
	Maryland Riparian	local	---	-0.3169	0.0278	---	---	---
	Global Wetland	global	0.2876	-0.0248	0.1553	-0.002	---	0.1652
	Maryland Agriculture	local	---	---	---	---	---	-0.0364
Total Abundance Corrected Estimate	Maryland Wetland	local	0.3346	0.1095	0.8233	0.7415	0.7713	---
	Urban	global	0.7979	---	---	---	---	---
	Industrial manure	manure	0.6671	0.8480	-0.1080	0.3057	0.1202	---
	Maryland Riparian	local	---	0.0409	0.3234	---	---	---
	Global Wetland	global	0.4330	0.3692	0.5242	0.5632	---	0.5996
	Maryland Agriculture	local	---	---	---	---	---	0.0347

Table 3.5: Correlation coefficients for PICRUSt2 / EC before and after correcting for taxon abundance.

Correction	DataSet	Setting	pmoA-amoA	nirK	nirS	nosZ	mcrA	nifH
Native Estimate	**Maryland Wetland**	local	0.3240	0.2807	-0.3883	0.4695	0.3879	---
	Urban	global	0.7411	---	---	---	---	---
	Industrial manure	local	0.1239	0.3866	-0.1303	0.1239	-0.416	---
	Maryland Riparian	local	---	-0.3219	0.3157	---	---	---
	Global Wetland	global	0.2881	-0.0304	0.0183	-0.004	---	NA
	Maryland Agriculture	local	---	---	---	---	---	NA
Total Abundance Corrected Estimate	**Maryland Wetland**	local	0.3400	0.1543	0.7639	0.7363	0.7711	---
	Urban	global	0.7949	---	---	---	---	---
	Industrial manure	manure	0.6670	0.8706	-0.133	0.3060	-0.049	---
	Maryland Riparian	local	---	0.0549	0.5075	---	---	---
	Global Wetland	global	0.4353	0.3904	0.4288	0.5699	---	NA
	Maryland Agriculture	local	---	---	---	---	---	NA

Tax4Fun2 requires the user to select a reference dataset and a match percent to the reference. In general, low match percentages were more accurate than high match percentages (Figure 3.6 & 3.7). A notable exception was pmoA which was most accurately predicted at 100% match to 100NR. nosZ matches had the worst correlations at 99% match to either 99NR or 100NR reference sets; however, this was driven by a few outliers. As match % increased fewer sites contained nosZ sequences and therefore more samples were excluded. At 100% only a few remaining sites had predicted nosZ genes (Figure S3.2).

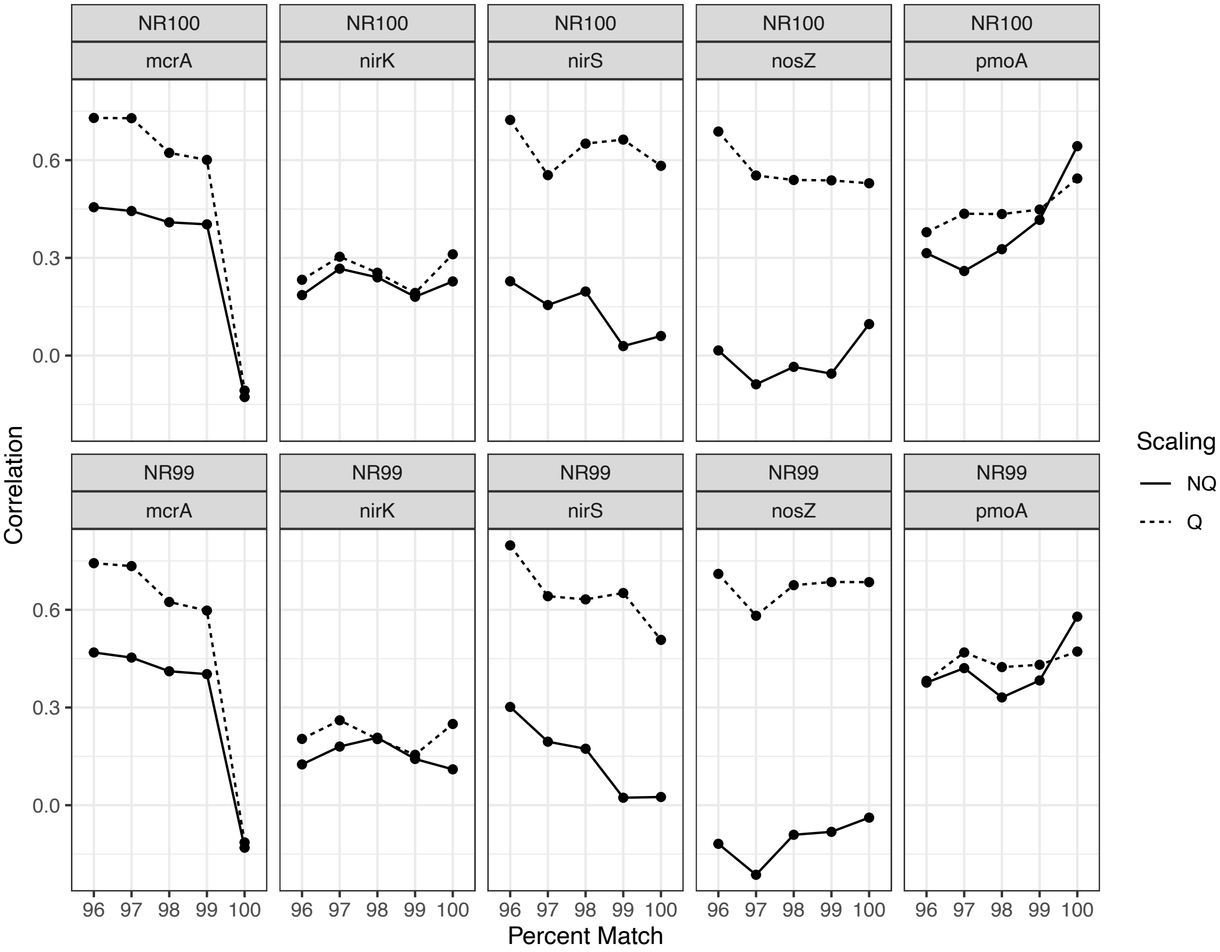

Figure 3.6: Tax4Fun2 correlation for local wetlands. Blue lines represent the correlation accuracy for Picrust2. Solid lines are the native output, while dashed lines represent abundance normalized output correlation with QPCR values. Tax4Fun2 allows a range of parameters to be chosen; 100NR and 99NR refers to the accuracy of the reference clustering (either clustered at 100% or 99% similarity); whereas the range on the X axis (96-100) represents the match similarity of the sample sequence to the reference. Thus, variation in the precision of the reference and the match to the reference are represented in this figure.

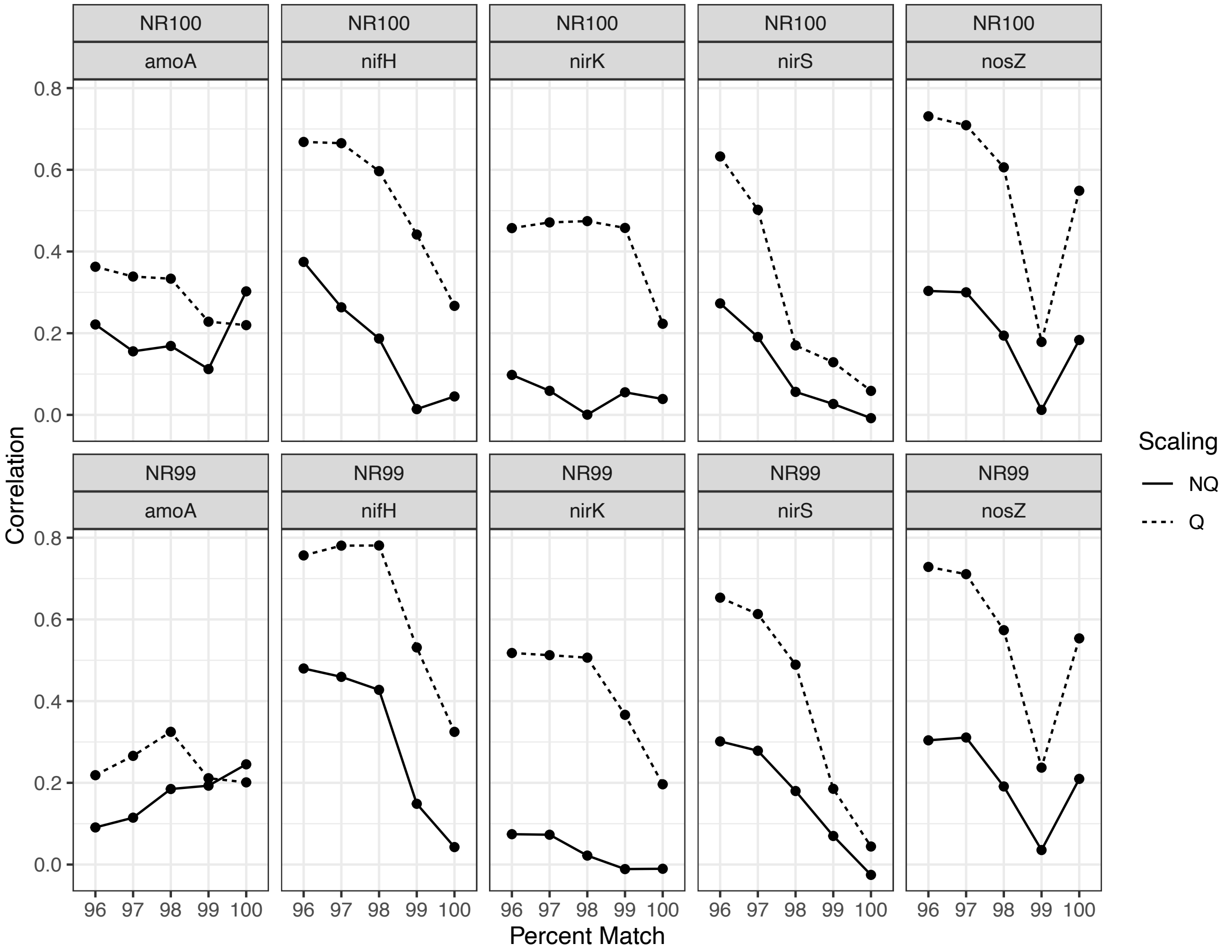

Figure 3.7: Tax4Fun2 correlations for global wetlands. Blue lines represent the correlation accuracy for Picrust2. Solid lines are the native output, while dashed lines represent abundance normalized output correlation with QPCR values. Tax4Fun2 allows a range of parameters to be chosen; 100NR and 99NR refers to the accuracy of the reference clustering (either clustered at 100% or 99% similarity); whereas the range on the X axis (96-100) represents the match similarity of the sample sequence to the reference. Thus, variation in the precision of the reference and the match to the reference are represented in this figure.

3.4 Discussion

Our hypothesis that total abundance correction would improve predictions made by PICRUSt2 was supported. The accuracy to predict gene abundance varied among genes and studies, but was almost always improved by using 16S rRNA QPCR to scale the total abundance, regardless of whether EC or KO values were used (Table 4; Chi-Square test P < 0.001, each). The second hypothesis that specific genes would be more reliably predicted was not supported. Rather, the accuracy of gene prediction depended on both the gene identity and the dataset. For example, the predicted nirS quantity was highly correlated to measured quantity in Maryland wetlands (Table 3.5, EC R = 0.82; Table 3.4, KO R = 0.76), but not in the industrial manure (Table 3.5, EC R = -0.13; Table 3.4 KO R = -0.10). In contrast predicted nirK quantity was better correlated to measured quantity in industrial manure (Table 3.5, EC R = 0.87; Table 3.4, KO R = 0.84), but not Maryland wetlands (Table 3.5, EC R = 0.15, Table 3.4 KO R = 0.11). These two studies have very different microbial communities; the environment in Maryland Wetlands consisted of soil sampled from temperate freshwater tidal wetlands and for four distinct wetland plant communities, whereas the industrial manure samples were dairy manure aged and dried. The fact that both datasets produced high correlations among predicted and measured gene quantities for at least one gene suggests that differences in the correlation of specific genes among the datasets is not related to laboratory technique. The differences between studies were likely related to how well the distribution of nir genes is understood for the sampled communities. For example, nirS genes in sequenced microbial genomes tend to be isolated from soil and extreme environments, while nirK genes tend to be found in microbial genomes isolated from plant and animal hosts (Graf et al., 2014). These two studies also used different primers to quantify nirS and nirK genes (Table 3.2), so the pattern could equally reflect the degree to which the primers accurately capture the abundance of the gene. This highlights the need within molecular ecology research for continued work building open-source databases of environmental genetic data, and developing tools such as marker gene and primer selection repositories.

Although there were strong correlations between predicted and measured gene abundance for most genes in at least one study included in this synthesis, in almost all cases PICRUSt2 underestimated gene abundance by several orders of magnitude (Figures 3.2-3.4). There is clear evidence for some genes that reference genomes do not sufficiently represent the diversity of samples in the study. For example, PICRUSt2 erroneously predicted zero nirS gene abundance in several samples from 3 out of 4 of the studies that measured nirS abundance. The most accurate nirS prediction was in wetland soils; whereas nirK was most accurately measured in manure (Tables 3.4 & 3.5). This makes sense given that previous research found nirK more often in animal associated denitrify bacteria (Graf et al., 2014). Both nirS and nirK gene abundance were poorly predicted in agricultural and urban soil environments.

Gene copy numbers in reference genomes do not explain underestimation of predicted gene abundance. Genomes generally contain only one copy of nosZ (Palmer et al., 2009). Some cyanobacteria (Thiel and Pratte, 2014) and rhizobium (Morett et al., 1988) have been shown to contain 3 nifH copies per cell; but recently the discovery of high copy numbers of pseudo-nifH genes has led to the suggestion that other *nif* genes should be used instead to quantify N fixation capacity in bacteria (Mise et al., 2021). Ammonia oxidizing bacteria have 2-3 copies of amoA per genome (Norton et al., 2002). Ammonia oxidizing archaea in marine environments are estimated to have anywhere from three (Christman et al., 2011) to eight (Galand et al., 2009) copies of amoA. Aerobic methane oxidizers are estimated to have two copies of pmoA (Kolb et al., 2003). If PICRUSt2 were underestimating gene abundance because it predicts too few gene copies per genome, then we'd expect to see more underestimation when there are higher gene copy numbers. In this case, we'd expect to see the greatest underestimation in amoA, and the highest accuracy in nosZ. Rather the pattern is the opposite, suggesting that either underestimating gene copy number is not the cause of this systemic estimation error, or the literature values do not represent the actual gene copy number in environmental microbes.

Previous papers have posited that functional niche adaptations are sometimes decoupled from taxonomy (Louca et al., 2016b, 2016a; Philippot et al., 2010). These findings are supported by a recent study of 19 soil bacterial genomes that found evidence of recombination in as much as 35% of gene families (Crits-Christoph et al., 2020). This finding is further corroborated by other research finding strain-specific adaptation by bacteria to soil conditions (Chase et al., 2021). In general, genomic content in bacteria is dominated by locally adapted genes (Coelho et al., 2022), especially in soils (Mantri et al., 2021). There is plentiful evidence that nifH and denitrification genes often horizontally gene transfer (Alvarez et al., 2011;

Bolhuis et al., 2010; Gaby and Buckley, 2014; Jang et al., 2019). For example, inconsistencies between nir gene phylogenies and 16S rRNA phylogeny prompted early researchers to suggest wide-spread horizontal gene transfer (Delorme et al., 2003; Heylen et al., 2006; Horn et al., 2006). Recently, the nirK gene has been found on plasmids in *Azospirillum brasilense* multiple times (Petrova et al., 2010; Pothier et al., 2008); and has been found alongside a nir-nor gene cluster within a transposon structure, indicating that the genes were likely acquired through horizontal gene transfer (Jang et al., 2019). And nirS genes have been shown to be horizontally transferred among *Thermus thermophilus* (Alvarez et al., 2011). These findings lend support to the notion that horizontal gene transfer of nirK and nirS may be occurring in at least some lineages, even though other evolutionary mechanisms such as gene duplication and lineage sorting may also explain some of the incongruencies among 16S rRNA and denitrification gene phylogenies (Jones et al., 2008).

Although horizontal gene transfer is a widespread mechanism for adaptation in soil bacteria, current evidence suggests that certain gene families are less likely to be transferred. For example, a high degree of correlation between amoA and 16S rRNA phylogeny indicates that horizontal gene transfer is rare (Alves et al., 2018; Pester et al., 2012; Wang et al., 2016). Thus, gene families are very likely to differ in the degree to which reference genomes provide sufficient evidence to predict the presence of family members. It is also possible that differences in the ecology of gene homologues also result in different ecological behavior reflective in their phylogeny. For example, the production of nitrous oxide by bacteria may be used during nodule development to induce lateral root formation as part of the nodulation process (Pii et al., 2007; Signorelli et al., 2020). A large proportion of plant associated taxa harbor only a nirK gene, which produces nitrous oxide, with no other genes for denitrification (Graf et al., 2014) indicating nirK likely has an important role for host-bacteria signaling. Thus, multiple ecological functions may drive different evolutionary mechanisms and divergent phylogenetic signals among nir lineages; it may be necessary to distinguish these functions to make more accurate prediction of nir gene abundance in the environment.

Taxa detected in datasets are often poorly matched to reference databases. This can be clearly seen by using Tax4Fun2 to vary the match percent of sample to reference dataset (Figure 3.6 & 3.7). Tax4Fun2 predicts genes on an OTU basis matching OTUs at a defined match threshold to reference genomes. Almost all genes in all datasets achieve the best match when the match threshold was low (96%). The pmoA and amoA genes were an exception, but this may be because two distinct genes and functions are confounded in the pmoA-amoA KO number (Table 3.2). In this case, higher percent identity match (100%) leads to more accurate predictions for pmoA but not amoA (pmoA: Figure 3.6; amoA: Figure 3.7). A lower threshold allows a larger portion of the microbial community to be annotated in each sample. This is important because there are very often relatively few taxa that are reliably assigned taxonomy at the Order level in each sample (Supplemental Figure 3.1), making accurate functional predictions for the whole community challenging. This is a consistent challenge across prediction software platforms. For example, the metagenome-inference software Piphillin (Iwai et al., 2016; Narayan et al., 2020) produced the highest correlation between predicted metagenome and the reference metagenome across several animal microbiome datasets when the sample sequence to reference sequence identity match threshold was between 0.75 and 0.99; and all datasets exhibited a pattern where matching above a certain threshold (ranging from 0.96-0.99) resulted in a decrease in correlation to the reference metagenome (Narayan et al., 2020). Very similar microbiomes may vary substantially in the accuracy of metagenome prediction (Narayan et al., 2020). Most likely the high diversity of unknown and undescribed organisms and horizontal gene transfer events both contribute at some level to the systemic underestimation bias in PICRUSt2 data.

Finally, QPCR often provides a narrower context for quantifying gene abundance than KO numbers. Ammonia monooxygenase subunit A (amoA) is used to measure the abundance of ammonia oxidizers, but separate primers are used to detect archaeal (i.e. AOA: Tourna et al., 2008) versus bacterial (i.e. AOB: Rotthauwe et al., 1997) amoA genes. Although the particulate methane monooxygenase subunit A (pmoA) gene is used to measure methane oxidizers (Im et al., 2011), both pmoA and amoA are designated by the same Kegg Ontology reference number (Table 3.2). They are, however, distinguished by EC number (Table 3.2). This ambiguity means that whether amoA/pmoA quantity is accurately predicted may depend significantly on both the primers used for quantification and which group happens to be dominant in the sample. This is also an issue for nirS and nirK, which share EC numbers (Table 3.2). Annotation using EC numbers in almost all cases improved the prediction accuracy (Tables 3.4 & 3.5); one exception was for nifH which does not have an EC number and so using EC nomenclature would erroneously predict no nifH genes. Therefore, EC nomenclature will not always be a better choice than KO number for predicting gene abundance. And while EC nomenclature tended to improve the accuracy of gene quantity prediction across datasets, the correlations within datasets were comparable to KO numbers (Tables 3.4 & 3.5).

3.5 Conclusion

Using quantitative scaling methods reliably improves the accuracy of functional gene prediction using the PICRUSt2 software. And, under certain conditions, gene quantity estimates were highly correlated to measured gene abundance. However, it is not yet feasible to both accurately and reliably predict the quantity of functional genes in soils from 16S amplicon sequence data. Predicted gene abundance is consistently low compared to measured gene abundance. And while certain genes are more accurately estimated under certain conditions, for all genes the majority of predicted gene quantities were poorly correlated to measured gene quantities in the majority of environments. Gene quantity estimation will likely be improved as more reference genomes are sequenced and annotated. But because horizontal gene transfer is an important mode of adaptation in bacteria, it is likely that for certain genes quantity estimation may be significantly improved by including model parameterization for the environmental context of the sample.

3.6 Literature Cited:

Alvarez, L., Bricio, C., José Gómez, M., Berenguer, J., 2011. Lateral Transfer of the Denitrification Pathway Genes among Thermus thermophilus Strains. Appl Environ Microbiol 77, 1352–1358. https://doi.org/10.1128/AEM.02048-10

Alves, R.J.E., Minh, B.Q., Urich, T., von Haeseler, A., Schleper, C., 2018. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. Nat Commun 9, 1517. https://doi.org/10.1038/s41467-018-03861-1

Angel, R., Matthies, D., Conrad, R., 2011. Activation of Methanogenesis in Arid Biological Soil Crusts Despite the Presence of Oxygen. PLOS ONE 6, e20453. https://doi.org/10.1371/journal.pone.0020453

Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., Pawlowski, J., 2017. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Molecular Ecology Resources 17, 1231–1242. https://doi.org/10.1111/1755-0998.12668

Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31, 2882–2884. https://doi.org/10.1093/bioinformatics/btv287

Bahram, M., Espenberg, M., Pärn, J., Lehtovirta-Morley, L., Anslan, S., Kasak, K., Kõljalg, U., Liira, J., Maddison, M., Moora, M., Niinemets, Ü., Öpik, M., Pärtel, M., Soosaar, K., Zobel, M., Hildebrand, F., Tedersoo, L., Mander, Ü., 2022. Structure and function of the soil microbiome underlying N2O emissions from global wetlands. Nat Commun 13, 1430. https://doi.org/10.1038/s41467-022-29161-3

Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. Nature 515, 505–511. https://doi.org/10.1038/nature13855

Barlow, J.T., Bogatyrev, S.R., Ismagilov, R.F., 2020. A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. Nature Communications 11, 2590. https://doi.org/10.1038/s41467-020-16224-6

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., Gregory Caporaso, J., 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6, 90. https://doi.org/10.1186/s40168-018-0470-z

Bolhuis, H., Severin, I., Confurius-Guns, V., Wollenzien, U.I.A., Stal, L.J., 2010. Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium Microcoleus chthonoplastes. ISME J 4, 121–130. https://doi.org/10.1038/ismej.2009.99

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., Behrenfeld, M.J., Boetius, A., Boyd, P.W., Classen, A.T., Crowther, T.W., Danovaro, R., Foreman, C.M., Huisman, J., Hutchins, D.A., Jansson, J.K., Karl, D.M., Koskella, B., Mark Welch, D.B., Martiny, J.B.H., Moran, M.A., Orphan, V.J., Reay, D.S., Remais, J.V., Rich, V.I., Singh, B.K., Stein, L.Y., Stewart, F.J., Sullivan, M.B., van Oppen, M.J.H., Weaver, S.C., Webb, E.A., Webster, N.S., 2019. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol 17, 569–586. https://doi.org/10.1038/s41579-019-0222-5

Chase, A.B., Weihe, C., Martiny, J.B.H., 2021. Adaptive differentiation and rapid evolution of a soil bacterium along a climate gradient. PNAS 118. https://doi.org/10.1073/pnas.2101254118

Christman, G.D., Cottrell, M.T., Popp, B.N., Gier, E., Kirchman, D.L., 2011. Abundance, Diversity, and Activity of Ammonia-Oxidizing Prokaryotes in the Coastal Arctic Ocean in Summer and Winter. Applied and Environmental Microbiology 77, 2026–2034. https://doi.org/10.1128/AEM.01907-10

Coelho, L.P., Alves, R., del Río, Á.R., Myers, P.N., Cantalapiedra, C.P., Giner-Lamia, J., Schmidt, T.S., Mende, D.R., Orakov, A., Letunic, I., Hildebrand, F., Van Rossum, T., Forslund, S.K., Khedkar, S., Maistrenko, O.M., Pan, S., Jia, L., Ferretti, P., Sunagawa, S., Zhao, X.-M., Nielsen, H.B., Huerta-Cepas, J., Bork, P., 2022. Towards the biogeography of prokaryotic genes. Nature 601, 252–256. https://doi.org/10.1038/s41586-021-04233-4

Cordier, T., Forster, D., Dufresne, Y., Martins, C.I.M., Stoeck, T., Pawlowski, J., 2018. Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. Molecular Ecology Resources 18, 1381–1391. https://doi.org/10.1111/1755-0998.12926

Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., Pawlowski, J., 2019. Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring. Trends in Microbiology 27, 387–397. https://doi.org/10.1016/j.tim.2018.10.012

Crits-Christoph, A., Olm, M.R., Diamond, S., Bouma-Gregson, K., Banfield, J.F., 2020. Soil bacterial populations are shaped by recombination and gene-specific selection across a grassland meadow. ISME J 14, 1834–1846. https://doi.org/10.1038/s41396-020-0655-x

Delorme, S., Philippot, L., Edel-Hermann, V., Deulvot, C., Mougel, C., Lemanceau, P., 2003. Comparative Genetic Diversity of the narG, nosZ, and 16S rRNA Genes in Fluorescent Pseudomonads. Applied and Environmental Microbiology 69, 1004–1012. https://doi.org/10.1128/AEM.69.2.1004-1012.2003

Djemiel, C., Maron, P.-A., Terrat, S., Dequiedt, S., Cottin, A., Ranjard, L., 2022. Inferring microbiota functions from taxonomic genes: a review. GigaScience 11, giab090. https://doi.org/10.1093/gigascience/giab090

Douglas, G.M., Maffei, V.J., Zaneveld, J.R., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., Langille, M.G.I., 2020. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol 38, 685–688. https://doi.org/10.1038/s41587-020-0548-6

Epp Schmidt, D., Dlott, G., Cavigelli, M., Yarwood, S., Maul, J.E., 2022. Soil microbiomes in three farming systems more affected by depth than farming system. Applied Soil Ecology 173, 104396. https://doi.org/10.1016/j.apsoil.2022.104396

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

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. PNAS 103, 626–631. https://doi.org/10.1073/pnas.0507535103

Fierer, N., Wood, S.A., Bueno de Mesquita, C.P., 2021. How microbes can, and cannot, be used to assess soil health. Soil Biology and Biochemistry 153, 108111. https://doi.org/10.1016/j.soilbio.2020.108111

Gaby, J.C., Buckley, D.H., 2014. A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. Database (Oxford) 2014. https://doi.org/10.1093/database/bau001

Galand, P.E., Lovejoy, C., Hamilton, A.K., Ingram, R.G., Pedneault, E., Carmack, E.C., 2009. Archaeal diversity and a gene for ammonia oxidation are coupled to oceanic circulation. Environmental Microbiology 11, 971–980. https://doi.org/10.1111/j.1462-2920.2008.01822.x

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 0. https://doi.org/10.3389/fmicb.2017.02224

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., Egozcue, J.J., 2016. It's all relative: analyzing microbiome data as compositions. Ann Epidemiol 26, 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003

Graf, D.R.H., Jones, C.M., Hallin, S., 2014. Intergenomic Comparisons Highlight Modularity of the Denitrification Pathway and Underpin the Importance of Community Structure for N2O Emissions. PLOS ONE 9, e114118. https://doi.org/10.1371/journal.pone.0114118

Henry, S., Baudoin, E., López-Gutiérrez, J.C., Martin-Laurent, F., Brauman, A., Philippot, L., 2004. Quantification of denitrifying bacteria in soils by nirK gene targeted real-time PCR. Journal of Microbiological Methods 59, 327–335. https://doi.org/10.1016/j.mimet.2004.07.002

Henry, S., Bru, D., Stres, B., Hallet, S., Philippot, L., 2006. Quantitative detection of the nosZ gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, narG, nirK, and nosZ genes in soils. Appl Environ Microbiol 72, 5181–5189. https://doi.org/10.1128/AEM.00231-06

Heylen, K., Gevers, D., Vanparys, B., Wittebolle, L., Geets, J., Boon, N., De Vos, P., 2006. The incidence of nirS and nirK and their genetic heterogeneity in cultivated denitrifiers. Environmental Microbiology 8, 2012–2021. https://doi.org/10.1111/j.1462-2920.2006.01081.x

Horn, M.A., Drake, H.L., Schramm, A., 2006. Nitrous Oxide Reductase Genes (nosZ) of Denitrifying Microbial Populations in Soil and the Earthworm Gut Are Phylogenetically Similar. Appl. Environ. Microbiol. 72, 1019–1026. https://doi.org/10.1128/AEM.72.2.1019-1026.2006

Im, J., Lee, S.-W., Bodrossy, L., Barcelona, M.J., Semrau, J.D., 2011. Field application of nitrogen and phenylacetylene to mitigate greenhouse gas emissions from landfill cover soils: effects on microbial community structure. Appl Microbiol Biotechnol 89, 189–200. https://doi.org/10.1007/s00253-010-2811-0

Iwai, S., Weinmaier, T., Schmidt, B.L., Albertson, D.G., Poloso, N.J., Dabbagh, K., DeSantis, T.Z., 2016. Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. PLOS ONE 11, e0166104. https://doi.org/10.1371/journal.pone.0166104

Jang, J., Sakai, Y., Senoo, K., Ishii, S., 2019. Potentially Mobile Denitrification Genes Identified in Azospirillum sp. Strain TSH58. Applied and Environmental Microbiology 85, e02474-18. https://doi.org/10.1128/AEM.02474-18

Jian, C., Luukkonen, P., Yki-Jarvinen, H., Salonen, A., Korpela, K., 2018. Quantitative PCR provides a simple and accessible method for quantitative microbiome profiling. bioRxiv 478685. https://doi.org/10.1101/478685

Jones, C.M., Graf, D.R., Bru, D., Philippot, L., Hallin, S., 2013. The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink. ISME J 7, 417–426. https://doi.org/10.1038/ismej.2012.125

Jones, C.M., Stres, B., Rosenquist, M., Hallin, S., 2008. Phylogenetic Analysis of Nitrite, Nitric Oxide, and Nitrous Oxide Respiratory Enzymes Reveal a Complex Evolutionary History for Denitrification. Molecular Biology and Evolution 25, 1955–1966. https://doi.org/10.1093/molbev/msn146

Jun, S.-R., Robeson, M.S., Hauser, L.J., Schadt, C.W., Gorin, A.A., 2015. PanFP: pangenome-based functional profiles for microbial communities. BMC Research Notes 8, 479. https://doi.org/10.1186/s13104-015-1462-8

Kandeler, E., Deiglmayr, K., Tscherko, D., Bru, D., Philippot, L., 2006. Abundance of narG, nirS, nirK, and nosZ Genes of Denitrifying Bacteria during Primary Successions of a Glacier Foreland. Applied and Environmental Microbiology 72, 5957–5962. https://doi.org/10.1128/AEM.00439-06

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016a. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44, D457-462. https://doi.org/10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., Morishima, K., 2016b. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428, 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

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

Louca, S., Jacques, S.M.S., Pires, A.P.F., Leal, J.S., Srivastava, D.S., Parfrey, L.W., Farjalla, V.F., Doebeli, M., 2016a. High taxonomic variability despite stable functional structure across microbial communities. Nat Ecol Evol 1, 1–12. https://doi.org/10.1038/s41559-016-0015

Louca, S., Parfrey, L.W., Doebeli, M., 2016b. Decoupling function and taxonomy in the global ocean microbiome. Science 353, 1272–1277. https://doi.org/10.1126/science.aaf4507

Mantri, S.S., Negri, T., Sales-Ortells, H., Angelov, A., Peter, S., Neidhardt, H., Oelmann, Y., Ziemert, N., 2021. Metagenomic Sequencing of Multiple Soil Horizons and Sites in Close Vicinity Revealed Novel Secondary Metabolite Diversity. mSystems. https://doi.org/10.1128/mSystems.01018-21

Meinhardt, K.A., Bertagnolli, A., Pannu, M.W., Strand, S.E., Brown, S.L., Stahl, D.A., 2015. Evaluation of revised polymerase chain reaction primers for more inclusive quantification of ammonia-oxidizing archaea and bacteria. Environmental Microbiology Reports 7, 354–363. https://doi.org/10.1111/1758-2229.12259

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., DeLong, E.F., 2007. Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. Environmental Microbiology 9, 1162–1175. https://doi.org/10.1111/j.1462-2920.2007.01239.x

Mise, K., Masuda, Y., Senoo, K., Itoh, H., 2021. Undervalued Pseudo-nifH Sequences in Public Databases Distort Metagenomic Insights into Biological Nitrogen Fixers. mSphere 6, e00785-21. https://doi.org/10.1128/msphere.00785-21

Morett, E., Moreno, S., Espin, G., 1988. Transcription analysis of the three nifH genes of Rhizobium phaseoli with gene fusions. Mol Gen Genet 213, 499–504. https://doi.org/10.1007/BF00339622

Narayan, N.R., Weinmaier, T., Laserna-Mendieta, E.J., Claesson, M.J., Shanahan, F., Dabbagh, K., Iwai, S., DeSantis, T.Z., 2020. Piphillin predicts metagenomic composition and dynamics from DADA2-corrected 16S rDNA sequences. BMC Genomics 21, 56. https://doi.org/10.1186/s12864-019-6427-1

Norton, J.M., Alzerreca, J.J., Suwa, Y., Klotz, M.G., 2002. Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. Arch Microbiol 177, 139–149. https://doi.org/10.1007/s00203-001-0369-z

Palmer, K., Drake, H.L., Horn, M.A., 2009. Genome-Derived Criteria for Assigning Environmental narG and nosZ Sequences to Operational Taxonomic Units of Nitrate Reducers. Applied and Environmental Microbiology 75, 5170–5174. https://doi.org/10.1128/AEM.00254-09

Pereyra, L.P., Hiibel, S.R., Prieto Riquelme, M.V., Reardon, K.F., Pruden, A., 2010. Detection and quantification of functional genes of cellulose-degrading, fermentative, and sulfate-reducing bacteria and methanogenic archaea. Appl Environ Microbiol 76, 2192–2202. https://doi.org/10.1128/AEM.01285-09

Pester, M., Rattei, T., Flechl, S., Gröngröft, A., Richter, A., Overmann, J., Reinhold-Hurek, B., Loy, A., Wagner, M., 2012. amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions. Environmental Microbiology 14, 525–539. https://doi.org/10.1111/j.1462-2920.2011.02666.x

Petrova, L.P., Varshalomidze, O.E., Shelud'ko, A.V., Katsy, E.I., 2010. Localization of denitrification genes in plasmid DNA of bacteria

Azospirillum brasilense. Russ J Genet 46, 801–807. https://doi.org/10.1134/S1022795410070045

Philippot, L., Andersson, S.G.E., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B., Hallin, S., 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Micro 8, 523–529. https://doi.org/10.1038/nrmicro2367

Pii, Y., Crimi, M., Cremonese, G., Spena, A., Pandolfini, T., 2007. Auxin and nitric oxide control indeterminate nodule formation. BMC Plant Biology 7, 21. https://doi.org/10.1186/1471-2229-7-21

Porter, T.M., Hajibabaei, M., 2018. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Molecular Ecology 27, 313–338. https://doi.org/10.1111/mec.14478

Pothier, J.F., Prigent-Combaret, C., Haurat, J., Moënne-Loccoz, Y., Wisniewski-Dyé, F., 2008. Duplication of Plasmid-Borne Nitrite Reductase Gene nirK in the Wheat-Associated Plant Growth–Promoting Rhizobacterium Azospirillum brasilense Sp245. MPMI 21, 831–842. https://doi.org/10.1094/MPMI-21-6-0831

Quinn, T.P., Erb, I., Gloor, G., Notredame, C., Richardson, M.F., Crowley, T.M., 2019. A field guide for the compositional analysis of any-omics data. Gigascience 8, giz107. https://doi.org/10.1093/gigascience/giz107

Rösch, C., Mergel, A., Bothe, H., 2002. Biodiversity of denitrifying and dinitrogen-fixing bacteria in an acid forest soil. Appl Environ Microbiol 68, 3818–3829. https://doi.org/10.1128/AEM.68.8.3818-3829.2002

Rotthauwe, J.H., Witzel, K.P., Liesack, W., 1997. The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. Appl. Environ. Microbiol. 63, 4704–4712.

Sansupa, C., Wahdan, S.F.M., Hossen, S., Disayathanoowat, T., Wubet, T., Purahong, W., 2021. Can we use functional annotation of prokaryotic taxa (FAPROTAX) to assign the ecological functions of soil bacteria? Applied Sciences 11, 688. https://doi.org/10.3390/app11020688

Signorelli, S., Sainz, M., Tabares-da Rosa, S., Monza, J., 2020. The Role of Nitric Oxide in Nitrogen Fixation by Legumes. Frontiers in Plant Science 11.

Suman, J., Rakshit, A., Ogireddy, S.D., Singh, S., Gupta, C., Chandrakala, J., 2022. Microbiome as a Key Player in Sustainable Agriculture and Human Health. Frontiers in Soil Science 2.

Thiel, T., Pratte, B.S., 2014. Regulation of Three Nitrogenase Gene Clusters in the Cyanobacterium Anabaena variabilis ATCC 29413. Life 4, 944–967. https://doi.org/10.3390/life4040944

Throbäck, I.N., Enwall, K., Jarvis, A., Hallin, S., 2004. Reassessing PCR primers targeting nirS, nirK and nosZ genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiol Ecol 49, 401–417. https://doi.org/10.1016/j.femsec.2004.04.011

Tourna, M., Freitag, T.E., Nicol, G.W., Prosser, J.I., 2008. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. Environmental Microbiology 10, 1357–1364. https://doi.org/10.1111/j.1462-2920.2007.01563.x

Wang, H., Bagnoud, A., Ponce-Toledo, R.I., Kerou, M., Weil, M., Schleper, C., Urich, T., 2016. Linking 16S rRNA Gene Classification to amoA Gene Taxonomy Reveals Environmental Distribution of Ammonia-Oxidizing Archaeal Clades in Peatland Soils. mSystems 6, e00546-21. https://doi.org/10.1128/mSystems.00546-21

Wei, W., Isobe, K., Nishizawa, T., Zhu, L., Shiratori, Y., Ohte, N., Koba, K., Otsuka, S., Senoo, K., 2015. Higher diversity and abundance of denitrifying microorganisms in environments than considered previously. ISME J 9, 1954–1965. https://doi.org/10.1038/ismej.2015.9

Wemheuer, F., Taylor, J.A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T., Wemheuer, B., 2020. Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. Environmental Microbiome 15, 11. https://doi.org/10.1186/s40793-020-00358-7

Wemheuer, F., Taylor, J.A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T., Wemheuer, B., 2018. Tax4Fun2: a R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene marker gene sequences (preprint). Bioinformatics. https://doi.org/10.1101/490037

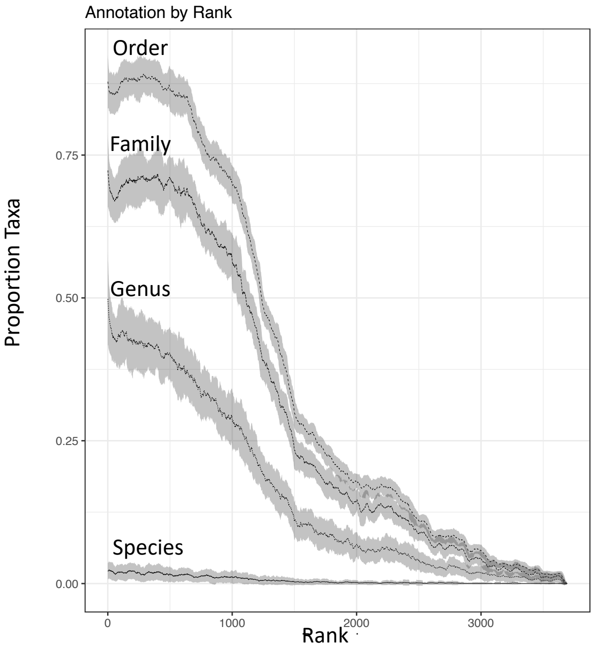

Figure S3.1: Proportion of taxa from GLU (Epp Schmidt et al 2017, 2019) that are annotated to a given level of taxonomy (Species – Order) as a function of rank abundance. Less than 50 percent of the most abundant taxa have genus or species annotations.

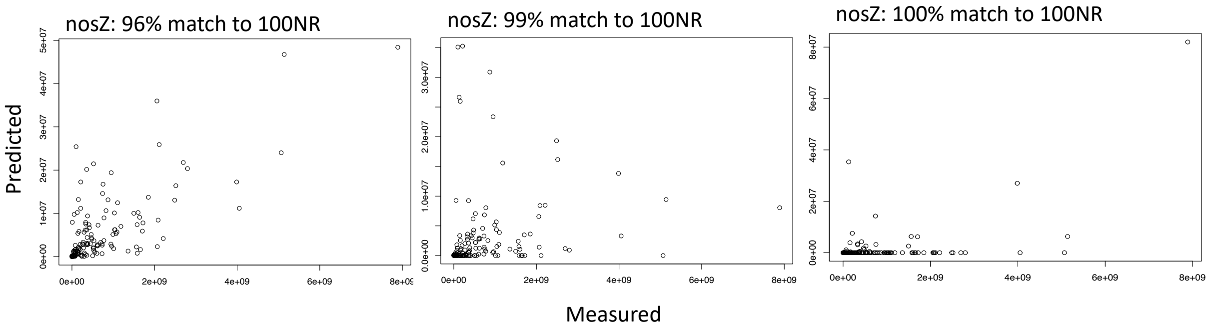

Figure S3.2: Tax4Fun2 plots of predicted nosZ abundance versus measured from Global Wetlands (Bahram et al., 2022) using different match percent. Demonstrates lower rate of nosZ prediction as match threshold becomes more stringent.

Published in Applied Soil Ecology

Citation:
Epp Schmidt, D., Dlott, G., Cavigelli, M., Yarwood, S., Maul, J.E. (2022). Soil microbiomes in three farming systems more affected by depth than farming system. Applied Soil Ecology 173. DOI: 10.1016/j.apsoil.2022.104396

Chapter 4: Soil microbiomes in three farming systems more affected by depth than farming system

Authors/Affiliations : Dietrich Epp Schmidt[a], Glade Dlott[b], Michel Cavigelli[c], Stephanie Yarwood[a], Jude E. Maul[c].

[a] Department of Environmental Science and Technology, University of Maryland, College Park, MD 20742, USA
[b] Kellogg Biological Station, Michigan State University, Hickory Corners, MI 49060, USA
[c] Sustainable Agricultural Systems Laboratory, USDA Agricultural Research Service, Beltsville, MD 20705, USA

Abstract

Agriculture, by its intentional design, manipulates the ecological functions of soils. It does so by altering carbon and nutrient inputs; by controlling the plant community; and in the case of tillage, by physically disrupting and redistributing the soil within the soil profile. While there are many studies that contrast soil microbiomes across farming systems, few studies have examined the effect of farming system on the vertical organization of taxa in the soil profile. We hypothesized that large effects of farming systems on edaphic factors would lead to large impacts on the microbial community that would reflect the underlying life history strategy of the microbes. For example, that tillage would increase the proportion of unicellular fungi. Our study compared farming systems that had been in place for 13 years in the mid-Atlantic region of the United States. Each system is a 3-year rotation of corn (*Zea mays*), soybean (*Glycine max*), and wheat (*Triticum aestivum*), managed with either conventional no-till, conventional chisel-till, or using organic methods. We determined the relative effect of long-term farming system management on edaphic factors and the soil microbial community with depth structured sampling (0-5, 5-10, 10-Ap, Ap-30 cm). We found relatively small effects of farming system but substantial effects of depth on edaphic factors and microbial communities. For example, differences in management resulted in subtle differences in %C above the Ap horizon. Fungal gene abundance increased in the organic system relative to the no-till system, although neither fungal nor bacterial richness differed across farming systems. Farming system effects on microbial community composition were greatest in the top 10 cm but did not affect the abundance of unicellular fungi. Several groups including *Pseudomonas* and Mortierellomycota appear to be sensitive to redistribution by tillage in the organic system, but it is not clear to what extent this effect is due to legacy DNA from tillage.

4.1 Introduction

Soil microbial communities are critical to soil function and contribute to healthy soils. As interest in assessing and measuring soil health has increased, so has a debate about whether soil microbial community metrics can be and should be included in soil health tests (Bach et al., 2020; Fierer et al., 2021). As noted in the recent NRCS technical note on soil health, current tests measure a variety of chemical and physical parameters and typically do so on a bulked sample of the upper 10 to 30 cm of soil (Stott, 2019). For microbial measures, the NRCS technical note suggests determining microbial composition based on fatty acid profiles and archiving soils for future analysis. Stott (2019) notes that currently there is not a standard method for measuring the microbial community and there is a lack of data on farming system effects on microbial community metrics. Addressing these limitations requires studies that assess the microbial quantity and composition in farming systems that control for soil type, tillage, crop species, and ideally have been managed for several years to allow for potential changes in the microbial community.

Conversion from tilled to no-till systems have been observed to change some microbial community metrics. For example, microbial biomass and enzyme activity were consistently higher in no-till compared to tilled systems across several studies (Zuber and Villamil, 2016). The effect of tillage on microbial community composition is mixed, however. For example, a recent meta-analysis compared 60 studies and observed a general increase in both bacterial and fungal biomass under no-till, but not specific trends in fungal to bacterial ratios (Chen et al., 2020). The relative effect size varied based on soil texture, with an increase in fungal biomass under no-till management in clayey soils. This underscores the importance of evaluating farming practices on the same soil type and conducting similar experiments across different soil types and climates. Increases in the fungal to bacterial ratios have sometimes been pointed to as an important soil health metric, likely due to the assumption that increased fungal biomass is driven by an increase in arbuscular mycorrhizal fungi (AMF). The amount of saprophytic fungi can also change however, masking AMF differences (Mbuthia et al., 2015). No-till has been shown to increase AMF phospholipid fatty acid (PLFA) biomarkers (Helgason et al., 2010; Mbuthia et al., 2015).

Outside of comparisons between no-till and tilled systems there is a broad spectrum of practices used in agriculture. Organic farming, for instance, typically requires tillage for weed control and residue management, but also includes diversified soil C inputs via the use of green and animal manures and crop rotational diversity (Teasdale et al., 2007). Recent studies have observed that bacterial and fungal biomass are higher under organic management compared to no-till or conventional (Wu et al., 2015) and AMF PLFA markers increased with cover crop additions (Finney et al., 2017). King and Hofmockel (2017) observed that diversified cropping systems supported higher microbial biomass and increased microbial activity in the top 20 cm of soil.

Aside from AMF, most differences in microbial community structure cannot be directly related to function. For example, the phylum Acidobacteria has been observed to decrease under no-till (Dong et al. 2017) but increase under organic management (Chen et al., 2020), suggesting that these organisms are sensitive to management, but little is known about the role they play in soils aside from being aerobic heterotrophs (Kielak et al., 2016). More studies are needed to substantiate broad patterns in shifts in microbial community composition under different farming systems and elucidate their underlying drivers.

One hypothesis suggests that the form of the microbe –whether it is filamentous or unicellular growth habit– fundamentally constrains their life-history traits, and therefore their ecological behavior (Daws et al., 2020). Some key attributes that may distinguish filamentous microbes from single-cell microbes include a greater ability to compete for resources, a greater ability to disperse through soils, but slower recovery from some types of stressors. Some authors have therefore posited a general trade-off between growth-oriented and competition- oriented strategies even in filamentous fungi (Malik et al., 2020; Treseder and Lennon, 2015). Interestingly, some filamentous bacteria, such as Actinobacteria, appear to behave more similarly to small fungi (Griffin, 1985), especially when pore size distribution and moisture content create conditions that limit the dispersal of unicellular bacteria (Wolf et al., 2013). If these hypotheses are accurate, then we would expect tillage to drive higher abundance of smaller, faster growing fungi that are tolerant of a greater diversity of environmental conditions (Egidi et al., 2019); or, alternatively, that lack of tillage would promote taxa that form larger, more complex structures.

In the current study, we characterized the soil microbial community of three corn-soybean--wheat crop rotations managed using conventional no-till (NT), conventional tillage (CT) and organic (ORG) farming practices in the

Mid-Atlantic, USA. Soils were collected from the long-term USDA-ARS Farming Systems Project. At the time of sampling, the experimental treatments had been in place for at least 13 years. Soils were sampled at 0-5 cm, 5-10 cm, 10cm – plow layer (Ap horizon), and Ap-30 cm depths. Ap depth varied by plot but was between 12 and 26 cm (Figure S1). Amplicon sequencing was used to determine the composition of the bacterial/archaeal and fungal communities, and qPCR was conducted to quantify microbial gene abundances. Additional soil physical and chemical parameters were measured including pH, texture, C and N content, and mineral N concentrations. We tested three hypotheses: 1) long-term organic practices increase the gene copy numbers of bacteria and fungi because of more diverse and possibly greater C inputs in ORG compared to NT or CT; 2) no-till treatments have a different vertical distribution of bacteria and fungi above the Ap horizon compared to tilled treatments because the lack of tillage results in more concentrated resources and organisms at the surface of the soil; and 3) because of differing quality and quantity of C inputs, as well as differences in their incorporation within the plow layer, the composition of bacterial and fungal communities will differ above the plow layer among the three farming systems, diverging in particular with respect to differences in putative growth habit.

4.2 Materials and Methods

Site Description

Soils were sampled in August 2013 from the Farming Systems Project (FSP) in Beltsville, MD (39.03° N, 76.90° W). The climate is humid subtropical bordering humid continental, with an average annual precipitation of approximately 1110 mm, and average temperature of 12.8 °C. Soil series mapped at the study site include Christiana (fine, kaolinitic, mesic Typic Paleudults), Keyport (fine, mixed, semiactive, mesic Aquic Hapludults), Matapeake (fine-silty, mixed, semiactive, mesic Typic Hapludults), and Mattapex (fine-silty, mixed, active, mesic Aquic Hapludults) silt loams (Spargo et al., 2011). All soils are designated as Ultisols with variable amounts of clay in illuvial horizons, and variable average water table depths (Soil Survey Staff, 2014).

The FSP is a long-term experiment testing the effects of agricultural management and crop rotation length on agricultural sustainability. Soils were sampled from chisel-till (CT), no-till (NT), and organic (ORG) plots in each of four experimental blocks. Only plots in the wheat phase of the crop rotation were sampled (Table 4.1). Further details of the agricultural management of the FSP can be found in Spargo et al. (2011) and White et al. (2019). In the case of the CT and NT treatments, the plots we sampled were first managed as part of these treatments in 2000 as described in Cavigelli et al. (2008). Thus, they had been under consistent management for 13 years when sampled.

Table 4.1: Agricultural management practices of the Farming Systems Project (adapted from Spargo et al., 2011).

Management Practice	**Cropping System**		
	No Till (NT)	Chisel Till (CT)	Organic (ORG)
Crop rotation †	C-r-S-W/S	C-r-S-W/S	C-r-S-W-v
Primary tillage ‡	None	Ch	D, MB, or Ch
Weed control §	Herbicides	Herbicides	RH, RC
Fertility ¶	N, P, K	N, P, K	GM, AM, K

†C=corn, S=soybean, W=winter wheat, W/S=winter wheat followed by double-cropped soybean; r=rye cover crop, v=hairy vetch cover crop. NT and CT followed a 2-year C-W/S rotation from 1996 to 2000.

‡D=disk, MB=moldboard plow, Ch=chisel plow

§RH=rotary hoe, RC=between-row cultivator

¶N=synthetic nitrogen fertilizer; P=acidulated phosphorus fertilizer; K=potassium sulfate fertilizer; GM=green manure; AM=animal manure

Soil Sampling

Nine cores (2.5 cm diameter, 100 cm depth) within plastic sleeves were taken from each of 12 plots using a truck-mounted hydraulic probe. Within each plot, 3 samples were taken at 3 m intervals along a transect that followed the topographical contour of the land, to control for topography. At each location three cores were sampled at 0°, 90°, and 180° orientations relative to the long axis of the truck. An additional 10 shallower cores (2 cm diameter, 20 cm depth) were taken at random points near the deeper samples to ensure enough soil for all analyses in the 0-5 cm and 5-10 cm depth increments.

The plastic sleeves containing the 100 cm long soil cores were capped at either end and transported to a nearby lab where they were immediately extruded from plastic sleeves and divided into 0-5 cm, 5-10 cm, 10 cm-Ap Horizon, Ap Horizon-30 cm, and 30-60 cm depth fractions in sterile conditions to eliminate surface contamination. All cores from a given plot were lined up relative to an abrupt color change that indicated the Ap horizon. The depth of the Ap horizon varied by <5cm for any individual plot but varied by 14 cm across all plots and treatments. Soil from all 9 cores for a given depth increment collected from a given plot was combined in a sterile container. The soil from the 0-5 cm and 5-10 cm shallower cores were combined with the appropriate depth increments from the deeper cores. Combined core fractions were homogenized, sieved (4 mm) and stored for further analysis. All sampling materials were cleaned with 70% ethanol between samples to limit cross contamination. Samples to be used for DNA extraction were frozen at -20 °C, for NO_3^- and NH_4^+ extraction were stored at 4 °C, for moisture content and particle size analysis were oven-dried at 105 °C for 24 hours, and for all other analyses were air-dried at room temperature for three days.

Soil Physiochemical Properties

Bulk density was calculated by dividing the oven-dry corrected mass of aggregated soil depth fractions from each plot by the calculated volume of the soil cores. Soil texture was determined using the pipette method of particle size analysis beginning with 20 g of air dried soil (Gee and Bauder, 1986).

Suspensions of 5.0 g of soil in 10.0 mL of deionized H_2O were allowed to equilibrate for 30 minutes, after which pH was measured using an Accumet AB15 soil pH meter with a glass electrode (Fisher Scientific, Waltham, MA). Percent C and N were measured from air-dry soil samples (0.20 g) using dry combustion with a LECO CHN 2000 analyzer (LECO Corporation, Lakeville, MI). Mineral N was extracted from 5 g of field-moist soil using 25 mL of 2 M KCl. Soil suspensions were shaken for 1 hour on a reciprocating shaker and filtered through Whatman #2 filters into scintillation vials. Extracts were frozen for ~3 months, thawed, and NO_3^- and NH_4^+ were quantified using an autoanalyzer (Lachat, Loveland, CO).

Microbial Community Analysis

DNA was extracted using MoBio “Powerlyzer” Powersoil DNA isolation kits (MoBio Laboratories, Carlsbad, CA) according to manufacturer’s instructions. The Powersoil extraction kits were used according to manufacturer’s instructions with the following alteration: 0.25 g of frozen soils were added to bead-beating tubes and shaken using a FastPrep-24 Instrument (MP Biomedical, Solon, OH) set at 5.5 m s^{-1} for 45 seconds. DNA was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA). Soils from the 30-60 cm depths were excluded from microbial analysis because DNA yields from these samples were inconsistent and sometimes too low for amplification. Extracts were diluted to 2.0 ng μL^{-1} for amplification and sequencing. The V4-V5 region of the 16S rRNA gene was amplified using 505F – 806R universal primers (Caporaso et al., 2012), and a region of the fungal ITS gene was amplified with modified ITS1f - ITS2 fungal primers (Smith and Peay, 2014), both amplified according to published amplification parameters. Both sets of primers had Illumina adaptor overhang sequences: 5’ TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG – Forward Primer – 3’, and 5’ GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G – Reverse Primer – 3’. Amplified products were cleaned and prepared for Illumina sequencing using the 16S Metagenomic Sequencing Library Preparation protocol (Part # 15044223 Rev. B, support.illumina.com). Both 16S and ITS PCR products were cleaned using AMPure XP beads (Beckman Coulter, Pasadena, CA). The 8-nucleotide indexes and Illumina sequencing indices were attached using a Nextera XT Index Kit (Illumina, San Diego, CA), and final products were re-cleaned using AMPure XP beads. Two µL of each cleaned final amplified product were combined into one pooled sample for each target (16S or ITS), which was quantified in triplicate using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA) before being diluted into two aliquots of 10 nM and 1.0 ng μL^{-1}. The 1.0 ng μL^{-1} dilution of combined 16S and ITS samples was analyzed using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) to test for product purity and appropriate amplicon length (variable for ITS sequences) prior to sequencing. Sequencing of the 10 nM dilutions was performed using a MiSeq Desktop Sequencer (Illumina, San Diego, CA) at the Center for Genome Research and Biocomputing at Oregon State University. Spikes of 5% and 10% PhiX were included on 16S and ITS plates, respectively.

Quantitative PCR was conducted using the same primers as for sequencing, but without the Illumina adapters. The minimum standard efficiency threshold was 0.95. The master mix for each Q-PCR reaction was as follows: 16 µl of Quick-start Sybr Green Readymix with Rox (Sigma), 1.0 µl of 10 µM forward and reverse primers, and 2 µl DNA diluted to 1.25 ng ml^{-1}. Our standard was a 1:10 serial dilution of 2.5 µg ml^{-1} linearized plasmid DNA from *E. coli K-12* and *Saccharomyces cerevisiae S288c* bacterial and fungal strains respectively.

Bioinformatics

Paired-end sequences (250 bp) were demultiplexed and then joined. They were then processed using the DADA2 pipeline (Callahan et al., 2016, p. 2). Sequences were separated into unique ASVs, all of which were used for alpha and beta diversity, except 16S sequences that matched mitochondria or chlorophyll sequences.

Fungal taxonomy was assigned in DADA2, referencing the UNITE database (Cole et al., 2009; Nilsson et al., 2019). Bacterial taxonomy was assigned in DADA2 referencing the SILVA database (Quast et al., 2013; Yarza et al., 2014). Fungal ASVs that were assigned to a genus but not species were assigned a species label of “undefined.” Subsequently, ASV counts were aggregated at the species level using the software phyloseq (McMurdie and Holmes, 2013). Prokaryote taxa were aggregated at the genus level. The aggregated taxonomy was not used in diversity testing (alpha or beta), but rather to examine the associations of the taxonomic groups to experimental conditions. Fungal taxa were assigned a putative functional group using the FungalTraits database (Põlme et al., 2020).

Quantitative sequencing of the count abundance was used to correct sample-wise counts for differences in total bacterial fungal abundance (Epp Schmidt et al., 2017; Jian et al., 2018). It has been well-established that sequence data alone provides only compositional information – that is only information on the abundance of taxa relative to one another (Gloor et al., 2017). However, when combined with other quantitative data, sequence abundance distributions may be transformed into estimated abundances that provides information on both the relative and total abundance of taxa (Jian et al., 2020), removing the constraints of compositionality. This approach has been called “Quantitative Sequencing;” hereafter we refer to this method as “QSeq.” QSeq works in two steps: the raw count data from the DADA2 pipeline are transformed into relative abundance, then the percentages were multiplied by the sample-wise total abundance, as measured using Quantitative PCR of the 16S or ITS genes. QPCR has been previously used in a QSeq approach for soils with rarefaction (Epp Schmidt et al, 2017); however in the current work we follow the recommendations of McMurdie and Holmes (2014), and implement the QSeq approach without rarefaction to maintain the highest degree of information about the species abundance distribution we have in each sample. This created an estimate of the total abundance of each ASV. Additionally, for the indicator species analysis, taxa that did not meet a prevalence threshold (at least 3 counts across 10 samples) were removed to ensure that the only taxa included in this dataset had sufficient information for the indicator species analysis.

Statistical Analysis

We use alpha diversity to refer to the total number of unique taxa in each sample (i.e. richness). Instead of using an alpha diversity index, we use the method of Tedersoo et al. (2014) to model the relationship between sequencing depth and detected taxa. We use the residuals of the model to estimate the treatment effect on number of ASVs detected; by using the residuals of the model we account for sequencing depth. Thus, we applied an ANOVA to the residuals of this model to determine if farming system or depth had a higher or lower number of unique taxa than expected given the sequencing depth of the sample.

Beta diversity was measured using a Jaccard distance metric and tested with a PERMANOVA analysis from the vegan package (Oksanen et al., 2016) in R (R Core Team, 2015) . The resulting output was used to test the effect of experimental and environmental factors on sample similarity. The experimental factors were soil depth and farming system. Additional covariates included: soil pH, percent soil C, percent soil N, soil NH_4^+-N and NO_3^- -N, C:N ratio, percent sand, silt, and clay, and bulk density. Differences in fungal and prokaryote beta diversity were visualized using a NMDS (Non-metric multidimensional scaling) ordination on the most limited fungal dataset

(ASVs aggregated into species, filtered to remove low prevalence and poorly annotated taxa). The package indicspecies was used to identify the association of taxa to each experimental treatment (Cáceres and Legendre, 2009). The R package igraph was used to construct a bipartite network of strong indicators (from indicspecies package that met the alpha threshold of 0.05) for both depth and farming system categories (Csárdi and Nepusz, 2006). For each dataset (ITS and 16S), additional bipartite networks were made that included only those taxa that were strongly associated with each farming system to visualize the relative overlap of depth and farming system indicators. Linear modeling on the quantitatively scaled community was used to evaluate the main and interacting effects of farming system and depth on individual taxa.

4.3 Results

Physiochemical Properties

Of the edaphic factors measured, only %C (ANOVA: F = 2.34, P = 0.052) and %N (ANOVA: F = 2.75, P = 0.027) had a significant interaction between farming system and depth (Figure 4.1A and 4.1C). Both %C and %N were significantly higher in the 0-5 cm increment and decreased rapidly below 5 cm, presumably due to surface residue impacts. Percent C was higher and %N trended higher in the 0-5 cm increment in NT and ORG, compared to CT (Figure 4.1A and 4.1C). The %C differed significantly among farming systems across all depths (ANOVA: F = 4.26, p = 0.022) with the highest concentrations in the ORG plots and similar values in NT and CT. The same differences were observed for %N (ANOVA: F = 4.06, P = 0.026). Regardless of farming system, %C and %N were substantially lower in soils below than above the plow layer. Soil NO_3^--N was different among farming systems (ANOVA: F = 2.47, p = 0.099), being lower in NT compared to CT and ORG (Figure 4.1E). Bulk density followed the same pattern (ANOVA: F = 3.52, p = 0.040) (Figure1D). Meanwhile, pH increased with depth in all treatments, but was higher in the ORG plots (ANOVA: F = 7.308, p = 0.002; Figure 4.1B). The extractable NH_4^+ values did not differ with farming system or depth and, as expected, the tilled systems, CT and ORG, had greater heterogeneity, possibly as an effect of incomplete mixing of plant residues and poultry litter (ORG) during tillage prior to planting wheat the previous fall (Figure 4.1F).

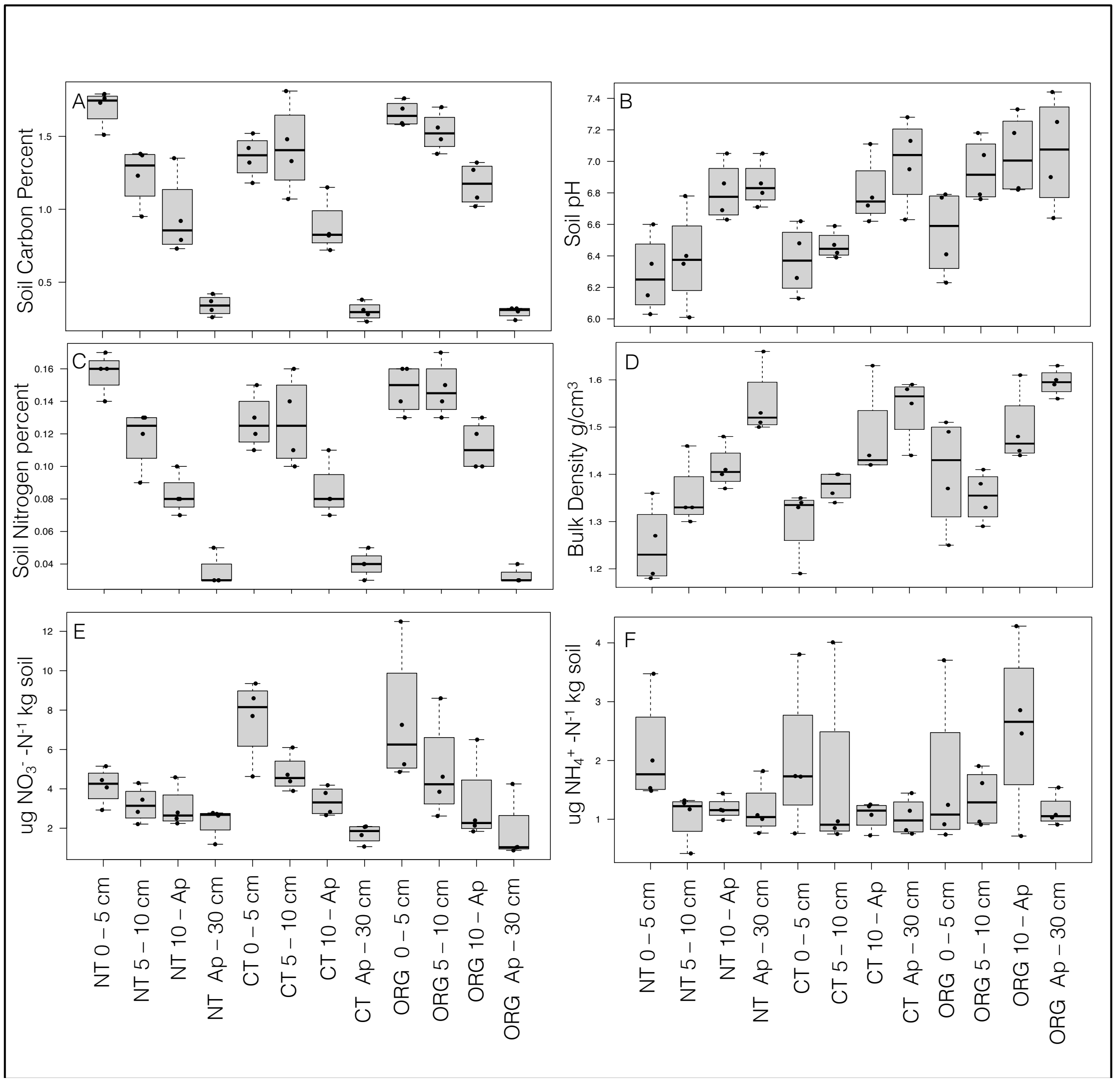

Figure 4.1: Soil physiochemical properties for plots within farming system and by depth: A) C mass as a percent of soil mass; B) soil pH; C) N mass as a percent of soil mass; D) bulk density of soil in g/cm^3; E) mass of nitrate per kg of soil; and F) mass of ammonium per kg of soil.

Microbial Quantity and Alpha Diversity

Total ITS gene copy numbers were greater in the ORG relative to both NT and CT systems (Figure 4.2A: ANOVA: F = 8.99, P < 0.001), though depth rather than farming system was a much larger influence on gene copy numbers (ANOVA: F = 138.7, P < 0.001). The ITS gene copy numbers ranged from 3.6 x 10^5 to 4.0 x 10^7 copies g^{-1} of dry soil. The 16S rRNA gene copy numbers did not differ among farming systems (Figure 4.2B, ANOVA: F = 0.24, p = 0.79), but ranged from 1.4 x 10^{11} to 4.6 x10^{13} gene copies g^{-1} of dry soil. A single ORG core was omitted from this analysis because the calculated gene values were several orders of magnitude below normal soil 16S gene abundances (see Figure S4.2). The 16S rRNA gene copy numbers decreased with depth in all farming systems (Figure 4.2B, F = 32.9, P < 0.001).

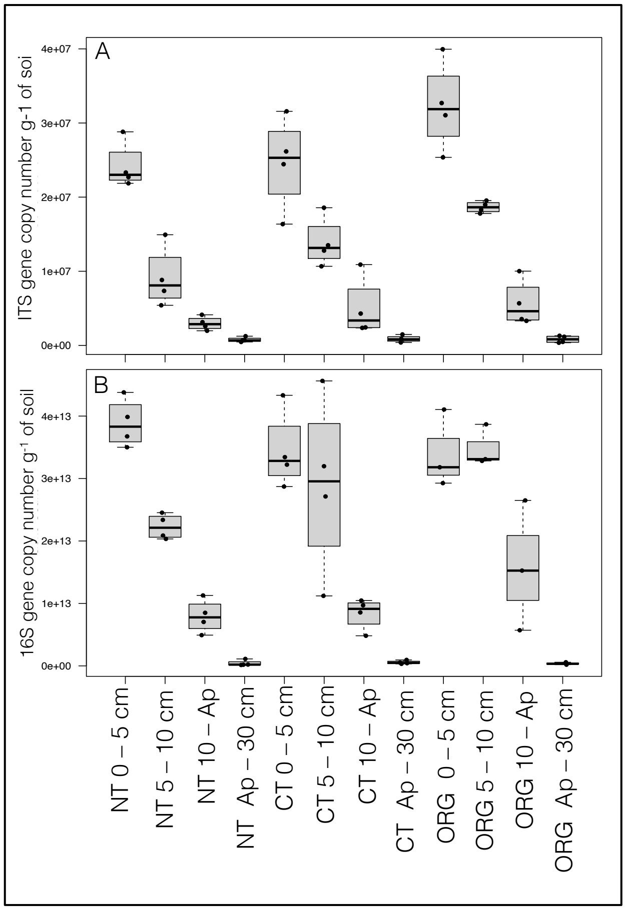

Figure 4.2: Boxplots of gene copy numbers per gram of soil for A) ITS and B) 16S rRNA gene loci. Copy numbers were determined using QPCR.

Farming system interacted with depth in the case of ITS copy numbers (Figure 4.2A; F = 2.11, P = 0.08). Farming system and depth did not interact in the case of 16S gene abundance even though plotted data would suggest an interaction (Figure 4.2B). ITS and 16S rRNA gene copy numbers were correlated across farming systems (Figure S4.3: Pearson's r = 0.86). The correlation was strongest under NT (Pearson r = 0.94) compared to CT (Pearson's r = 0.81) and ORG (Pearson's r = 0.84). Profile depth, but not farming system, influenced the ratio of ITS to 16S rRNA gene copy numbers (Figure S4.3; F = 23.3, P < 0.001). The %C was correlated with both ITS and 16S rRNA gene copy numbers g^{-1} soil (Pearson's r = 0.76; r = 0.80, respectively) but there was no correlation between the F:B ratio and the %C (Pearson's r = 0.08; data not shown). There was little or no relationship between the F:B ratio and the C:N ratio in any system (Pearson's r NT = -0.033; CT = 0.26; ORG = -0.006).

Farming systems did not differ in taxa richness for fungal or prokaryote datasets (Figure 4.3A, 4.3B; ANOVA: F < 1.3, p > 0.1) but fungal and prokaryote richness decreased with depth (Figure 4.3A,4.3B, F > 30; P < 0.001). The number of fungal taxa detected ranged from 322 to 1048 ASVs per sample. The number of detected bacterial taxa ranged from 4419 to 9576 ASVs per sample. There was no interaction between farming system and depth on richness (ANOVA, P > 0.5). The evenness of fungal taxa was similar across farming systems and depths (Figure 4.3C; ANOVA: F < 1.5, P > 0.25), but there was a pronounced decrease in the prokaryote evenness below 5 cm in all farming systems (Figure 4.3D: ANOVA, F = 40.7, P < 0.001).

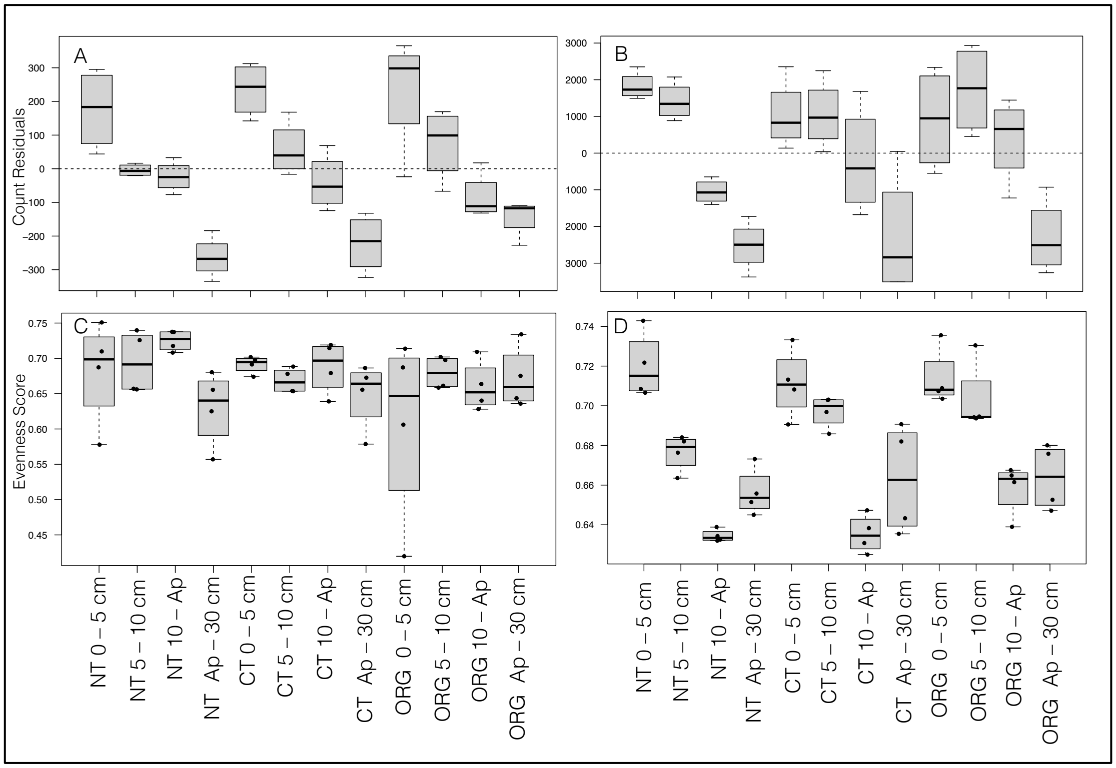

Figure 4.3: Boxplots of relative ASV richness effect across farming system and soil depth for A) ITS and B) 16S rRNA gene loci. ASV richness is defined as the number of taxa detected in a sample. Because this value is confounded with sequencing depth (more sequencing effort results in more ASVs detected), a linear model that relates sequencing depth to ASV richness was constructed, and the residual values from this model for each sample are used to determine the effect of the experimental treatments on the detection rate of unique ASVs. Residual values above zero can be interpreted as an increase in sample-wise ASV richness above the expected value for the sequencing effort of a given sample. Similarly, residual values below the line are interpreted as a decrease in ASVs richness below the expected value of a sample given the sequencing effort of that sample. The effect size between treatment conditions is the absolute distance between average residuals of each condition. These plots represent the sample residuals grouped by treatment condition and depth; the position of the boxplot along the Y axis can be interpreted as the effect of the treatment condition on the ASV detection rate. Treatment conditions with higher detection rates are interpreted as having higher richness. Evenness scores (Pielou's J) were also calculated at each farming system and depth category for C) ITS and D) 16S rRNA gene loci. Higher values indicate higher evenness score.

Microbial Beta Diversity

After 13 years, the microbial community composition did not differ below the Ap horizon among farming systems but did differ at shallower depths, in support of our hypothesis. Points on an NMDS of fungal ASVs clustered based on farming system and with depth (Figure 4.4A). Depth explained 24% of the variation in the fungal community composition (PERMANOVA: F = 5.12, p = 0.001) and farming system explained 7.5% of the variation in the fungal community composition (Figure 4.4; PERMANOVA: F = 2.46, p = 0.001). Importantly, there was an interaction between farming system and depth increment that explained 12.7% of variation in the fungal community composition (Figure 4.4C, PERMANOVA: F = 1.38, P = 0.004). This interaction is evident on the NMDS ordination plots, where points below the Ap in general did not cluster as tightly and were not as clearly clustered by farming system as for shallower depths. Similarly, the points on the prokaryote NMDS ordination clustered by farming system and depth, although farming system patterns were not as clear as for the fungi (Figure 4.4B). Both depth (PERMANOVA: F = 7.6412, P = 0.001) and farming system (PERMANOVA: F = 1.9259, P = 0.015) had a significant effect, but there was no interaction between depth and farming system for prokaryotes (PERMANOVA: F = 1.1057, P = 0.273). For both bacterial and fungal communities, the Ap-30 cm depth was least affected by farming system (Figure 4.4C).

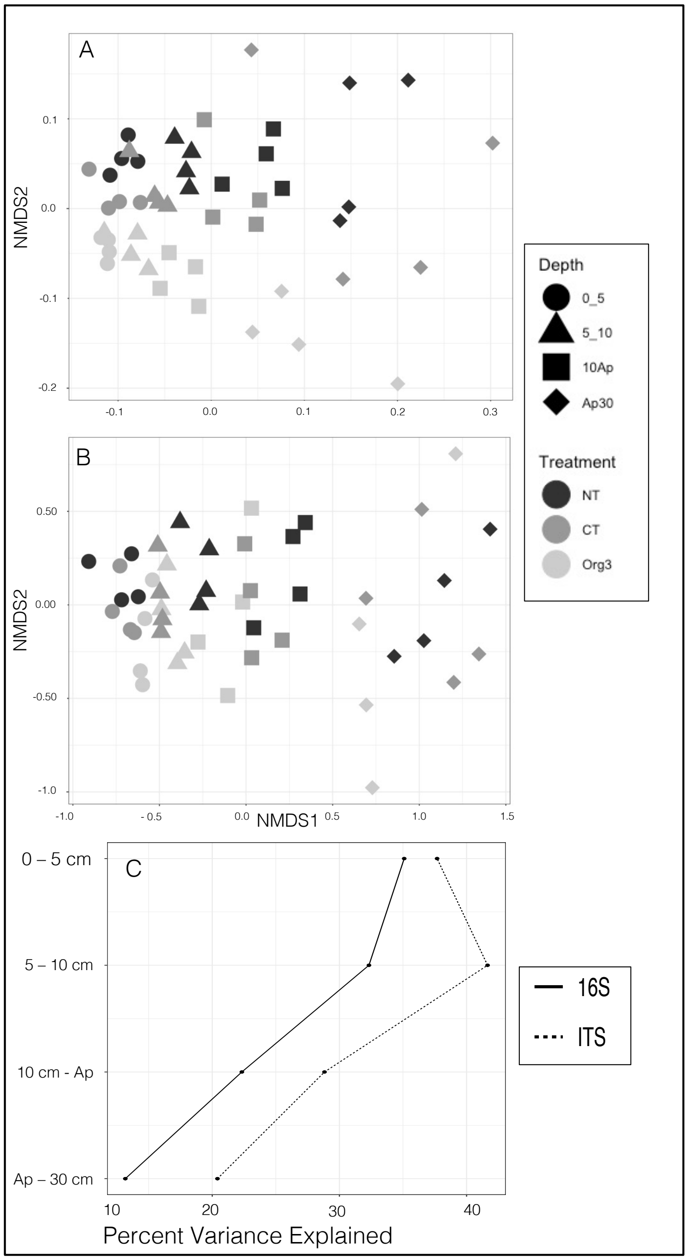

Figure 4.4: NMDS ordinations based on Jaccard distance of A) ITS and B) 16S rRNA gene ASVs by farming systems and soil depth increment; and C) percent of variation in 16S and ITS by soil depth increment.

Indicator species analysis was used to determine what taxa were associated with depth and farming system (Figure 4.5). A greater number of taxa were uniquely associated with a depth increment than a given farming system. In total, 168 fungal taxa were associated with the 0-5 cm depth; 134

were associated with the 5-10 cm depth; 54 were associated with the 10 cm – Ap depth, and 4 taxa were associated with the Ap – 30 cm depth (Figure 4.5). A total of 352 prokaryote taxa were associated with the 0-5 cm depth; 320 were associated with the 5-10 cm depth; 180 were associated with the 10 cm – Ap depth, and 16 taxa were associated with the Ap – 30 cm depth (Figure 4.5). The decrease in indicator species with depth was consistent with decreases in gene copy numbers (Figure 4.2) and species richness (Figure 4.3A and 4.3B). Relatively few taxa were uniquely associated with a particular farming system (Table 4.2). In the fungal community, there were 8 fungal taxa that uniquely associated with NT; 9 fungal taxa that uniquely associated with CT; and 22 fungal taxa that uniquely associated with ORG. Those taxa that were associated with a specific farming system and a particular depth were typically associated with the top 10 cm, with a few exceptions in ORG (see below). For both NT and CT, one taxon was associated with the 0-5 cm depth and one taxon was associated with the 5-10 cm depth.

Table 4.2. Putatively identified fungal and prokaryote taxa indicative of farming systems. Lines with parenthetical numbers denote multiple indicator taxa matching that group. Fungal growth form annotations compiled from the FungalTraits database.

	Phylum	Genus	Growth Form
ITS			
NT Indicators	*Ascomycota*	*Sclerostagonospora*	Filamentous
		Trichoderma	Filamentous
		Paecilomyces	Filamentous
		Paraconiothyrium	Filamentous
		Zopfiella	Filamentous
		Magnaporthiopsis	Filamentous
		Pseudorobillarda	Filamentous
	Glomeromycota		Filamentous
CT Indicators	*Ascomycota*	*Myxocephala*	Filamentous
		Candida	Yeast
		Pseudeurotium	Filamentous
		Corynespora	Filamentous
		Penicilium	Filamentous
		Coniochaeta	Filamentous
	Glomeromycota	*Gigaspora (2)*	Filamentous
	Mortierellomycota	*Mortierella*	Filamentous
ORG Indicators	*Ascomycota*	*Preusssia*	Filamentous
		Leucothecium	Filamentous
		Lecythophora	Filamentous
		Volutella	Filamentous
		Humicola	Filamentous
		Schizothecium	Filamentous
		Atractium	Filamentous
		Neosetophoma	Filamentous
		Phaeosphaeria	Filamentous
		Chrysosporium	Filamentous
	Basidiomycota	*Coniophora*	Filamentous
		Thecaphora	Yeast
	Mortierellomycota	*Mortierella (2)*	Filamentous
16S rRNA			
NT Indicators	*Proteobacteria*		
	Actinobacteriota		
	Verrocomicrobiota		
CT Indicators	Cyanobacteria (2)		
	Methylomirabilota		
ORG Indicators	Actinobacteriota (2)		
	Bacteroidota (2)		
	Bdellovibrionota		
	Chloroflexi		
	Firmicutes (4)		
	Planctomycetota		
	Proteobacteria (5)		
	Verrumicrobiota		

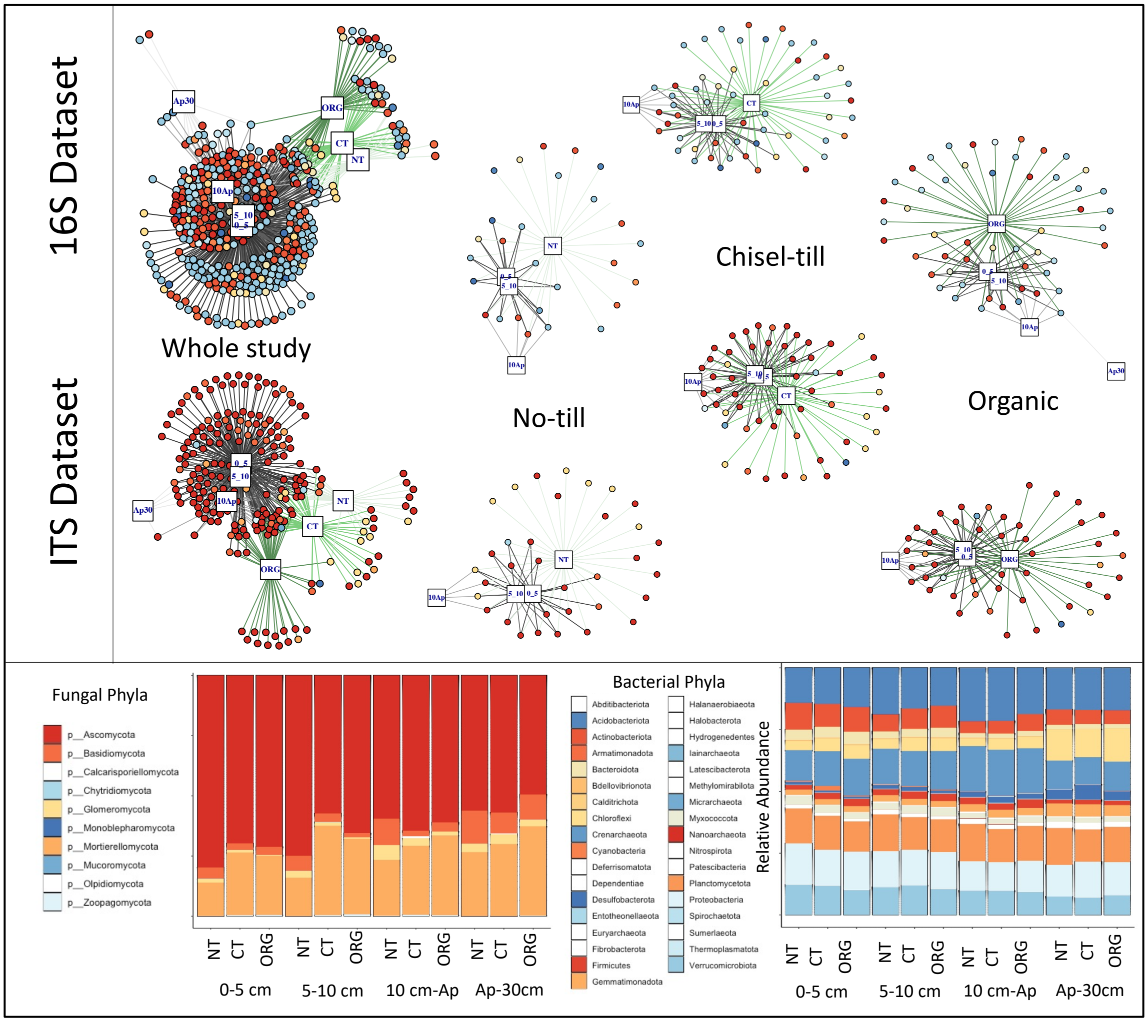

Figure 4.5. Bipartite networks of fungal and prokaryote indicator taxa corresponding to each farming system and depth category, colored by phylum. For each dataset, we show a complete network of all indicator taxa, and a network for each farming system that includes only the indicators for each farming system. Indicator taxa are mostly associated with the top 10 cm, with some differentiation of the 10cm-Ap depth category and almost no taxa strongly associated with depth below the Ap horizon. The stacked bar charts represent the relative abundance of each phylum for each farming system and depth category. Bar chart and network colors match; white colors in the bar chart represent phyla that have no indicator taxa in the networks.

The putative identification of fungal taxa can provide some indication of their function. Seven NT-associated taxa within phylum Ascomycota were putatively assigned genus-level identification, and one taxon was assigned to phylum Glomeromycota (Table 4.2). Taxa associated with CT included members of Glomeromycota belonging to the genus *Gigaspora* (Table 4.2); the Mortierellomycota taxon belonged to the genus *Mortierella*. The fungal indicator taxa for ORG were also associated with three depths. Ten taxa were associated with the 0-5 cm depth; seven were associated with the 5-10 cm depth, and three were associated with the 10 cm-Ap depth. Fifteen of the ORG associated taxa were Ascomycota, two indicator taxa belonged to Basidiomycota and two were Mortierellomycota (Table 4.2, Figure 4.5).

There were 35 fungal taxa that were only associated with the top 5 cm of soil and no fungal taxa were uniquely associated with any other depth increment. The 0-5 cm fungal indicator taxa belonged to Basidiomycota (8) or Ascomycota (27). Of the basidiomycetes, three taxa belonged to the genus *Hannaella*, two belonged to *Saitozyma*, and three belonged to the genera *Thanatephorus*, *Limonomyces* and *Coniophora*. The ascomycetes belonged primarily to the orders *Pleosporales* (11), *Hypocreales* (6), and *Tremellales* (5).

For the putatively identified prokaryote indicators three taxa were indicators of NT; four taxa were indicators of CT; and 17 taxa were indicators of ORG (Table 4.2). Two of the three CT taxa belonged to phylum Cyanobacteria.

Within the ORG, five prokaryote taxa were strongly associated with the 0-5 cm depth increment (Chloroflexi, Firmicutes, Actinobacteriota (2), and Bdellovibrionota. Within the ORG, six taxa were strongly associated with the 5-10 cm depth increment (Proteobacteria, Chloroflexi, Firmicutes, Actinobacteriota (2), Bdellovibrionota), and four were strongly associated with the 10 cm -Ap depth increment (Proteobacteria, Firmicutes, Bdellovibrionota, Actinobacteria), and only one was associated with the Ap – 30 depth increment (Proteobacteria). The Proteobacteria below the Ap horizon belonged to the genus *Propionvibrio*.

There were no prokaryote taxa that uniquely associated with NT or CT that also strongly associated with a depth category below 10 cm. There were 41 taxa that were uniquely associated with the top 5 cm (Figure 4.5), but no unique indicator taxa in either the 5-10 cm or 10 cm – Ap depth increments. Two prokaryote taxa belonging to Chloroflexi (family *Ktedonobacteraceae*) and Proteobacteria were strongly associated with the Ap – 30 depth increment (but with no particular farming system) (Figure 4.5). The Proteobacteria belong to the *Elsterales* family.

We found that the bacterial order *Pseudomonales* was strongly impacted by the interacting effect of farming system and depth (Figure 4.6; F = 2.61, P = 0.03), but there was no main effect of farming system on total *Pseudomonales* (Figure 4.6A). The fungal phylum Mortierellomycota abundance was significantly affected by an interaction between farming system and depth (Figure 4.6B; F = 3.12, P = 0.014), and was more abundant in CT and ORG. Other known important rhizosphere groups, such as genera *Burkholderia* and *Rhizobium,* and phyla Firmicutes, Ascomycota, and Glomeromycota did not respond to farming system or have a depth interaction with farming system (data not shown).

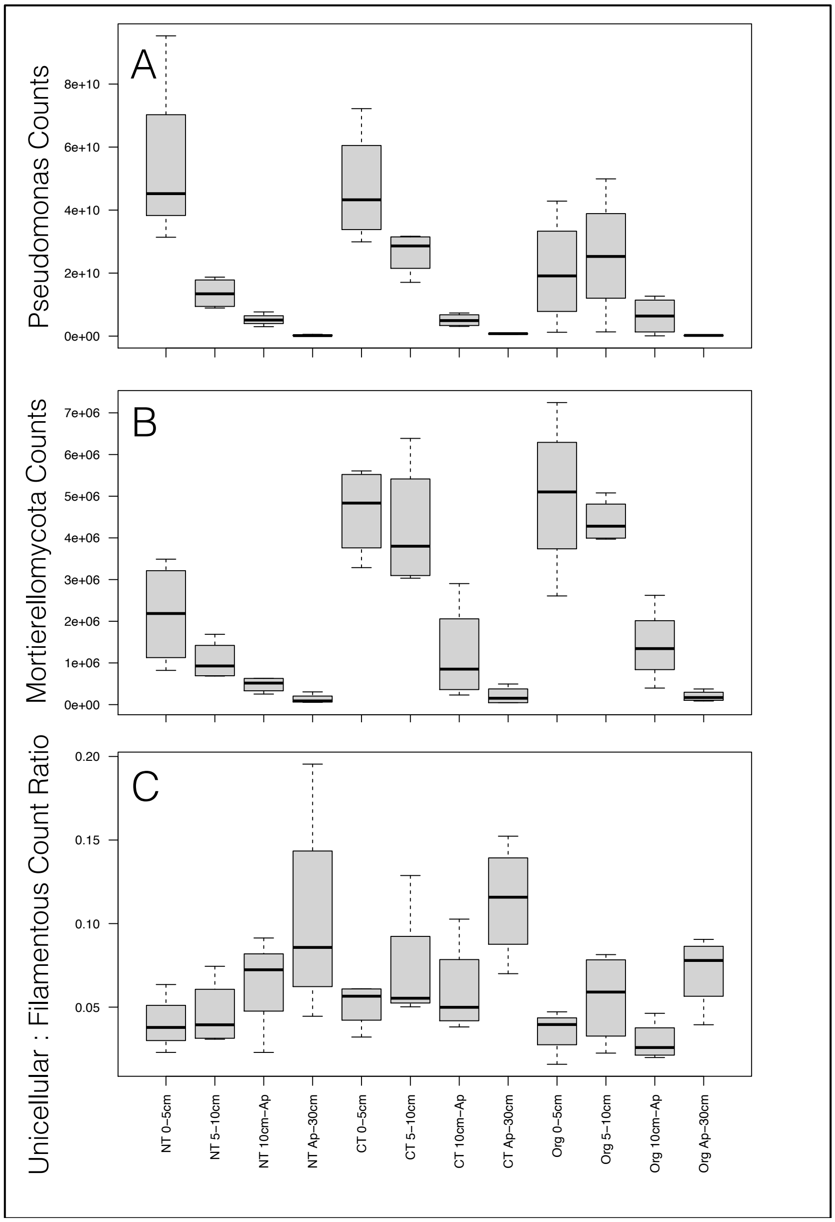

Figure 4.6. Boxplots of estimated abundance of a) Pseudomonas and b) Mortierellomycota phyla by farming system and depth increment. C) Boxplot of ratio of single-cell fungi to filamentous fungi.

Growth Strategies

We found that putatively identified filamentous fungi increased slightly in abundance in ORG relative to NT and CT (Figure 4.6C; ANOVA: F = 5.2, P = 0.01); and that there was no effect of farming system on the abundance of single-celled fungi (ANOVA: F = 2.2, P = 0 .12). Thus, the ratio of single-celled fungi to filamentous fungi was driven by the abundance of putative filamentous fungi, and was lowest in ORG compared to NT and CT (ANOVA: F = 3.1, P = 0.06), and generally increased with depth (Figure 4.6C; ANOVA: F = 7.2, P = 0.001).

4.4 Discussion

Microbial Abundance and Diversity

We hypothesized that higher C in the ORG system would drive higher microbial biomass in ORG than NT and CT. Our data indicate that both prokaryote and fungal abundance is highly correlated with soil %C and %N; each decreased with depth (Figure 4.1 and 4.2). Our data agree with previous studies that observed an exponential drop in bacterial gene abundance (Hao et al., 2021) and cell density (Fierer et al., 2003) with increasing distance from the soil surface (Figure 4.2). Prokaryote but not fungal evenness decreased with depth (Figure 4.3), and most taxa were strongly associated with surface horizons (Figure 4.5). The strong association between %C and prokaryote and fungal gene abundance is consistent with C availability being a primary driver of soil microbial metabolism (Kuzyakov and Blagodatskaya, 2015).

In partial support of our first hypothesis, fungi but not prokaryotes increased in abundance in ORG relative to the other farming systems (Figure 4.2). This difference in response between fungi and prokaryotes did not lead to an overall difference in gene ratios between fungi and prokaryotes, supporting the findings of Chen et al. (2020) that conservation tillage does not alter the fungal-to-bacterial ratios. Although prokaryote and fungal abundance were strongly positively correlated, our data also showed that the correlation was weakest in the ORG system, especially in the top 5 cm (Section 3.2; Figure S4.3). We found no effect of farming system on ASV richness (Figure 4.3), which contrasts with some studies that have observed increased soil microbial diversity in organic compared to conventional systems. For example, a meta-analysis that included 32 studies detected a trend towards increased bacterial and fungal diversity within organic farming systems but also found a wide margin of error (de Graaff et al., 2019).

Although our hypothesis regarding the abundance of prokaryotes and fungi due to farming system was only partially supported, we did identify both fungal and prokaryote taxa that were associated with particular farming systems. Indicators for NT include members of genus *Trichoderma*, which are commonly considered beneficial agricultural fungi and include several successful biocontrol agents (Sood et al., 2020). Recently, *Trichoderma* were used to control the maize pathogen *Magnaporthiopsis maydis* that causes late blight (Degani et al., 2021). The *Magnaporthiopsis* genus, which is composed of necrotrophic parasites of grass roots (Luo and Zhang, 2013), was also a strong indicator of NT farming systems.

Indicators of CT included *Glomeromycota gigaspora*, a common group of AM that was also observed to make up 60% of AM spores in Pennsylvania corn plots under both conventional and low-input agriculture (Franke-Snyder et al., 2001). Cyanobacteria were also strongly associated with CT soils, which might reflect that CT soils are exposed to more solar radiation than NT soils since CT soils have no residue cover for various portions of the rotation while NT generally has some residue covering at least a portion of the soil continuously (Table 4.2, Figure 4.5). In ORG, less sun might hit the surface of the soil during various parts of the rotation than in CT given the greater weed density and the fact that wheat residue is not harvested as in CT. Others have found that cyanobacterial biofilms quickly establish on disturbed agricultural soils that are exposed to sunlight (Peng and Bruns, 2019a), and may fix up to 2 kg N ha^{-1} per year (Peng and Bruns, 2019b).

A strong indicator of both CT and ORG from the fungal community was the *Mortierella* genus, which is considered to be primarily a saprotrophic genus associated with manures (Põlme et al., 2020). However it has also been shown to have facultative plant-fungal and fungal-bacterial mutualisms (Liao et al., 2019; Uehling et al., 2019). One of the ORG associated Mortierellamycota was *Mortierella capitata*, which has been found to exert a strong influence on rhizosphere microbiomes, supporting plant health (Li et al., 2020). A bacterial indicator of ORG, Bdellovibrionota, contains known bacteriovores (Iebba et al., 2013).

Depth Structure

The effect of farming system on community composition attenuated with depth (Figure 4.4C), with a significant decrease in diversity across depths (Figure 4.2, 4.4A & 4.4B). Together these results are consistent with research that found a turnover in community composition with soil depth is associated with a loss of diversity as distance to surface increases (Leeuwen et al., 2017; Schmidt et al., 2018). Further, there was a strong tendency for indicator taxa of a given farming system to primarily be indicators of the surface depth increments within those systems (Figure 4.5). In general, our data support the notion that treatments applied at the soil surface have decreasing influence with depth (Zeng et al., 2016). Stone et al., 2014) found that different soils may experience similar shifts in community composition with depth, to such a degree that depth effects could explain more community variance than soil types. In our dataset, depth increment was associated with large differences in several important edaphic factors (Figure 4.1). Generally, %C, pH, %N and nitrate were all important determinants for the abundance of many individual taxa (Figure S4.4); and thus, were also important covariates explaining microbial community turnover. Soil pH, %C, and %N have also been found to strongly predict beta diversity patterns at the landscape scale (Fierer et al., 2009; Fierer and Jackson, 2006). Depth may be an important predictor of changes in community composition in our dataset because these important edaphic factors covary along the soil profile.

We found two bacterial taxa in the *Elsterales* order were strong indicators of the Ap-30 cm depth increment; this family has been associated with phosphorus-limited soil communities (Oliverio et al., 2020). Previous research using PLFA and sampling explicitly by soil horizon found that the surface horizon, defined as 0 – Ap, differed substantially in microbial community composition compared to adjacent horizons (Steenwerth et al., 2008). The only taxa in our study that were strongly associated with both deep soils and a particular farming system (in this case, ORG) belonged in the *Propionvibrio* genus, which includes anaerobic bacteria specializing in the fermentation of aromatic compounds (Brune et al., 2002). Chloroflexi were important bacterial indicators of ORG, and are predominantly filamentous (Speirs et al., 2019). Chloroflexi are known to engage in a diversity of metabolic strategies, including sugar respiration, fermentation, C fixation, acetogenesis, and obligate organohalide respiration (Hug et al., 2013). Together, these Cholorflexi and *Propionvibrio* taxa could indicate more reducing conditions or higher availability of aromatic compounds for decomposition below the Ap horizon under ORG.

We hypothesized that differences in tillage across farming systems would result in different distributions of environmental conditions and taxa across the soil profile in each farming system. Our study provides support for this hypothesis in particular with regards to changes in %C; we found that after 13 years the effect of farm management impacts % C primarily in the top 10 cm (Figure 4.1A). Our study agrees with previous research showing that no-till systems accumulate C at the soil surface (Mestelan et al., 2021; Ogle et al., 2019). However, %C across all depths was not different between NT and CT and was greater in ORG than the two conventional systems. Further investigation of these data in the context of the soil C sequestration debate (e.g. (Ogle et al., 2019) is warranted—and should include comparing samples collected in 2013 with those collected when the experiment was initiated in 1996 and including data to 50 cm depth—but is beyond the scope of this paper.

Sequence data are inherently compositional, hampering our ability to decipher the abundance of individual taxa (Gloor et al., 2017, 2016). In order to alleviate this limitation and make inferences about the true abundance of taxa among farming systems and across the depth profile, we employed a QSeq (quantitative sequencing) approach that combines sequence profiles with quantitative PCR (Epp Schmidt et al., 2017; Jian et al., 2020). This allowed us to illustrate that the vertical distribution of certain taxa in the profile changed as a function of the farming system. For example, Mortierellomycota was more abundant in CT and ORG treatments (Figure 4.6B); while Pseudomonales were not affected overall by farming systems, but their vertical distribution was changed somewhat between farming systems (Figure 4.6A). In each of these cases, the interaction among farming systems occurs primarily in the top two (0-10 cm) depth categories. These data, combined with the strong association of farming system indicators with the top 10 cm (Figure 4.5) suggests that in our row-cropping agricultural systems, soil microbial community responses to management are primarily found in the top 10 cm.

Growth Habit and Tillage

Our data do not support the hypothesis that tillage leads to a decrease in competitiveness of filamentous fungi (Figure 4.6C); we found higher abundance of putative filamentous fungi in ORG under tillage. We also found that the abundance of single-cell-fungi were not affected by farming system; rather there was a slight increase in relative abundance of single-celled fungi as depth increased, especially below the Ap horizon. Pore size distribution and soil moisture are important determinants of diversity for both bacteria and single-celled fungi (Wolf et al., 2013). The distribution of pore size at different soil depths, as well as the general availability of C resources, may lead to a shift in relative advantage between single-celled versus filamentous microbes. If it is true that single-cell fungal dispersal is limited by similar conditions to bacteria, then dynamics in the bacterial community may provide insight into the factors affecting single-celled fungi. It has been suggested that soil bacterial evenness at large geographic scales is in part mediated by the connectivity of aqueous-phase habitat in soils; where low connectivity of aqueous-phase pore space translates into a barrier to bacterial dispersal within the soil profile, thereby reducing competition and resulting in greater evenness (Bickel and Or, 2020). If this pattern holds true with soil depth, then our data could imply that on average the top 5 cm of our soils are drier than the 5-30 cm depths, leading to higher evenness at the surface than below 5 cm. We found that after 13 years, differences among farming systems were not sufficient to cause differences in evenness of the bacterial community in

ORG soils compared to CT and NT, whereas there were large differences by depth increment (Figure 4.3D).

Some studies have observed that tillage promoted a community that was more adapted to rapid growth than communities without tillage (Schmidt et al., 2018). We did not find strong support for this trend in our data. We note that though there are many exceptions, Ascomycetes are sometimes considered to be faster growing compared to Basidiomycetes (Egidi et al., 2019). Basidiomycetes are not well-represented in our indicator analysis, being primarily associated with the ORG treatment (a high-tillage treatment); whereas Ascomycetes are well represented as indicators across farming systems (Figure 4.5). It is likely the case that all three farming systems fall on the side of the spectrum that favors faster growing organisms when taken in context of the full diversity of soil ecosystems.

Land Use Legacies

Our edaphic data also demonstrated that soil factors may reflect a long-term legacy of previous land-use (Figure 4.1), which should be accounted for to the maximum extent possible in agricultural experiments. For example, soils in the coastal plain of the mid-Atlantic US region typically decrease in pH with depth, with subsoils often having a pH of around 5 (Soil Survey Staff, 2022). However, pH in our soils increases with depth (Figure 4.1). The USDA-ARS field in which this study was established has been under USDA management since 1910, and immediately prior to this study was cropped in alfalfa for eight years, ending in 1992. Since alfalfa prefers more alkaline soils than the field crops grown after experimental plots were established, we believe the soil pH pattern across all plots reflects earlier management when regular lime and cow manure applications were common. Over time, leaching of these materials could have increased pH to 30 cm depths. Soil samples collected and analyzed in fall 2012 showed that pH for the 0-25 cm depth averaged 6.5 for CT and NT treatments (data not shown), which is consistent with the data presented in Fig. 4.1B. At deeper soil depths, pH values for cores collected in the current study were more typical of the soil series (Soil Survey Staff, 2022) in the region, averaging 5.8 for the 30-60 cm depth and 4.6 to 4.9 for the 60-100 cm depth. Contemporary poultry litter additions to ORG are likely to also contribute to the higher pH in ORG than in NT and CT plots (Duruigbo et al., 2007).

4.5 Conclusions

There is considerable public interest in farming methods that may improve soil health and lead to improved ecological outcomes. After 13 years under different farming systems, the effect of system on the microbial community and edaphic factors were strongly associated with the top 10 cm of the soil profile. Depth explained more variance in the microbial community than did farming system, reflecting the covariance of important edaphic factors along the depth gradient. We found no evidence that tillage reduces the number of filamentous fungi when compared to NT or ORG treatments. The richness of prokaryotes and fungi within these farming systems were not statistically distinguishable after 13 years. Considering these findings, we conclude that the ORG system offers a slight advantage in terms of %C and fungal abundance, but not richness of taxa within the first decade of organic management.

4.6 Acknowledgements

We thank the reviewers for their thoughtful feedback that substantially improved this paper. This research was funded though the Department of Environmental Science and Technology at the University of Maryland, as well as a USDA NIFA Organic Transitions Grant (USDA-NIFA-ICGP-003448). We would like to thank Kreshnik Bejleri, Stanley Schlosnagle, Sarah Emche and Stanley Tesch for their technical support; and Eni Baballari, Stephanie Jamis, Sara Elbeheiry, Matthew Spielman, Nick LaPointe, and Dr. Steven Mirsky for assistance in soil sampling and processing; and Harry Schomberg for helpful discussions in interpreting soils data. Dietrich Epp Schmidt was supported by NRT-INFEWS: UMD Global STEWARDS (STEM Training at the Nexus of Energy, WAter Reuse and FooD Systems) that was awarded to the University of Maryland School of Public Health by the National Science Foundation National Research Traineeship Program, Grant number 1828910.

4.7 Literature Cited

Bach, E.M., Ramirez, K.S., Fraser, T.D., Wall, D.H., 2020. Soil Biodiversity Integrates Solutions for a Sustainable Future. Sustainability 12, 2662. https://doi.org/10.3390/su12072662

Bickel, S., Or, D., 2020. Soil bacterial diversity mediated by microscale aqueous-phase processes across biomes. Nat Commun 11, 116. https://doi.org/10.1038/s41467-019-13966-w

Brune, A., Ludwig, W., Schink, B., 2002. Propionivibrio limicola sp. nov., a fermentative bacterium specialized in the degradation of hydroaromatic compounds, reclassification of Propionibacter pelophilus as Propionivibrio pelophilus comb. nov. and amended description of the genus Propionivibrio. Int J Syst Evol Microbiol 52, 441–444. https://doi.org/10.1099/00207713-52-2-441

Cáceres, M.D., Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90, 3566–3574. https://doi.org/10.1890/08-1823.1

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

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

Cavigelli, M., Teasdale, J.R., Conklin, A.E., 2008. Long-Term Agronomic Performance of Organic and Conventional Field Crops in the Mid-Atlantic Region. Agron. J. 100. https://doi.org/10.2134/agronj2006.0373

Chen, H., Dai, Z., Veach, A.M., Zheng, J., Xu, J., Schadt, C.W., 2020. Global meta-analyses show that conservation tillage practices promote soil fungal and bacterial biomass. Agriculture, Ecosystems & Environment 293, 106841. https://doi.org/10.1016/j.agee.2020.106841

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., Tiedje, J.M., 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37, D141–D145. https://doi.org/10.1093/nar/gkn879

Csárdi, G., Nepusz, T., 2006. The igraph software package for complex network research. undefined.

Daws, S.C., Cline, L.A., Rotenberry, J., Sadowsky, M.J., Staley, C., Dalzell, B., Kennedy, P.G., 2020. Do shared traits create the same fates?

Examining the link between morphological type and the biogeography of fungal and bacterial communities. Fungal Ecology, Trait-based approaches to Fungal Ecology 46, 100948. https://doi.org/10.1016/j.funeco.2020.100948

de Graaff, M.-A., Hornslein, N., Throop, H.L., Kardol, P., van Diepen, L.T.A., 2019. Chapter One - Effects of agricultural intensification on soil biodiversity and implications for ecosystem functioning: A meta-analysis, in: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 1–44. https://doi.org/10.1016/bs.agron.2019.01.001

Degani, O., Khatib, S., Becher, P., Gordani, A., Harris, R., 2021. Trichoderma asperellum Secreted 6-Pentyl-α-Pyrone to Control Magnaporthiopsis maydis, the Maize Late Wilt Disease Agent. Biology 10, 897. https://doi.org/10.3390/biology10090897

Dong, W., Liu, E., Yan, C., Tian, J., Zhang, H., Zhang, Y., 2017. Impact of no tillage vs. conventional tillage on the soil bacterial community structure in a winter wheat cropping succession in northern China. European Journal of Soil Biology 80, 35–42. https://doi.org/10.1016/j.ejsobi.2017.03.001

Duruigbo, C.I., Obiefuna, J.C., Onweremadu, E.U., 2007. Effect of Poultry Manure Rates on Soil Acidity in an Ultisol. International Journal of Soil Science. https://doi.org/10.3923/ijss.2007.154.158

Egidi, E., Delgado-Baquerizo, M., Plett, J.M., Wang, J., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., 2019. A few Ascomycota taxa dominate soil fungal communities worldwide. Nat Commun 10, 2369. https://doi.org/10.1038/s41467-019-10373-z

Epp Schmidt, D.J., Pouyat, R., Szlavecz, K., Setälä, H., Kotze, D.J., Yesilonis, I., Cilliers, S., Hornung, E., Dombos, M., Yarwood, S.A., 2017. Urbanization erodes ectomycorrhizal fungal diversity and may cause microbial communities to converge. Nature Ecology & Evolution 1, 0123. https://doi.org/10.1038/s41559-017-0123

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. PNAS 103, 626–631. https://doi.org/10.1073/pnas.0507535103

Fierer, N., Schimel, J.P., Holden, P.A., 2003. Variations in microbial community composition through two soil depth profiles. Soil Biology and Biochemistry 35, 167–176. https://doi.org/10.1016/S0038-0717(02)00251-1

Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A., Cleveland, C.C., 2009. Global patterns in belowground communities. Ecology Letters 12, 1238–1249. https://doi.org/10.1111/j.1461-0248.2009.01360.x

Fierer, N., Wood, S.A., Bueno de Mesquita, C.P., 2021. How microbes can, and cannot, be used to assess soil health. Soil Biology and Biochemistry 153, 108111. https://doi.org/10.1016/j.soilbio.2020.108111

Finney, D.M., Buyer, J.S., Kaye, J.P., 2017. Living cover crops have immediate impacts on soil microbial community structure and function. Journal of Soil and Water Conservation 72, 361–373. https://doi.org/10.2489/jswc.72.4.361

Franke-Snyder, M., Douds, D.D., Galvez, L., Phillips, J.G., Wagoner, P., Drinkwater, L., Morton, J.B., 2001. Diversity of communities of arbuscular mycorrhizal (AM) fungi present in conventional versus low-input agricultural sites in eastern Pennsylvania, USA. Applied Soil Ecology 16, 35–48. https://doi.org/10.1016/S0929-1393(00)00100-1

Gee, G., Bauder, J., 1986. 1986, Particle-size analysis. p. 383-411. In A. Klute(ed.) Methods of soil analysis, Part 1. Agron. Monogr. 9. ASA and SSSA, Madison, WI.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 0. https://doi.org/10.3389/fmicb.2017.02224

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., Egozcue, J.J., 2016. It's all relative: analyzing microbiome data as compositions. Ann Epidemiol 26, 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003

Griffin, D.M., 1985. A Comparison of the Roles of Bacteria and Fungi, in: Leadbetter, E.R., Poindexter, J.S. (Eds.), Bacteria in Nature: Volume 1: Bacterial Activities in Perspective, Bacteria in Nature. Springer US, Boston, MA, pp. 221–255. https://doi.org/10.1007/978-1-4615-6511-6_8

Hao, J., Chai, Y.N., Lopes, L.D., Ordóñez, R.A., Wright, E.E., Archontoulis, S., Schachtman, D.P., 2021. The Effects of Soil Depth on the Structure of Microbial Communities in Agricultural Soils in Iowa (United States). Applied and Environmental Microbiology 87, e02673-20. https://doi.org/10.1128/AEM.02673-20

Helgason, B.L., Walley, F.L., Germida, J.J., 2010. No-till soil management increases microbial biomass and alters community profiles in soil aggregates. Applied Soil Ecology 46, 390–397. https://doi.org/10.1016/j.apsoil.2010.10.002

Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R., Williams, K.H., Tringe, S.G., Banfield, J.F., 2013. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome 1, 22. https://doi.org/10.1186/2049-2618-1-22

Iebba, V., Santangelo, F., Totino, V., Nicoletti, M., Gagliardi, A., Biase, R.V.D., Cucchiara, S., Nencioni, L., Conte, M.P., Schippa, S., 2013. Higher Prevalence and Abundance of Bdellovibrio bacteriovorus in the Human Gut of Healthy Subjects. PLOS ONE 8, e61608. https://doi.org/10.1371/journal.pone.0061608

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., Korpela, K., 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLOS ONE 15, e0227285. https://doi.org/10.1371/journal.pone.0227285

Jian, C., Luukkonen, P., Yki-Jarvinen, H., Salonen, A., Korpela, K., 2018. Quantitative PCR provides a simple and accessible method for quantitative microbiome profiling. bioRxiv 478685. https://doi.org/10.1101/478685

Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., van Veen, J.A., Kuramae, E.E., 2016. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. Frontiers in Microbiology 7, 744. https://doi.org/10.3389/fmicb.2016.00744

King, A.E., Hofmockel, K.S., 2017. Diversified cropping systems support greater microbial cycling and retention of carbon and nitrogen. Agriculture, Ecosystems & Environment 240, 66–76. https://doi.org/10.1016/j.agee.2017.01.040

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: Concept & review. Soil Biology and Biochemistry 83, 184–199. https://doi.org/10.1016/j.soilbio.2015.01.025

Leeuwen, J.P. van, Djukic, I., Bloem, J., Lehtinen, T., Hemerik, L., Ruiter, P.C. de, Lair, G.J., 2017. Effects of land use on soil microbial biomass, activity and community structure at different soil depths in the Danube floodplain. European Journal of Soil Biology 79, 14–20. https://doi.org/10.1016/j.ejsobi.2017.02.001

Li, F., Zhang, S., Wang, Y., Li, Y., Li, P., Chen, L., Jie, X., Hu, D., Feng, B., Yue, K., Han, Y., 2020. Rare fungus, Mortierella capitata, promotes crop growth by stimulating primary metabolisms related genes and reshaping rhizosphere bacterial community. Soil Biology and Biochemistry 151, 108017. https://doi.org/10.1016/j.soilbio.2020.108017

Liao, H.-L., Bonito, G., Rojas, J.A., Hameed, K., Wu, S., Schadt, C.W., Labbé, J., Tuskan, G.A., Martin, F., Grigoriev, I.V., Vilgalys, R., 2019. Fungal Endophytes of Populus trichocarpa Alter Host Phenotype, Gene Expression, and Rhizobiome Composition. MPMI 32, 853–864. https://doi.org/10.1094/MPMI-05-18-0133-R

Luo, J., Zhang, N., 2013. Magnaporthiopsis, a new genus in Magnaporthaceae (Ascomycota). Mycologia 105, 1019–1029. https://doi.org/10.3852/12-359

Malik, A.A., Martiny, J.B.H., Brodie, E.L., Martiny, A.C., Treseder, K.K., Allison, S.D., 2020. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. ISME J 14, 1–9. https://doi.org/10.1038/s41396-019-0510-0

Mbuthia, L.W., Acosta-Martínez, V., DeBruyn, J., Schaeffer, S., Tyler, D., Odoi, E., Mpheshea, M., Walker, F., Eash, N., 2015. Long term tillage, cover crop, and fertilization effects on microbial community structure, activity: Implications for soil quality. Soil Biology and Biochemistry 89, 24–34. https://doi.org/10.1016/j.soilbio.2015.06.016

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Mestelan, S., Smeck, N., Sprunger, C., Dyck, A., Dick, W., 2021. Four decades of continuously applied tillage or no-tillage on soil properties and soil morphology. Agrosystems, Geosciences & Environment 4, e20195. https://doi.org/10.1002/agg2.20195

Nilsson, R.H., Larsson, K.-H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., Saar, I., Kõljalg, U., Abarenkov, K., 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research 47, D259–D264. https://doi.org/10.1093/nar/gky1022

Ogle, S.M., Alsaker, C., Baldock, J., Bernoux, M., Breidt, F.J., McConkey, B., Regina, K., Vazquez-Amabile, G.G., 2019. Climate and Soil Characteristics Determine Where No-Till Management Can Store Carbon in Soils and Mitigate Greenhouse Gas Emissions. Sci Rep 9, 11665. https://doi.org/10.1038/s41598-019-47861-7

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O’Hara, R.B., Simpson, G., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Wagner, H., 2016. vegan: Community Ecology Package version 2.4-0.

Oliverio, A.M., Bissett, A., McGuire, K., Saltonstall, K., Turner, B.L., Fierer, N., 2020. The Role of Phosphorus Limitation in Shaping Soil Bacterial Communities and Their Metabolic Capabilities. mBio 11, e01718-20. https://doi.org/10.1128/mBio.01718-20

Peng, X., Bruns, M.A., 2019a. Development of a nitrogen-fixing cyanobacterial consortium for surface stabilization of agricultural soils. Journal of Applied Phycology 31, 1047–1056. https://doi.org/10.1007/s10811-018-1597-9

Peng, X., Bruns, M.A., 2019b. Cyanobacterial Soil Surface Consortia Mediate N Cycle Processes in Agroecosystems. Frontiers in Environmental Science 6, 156. https://doi.org/10.3389/fenvs.2018.00156

Põlme, S., Abarenkov, K., Henrik Nilsson, R., Lindahl, B.D., Clemmensen, K.E., Kauserud, H., Nguyen, N., Kjøller, R., Bates, S.T., Baldrian, P., Frøslev, T.G., Adojaan, K., Vizzini, A., Suija, A., Pfister, D., Baral, H.-O., Järv, H., Madrid, H., Nordén, J., Liu, J.-K., Pawlowska, J., Põldmaa, K., Pärtel, K., Runnel, K., Hansen, K., Larsson, K.-H., Hyde, K.D., Sandoval-Denis, M., Smith, M.E., Toome-Heller, M., Wijayawardene, N.N., Menolli, N., Reynolds, N.K., Drenkhan, R., Maharachchikumbura, S.S.N., Gibertoni, T.B., Læssøe, T., Davis, W., Tokarev, Y., Corrales, A., Soares, A.M., Agan, A., Machado, A.R., Argüelles-Moyao, A., Detheridge, A., de Meiras-Ottoni, A., Verbeken, A., Dutta, A.K., Cui, B.-K., Pradeep, C.K., Marín, C., Stanton, D., Gohar, D., Wanasinghe, D.N., Otsing, E., Aslani, F., Griffith, G.W., Lumbsch, T.H., Grossart, H.-P., Masigol, H., Timling, I., Hiiesalu, I., Oja, J., Kupagme, J.Y., Geml, J., Alvarez-Manjarrez, J., Ilves, K., Loit, K., Adamson, K., Nara, K., Küngas, K., Rojas-Jimenez, K., Bitenieks, K., Irinyi, L., Nagy, L.L., Soonvald, L., Zhou, L.-W., Wagner, L., Aime, M.C., Öpik, M., Mujica, M.I., Metsoja, M., Ryberg, M., Vasar, M., Murata, M., Nelsen, M.P., Cleary, M., Samarakoon, M.C., Doilom, M., Bahram, M., Hagh-Doust, N., Dulya, O., Johnston, P., Kohout, P., Chen, Q., Tian, Q., Nandi, R., Amiri, R., Perera, R.H., dos Santos Chikowski, R., Mendes-Alvarenga, R.L., Garibay-Orijel, R., Gielen, R., Phookamsak, R., Jayawardena, R.S., Rahimlou, S., Karunarathna, S.C., Tibpromma, S., Brown, S.P., Sepp, S.-K., Mundra, S., Luo, Z.-H., Bose, T., Vahter, T., Netherway, T., Yang, T., May, T., Varga, T., Li, W., Coimbra, V.R.M., de Oliveira, V.R.T., de Lima, V.X., Mikryukov, V.S., Lu, Y., Matsuda, Y., Miyamoto, Y., Kõljalg, U., Tedersoo, L., 2020. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. Fungal Diversity 105, 1–16. https://doi.org/10.1007/s13225-020-00466-2

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41, D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team, 2015. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schmidt, R., Gravuer, K., Bossange, A.V., Mitchell, J., Scow, K., 2018. Long-term use of cover crops and no-till shift soil microbial community life strategies in agricultural soil. PLOS ONE 13, e0192953. https://doi.org/10.1371/journal.pone.0192953

Smith, D.P., Peay, K.G., 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLoS ONE 9, e90234. https://doi.org/10.1371/journal.pone.0090234

Soil Survey Staff, 2022. Web Soil Survey.

Soil Survey Staff, 2014. Keys to Soil Taxonomy. USDA-NRCS 12.

Sood, M., Kapoor, D., Kumar, V., Sheteiwy, M.S., Ramakrishnan, M., Landi, M., Araniti, F., Sharma, A., 2020. Trichoderma: The “Secrets” of a Multitalented Biocontrol Agent. Plants 9, 762. https://doi.org/10.3390/plants9060762

Spargo, J.T., Cavigelli, M.A., Mirsky, S.B., Maul, J.E., Meisinger, J.J., 2011. Mineralizable soil nitrogen and labile soil organic matter in diverse long-term cropping systems. Nutrient Cycling in Agroecosystems 90, 253–266. https://doi.org/10.1007/s10705-011-9426-4

Speirs, L.B.M., Rice, D.T.F., Petrovski, S., Seviour, R.J., 2019. The Phylogeny, Biodiversity, and Ecology of the Chloroflexi in Activated

Sludge. Frontiers in Microbiology 10, 2015. https://doi.org/10.3389/fmicb.2019.02015

Steenwerth, K., Drenovsky, R., Lambert, J.J., Kluepfel, D., Scow, K., Smart, D., 2008. Soil morphology, depth and grapevine root frequency influence microbial communities in a Pinot noir vineyard. Soil Biology and Biochemistry 40. https://doi.org/10.1016/j.soilbio.2007.04.031

Stone, M.M., DeForest, J.L., Plante, A.F., 2014. Changes in extracellular enzyme activity and microbial community structure with soil depth at the Luquillo Critical Zone Observatory. Soil Biology and Biochemistry Complete, 237–247. https://doi.org/10.1016/j.soilbio.2014.04.017

Stott, D., 2019. Recommended Soil Health Indicators and Associated Laboratory Procedures (Technical Note No. 450– 03), Soil Health. USDA-NRCS.

Teasdale, J.R., Coffman, C.B., Mangum, R.W., 2007. Potential Long-Term Benefits of No-Tillage and Organic Cropping Systems for Grain Production and Soil Improvement. Agronomy Journal 99, 1297–1305. https://doi.org/10.2134/agronj2006.0362

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., Piepenbring, M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A.L., Nilsson, R.H., Morgado, L.N., Mayor, J., May, T.W., Majuakim, L., Lodge, D.J., Lee, S.S., Larsson, K.-H., Kohout, P., Hosaka, K., Hiiesalu, I., Henkel, T.W., Harend, H., Guo, L., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan, R., Dearnaley, J., Kesel, A.D., Dang, T., Chen, X., Buegger, F., Brearley, F.Q., Bonito, G., Anslan, S., Abell, S., Abarenkov, K., 2014. Global diversity and geography of soil fungi. Science. https://doi.org/10.1126/science.1256688

Treseder, K.K., Lennon, J.T., 2015. Fungal Traits That Drive Ecosystem Dynamics on Land. Microbiology and Molecular Biology Reviews 79, 243–262. https://doi.org/10.1128/MMBR.00001-15

Uehling, J.K., Entler, M.R., Meredith, H.R., Millet, L.J., Timm, C.M., Aufrecht, J.A., Bonito, G.M., Engle, N.L., Labbé, J.L., Doktycz, M.J., Retterer, S.T., Spatafora, J.W., Stajich, J.E., Tschaplinski, T.J., Vilgalys, R.J., 2019. Microfluidics and Metabolomics Reveal Symbiotic Bacterial–Fungal Interactions Between Mortierella elongata and Burkholderia Include Metabolite Exchange. Frontiers in Microbiology 10, 2163. https://doi.org/10.3389/fmicb.2019.02163

White, K.E., Cavigelli, M.A., Conklin, A.E., Rasmann, C., 2019. Economic Performance of Long-term Organic and Conventional Crop Rotations in the Mid-Atlantic. Agronomy Journal 111, 1358–1370. https://doi.org/10.2134/agronj2018.09.0604

Wolf, A.B., Vos, M., Boer, W. de, Kowalchuk, G.A., 2013. Impact of Matric Potential and Pore Size Distribution on Growth Dynamics of Filamentous and Non-Filamentous Soil Bacteria. PLOS ONE 8, e83661. https://doi.org/10.1371/journal.pone.0083661

Wu, T., Milner, H., Díaz-Pérez, J.C., Ji, P., 2015. Effects of soil management practices on soil microbial communities and development of southern blight in vegetable production. Applied Soil Ecology 91, 58–67. https://doi.org/10.1016/j.apsoil.2015.02.011

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R., 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Micro 12, 635–645. https://doi.org/10.1038/nrmicro3330

Zeng, J., Liu, X., Song, L., Lin, X., Zhang, H., Shen, C., Chu, H., 2016. Nitrogen fertilization directly affects soil bacterial diversity and indirectly affects bacterial community composition. Soil Biology and Biochemistry 92, 41–49. https://doi.org/10.1016/j.soilbio.2015.09.018

Zuber, S.M., Villamil, M.B., 2016. Meta-analysis approach to assess effect of tillage on microbial biomass and enzyme activities. Soil Biology and Biochemistry 97, 176–187. https://doi.org/10.1016/j.soilbio.2016.03.011

Supplemental Information

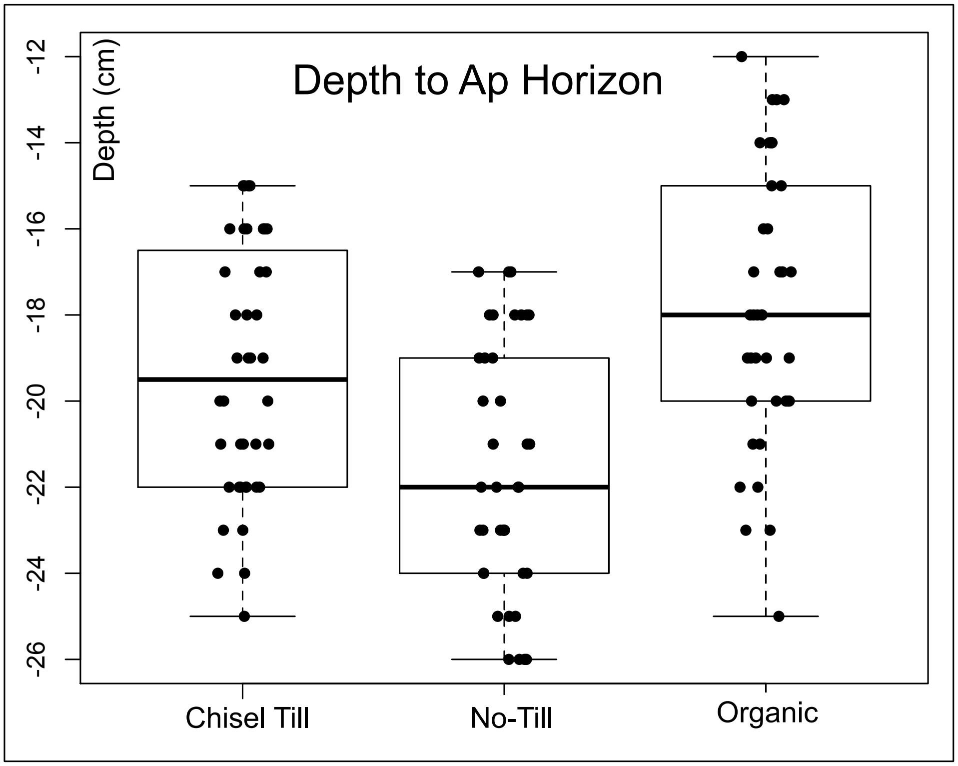

Figure S4.1: Box plot showing the depth of the Ap horizon as determined by abrupt color change within individual cores taken from soils under CT, NT, or ORG management. For each management regime, four replicate field plots were sampled. Within each plot, 9 cores (100 cm deep, 2.5 cm diameter) were removed.

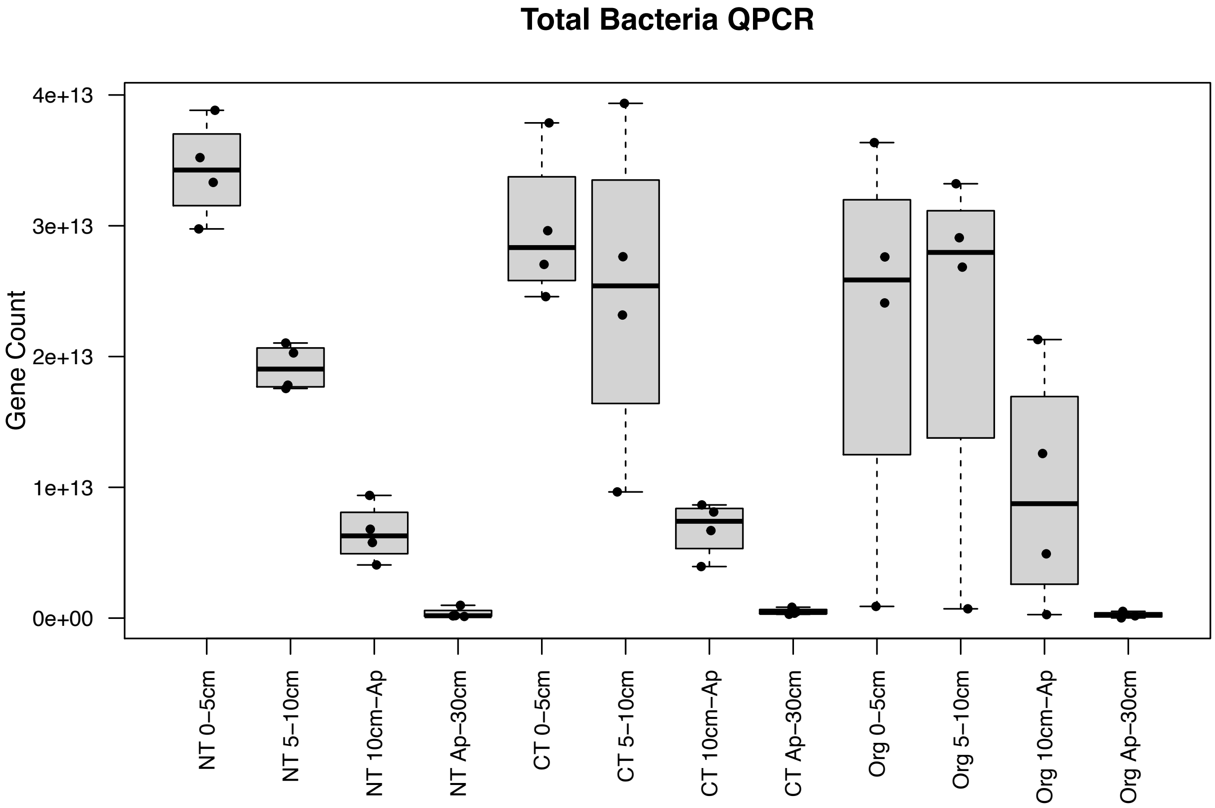

Figure S4.2: Boxplot of 16S gene abundance including the organic core that had anomalously low abundance.

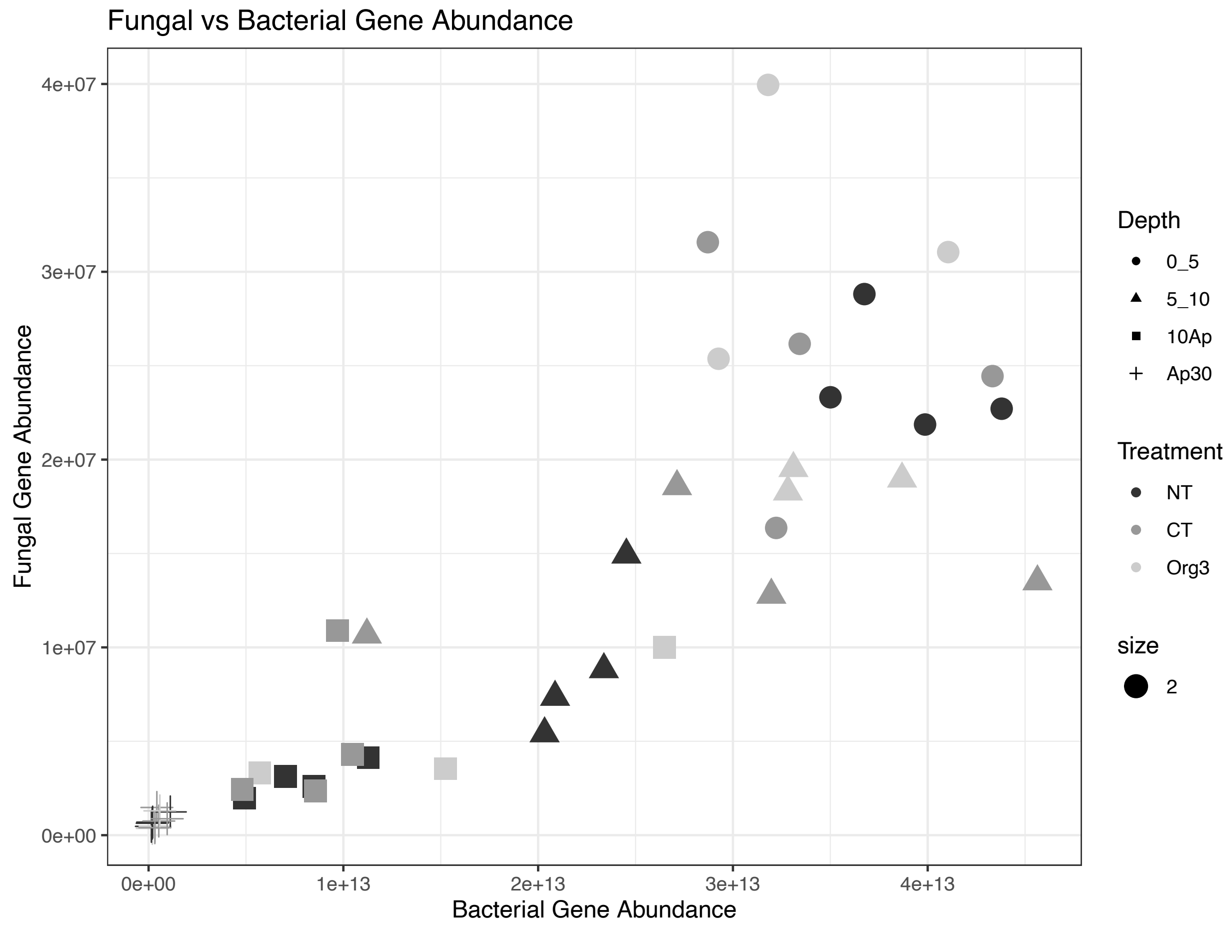

Figure S4.3: Scatterplot of ITS gene abundance versus 16S gene abundance.

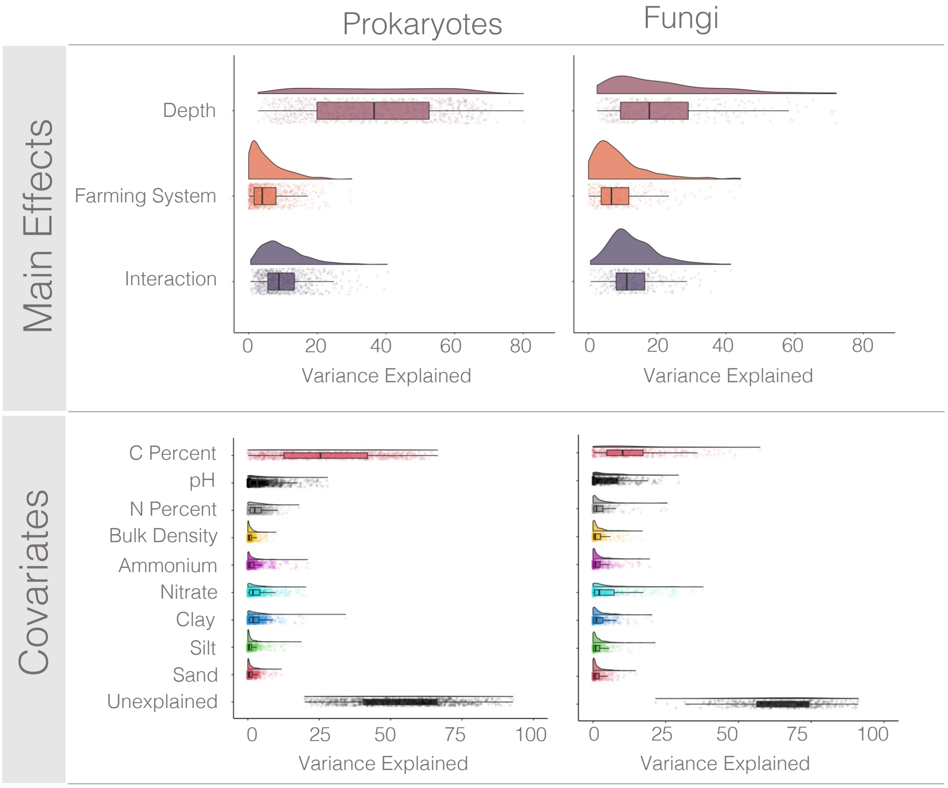

Figure S4.4: Taxon-wise variance explained by experimental factors, and edaphic covariates for prokaryotes and fungi. CSV data files with tabulated coefficients and variance explained from each taxon are available at github.com/djeppschmidt/FSP1/Analysis/ under the file names Fungi_modeled_abundance.csv (Experimental treatment coefficients and variance explained); Fungi_modeledCovariates.csv (edaphic covariates); Bacterial_modeled_abundance.csv (Experimental treatment and variance explained); Bacteria_modeledCovariates.csv (edaphic covariates).

Submitted to Applied Soil Ecology

Chapter 5: Components of N-fixation system response to glyphosate are un-coupled in Agroecosystems

Dietrich Epp Schmidt[1]., Michel A Cavigelli[2], Jude E Maul[2], Harry H Schomberg[2], Stephanie A Yarwood[1]

University of Maryland, College Park[1]
USDA-ARS, Beltsville[2]

Abstract

Glyphosate, introduced 50 years ago, is one of the most widely used herbicides. There are lingering concerns over the potential environmental impacts of glyphosate, prompting many studies. Many of these studies have isolated specific types of glyphosate effects – such as on N-fixation rate, or microbial community dynamics – but have not examined how glyphosate effects behave in relationship to one another. In this study, we were interested in the underlying ecological mechanisms associated with effects of glyphosate on N fixing components of the agroecosystem; we evaluated the effect of glyphosate on biological N-fixation in soybean, soil microbial community dynamics in the rhizosphere, and soil nifH gene quantity within and among four types of crop management systems. Our hypothesis was that if farm management mediates the effects of glyphosate, then the effects of glyphosate on different N-fixing components of crop management systems should have coordinated effects; when one component is impacted, the

others should be as well. We measured % N derived from the atmosphere (N_{dfa}) in soybean as a proxy for fixation, the quantity of 16S rRNA marker genes, and the nifH gene in rhizoplane (soil attached to roots) and rhizosphere (soil unattached, but under root influence) samples. We determined the response of free-living putative diazotrophs to glyphosate in the rhizoplane and rhizosphere using a combination of Quantitative-Polymerase Chain Reaction (QPCR) and 16S gene amplicon sequencing. The crop management systems included two under conventional management (NT is No-till and CT is Chisel till) and two under organic rotations (O3 is a 3-year rotation, and O6 is a 6-year rotation). We found that glyphosate led to changes in some components, but that these effects were likely governed by separate mechanisms and did not behave in a coordinated manner. Glyphosate caused a decrease in N_{dfa} under CT; the effect of glyphosate on N_{dfa} across systems was weakly negatively correlated to the amount of soil organic matter in the crop management systems. The 16S rRNA gene abundance was not affected by glyphosate. Glyphosate caused a 217% increase in nifH gene abundance in the rhizosphere, but this effect was not affected by crop management system. Glyphosate treatment affected community composition only in the rhizoplane of O6. The assemblage of taxa that responded to glyphosate in each system differed and it is possible that dynamics in the population of 16S sequence variants is decoupled from the selective pressure of glyphosate treatment in soils.

5.1 Introduction

Glyphosate is a widely used herbicide that inhibits the shikimate biosynthesis pathway by inhibiting 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) (Steinrücken and Amrhein, 1980). EPSPS is found in plants and bacteria (Schönbrunn et al., 2001). The shikimate biosynthesis pathway is responsible for producing aromatic amino acids, including phenylalanine, tyrosine, and tryptophan. In turn, these amino acids are necessary for plant growth and immune function (Tzin and Galili, 2010). Glyphosate tolerance has been engineered into several crops, including soybean, by introducing a bacterial EPSPS gene resistant to glyphosate. While this new enzyme allows soybean to tolerate glyphosate treatment, glyphosate application may still impact important physiological functions of the engineered soybean. For example, nitrogen fixation in soybean is a product of symbiotic bacteria that live in root nodules (Delves et al., 1986). These bacteria may be susceptible to glyphosate (Rainio et al., 2021). When glyphosate is applied to plants as a foliar spray, it is translocated throughout the plant and is detectable in root exudates within a day of foliar application (Coupland and Caseley, 1979; Laitinen et al., 2007). Therefore, the ability of soybean to fix nitrogen depends both on the susceptibility of the plant, and the susceptibility of specific members of their microbiome (bacterial symbionts) to glyphosate. Studies on bacteria show a wide range of resistance, tolerance and susceptibility to glyphosate; and the EPSPS enzyme appears to have isomers within and among Kingdoms that vary in its response to glyphosate (Grube et al., 2019; Hao et al., 2019). At recommended application rates, glyphosate has shown a small impact on biological N fixation in soybean, but can have a large impact on biological N fixation at higher application rates (Zablotowicz and Reddy, 2007; Zobiole et al., 2011). Even low glyphosate application rates have been documented to substantially decrease nodule number and mass in glyphosate resistant cultivars (Zobiole et al., 2012), indicating that impacts on the plant-microbiome interactions may play a role in the glyphosate effect on biological N fixation.

The effect of glyphosate on N fixation is mediated by water stress in a short window after application (Zablotowicz and Reddy, 2007), indicating that crop management system conditions can be important in mediating the interaction between glyphosate, soybean, and their microbiome. There are direct and indirect mechanisms by which glyphosate may impact the rhizosphere microbiome: glyphosate application has been shown to alter the root exudates of soybean, primarily by increasing the rate of amino acid exudation, and slightly increasing the amount of labile carbohydrates exuded (Kremer et al., 2005). Therefore, the rhizosphere community of glyphosate resistant plants may be impacted indirectly via the altered composition of root exudates; or directly when glyphosate is excreted in root exudates, potentially impacting glyphosate sensitive bacteria (Coupland and Caseley, 1979; Laitinen et al., 2007).

Conceptual models of the rhizosphere indicate that root physiology imposes strong selective gradients on the microbes across very short distances (Kuzyakov and Blagodatskaya, 2015; Kuzyakov and Razavi, 2019) and data support the notion that different processes drive the assembly of bacterial communities in different components of the rhizosphere niche-space (Kuzyakov and Razavi, 2019; Reinhold et al., 2015; Vetterlein et al., 2020). So, it is reasonable to conjecture that an experimental treatment that is mediated by the plant root might present differently depending on the rhizosphere component that is sampled. Studies of the rhizosphere tend to find low (Schlatter et al., 2017) or no (Kepler et al., 2019) effect of glyphosate on soil community composition when it is applied as a foliar spray at recommended application levels. Studies also have found inconsistent impacts of glyphosate on the activity of diverse microbially-derived soil enzymes (Lupwayi and Blackshaw, 2012) and context-dependent effects on total soil microbial biomass (Nguyen et al., 2016). However, operational definitions of “rhizosphere” soil may differ across studies (Reinhold et al., 2015) and studies of glyphosate in the rhizosphere have generally restricted sampling to only small particles of soil that are strongly adhered to the root-i.e. the rhizoplane compartment-to represent the rhizosphere effect (Kepler et al., 2019; Lupwayi and Blackshaw, 2012; Schlatter et al., 2017; Zobiole et al., 2010). For this study, we distinguish between rhizoplane and rhizosphere soil. We define rhizoplane soil as the fraction of the total rhizosphere that is tightly adhered to the root surface. We acknowledge that more generally the term “rhizosphere” is meant to capture the total soil volume under root influence; however, for convenience, in the remainder of this paper we will use “rhizosphere” to refer to the fraction of soil that is under root influence but that is not rhizoplane (i.e., soil not tightly adhered to the root).

Our goal in this research was to elucidate ecological mechanisms that may cause the effects of glyphosate to differ depending on the farm management. Thus, we planted genetically modified soybean, and applied glyphosate within the contexts of several different farming systems – some of which were under organic management. While both the soybean cultivar and the glyphosate treatments are not normally a component of organic farming systems, their inclusion in this study provides us with a broader range of contexts to understand the impacts and mechanisms of the effect of glyphosate on the ecosystem components involved in the nitrogen fixation function. We hypothesized that different components of the nitrogen fixation system (plant N derived from the atmosphere, soil nifH gene abundance, and soil diazotroph composition) would respond to glyphosate treatment in a coordinated manner across cropping systems. For example, cropping systems that buffer the effect of glyphosate on % N derived from the atmosphere in the soybean biomass will also buffer the effect of glyphosate on microbial community composition and nifH gene abundance; or vice versa – that when glyphosate has a large impact on diazotroph community composition, it will also have an impact on nifH gene abundance and on the percent plant N derived from the atmosphere.

5.2 Materials and Methods

Experimental Design

This study was conducted in 2014 at the US Department of Agriculture, Agricultural Research Service (USDA-ARS) Sustainable Agricultural Systems Laboratory in Beltsville, MD. This study, the Farming Systems Project, is managed as part of the Lower Chesapeake Bay Long-Term Agroecological Research site and includes cropping systems typical to the mid-Atlantic region of the USA. The FSP is arranged in a split-split-plot design where the main plots are crop management systems replicated four times, the first split is crop rotation phase, and the glyphosate experimental plots described herein were split-plots within the soybean phase of the rotation. The glyphosate experiment is a partial factorial with genotype (glyphosate resistant and glyphosate sensitive isolines) and glyphosate treatment (+/-) with the + treatment applied only in the glyphosate resistant isoline. Hereafter, we will use “plot” to refer to each replicate of the soybean crop for each management system (N = 4); and “microplot” to refer to the genotype X glyphosate treatment split-plots within each replicate of the soybean crop.

The current study includes four of the five FSP cropping systems: no-till management (NT) and chisel-till management (CT), each three-year crop rotations, and two organic systems a three-year rotation (O3) and six-year rotation (O6). CT and NT follow a rotation of corn (*Zea mays* L.), rye cover crop (*Secale cereale* L.), soybean (*Glycine max* (L.) Merr.), and winter wheat (*Triticum aestivum* L.)/soybean. Soybean planted after wheat harvest is referred to as a double-cropped soybean. O3 replaces the double-cropped soybean phase with a hairy vetch (*Vicia villosa,* Roth) crop. O6 replaces the vetch in O3 with a 3-year alfalfa (*Medicago sativa*, L.) crop. Additional management details can be found in (Cavigelli et al., 2008; Kepler et al., 2019). The conventional systems receive mineral nitrogen, phosphate and potassium fertilizers, and herbicides. NT and CT differ in that a chisel plow is used in CT. The organic systems use poultry litter, cover crop biomass, and K_2SO_4 (potassium sulfate) as nutrient inputs, and tillage for weed suppression.

The experiment was conducted in the 2014 full-season soybean phase of the rotation in the four cropping systems. Glyphosate treatments were applied in 4.6 m x 6.1 m microplots adjacent to each other within each plot. The microplot treatments were genotype (Allen is a glyphosate-resistant genotype, USG5601 is a glyphosate-sensitive isoline), and treatments with or without glyphosate (Allen (+), Allen (-)). The glyphosate-sensitive genotype did not receive glyphosate. Thus, the treatments were genotype-negative control (UGS 5601, no glyphosate), genotype-positive control (Allen (-)), and glyphosate treatment (Allen (+)). These cultivars were planted at a rate of 516,400 seeds ha^{-1} in May. Each of the non-glyphosate treated subplots were kept weed free with hand-hoeing as necessary. Glyphosate was applied to the (+) treatment 4 weeks after soybean planting. Soil samples for DNA analysis were collected in each subplot one day prior to the glyphosate application and 20 days after the glyphosate application.

Soil Conditions

Soils in this study are silt loam Ultisols that consist primarily of Christiana, Keyport, Matapeake, and Mattapex soil map units. We collected soil samples to a depth of 15 cm from each plot in November 2014 for analysis at the Cornell Soil Health Lab (Moebius-Clune et al., 2016). The Comprehensive Assessment of Soil Health provides a number of soil health metrics, including: organic matter (OM), pH, and extractable nutrients (Moebius et al., 2007; Moebius-Clune et al., 2016). Briefly: percent OM is determined by mass loss on ignition following (Broadbent, 1965). Soil pH was measured using a 2:1 solution of water to soil by volume. Plant-available nutrients were extracted with an ammonium acetate and acetic acid solution buffered at a pH of 4.8 (Moebius-Clune et al., 2016).

Soil samples for inorganic N extraction were taken at the same time as those taken for soil health from 0-15 cm and 15-30 cm depth increments. They were sieved through a 2-mm sieve and stored in sealed 50 mL Falcon tubes in a freezer until the extraction was conducted. Moisture content for each sample was determined by drying a sub-sample at 100 C for 2 days. Soil extraction was conducted following (Bremner and Keeney, 1965). Briefly, samples were removed from the freezer and allowed to warm to room temperature over the course of 4 h. 3.0 g of soil and 30 mL of 1 M KCl were added to a 50 mL Falcon tube, and placed on a reciprocating shaker for 1 hr. After one hour, soils were removed and allowed to settle for 10 min. 10 mL of supernatant from each sample were filtered using a vacuum manifold and Millipore 0.45 um custom cut filter paper. Filtered samples were collected in culture tubes that were capped and frozen until they could be analyzed. The filtrate was analyzed for NO_3–N (copper-cadmium reduction method G-200-97 Rev. 6) and NH_3–N (salicylate method G-102-93 Rev. 7) using the SEAL AutoAnalyzer 3 HR (SEAL Analytical Inc. Mequon, Wisconsin 53092).

Biological Nitrogen Fixation

We used natural isotope ratios to measure the rate of biological nitrogen fixation. In addition to the two soybean varieties listed above we also planted a non-nodulating variety (Williams 82 variant) in microplots within the soybean phase of each cropping system. The non-nodulating variety was used as a background control to establish the ^{15}N enrichment (relative to the atmosphere) of the soil N pool in each plot. The natural abundance δ^{15}N was measured by placing samples in a PN150 autosampler (Costech, CA, USA), which feeds into a Carlo Erba NC2500 (Italy) elemental analyzer for combustion analysis, which was then isotopically analyzed using a Thermo Scientific Delta V Advantage Isotope Ratio Mass Spectrometers (IRMS).

The proportion of N fixed from the atmosphere was calculated as follows:

$$Ndfa = \left(\frac{\delta^{15}N_{ref} - \delta^{15}N_{fix}}{\delta^{15}N_{ref-B}}\right) * 100,$$

where $\delta^{15}N_{ref}$ for each sample is the value for the non-nodulating variety (NN82) sampled from the same plot; $\delta^{15}N_{fix}$ is the value for Allen; and B is the literature value for N enrichment in soybean when only atmospheric N is available (B = - 1.75; Balboa and Ciampitti (2020).

To determine if there was an effect of glyphosate treatment on N_{dfa} within each cropping system, we conducted a one-way ANOVA using a mixed model of N_{dfa} value against glyphosate treatment, with crop system replicate plot ID as a random intercept to account for spatial coupling of treatments within each plot. To determine if the effect size of glyphosate treatment was predicted by soil OM, we first calculated the effect size in each plot by subtracting the untreated N_{dfa} from the treated N_{dfa} in each plot. Then we tested the correlation between the effect size and soil OM in each plot with both a linear regression and a Pearson correlation.

Soil Sampling for DNA and Sequencing

Six plants were excavated from each microplot one day prior to glyphosate treatment in each microplot, and again 14 d after glyphosate application. Glyphosate was applied 4 weeks after planting (V3-V4 growth stage). For each plant a 15 cm deep soil monolith of 30 cm diameter, centered on the plant, was excavated using surface-sterilized sharpshooter shovels. The soil monoliths, including the plant, were placed on a 2-mm sieve, which we shook gently. Soil that was removed in this manner was separated as the “rhizosphere” soils. Remaining soil that was adhered to the roots was removed using a camel hairbrush and collected as the “rhizoplane” sample. Rhizoplane and rhizosphere soils from the six plants were combined respectively to make representative “rhizoplane” and “rhizosphere” samples for each microplot. Five grams of soil were sampled from each of the rhizoplane and rhizosphere samples and preserved in separate 15 mL Falcon tubes that contained 19 mL of Mo Bio LifeGuard nucleic acid preservation solution. The tubes were agitated to thoroughly mix the preservation solution with the soil sample; then the samples were placed in a -80 C freezer until DNA was extracted.

Soil samples were thawed and 800 mL of soil slurry were used for DNA isolation using a PowerSoil htp 96-well soil DNA isolation kit (Mo Bio Laboratories, inc. Solana Beach, CA), following the kit instructions. The resulting DNA solution was quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Pittsburgh, PA). Bacterial 16S rRNA gene amplicons were generated with the primers f515 and r806 (Caporaso et al., 2012). 16S gene amplicon indexing was conducted following the Illumina protocol library preparation manual (part number 1504423; rev B; Illumina Inc.). For each 96-well plate, two microliter aliquots of each sample were pooled to construct a library for Illumina sequencing. For each library, 100 mL of 10 nM library solution were frozen and shipped on dry ice to the Center for Genome Research and Bioinformatics (CGRB), Oregon State University, Corvallis, OR. Libraries were sequenced on an Illumina Miseq system.

Bacterial quantitative PCR was done using f515 and r806 primers, using 2.5 ng DNA input. The nifH gene QPCR was done using nifH-F and nifH-R primers (Rosch et al., 2002). The QPCR reaction conditions were: heat inactivation at 95 C for 2 minutes, followed by 40 cycles of 95 C for 5 seconds, 65 C for 10 seconds, and 72 C for 10 seconds. QPCR runs were accepted if the efficiency fell between 95-105 %. QPCR reactions were run on a StepOne Real-Time PCR System (ThermoFisher Scientific).

Bioinformatics and Community Analysis

Initial quality control and adapter trimming were conducted at the CGRB. Subsequently, forward and reverse primers were removed using Cutadapt (version 1.8.3), and sequences were quality trimmed (-q 22). Reads were discarded if they lacked a primer sequence or were shorter than 75 bp before trimming. Additional quality filtering using the DADA2 package (Callahan et al., 2016) included removing any sequences with an expected error rate >2, and any sequences containing “N” values (unreadable bases). Then, DADA2 was further used to dereplicate, denoise, merge forward and reverse reads, and remove chimeras. Taxonomy was assigned to the 16S reads using the SILVA database (Quast et al., 2013; Yilmaz et al., 2014), using the RDP classifier as implemented in DADA2. “Quantitative Sequencing” normalization was implemented by first converting the sequence counts into relative abundance, then scaling the counts in each sample by the QPCR determined sample-wise gene copy number as described in (Jian et al., 2020). Finally, sequences were removed if they were not assigned to a genus known to contain members that fix nitrogen. We used the recommended list of diazotroph genera offered by (Sansupa et al., 2021). This is an augmented version of diazotrophs from FAPROTAX (Louca et al., 2016b), that additionally includes taxa that are known N-fixing bacteria in soils. Jaccard dissimilarity index was calculated, and detrended correspondence analysis was applied to the Jaccard distance using the vegan package v 2.5-7. PERMANOVA was implemented using the adonis2 function from vegan to determine the effect of the experimental treatments, using the plot ID to stratify randomization. We also conducted a PERMANOVA on each of the soil compartments independently (Rhizoplane vs Rhizosphere); and to evaluate the relative effect of the abundance and occurrence of taxa on the PERMANOVA results, we also transformed the community data into presence-absence and ran the PERMANOVA analysis.

We reasoned that if glyphosate alters the microbial community, then the Allen (+) and Allen (-) treatments within each plot should become less similar after treatment. To determine if glyphosate caused a divergence, we calculated pairwise distance between the Allen (+) and Allen (-) treatments within each plot before and after the glyphosate application. We conducted ANOVAs to determine if there was a significant difference in Jaccard distance in each cropping system between the pre-treatment and post-treatment sampling dates. Only an increase in distance was considered evidence of a glyphosate effect because if glyphosate impacts the community composition, then the communities should become less similar through time.

Statistics were conducted in R v4.1.2 (R Core Team, 2020). We conducted a principal components analysis (PCA) to determine the major differences in soil health metrics among cropping systems using the R package factoextra v1.0.7. We used the lmer function from the lme4 v1.1-27.1 package(Bates et al., 2015) to construct our mixed effects models. We used linear mixed effects models to separately determine the effects of sampling date, cropping system, soil zone, glyphosate application, and genotype on the abundances of 16S and nifH gene copy numbers. The dataset is unbalanced with respect to both genotype and glyphosate treatment. To account for this, for the genotype models, we excluded the glyphosate treated samples such that only Allen (-) was tested against the isoline, USG5601. Similarly, we excluded USG5601 (not sprayed with glyphosate) from the statistical test for glyphosate effects such that only Allen (+) and Allen (-) were considered. To control spatial and temporal variability, in all mixed models we included plot ID as a random intercept, and sampling date as random slope. For the glyphosate model, we only tested an interaction between glyphosate and sampling date because 1) there was no glyphosate applied in the first sampling date, and 2) we expected that gene abundances should naturally change through time even without glyphosate effects – our interest was in determining how glyphosate impacted these existing dynamics.

We tested the response of individual bacterial amplicon sequence variants (ASVs) to glyphosate application using a mixed effects model with cropping system, genotype, soil zone, and a sampling date – glyphosate application interaction as explanatory variables. Plot ID was set as the random intercept for the model, and sampling date was set as the random slope. The sampling date – glyphosate application interaction term was used to determine if the ASV responded to glyphosate application at an alpha of 0.1.

5.3 Results

Soil Properties

Soil factors clustered distinctly between conventionally managed (CT and NT) and organically managed (O3 and O6) cropping systems in a PCA biplot (Figure 5.1). Aggregate stability was lowest in O3 (ANOVA: F = 3.70, P = 0.043) and highest in NT, but did not differ between NT, CT, or O6 (TukeyHSD, P > 0.31). Soil OM ranged from 1.85 to 3.04 % and was strongly affected by crop management system (ANOVA, F = 3.696, P = 0.043), with the greatest differences being between CT and O6 (TukeyHSD: P = 0.049). Potassium was most abundant under O3 and was lower in CT and NT than O3 and O6 (Figure S5.1; ANOVA: F = 22.98, P < 0.001). Zinc abundance closely matched potassium abundance (Figure S5.1; ANOVA: F = 4.44, P = 0.026). Soil OM was weakly correlated with soil pH (Pearson R = 0.51), but soil pH did not differ among cropping systems (ANOVA: F = 1.51, P = 0.262). In the top 15 cm, available nitrate-N was lowest in NT and highest in O6 (Figure 5.2; ANOVA: F = 2.78, P = 0.063). Nitrate-N was particularly elevated at the 15 to 30 cm depth in O6 relative to other cropping systems (Figure

5.2B; ANOVA: F = 6.00, P = 0.0034). There were no cropping system effects on ammonium-N at either depth interval (Figure 5.2C&D; ANOVA: F < 1.353, P > 0.28).

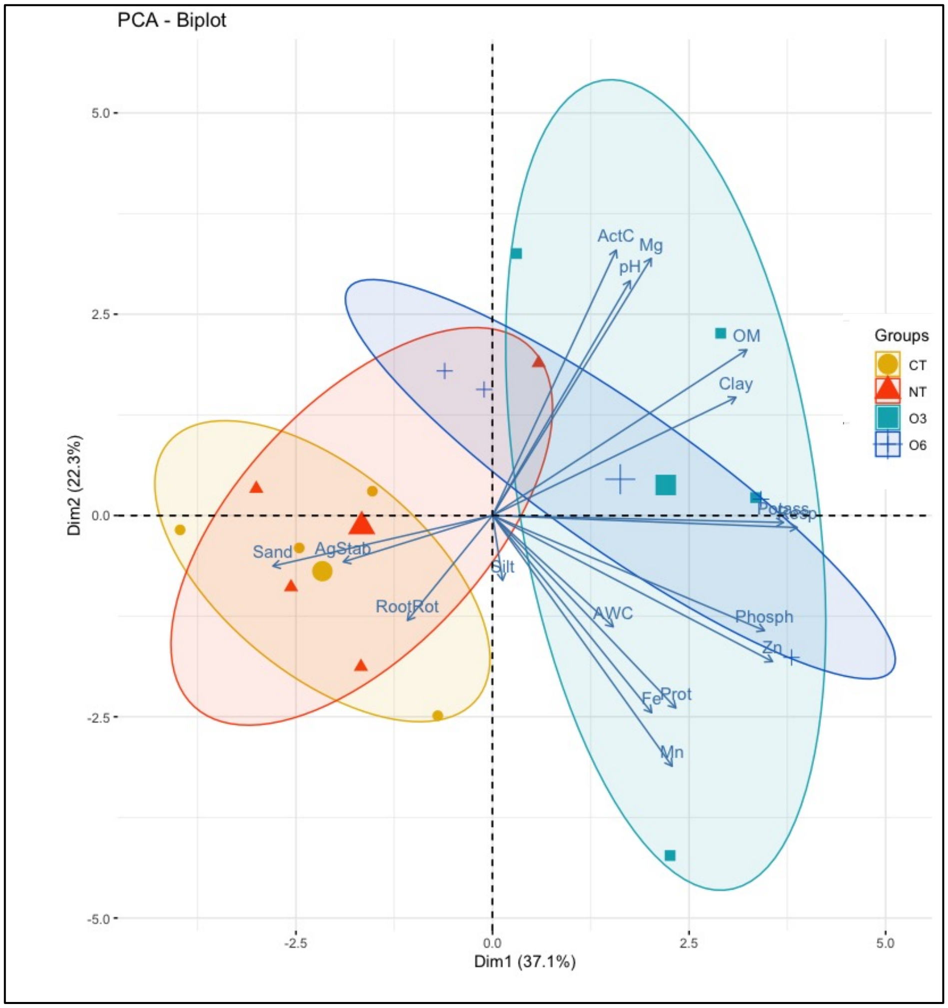

Figure 5.1: PCA biplot of soil health factors from organic (three-year (O3), six-year (O6) rotations), conventional tilled (CT), and conventional no-till (NT) systems. Axis 1 explained 37.1% of variation; Axis 2 explained 22.3% of variation of soil health factors.

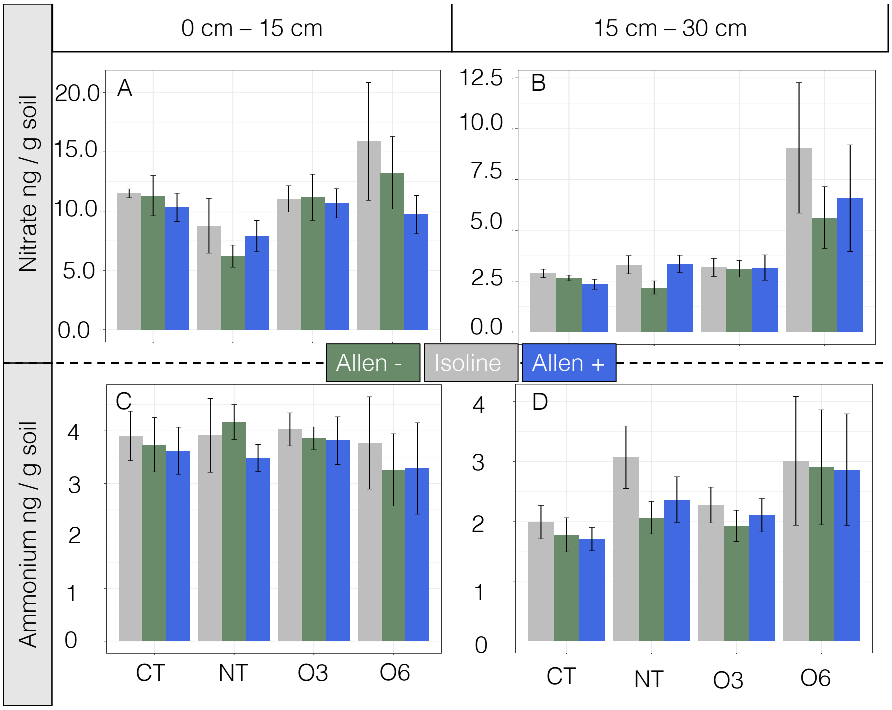

Figure 5.2: Soil mineralizable nitrate-N and ammonium-N at 0-15 and 15-30 cm depth intervals for Allen soybean variety treated (+) or not treated (-) with glyphosate and for a non-N-fixing soybean isoline not treated with glyphosate planted in four different cropping systems. The cropping systems included two organic systems (three-year (O3), six-year (O6) rotations), conventional tilled (CT), and conventional no-till (NT) systems.

N Fixation

Soybean % N derived from the atmosphere was highest in NT and lowest in CT (Figure 5.3; ANOVA: F = 5.76, P = 0.011). Our statistical model for all cropping systems found no support for an interaction between cropping system and glyphosate treatment on N_{dfa} (ANOVA: F = 0.69, P = 0.17) and weak support for a main effect of glyphosate treatment on N_{dfa} (ANOVA: F = 4.56 , P = 0.054); however, only NT had a main effect of glyphosate when cropping systems were analyzed individually (ANOVA: F = 44.15, P = 0.007; CT ANOVA: F = 2.81, P = 0.19; O3 & O6 ANOVA: F < 0.56, P > 0.51). We calculated pairwise differences between N_{dfa} of treated and untreated samples (delta N_{dfa}; Figure 5.4B). Soil OM was weakly correlated with delta N_{dfa} (Pearson R = 0.43; ANOVA: F = 3.13, P = 0.099), but cropping system was not a good predictor of delta N_{dfa} (ANOVA: F = 1.98, P = 0.171). Thus, while average percent soil OM increased in order from CT, NT, O3, O6, the decrease in N_{dfa} values between treated and untreated decreased in the same order (Figure 5.4).

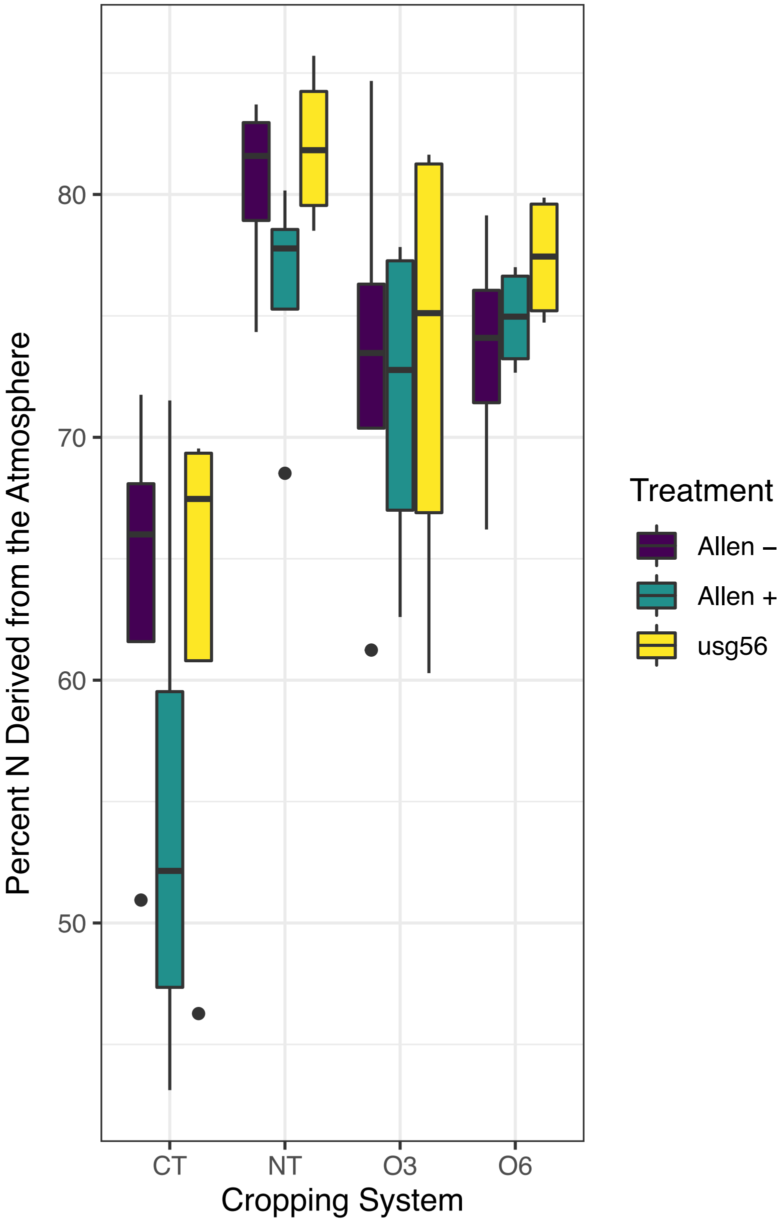

Figure 5.3: Percent N derived from the atmosphere (N_{dfa}) for Allen soybean variety treated (+) or not treated (-) with glyphosate and for a non-N-fixing soybean isoline not treated with glyphosate planted in four different cropping systems. The cropping systems included two organic systems (three-year (O3), six-year (O6) rotations), conventional tilled (CT), and conventional no-till (NT) systems.

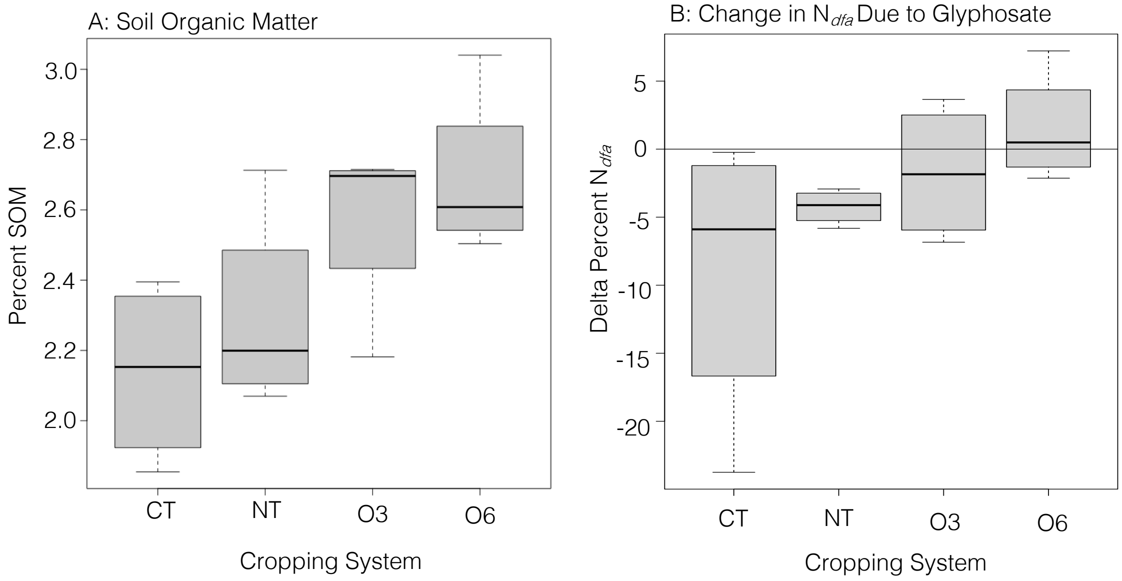

Figure 5.4: A) soil organic matter in four cropping system. B) effect of glyphosate on N_{dfa} in four cropping systems. The effect of glyphosate was calculated by subtracting the percent N_{dfa} in soybean treated without glyphosate from soybean treated with glyphosate in adjacent microplots. Thus, a large negative value indicates that the glyphosate treatment had much less N derived from the atmosphere than the treatment not receiving glyphosate. The cropping systems included two organic systems (three-year (O3), six-year (O6) rotations), conventional tilled (CT), and conventional no-till (NT) systems.

QPCR Gene Abundance

In all cropping systems, 16S rRNA gene abundance decreased on average by 42% from pre to post sampling dates (Figure 5.5 A&B; ANOVA: F = 19.35, P < 0.001). There was a strong interaction effect between cropping system and soil zone (ANOVA: F = 3.69, P = 0.011). Rhizoplane 16S rRNA gene abundance was 25% higher in CT than in O3 (Figure 5.5; ANOVA: F = 2.65, P = 0.059), but not affected by cropping system in rhizosphere samples (ANOVA: F = 0.35, P = 0.79). Genotype did not influence 16S rRNA gene abundance in the rhizoplane (Figure 5.5: ANOVA: F = 0.095, P = 0.76), but in the rhizosphere soils there was moderate support that 16S rRNA gene abundance decreased by 31.6 % in the post sampling date under Allen compared to USG5601 (ANOVA: F = 3.77, P = 0.052). We found moderate support for a glyphosate effect; application caused a 17.7 % increase in rhizosphere 16S rRNA genes abundance (Figure 5.5; ANOVA: F = 2.70, P = 0.067), but did not impact abundance in rhizoplane soils (ANOVA: F = 0.12, P = 0.89). The abundance of the nifH gene was not affected by cropping system overall (ANOVA, F = 0.72, P = 0.55).

There was no effect of cropping system on nifH gene abundance in rhizosphere soils (ANOVA: F = 0.55, P = 0.65) or in rhizoplane soils (ANOVA: F = 1.68, P = 0.21), nor was there an effect of genotype on nifH gene abundance in rhizosphere soils (ANOVA: F = 1.21, P = 0.27) or in the rhizoplane (ANOVA: F = 2.12, P = 0.15) (Figure 5.5). There was no change in nifH gene abundance from pre to post sampling in either soil fraction (ANOVA; F < 0.53, P > 0.47). However, across all cropping systems, glyphosate treatment increased the abundance of nifH in the rhizosphere soils by 218% (Figure 5.5D; ANOVA: F = 4.99, P = 0.007), but not in the rhizoplane (ANOVA: F = 1.02, P = 0.36).

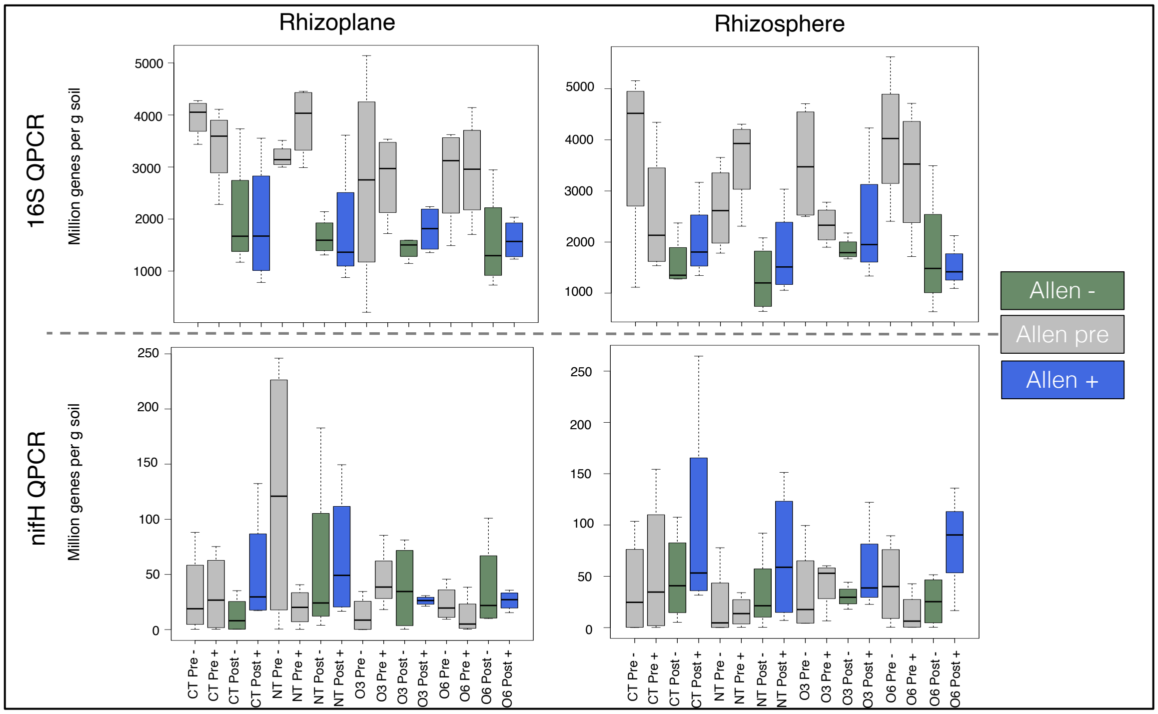

Figure 5.5: Abundance of 16S rRNA and nifH genes in rhizoplane and rhizosphere soils for Allen soybean variety treated (+) or not treated (-) with glyphosate and for a non-N-fixing soybean isoline not treated with glyphosate planted in four different cropping systems. The cropping systems included two organic systems (three year (O3), six year (O6) rotations), conventional tilled (CT), and conventional no-till (NT) systems. Grey represents Allen before treatment; green represents Allen after receiving no glyphosate; blue represents Allen 14 days after receiving glyphosate treatment.

Alpha Diversity

The detected ASV richness of putative diazotrophs differed by pre vs. post sampling (ANOVA: F = 5.068, P = 0.026), cropping system (ANOVA: F = 13.38, P < 0.001), glyphosate treatment (ANOVA: F = 3.40, P = 0.036), and showed an interaction between cropping system and glyphosate treatment (ANOVA F = 2.70, P = 0.017). O3 and O6 had the highest number of detected taxa (Figure 5.6). The greatest difference between soil zones in glyphosate response was in O6, but overall, there was no support for a system and soil zone interaction with glyphosate treatment (ANOVA, F= 0.49, P = 0.82).

Glyphosate affected the richness of detected putative diazotroph ASVs in CT (ANOVA: F =3.43, P = 0.066), O3 (ANOVA: F =4.01, P = 0.039), and O6 (ANOVA: F =5.90, P = 0.018). ASV richness increased in rhizosphere O6 and O3; and decreased in CT.

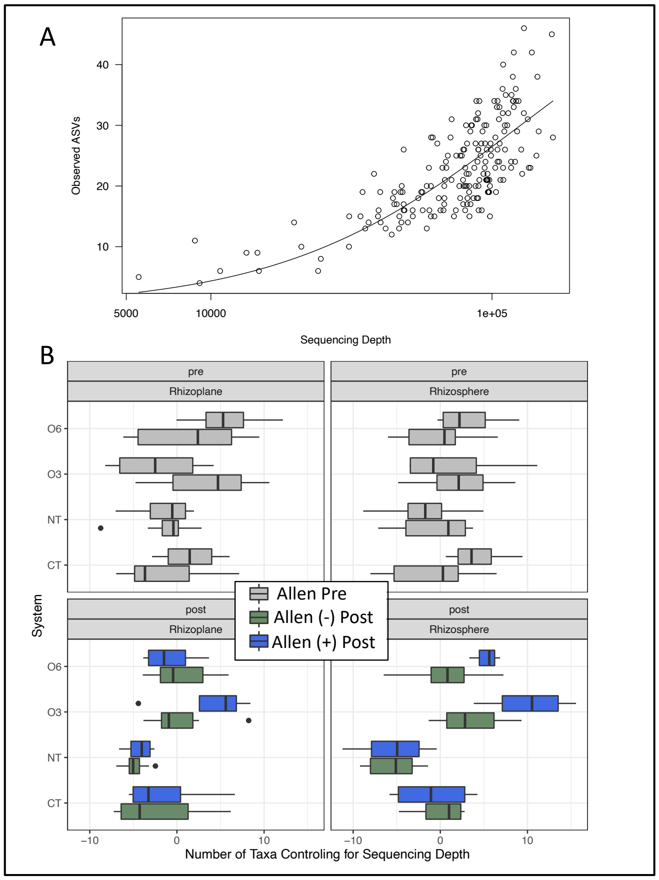

Figure 5.6: Model of Amplicon Sequence Variant (ASV) richness regressed against sequencing depth, and plot of residual ASV richness due to glyphosate treatment and sampling date in each cropping system. A residual value of zero represents the expected number of detected taxa given the sequencing depth of the sample; residual values above zero (points above the line in A) indicate the degree to which more taxa were detected than expected given the sequencing depth of the sample, and vice versa. When a treatment condition consistently has a positive or negative residual, we interpret it as having a higher or lower alpha diversity, respectively.

Community Data – Composition

Sampling date (pre vs. post glyphosate treatment) and cropping system explained the greatest amount of variation in the putative diazotroph subset of the microbial community (Table 5.1; PERMANOVA; date, system, and date-system interactions respectively: $R^2 = 0.067$, $P = 0.001$; $R^2 = 0.078$, $P = 0.001$; $R^2 = 0.028$, $P = 0.003$). The rhizoplane effect and an interaction between sampling date and genotype were each statistically significant at an alpha of 0.05 but explained 1% or less of variation (PERMANOVA; rhizoplane effect and date-genotype interaction respectively: $R^2 = 0.008$, $P = 0.022$; $R^2 = 0.001$, $P = 0.008$). Rhizosphere and rhizoplane effects were driven primarily by differences in abundance rather than occurrence of taxa (Table 5.2).

Table 5.1: PERMANOVA table evaluating the effects of experimental factors on the community composition of the putative diazotroph community subset across both depth categories (rhizoplane and rhizosphere).

	Df	SumsOfSqs	MeanSqs	F.Model	R2	Pr(>F)
Date (pre/post)	**1**	**3.133**	**3.13349**	**14.2773**	**0.0666**	**0.001**
Glyphosate (+/-)	1	0.204	0.20396	0.9293	0.00434	0.325
Cropping System (CT, NT, O3, O6)	**3**	**3.653**	**1.21776**	**5.5486**	**0.07765**	**0.001**
Genotype (Allen, USG5601)	1	0.266	0.26637	1.2137	0.00566	0.115
Compartment (rhizoplane / rhizosphere)	**1**	**0.397**	**0.39715**	**1.8096**	**0.00844**	**0.02**
Date:glyphosate	1	0.159	0.15922	0.7254	0.00338	0.66
Date:System	**3**	**1.337**	**0.44579**	**2.0312**	**0.02843**	**0.001**
Glyphosate:System	3	0.62	0.20673	0.942	0.01318	0.291
Date:Genotype	**1**	**0.461**	**0.46123**	**2.1015**	**0.0098**	**0.013**
System:Genotype	3	0.512	0.17081	0.7783	0.01089	0.621
Date:Compartment	1	0.254	0.25398	1.1572	0.0054	0.141
Glyphosate:Compartment	1	0.184	0.18421	0.8393	0.00392	0.404
System:Compartment	3	0.526	0.17518	0.7982	0.01117	0.581
Genotype:Compartment	1	0.143	0.14292	0.6512	0.00304	0.766
Date:Glyphosate:System	3	0.511	0.17021	0.7755	0.01085	0.653
Date:System:Genotype	3	0.63	0.21002	0.9569	0.01339	0.268
Date:Glyphosate:Compartment	1	0.146	0.14589	0.6647	0.0031	0.736
Date:System:Compartment	3	0.678	0.22616	1.0304	0.01442	0.193
Glyphosate:System:Compartment	3	0.453	0.1511	0.6885	0.00964	0.828
Date:Genotype:Compartment	1	0.151	0.15116	0.6887	0.00321	0.687
System:Genotype:Compartment	3	0.395	0.13177	0.6004	0.0084	0.966
Date:Glyphosate:System:Compartment	3	0.426	0.14215	0.6477	0.00906	0.913
Date:System:Genotype:Compartment	3	0.419	0.13964	0.6363	0.0089	0.919
Residuals	143	31.385	0.21947	0.6671		

Table 5.2: PERMANOVA R^2 values (P-value) for experimental factors explaining community composition of putative diazotrophs. Community data was subset into rhizoplane (Rp) and rhizosphere (Rs) categories for independent analysis; and the analysis was repeated on data that had been transformed to reflect only occurrence patterns. Bold indicates the values in the cell met an alpha of 0.05 for significance testing.

Effect	Rp Abundance	Rs Abundance	Rp Occurance	Rs Occurance
Date	**0.080 (0.001)**	**0.064 (0.001)**	**0.024 (0.001)**	**0.018 (0.001)**
glyphosate	0.007 (0.539)	0.009 (0.27)	0.010 (0.177)	0.010 (0.185)
system	**0.085 (0.001)**	**0.095 (0.003)**	0.143 (0.611)	0.142 (0.277)
genotype	0.009 (0.369)	0.009 (0.286)	0.007 (0.697)	**0.013 (0.013)**
Date:glyphosate	0.006 (0.861)	0.008 (0.452)	0.008 (0.54)	0.007 (0.675)
Date:system	**0.049 (0.004)**	**0.037 (0.046)**	0.023 (0.681)	0.028 (0.131)
glyphosate:system	0.020 (0.838)	0.026 (0.354)	0.020 (0.915)	0.022 (0.606)
Date:genotype	**0.017 (0.042)**	0.010 (0.258)	0.006 (0.928)	0.007 (0.569)
system:genotype	0.022 (0.718)	0.017 (0.965)	0.019 (0.951)	0.018 (0.964)
Date:glyphosate:system	0.019 (0.882)	0.0213 (0.686)	0.023 (0.668)	0.021 (0.803)
Date:system:genotype	0.024 (0.532)	0.0209 (0.758)	0.022 (0.748)	0.020 (0.89)
Residuals	0.6634	0.683	0.6964	0.6943

In order to control for the plot-wise variation in community composition, we calculated the Jaccard distance between treatment pairs (Allen (+) and Allen (-)) at each time point, under the assumption that if glyphosate affects community composition, then the distance among pairs should increase after the treatment. We analyzed each cropping system and soil compartment separately; in the rhizoplane we found that NT and O3 samples decreased in Jaccard distance (Figure 5.7; ANOVA: F > 3.9, P <0.088), while only O6 increased in Jaccard distance due to the glyphosate treatment (ANOVA: F = 9.60, P = 0.021). In rhizosphere samples, only CT differed (ANOVA: F = 6.77, P = 0.04), but as with the rhizoplane NT and O3 systems, the effect was a decrease in distance.

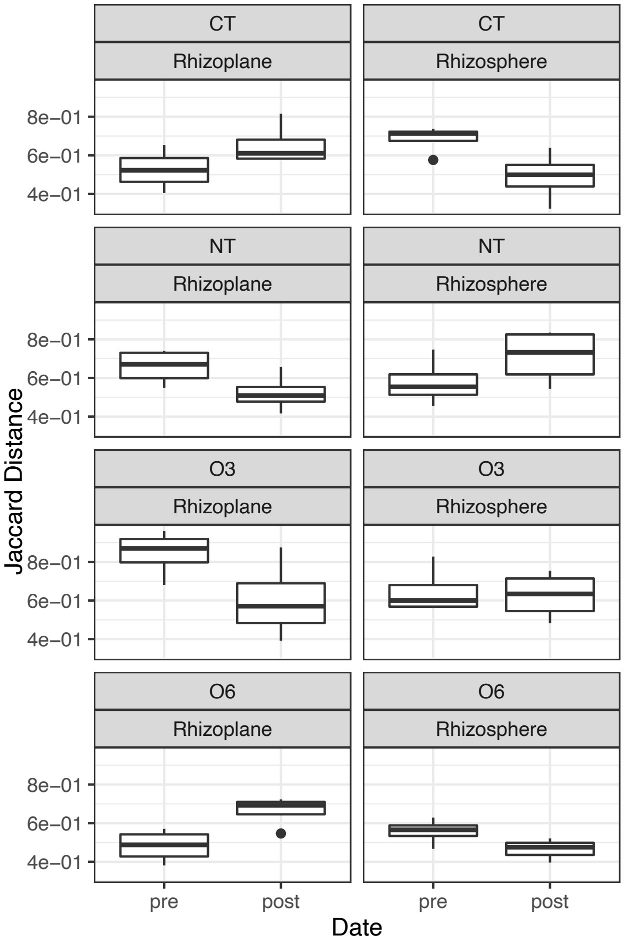

Figure 5.7: Community composition dissimilarity (Jaccard distance) due to glyphosate treatment pairs within each plot and soil compartment (Allen + / Allen -). We reasoned that if glyphosate selects against certain taxa, then it should increase the dissimilarity between treated and untreated samples; therefore, we interpret Increased distance from the pre to post treatment as a glyphosate effect on community composition; whereas a decrease in distance is not. Only O6 rhizoplane increased distance due to glyphosate at an alpha of 0.05 (ANOVA: F = 9.6, P = 0.02).

Taxon Responses

Bradyrhizobium, *Bacillus*, *Mesorhizobium*, and *Celvibrio* were more abundant in the rhizoplane than the rhizosphere (Figure 5.8A; alpha = 0.1). *Bradyrhizobium* was more abundant in response to glyphosate in the rhizosphere in all cropping systems, while *Mesorhizobium* and *Sphingomonas* ASVs differed in their response to glyphosate in the rhizosphere (Figure 8B;

alpha = 0.1). However, when cropping systems were analyzed separately, different assemblages of putative diazotroph taxa responded in each cropping system (Figure 9; alpha = 0.1)

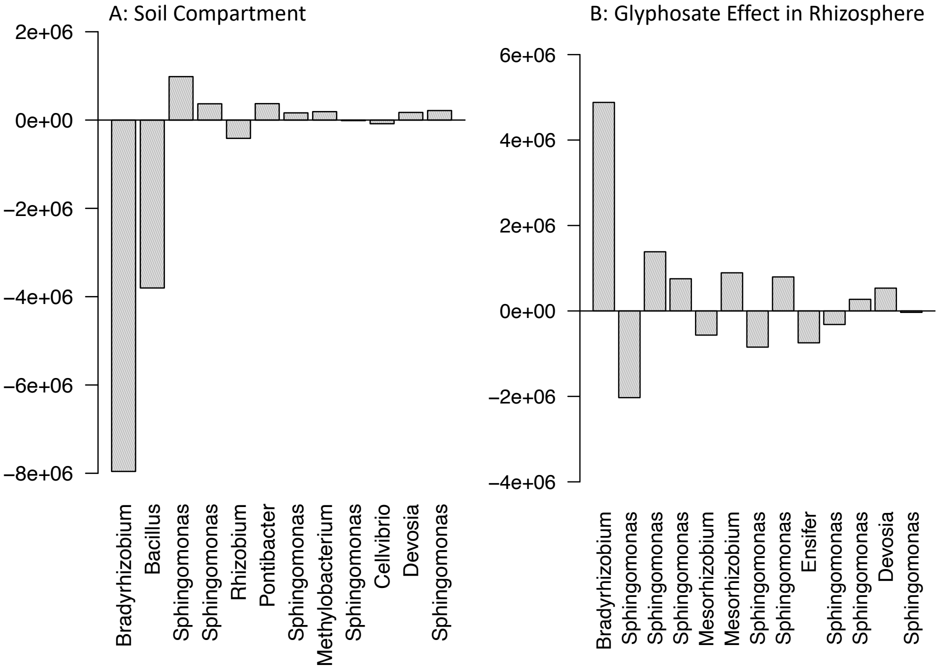

Figure 5.8: Model coefficients of putative diazotroph ASVs that respond at an alpha of 0.1 due to A) soil compartment and B) glyphosate effects within the rhizosphere. Positive values in A indicate the ASV is more abundant in the rhizosphere than the rhizoplane (and vice versa). Similarly, positive values in B indicate that the ASV is more abundant in the glyphosate treated rhizosphere soils than the non-treated rhizosphere soils.

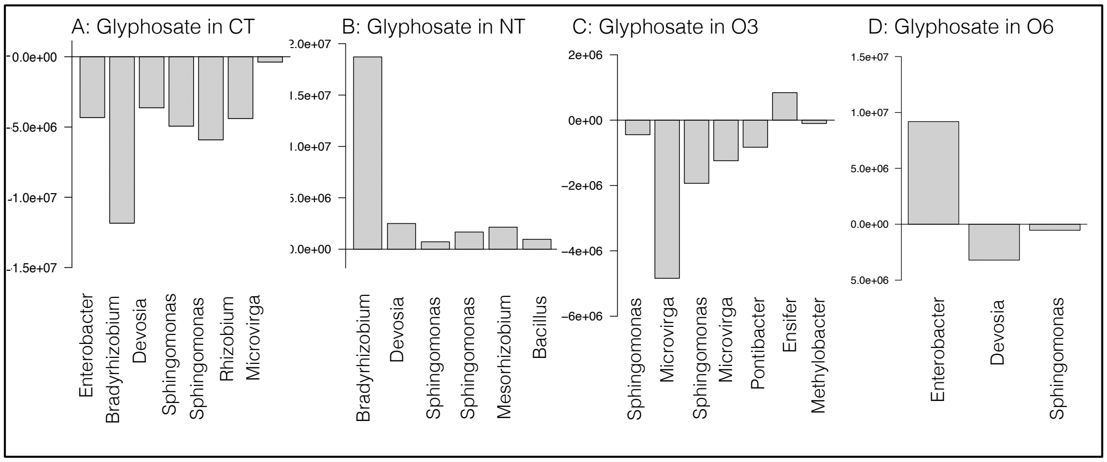

Figure 5.9: putative diazotroph ASVs in the rhizosphere compartment of each cropping system that respond to glyphosate treatment at an alpha of 0.1. Positive values indicate that the ASV increases in abundance due to glyphosate, whereas negative numbers indicate that the taxon decreases due to glyphosate.

5.4 Discussion

Our data did not support our hypothesis that glyphosate would impact different components of the N fixation system in a consistent manner. Cropping system did modulate the effect of glyphosate treatment on soybean N_{dfa} (Figures 5.3 & 5.4) but did not affect the response of nifH gene abundance to glyphosate (Figure 5.5). Cropping system did affect the glyphosate effect on community composition (Figure 5.7) but did so in a manner not consistent with N_{dfa} or nifH gene copy numbers. Specifically, changes in community composition can only be attributed to glyphosate in the rhizoplane of O6. If community composition were coordinated with the nifH gene abundance data, we would have expected community composition effects primarily in the rhizosphere rather than rhizoplane fractions. Similarly, if the community data were coordinated with the N_{dfa} response, we'd expect that the major effects would be primarily in CT or NT cropping systems. Therefore, the effects of glyphosate on each component of the cropping system did not exhibit a coordinated response and are likely governed by uncoupled processes.

Previous research has shown that the effect of glyphosate on N-fixation rate is largely mediated by water stress in a short window after glyphosate has been applied (Zablotowicz and Reddy, 2007). Increasing the amount of OM in soils leads to better moisture retention, and has been associated with reduced water stress in agricultural systems (Lepsch et al., 2019; Loveland and Webb, 2003); therefore, it is possible that greater OM could result in buffering against the glyphosate effect on N_{dfa}. To test this, we calculated the change in N_{dfa} within each plot due to the glyphosate treatment (Figure 5.4B). We found that there was weak statistical support (at an alpha of 0.1) for an association between OM and lower effect size of glyphosate on N_{dfa}; and average delta N_{dfa} across systems is consistent with the average increase in OM% across systems (Figure 5.4). The weak statistical support for this relationship in our dataset might be strengthened in a future study if the soil moisture content were measured during the period after glyphosate treatment in each of the plots.

We could find only one previous study addressing the effect of glyphosate on nifH gene quantity in the rhizosphere of glyphosate resistant soybean; that study found no effect of glyphosate on nifH quantity in the rhizoplane (Fan et al., 2017). Our study supports the finding that nifH quantity does not change due to glyphosate treatment in the rhizoplane subset of the rhizosphere, but further suggests that glyphosate treatment increases nifH quantity in the rhizosphere fraction beyond the rhizoplane (Figure 5.5). This suggests that something in the root exudates is responsible for the nifH gene response to glyphosate treatment. Allelochemical root exudates from soybean, including daidzein, are highest between V2 and R2 growth stages (Granato et al., 1983; Sugiyama et al., 2016). Soyasaponin Bb is exuded at a much higher rate between V5-R3 than other growth stages, and has been shown to disperse farther than daidzein, and to increase the proportion of potential beneficial bacterial groups in experimental treatments (Fujimatsu et al., 2020). In our experiment, soil samples were collected at V4 for the pre-spray sampling date and 20 days later for the post sampling date, at which time most soybean plants were between R1-R2. There was a marked increase in nifH genes in the post sampling date due to glyphosate treatment; our data support the notion that an alteration to the root exudate profile due to glyphosate treatment increased the quantity of nifH gene copies in the rhizosphere. This effect is consistent across cropping systems.

Our beta diversity analysis was consistent with previous research, finding little to no effect of glyphosate on bacterial community composition (Kepler et al., 2019; Schlatter et al., 2017). There are two mechanisms that may act to decouple 16S variant population dynamics from the effect of glyphosate. The prevailing model of bacterial gene evolution is that of nearly-neutral evolution with genetic draft (Rocha, 2018), where genetic draft is when changes in allele frequency of genes is due to selection on other genes that are nearby in the genome. This model predicts that the 16S gene variants are each not under direct selection by glyphosate, but instead are “drafting” along behind the EPSPS genes that are under selection. The size of the draft effect (i.e., the extent to which glyphosate treatment changes 16S variant frequencies in the community) is determined by how tightly the 16S gene locus is linked to the EPSPS gene locus; and how strong the selection pressure is on EPSPS. When strong selection on a single gene drives the entire bacterial genome to fixation, it’s considered a strong selective sweep; in contrast, a soft selective sweep is when strong selection only drives a few gene loci to fixation (Rocha, 2018). Evidence suggests that the dominant mechanism in environmental bacterial adaptation leads to soft selective sweeps. One mechanism that allows for soft selective sweeps is horizontal gene transfer; and, there is evidence that the EPSPS gene experiences horizontal gene transfer (Forlani et al., 2015). Second, existing EPSPS genes in bacteria have been documented to rapidly evolve resistance when exposed to glyphosate by modifying the enzyme’s active site (Rainio et al., 2021). When this rapid evolution occurs in parallel associated with many ASVs, it reduces the difference in selective pressure across ASVs, resulting in smaller effects of glyphosate on the composition of bacterial 16S rRNA sequence variants. These mechanisms suggest that it could be possible for glyphosate treatment to drive fixation of many resistant EPSPS alleles within a bacterial community without substantially altering the distribution of 16S sequence variants. And because EPSPS adaptation (and therefore selection) would be happening locally in each farming system, it offers a plausible mechanism for why different groups of taxa would respond to glyphosate in each cropping system (Figure 5.8).

In a similar pattern, the nifH quantity was not well-correlated to putatively identified diazotrophs. It is likely that taxonomy at the genus level is not a reliable predictor of the occurrence of nifH within the genome. This is in contrast to other genes, like amoA, which has been well correlated to putative ammonia oxidizing archaea in a QSeq dataset (see supplemental figure 2 from (Epp Schmidt et al., 2019)). The genomic content of soil bacteria containing the same 16S rRNA gene variant can shift considerably across a short distance (Crits-Christoph et al., 2020). There is a growing body of literature around microbial community dynamics that has found decoupling of at least certain functional niche adaptations from taxonomy (Louca et al., 2016b, 2016a; Philippot et al., 2010). The nifH operon is known to be subject to horizontal gene transfer (Bolhuis et al., 2010), to the extent that strains that contain nifH genes that are less than 3% divergent in their 16S rRNA gene may be as much as 23% divergent in the nifH gene locus (Gaby and Buckley, 2014). This may indicate additionally a higher chance that individual bacteria gain or lose the gene, reducing the relationship between nifH gene abundance and 16S rRNA gene taxonomy.

5.5 Conclusions

Although we observed that glyphosate does impact several components of the nitrogen fixation system, we found evidence that these responses are decoupled; glyphosate treatment affected soybean N_{dfa} but not the microbial community composition; and nifH gene quantity is affected by glyphosate application but is not well-correlated to the subset of the 16S rRNA genes putatively identified as diazotrophs, nor is it related to the soybean N_{dfa}. Overall, we found evidence supporting the notion that cropping systems with higher soil OM are more resilient to changes in function due to glyphosate treatments, likely because they buffer water stress. We also observed that the rhizosphere but not rhizoplane soils are likely to experience increased nifH gene quantity due to glyphosate treatment, but that 16S rRNA gene composition is unlikely to reflect this response even when restricting the analysis to ASVs likely to contain nifH genes.

5.6 Acknowledgements

This research was a contribution from the Long-Term Agroecosystem Research (LTAR) network. LTAR is supported by the United States Department of Agriculture. Dietrich Epp Schmidt was supported by NRT-INFEWS: UMD Global STEWARDS (STEM Training at the Nexus of Energy, WAter Reuseand FooD Systems) that was awarded to the University of Maryland School of Public Health by the National Science Foundation National Research Traineeship Program, Grant number 1828910.

5.7 Literature Cited

Balboa, G.R., Ciampitti, I.A., 2020. Estimating biological nitrogen fixation in field-grown soybeans: impact of B value. Plant Soil 446, 195–210. https://doi.org/10.1007/s11104-019-04317-1

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Bolhuis, H., Severin, I., Confurius-Guns, V., Wollenzien, U.I.A., Stal, L.J., 2010. Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium Microcoleus chthonoplastes. ISME J 4, 121–130. https://doi.org/10.1038/ismej.2009.99

Bremner, J.M., Keeney, D.R., 1965. Steam distillation methods for determination of ammonium, nitrate and nitrite. Analytica Chimica Acta. https://doi.org/10.1016/S0003-2670(00)88973-4

Broadbent, F.E., 1965. Part 1: physical and mineralogical methods; organic matter, in: Methods of Soil Analysis, Monograph. American Society of Agronomy, pp. 1397–1400.

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

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Cavigelli, M., Teasdale, J.R., Conklin, A.E., 2008. Long-term agronomic performance of organic and conventional field crops in the mid-atlantic region. Agron. J. 100. https://doi.org/10.2134/agronj2006.0373

Coupland, D., Caseley, J.C., 1979. Presence of 14c activity in root exudates and guttation fluid from Agropyron repens treated with 14c-labelled glyphosate. New Phytologist 83, 17–22. https://doi.org/10.1111/j.1469-8137.1979.tb00721.x

Crits-Christoph, A., Olm, M.R., Diamond, S., Bouma-Gregson, K., Banfield, J.F., 2020. Soil bacterial populations are shaped by recombination and gene-specific selection across a grassland meadow. ISME J 14, 1834–1846. https://doi.org/10.1038/s41396-020-0655-x

Delves, A.C., Mathews, A., Day, D.A., Carter, A.S., Carroll, B.J., Gresshoff, P.M., 1986. Regulation of the soybean-rhizobium nodule symbiosis by shoot and root factors 1. Plant Physiology 82, 588–590. https://doi.org/10.1104/pp.82.2.588

Epp Schmidt, D.J., Kotze, D.J., Hornung, E., Setälä, H., Yesilonis, I., Szlavecz, K., Dombos, M., Pouyat, R., Cilliers, S., Tóth, Z., Yarwood, S., 2019. Metagenomics reveals bacterial and archaeal adaptation to urban land-use: N catabolism, methanogenesis, and nutrient acquisition. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02330

Fan, L., Feng, Y., Weaver, D.B., Delaney, D.P., Wehtje, G.R., Wang, G., 2017. Glyphosate effects on symbiotic nitrogen fixation in glyphosate-resistant soybean. Applied Soil Ecology 121, 11–19. https://doi.org/10.1016/j.apsoil.2017.09.015

Forlani, G., Bertazzini, M., Barillaro, D., Rippka, R., 2015. Divergent properties and phylogeny of cyanobacterial 5-enol-pyruvyl-shikimate-3-phosphate synthases: evidence for horizontal gene transfer in the Nostocales. New Phytologist 205, 160–171. https://doi.org/10.1111/nph.13022

Fujimatsu, T., Endo, K., Yazaki, K., Sugiyama, A., 2020. Secretion dynamics of soyasaponins in soybean roots and effects to modify the bacterial composition. Plant Direct 4, e00259. https://doi.org/10.1002/pld3.259

Gaby, J.C., Buckley, D.H., 2014. A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. Database (Oxford) 2014. https://doi.org/10.1093/database/bau001

Granato, T.C., Banwart, W.L., Porter, P.M., Hassett, J.J., 1983. Effect of variety and stage of growth on potential allelochemic compounds in soybean roots. J Chem Ecol 9, 1281–1292. https://doi.org/10.1007/BF00982229

Grube, M., Kalnenieks, U., Muter, O., 2019. Metabolic response of bacteria to elevated concentrations of glyphosate-based herbicide. Ecotoxicology and Environmental Safety 173, 373–380. https://doi.org/10.1016/j.ecoenv.2019.02.045

Hao, Y., Chen, H., Xu, W., Gao, J., Yang, Y., Zhang, Y., Tao, L., 2019. Roundup® confers cytotoxicity through DNA damage and mitochondria-associated apoptosis induction. Environmental Pollution 252, 917–923. https://doi.org/10.1016/j.envpol.2019.05.128

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., Korpela, K., 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLOS ONE 15, e0227285. https://doi.org/10.1371/journal.pone.0227285

Kepler, R., Epp Schmidt, D., Yarwood, S., Cavigelli, M., Reddy, K., Duke, S., Bradley, C., Williams, M., Buyer, J., Maul, J., 2019. Soil microbial communities in diverse agroecosystems exposed to the herbicide glyphosate. Applied and Environmental Microbiology 86. https://doi.org/10.1128/AEM.01744-19

Kremer, R., Means, N., Kim, S., 2005. Glyphosate affects soybean root exudation and rhizosphere micro-organisms. International Journal of Environmental Analytical Chemistry 85, 1165–1174. https://doi.org/10.1080/03067310500273146

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: concept & review. Soil Biology and Biochemistry 83, 184–199. https://doi.org/10.1016/j.soilbio.2015.01.025

Kuzyakov, Y., Razavi, B.S., 2019. Rhizosphere size and shape: Temporal dynamics and spatial stationarity. Soil Biology and Biochemistry 135, 343–360. https://doi.org/10.1016/j.soilbio.2019.05.011

Laitinen, P., Rämö, S., Siimes, K., 2007. Glyphosate translocation from plants to soil – does this constitute a significant proportion of residues in soil? Plant Soil 300, 51–60. https://doi.org/10.1007/s11104-007-9387-1

Lepsch, H.C., Brown, P.H., Peterson, C.A., Gaudin, A.C.M., Khalsa, S.D.S., 2019. Impact of organic matter amendments on soil and tree water status in a California orchard. Agricultural Water Management 222, 204–212. https://doi.org/10.1016/j.agwat.2019.06.002

Louca, S., Jacques, S.M.S., Pires, A.P.F., Leal, J.S., Srivastava, D.S., Parfrey, L.W., Farjalla, V.F., Doebeli, M., 2016a. High taxonomic variability despite stable functional structure across microbial communities. Nat Ecol Evol 1, 1–12. https://doi.org/10.1038/s41559-016-0015

Louca, S., Parfrey, L.W., Doebeli, M., 2016b. Decoupling function and taxonomy in the global ocean microbiome. Science 353, 1272–1277. https://doi.org/10.1126/science.aaf4507

Loveland, P., Webb, J., 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. Soil and Tillage Research 70, 1–18. https://doi.org/10.1016/S0167-1987(02)00139-3

Lupwayi, N.Z., Blackshaw, R.E., 2012. Soil microbiology in glyphosate-resistant corn cropping systems. Agronomy Journal 104, 1041–1048. https://doi.org/10.2134/agronj2012.0054

Moebius, B.N., van Es, H.M., Schindelbeck, R.R., Idowu, O.J., Clune, D.J., Thies, J.E., 2007. Evaluation of laboratory-measured soil properties as indicators of soil physical quality. Soil Science 172, 895–912. https://doi.org/10.1097/ss.0b013e318154b520

Moebius-Clune, B.N., Moebius-Clune, D.J., Gugino, B.K., Idowu, O.J., Schindelbeck, R.R., Ristow, A.J., van Es, H.M., Thies, J.E., Shayler, H.A., McBride, M.B., Kurtz, K.S.M., Wolfe, D.W., Abawi, G.S., 2016. Comprehensive assessment of soil health: the Cornell framework manual.

Nguyen, D., Rose, M., Rose, T., Morris, S., Zwieten, L., 2016. Impact of glyphosate on soil microbial biomass and respiration: A meta-analysis. Soil Biology and Biochemistry 92, 50–57. https://doi.org/10.1016/j.soilbio.2015.09.014

Philippot, L., Andersson, S.G.E., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B., Hallin, S., 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Micro 8, 523–529. https://doi.org/10.1038/nrmicro2367

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41, D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rainio, M.J., Ruuskanen, S., Helander, M., Saikkonen, K., Saloniemi, I., Puigbò, P., 2021. Adaptation of bacteria to glyphosate: a microevolutionary perspective of the enzyme 5-enolpyruvylshikimate-3-phosphate synthase. Environmental Microbiology Reports 13, 309–316. https://doi.org/10.1111/1758-2229.12931

Reinhold, B., Bü, W., Burbano, C., Sabale, M., Hurek, T., 2015. Roots Shaping Their Microbiome: Global Hotspots for Microbial Activity.

Annual review of phytopathology 53, 403–24. https://doi.org/10.1146/annurev-phyto-082712-102342

Rocha, E.P.C., 2018. Neutral theory, microbial practice: challenges in bacterial population genetics. Molecular Biology and Evolution 35, 1338–1347. https://doi.org/10.1093/molbev/msy078

Sansupa, C., Wahdan, S.F.M., Hossen, S., Disayathanoowat, T., Wubet, T., Purahong, W., 2021. Can we use functional annotation of prokaryotic taxa (FAPROTAX) to assign the ecological functions of soil bacteria? Applied Sciences 11, 688. https://doi.org/10.3390/app11020688

Schlatter, D.C., Yin, C., Hulbert, S., Burke, I., Paulitz, T., 2017. Impacts of repeated glyphosate use on wheat-associated bacteria are small and depend on glyphosate use history. Appl. Environ. Microbiol. 83, e01354-17. https://doi.org/10.1128/AEM.01354-17

Schönbrunn, E., Eschenburg, S., Shuttleworth, W.A., Schloss, J.V., Amrhein, N., Evans, J.N.S., Kabsch, W., 2001. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proc Natl Acad Sci U S A 98, 1376–1380.

Steinrücken, H.C., Amrhein, N., 1980. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimic acid-3-phosphate synthase. Biochemical and Biophysical Research Communications 94, 1207–1212. https://doi.org/10.1016/0006-291X(80)90547-1

Sugiyama, A., Yamazaki, Y., Yamashita, K., Takahashi, S., Nakayama, T., Yazaki, K., 2016. Developmental and nutritional regulation of isoflavone secretion from soybean roots. Bioscience, Biotechnology, and Biochemistry 80, 89–94. https://doi.org/10.1080/09168451.2015.1062714

Tzin, V., Galili, G., 2010. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. Molecular Plant 3, 956–972. https://doi.org/10.1093/mp/ssq048

Vetterlein, D., Carminati, A., Kögel-Knabner, I., Bienert, G., Smalla, K., Oburger, E., Schnepf, A., Banitz, T., Tarkka, M., Schlüter, S., 2020. Rhizosphere spatiotemporal organization–a key to rhizosphere functions. Frontiers in Agronomy 2, 8. https://doi.org/10.3389/fagro.2020.00008

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Research 42, D643–D648. https://doi.org/10.1093/nar/gkt1209

Zablotowicz, R.M., Reddy, K.N., 2007. Nitrogenase activity, nitrogen content, and yield responses to glyphosate in glyphosate-resistant soybean. Crop Protection, Weed Science in Time of Transition 26, 370–376. https://doi.org/10.1016/j.cropro.2005.05.013

Zobiole, L.H.S., Kremer, R.J., de Oliveira Jr., R.S., Constantin, J., 2012. Glyphosate effects on photosynthesis, nutrient accumulation, and nodulation in glyphosate-resistant soybean. Journal of Plant Nutrition and Soil Science 175, 319–330. https://doi.org/10.1002/jpln.201000434

Zobiole, L.H.S., Kremer, R.J., Oliveira, R.S., Constantin, J., 2011. Glyphosate affects chlorophyll, nodulation and nutrient accumulation of “second generation” glyphosate-resistant soybean (Glycine max L.). Pesticide Biochemistry and Physiology 99, 53–60. https://doi.org/10.1016/j.pestbp.2010.10.005

Zobiole, L.H.S., Kremer, R.J., Oliveira, R.S., Constantin, J., 2010. Glyphosate affects micro-organisms in rhizospheres of glyphosate-resistant soybeans. Journal of Applied Microbiology 110, 118–127. https://doi.org/10.1111/j.1365-2672.2010.04864.x

Chapter 6: Conclusions

The actual abundance of taxa provides necessary information to accurately infer ecological drivers and community assembly processes that explain the existing patterns of community composition. The compositional nature of sequence data has limited the scope of inference for microbial ecologists that rely solely on sequence data to survey microbial communities. Recent development of quantitative sequencing has provided a means to alleviate the limitations of compositionality, and allow researchers to make good inferences about ecosystem processes that drive community assembly. These recent methodological advances are yet to become prevalent components of the community ecology analysis paradigm; this dissertation sought first to establish the conditions under which the quantitative sequencing method addressed known biases associated with compositional analysis; and then explore the utility of this method using several different experimental and analytic contexts.

6.1 Chapter 2 Summary

In chapter 2, we explore the boundary conditions that define when QSeq provides improved accuracy for inferring ecological processes. We used a combination of modeling, data synthesis, and literature review to evaluate QSeq methodologies. Using this information, we provided guidance for when QSeq should and should not be used. We also provide a software instance that can be used in R to transform microbiome data into QSeq data for analysis.

QSeq provides no conceptual advantage for alpha diversity estimation, and because it manipulates counts for each taxon, may actually reduce the accuracy of alpha diversity estimators. Beta diversity analysis may or may not benefit from QSeq depending on the question and similarity index being used. QSeq should not be used for beta diversity if the question is primarily focused on the relative dominance of taxa or the prevalence of taxa across sites. In contrast, QSeq offers distinct advantages for estimating the abundance of individual taxa; determining the effect of environment on the abundance of individual taxa; and determining the covariance among individual taxa across environmental gradients. These advantages become greater as the overlying gradient in total community abundance becomes greater; we identified that if sample abundance varies by more than 5 % of the mean, that QSeq provides more accurate estimation of taxon abundance and taxon-taxon covariance. We suggest that this 5% threshold is extremely low compared to the typical variation in total abundance of environmental datasets, and that therefore QSeq should be considered the default method for assessing taxon differential abundance and covariance in environmental microbiome datasets.

While this dissertation in general is focused primarily on soil communities from diverse environments, we have found multiple instances where QSeq is beneficial to the analysis of human and mouse microbiomes. Thus this method has broad applicability across all microbiome environments.

Current open-source microbiome analysis pipelines provide support only for compositional statistical inference methods, and no explicit support for QSeq transformations. We conclude that the field of microbiome research would benefit substantially if open source sequence analysis pipelines include support for the QSeq transformation and statistical inference solutions with assumptions that are appropriate for QSeq data.

6.2 Chapter 3 Summary

Chapter 3 evaluates whether scaling metagenome prediction by total bacterial abundance reliably improves the prediction of gene abundance. There are a large number of software platforms that have been developed to predict metagenomes from 16S sequence data. PICRUSt2 is generally held to be most accurate in soil systems. Across platforms, however, metagenomic inference produces compositional (relative abundance) gene data. In Chapters 1 and 2 we establish that relative abundance does not accurately represent true abundance; metagenomic datasets have strong correlations among gene families because certain functions (and therefore genes) must exist in all genomes, while others do not. This bias leads to high correlation in metagenomic prediction (0.6-0.8 Pearson correlation) because core maintenance functions are easy to predict. But this apparent correlation does not represent the accuracy with which the abundance of genes for specialized niche adaptation are predicted. In this chapter I test whether scaling by total abundance improves the accuracy of gene estimation by comparing gene abundance estimates to quantitative PCR of the same genes. My hypotheses were that 1) scaling by total abundance would reliably improve the accuracy of gene abundance estimates; and 2) that certain genes would be more reliably predicted because some genes regularly experience horizontal gene transfer while others do not. I used a data synthesis approach, drawing on several different datasets in order to address the potential variability in lab methods, skill, environments, and spatial scale.

My first hypothesis was supported, there was strong evidence that scaling by total abundance reliably improved the accuracy with which abundance of genes was predicted. However, my second hypothesis was not supported. High variability in the accuracy of genes within datasets, and high variability in the accuracy of genes among datasets resulted in no difference in the mean correlation between predicted gene abundance and measured gene abundance. Certain genes in each dataset were highly accurately predicted, suggesting that this variation is not due to different levels of skill in QPCR or sequencing. Datasets from different environments had different genes that were highly accurately predicted, suggesting that the variability is very likely due in part to how well the reference databases represents specific environmental contexts. No genes were always accurately predicted, indicating that there is a lot of improvement that is necessary before gene abundance predictions are reliable. Future replication of environments will need to be done to determine if the abundance of certain genes in certain environments can reliably be predicted.

6.3 Chapter 4 Summary

Chapter 4 evaluates the effect of farming system management on the distribution of microbial taxa along the soil depth gradient. Farming systems are known to influence the chemistry and distribution of organic matter in soils. In particular, organic practices rely more heavily on tillage, which actively homogenizes surface soils, disrupting the depth structure of soil organic matter accumulation. We evaluated three different farming systems that differed in their nutrient management (organic vs inorganic inputs), and tillage intensity. The three systems were no-till (NT), which had no tillage and relied primarily on inorganic nutrients for nutrient management; chisel-till (CT), which used chisel (reduced) tillage, and inorganic nutrients for nutrient management; and organic 3 (O3), which used tillage for weed control and organic nutrient inputs for nutrient management. All three systems were on a 3-year crop rotation, and sampled post-harvest in the wheat phase of the rotation. Our hypothesis was that organic farming systems would result in a redistribution of taxa in the soil profile.

The total abundance, and taxonomic diversity of taxa was exponentially higher at the surface than at depth, and these patterns were strongly correlated with soil C and N. This finding supports the notion that relative abundance of taxa in this dataset does not represent their true abundance across samples. Using the QSeq methodology, we were able to identify a set of taxa that changed abundance due to farming system; and also a small set of taxa whose abundance was redistributed in the soil profile because of tillage. For example, we found that members of mortierellomycota were more abundant in CT and O3 than in NT. And, we were able to show that tillage in O3 shifted members of pseudomonas down in the soil profile, but did not affect their overall abundance (Chapter 3, Figure 6).

We were surprised in general by the relative weakness of the farming system effects compared to depth effects. Farming systems had the greatest effect in the top 10 cm, where farming system explained as much as 30-40% of the variation in microbial community composition, depending on whether bacterial or fungal communities were being evaluated. Taxa that were strong indicators of farming system tended to also be found primarily in the top 10 cm of the soil profile.

These results illustrate that abundance and occurrence patterns both reflect important aspects of the community assembly process. Farming system primarily impacted the abundance and prevalence of taxa in the top 10 cm, but high tillage in Org3 did affect the distribution of pseudomonads in the soil profile. The conclusions of this chapter would not have been possible without quantitative measures of the abundance of individual taxa.

6.4 Chapter 5 Summary

Chapter 5 evaluates the effect of glyphosate on putative N-fixing bacterial community across four farming systems, and asks whether effects seen in the rhizosphere diazotroph community reflect other effects of glyphosate on N-fixation in soybean. In this chapter, we quantified N derived from the atmosphere in soybean, the quantity of nifH genes in rhizoplane and rhizosphere soil fractions, and the abundance and composition of the diazotroph community from rhizoplane and rhizosphere soil fractions. Our hypothesis was that if glyphosate decreased the amount of nitrogen fixed by soybean, then it would decrease the abundance of nifH genes, and alter the diazotroph community composition. Our data did not support this hypothesis; rather it indicated that the amount of nitrogen fixed by soybean, nifH gene abundance in rhizosphere soil fractions, and the community composition were all governed by different processes. We identified taxa that increased or decreased in abundance as a response to glyphosate in each of the soil fractions in each of the farming systems; however, the collection of taxa that responded were different in each farming system and each soil fraction. The abundance of these organisms would not have been reliably inferred without a quantitative sequencing method. Our results support the notions that local adaptation to glyphosate occurs rapidly in soil bacteria; that glyphosate sensitivity is decoupled from taxonomy; and that taxonomy may not reliably indicate the presence of the nifH gene.

6.5 Final Thoughts

Analysis in Chapters 2 and 4 indicate that gradients in total microbial abundance are likely to be large enough in most environmental datasets to warrant a QSeq approach. Therefore, measuring and adjusting sequence data to reflect the density of taxa in the environment is a necessary step to accurately infer the mechanisms driving community assembly of environmental microbiomes. Measuring total abundance should always be done in microbiome research; and open access bioinformatics pipelines should offer support for the QSeq transformation and statistical analysis of QSeq transformed data. Accounting for total abundance in metagenome prediction software almost always improves the accuracy of gene abundance estimates, indicating this concept has broad applicability. Yet the variability among prediction quality in datasets analyzed in Chapter 3, and the variability seen in the response of N-fixing taxa in Chapter 5, indicate that bacterial functions are not yet reliably predictable based on taxonomy. Some genes may never be reliably predictable based on taxonomy due to mechanisms like horizontal gene transfer that decouple the taxonomy of local bacteria from their actual ecological function.

It is likely that high throughput techniques for measuring functional gene abundance will become tractable more quickly than metagenome predictions will become accurate and reliable. For example, RNAseq costs are similar to amplicon sequencing, but give a snapshot into the genes that are actively being transcribed by the microbial community. Reference datasets that can be used to ascribe microbial taxonomy already exist for important functional genes (such as amoA). It is therefore conceivable that an analytic pipeline could be built for prokaryote and eukaryote RNAseq datasets to simultaneously describe the taxonomic, phylogenetic, and functional diversity of soil communities. These methods could easily be integrated into a QSeq approach that would allow soil researchers to connect environmental gradients to total expression levels for specific functional genes. Since a core challenge in soil microbial ecology is that many of the taxa are dormant most of the time, these tools could be incredibly powerful for illuminating which taxa are active and what they are doing.

Bibliography

CHAPTER 1:

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

Anthony, M.A., Crowther, T.W., Maynard, D., van den Hoogan, J., Averill, C., 2020. Distinct Assembly Processes and Microbial Communities Constrain Soil Organic Carbon Formation. One Earth 2, 349–360. https://doi.org/10.1016/j.oneear.2020.03.006

Auguet, J.-C., Barberan, A., Casamayor, E.O., 2009. Global ecological patterns in uncultured Archaea. ISME J 4, 182–190. https://doi.org/10.1038/ismej.2009.109

Averill, C., Hawkes, C.V., 2016. Ectomycorrhizal fungi slow soil carbon cycling. Ecol Lett 19, 937–947. https://doi.org/10.1111/ele.12631

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., Huerta-Cepas, J., Medema, M.H., Maltz, M.R., Mundra, S., Olsson, P.A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., Tedersoo, L., Bork, P., 2018. Structure and function of the global topsoil microbiome. Nature 560, 233–237. https://doi.org/10.1038/s41586-018-0386-6

Barlow, J.T., Bogatyrev, S.R., Ismagilov, R.F., 2020. A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. Nature Communications 11, 2590. https://doi.org/10.1038/s41467-020-16224-6

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., Fierer, N., 2011. Examining the global distribution of dominant archaeal populations in soil. ISME J 5, 908–917. https://doi.org/10.1038/ismej.2010.171

Baveye, P.C., 2021. Bypass and hyperbole in soil research: Worrisome practices critically reviewed through examples. European Journal of Soil Science 72, 1–20. https://doi.org/10.1111/ejss.12941

Blagodatsky, S., Smith, P., 2012. Soil physics meets soil biology: Towards better mechanistic prediction of greenhouse gas emissions from soil. Soil Biology and Biochemistry 47, 78–92. https://doi.org/10.1016/j.soilbio.2011.12.015

Blagodatsky, S.A., Yevdokimov, I.V., Larionova, A.A., Richter, J., 1998. Microbial growth in soil and nitrogen turnover: Model calibration with laboratory data. Soil Biology and Biochemistry 30, 1757–1764. https://doi.org/10.1016/S0038-0717(98)00029-7

Blankenship, R.E., 2010. Early Evolution of Photosynthesis1. Plant Physiol 154, 434–438. https://doi.org/10.1104/pp.110.161687

Blankenship, R.E., 2002. Molecular Mechanisms of Photosynthesis. John Wiley & Sons.

Bockheim, J.G., Gennadiyev, A.N., Hartemink, A.E., Brevik, E.C., 2014. Soil-forming factors and Soil Taxonomy. Geoderma 226–227, 231–237. https://doi.org/10.1016/j.geoderma.2014.02.016

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., Gregory Caporaso, J., 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6, 90. https://doi.org/10.1186/s40168-018-0470-z

Bokulich, N.A., Rideout, J.R., Mercurio, W.G., Wolfe, B., Maurice, C.F., Dutton, R.J., Turnbaugh, P.J., Knight, R., Caporaso, J.G., 2016. mockrobiota: a public resource for microbiome bioinformatics benchmarking (No. e2065v1). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.2065v1

Bokulich, N.A., Ziemski, M., Robeson, M.S., Kaehler, B.D., 2020. Measuring the microbiome: Best practices for developing and benchmarking microbiomics methods. Computational and Structural Biotechnology Journal 18, 4048–4062. https://doi.org/10.1016/j.csbj.2020.11.049

Bronner, I.F., Quail, M.A., Turner, D.J., Swerdlow, H., 2009. Improved Protocols for Illumina Sequencing. Curr Protoc Hum Genet 0 18, 10.1002/0471142905.hg1802s62. https://doi.org/10.1002/0471142905.hg1802s62

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. PNAS 108, 4516–4522. https://doi.org/10.1073/pnas.1000080107

Cardona, T., 2019. Thinking twice about the evolution of photosynthesis. Open Biol 9, 180246. https://doi.org/10.1098/rsob.180246

Cotrufo, M.F., Wallenstein, M.D., Boot, C.M., Denef, K., Paul, E., 2013. The Microbial Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition with soil organic matter stabilization: do labile plant inputs form stable soil organic matter? Global Change Biology 19, 988–995. https://doi.org/10.1111/gcb.12113

de Vries, F.T., Manning, P., Tallowin, J.R.B., Mortimer, S.R., Pilgrim, E.S., Harrison, K.A., Hobbs, P.J., Quirk, H., Shipley, B., Cornelissen, J.H.C., Kattge, J., Bardgett, R.D., 2012. Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities. Ecol Lett 15, 1230–1239. https://doi.org/10.1111/j.1461-0248.2012.01844.x

Eilers, K.G., Debenport, S., Anderson, S., Fierer, N., 2012. Digging deeper to find unique microbial communities: The strong effect of depth on the structure of bacterial and archaeal communities in soil. Soil Biology and Biochemistry 50, 58–65. https://doi.org/10.1016/j.soilbio.2012.03.011

Epp Schmidt, D.J., Kotze, D.J., Hornung, E., Setälä, H., Yesilonis, I., Szlavecz, K., Dombos, M., Pouyat, R., Cilliers, S., Tóth, Z., Yarwood, S., 2019. Metagenomics reveals bacterial and archaeal adaptation to urban land-use: N catabolism, methanogenesis, and nutrient acquisition. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02330

Ettwig, K.F., Butler, M.K., Le Paslier, D., Pelletier, E., Mangenot, S., Kuypers, M.M.M., Schreiber, F., Dutilh, B.E., Zedelius, J., de Beer, D., Gloerich, J., Wessels, H.J.C.T., van Alen, T., Luesken, F., Wu, M.L., van de Pas-Schoonen, K.T., Op den Camp, H.J.M., Janssen-Megens, E.M., Francoijs, K.-J., Stunnenberg, H., Weissenbach, J., Jetten, M.S.M., Strous, M., 2010. Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. Nature 464, 543–548. https://doi.org/10.1038/nature08883

Feng, X., Qin, S., Zhang, D., Chen, P., Hu, J., Wang, G., Liu, Y., Wei, B., Li, Q., Yang, Y., Chen, L., 2022. Nitrogen input enhances microbial carbon use efficiency by altering plant–microbe–mineral interactions. Global Change Biology 28, 4845–4860. https://doi.org/10.1111/gcb.16229

Fernandez, C.W., Kennedy, P.G., 2016. Revisiting the ‘Gadgil effect’: do interguild fungal interactions control carbon cycling in forest soils? New Phytologist 209, 1382–1394. https://doi.org/10.1111/nph.13648

Fierer, N., Bradford, M.A., Jackson, R.B., 2007. Toward an ecological classification of soil bacteria. Ecology 88, 1354–1364. https://doi.org/10.1890/05-1839

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. PNAS 103, 626–631. https://doi.org/10.1073/pnas.0507535103

Fischer, F., Zillig, W., Stetter, K.O., Schreiber, G., 1983. Chemolithoautotrophic metabolism of anaerobic extremely thermophilic archaebacteria. Nature 301, 511–513. https://doi.org/10.1038/301511a0

Gill, A.L., Finzi, A.C., 2016. Belowground carbon flux links biogeochemical cycles and resource-use efficiency at the global scale. Ecology Letters 19, 1419–1428. https://doi.org/10.1111/ele.12690

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 0. https://doi.org/10.3389/fmicb.2017.02224

Gloor, G.B., Reid, G., 2016. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can. J. Microbiol. 62, 692–703. https://doi.org/10.1139/cjm-2015-0821

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., Egozcue, J.J., 2016. It’s all relative: analyzing microbiome data as compositions. Ann Epidemiol 26, 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003

Guseva, K., Darcy, S., Simon, E., Alteio, L.V., Montesinos-Navarro, A., Kaiser, C., 2022. From diversity to complexity: Microbial networks in soils. Soil Biology and Biochemistry 169, 108604. https://doi.org/10.1016/j.soilbio.2022.108604

Hao, J., Chai, Y.N., Lopes, L.D., Ordóñez, R.A., Wright, E.E., Archontoulis, S., Schachtman, D.P., 2021. The Effects of Soil Depth on the Structure of Microbial Communities in Agricultural Soils in Iowa (United States). Applied and Environmental Microbiology 87, e02673-20. https://doi.org/10.1128/AEM.02673-20

He, L., Mazza Rodrigues, J.L., Soudzilovskaia, N.A., Barceló, M., Olsson, P.A., Song, C., Tedersoo, L., Yuan, Fenghui, Yuan, Fengming, Lipson, D.A., Xu, X., 2020. Global biogeography of fungal and bacterial biomass carbon in topsoil. Soil Biology and Biochemistry 151, 108024. https://doi.org/10.1016/j.soilbio.2020.108024

Hodge, A., Campbell, C., Fitter, A., 2001. An arbuscular mycorrhizal fungus accelerates decomposition and acquires nitrogen directly from organic material | Nature.

Hodge, A., Robinson, D., Fitter, A., 2000. Are microorganisms more effective than plants at competing for nitrogen? Trends in Plant Science 5, 304–308. https://doi.org/10.1016/S1360-1385(00)01656-3

Hornibrook, E.R.C., Longstaffe, F.J., Fyfe, W.S., 1997. Spatial distribution of microbial methane production pathways in temperate zone wetland soils: Stable carbon and hydrogen isotope evidence. Geochimica et Cosmochimica Acta 61, 745–753. https://doi.org/10.1016/S0016-7037(96)00368-7

Iijima, H., 2020. A Review of Wildlife Abundance Estimation Models: Comparison of Models for Correct Application. jmam 45, 177–188. https://doi.org/10.3106/ms2019-0082

Illumina, 2021. MiniSeq System Denature and Dilute Libraries Guide (1000000002697).

Jenny, H., 1941. Factors of Soil Formation, a System of Quantitative Pedology. Agronomy Journal 33, 857–858. https://doi.org/10.2134/agronj1941.00021962003300090016x

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., Korpela, K., 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLOS ONE 15, e0227285. https://doi.org/10.1371/journal.pone.0227285

Kurtz, Z.D., Müller, C.L., Miraldi, E.R., Littman, D.R., Blaser, M.J., Bonneau, R.A., 2015. Sparse and Compositionally Robust Inference of Microbial

Ecological Networks. PLOS Computational Biology 11, e1004226. https://doi.org/10.1371/journal.pcbi.1004226

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: concept & review. Soil Biology and Biochemistry 83, 184–199. https://doi.org/10.1016/j.soilbio.2015.01.025

Ladapo, J., Whitman, W.B., 1990. Method for isolation of auxotrophs in the methanogenic archaebacteria: role of the acetyl-CoA pathway of autotrophic CO2 fixation in Methanococcus maripaludis. Proceedings of the National Academy of Sciences 87, 5598–5602. https://doi.org/10.1073/pnas.87.15.5598

Lange, O.L., Belnap, J., Reichenberger, H., 1998. Photosynthesis of the cyanobacterial soil-crust lichen Collema tenax from arid lands in southern Utah, USA: role of water content on light and temperature responses of CO2 exchange. Functional Ecology 12, 195–202. https://doi.org/10.1046/j.1365-2435.1998.00192.x

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. Appl. Environ. Microbiol. 75, 5111–5120. https://doi.org/10.1128/AEM.00335-09

Leeuwen, J.P. van, Djukic, I., Bloem, J., Lehtinen, T., Hemerik, L., Ruiter, P.C. de, Lair, G.J., 2017. Effects of land use on soil microbial biomass, activity and community structure at different soil depths in the Danube floodplain. European Journal of Soil Biology 79, 14–20. https://doi.org/10.1016/j.ejsobi.2017.02.001

Lehtovirta-Morley, L.E., 2018. Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. FEMS Microbiology Letters 365, fny058. https://doi.org/10.1093/femsle/fny058

Lloréns-Rico, V., Vieira-Silva, S., Gonçalves, P.J., Falony, G., Raes, J., 2021. Benchmarking microbiome transformations favors experimental quantitative approaches to address compositionality and sampling depth biases. Nat Commun 12, 3562. https://doi.org/10.1038/s41467-021-23821-6

López-García, A., Pineda-Quiroga, C., Atxaerandio, R., Pérez, A., Hernández, I., García-Rodríguez, A., González-Recio, O., 2018. Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.03010

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

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Meyer, F., Lesker, T.-R., Koslicki, D., Fritz, A., Gurevich, A., Darling, A.E., Sczyrba, A., Bremges, A., McHardy, A.C., 2021. Tutorial: assessing metagenomics software with the CAMI benchmarking toolkit. Nat Protoc 16, 1785–1801. https://doi.org/10.1038/s41596-020-00480-3

Oono, R., 2017. A confidence interval analysis of sampling effort, sequencing depth, and taxonomic resolution of fungal community ecology in the era of high-throughput sequencing. PLOS ONE 12, e0189796. https://doi.org/10.1371/journal.pone.0189796

Op den Camp, H.J.M., Islam, T., Stott, M.B., Harhangi, H.R., Hynes, A., Schouten, S., Jetten, M.S.M., Birkeland, N.-K., Pol, A., Dunfield, P.F., 2009. Environmental, genomic and taxonomic perspectives on methanotrophic Verrucomicrobia. Environmental Microbiology Reports 1, 293–306. https://doi.org/10.1111/j.1758-2229.2009.00022.x

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., Wilding, T.A., 2014. Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Molecular Ecology Resources 14, 1129–1140. https://doi.org/10.1111/1755-0998.12261

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., Esling, P., 2016. Protist metabarcoding and environmental biomonitoring: Time for change. European Journal of Protistology, Current trends in protistology – results from the VII ECOP - ISOP Joint Meeting 2015 55, 12–25. https://doi.org/10.1016/j.ejop.2016.02.003

Peng, X., Bruns, M.A., 2019a. Development of a nitrogen-fixing cyanobacterial consortium for surface stabilization of agricultural soils. Journal of Applied Phycology 31, 1047–1056. https://doi.org/10.1007/s10811-018-1597-9

Peng, X., Bruns, M.A., 2019b. Cyanobacterial Soil Surface Consortia Mediate N Cycle Processes in Agroecosystems. Frontiers in Environmental Science 6, 156. https://doi.org/10.3389/fenvs.2018.00156

Pester, M., Knorr, K.-H., Friedrich, M., Wagner, M., Loy, A., 2012. Sulfate-reducing microorganisms in wetlands – fameless actors in carbon cycling and climate change. Frontiers in Microbiology 3.

Pouyat, R., Yesilonis, I., Dombos, M., Szlavecz, K., Setala, H., Cilliers, S., Hornung, E., Kotze, J., Yarwood, S., 2015. A global comparison of surface soil across five cities: a test of the urban ecosystem convergence hypothesis. Soil Science.

Pouyat, R.V., Szlavecz, K., Yesilonis, I.D., Groffman, P.M., Schwarz, K., 2010. Chemical, Physical, and Biological Characteristics of Urban Soils, in: Urban Ecosystem Ecology, Agronomy Monographs. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, pp. 119–152.

Pouyat, R.V., Yesilonis, I.D., Russell-Anelli, J., Neerchal, N.K., 2007. Soil Chemical and Physical Properties That Differentiate Urban Land-Use and Cover Types. Soil Science Society of America Journal 71, 1010. https://doi.org/10.2136/sssaj2006.0164

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47–e47. https://doi.org/10.1093/nar/gkv007

Schimel, J.P., Weintraub, M.N., 2003. The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. Soil Biology and Biochemistry 35, 549–563. https://doi.org/10.1016/S0038-0717(03)00015-4

Schloss, P.D., 2020. Reintroducing mothur: 10 Years Later. Appl Environ Microbiol 86. https://doi.org/10.1128/AEM.02343-19

Selosse, M.-A., Charpin, M., Not, F., 2017. Mixotrophy everywhere on land and in water: the grand écart hypothesis. Ecology Letters 20, 246–263. https://doi.org/10.1111/ele.12714

Singer, M.B., Asfaw, D.T., Rosolem, R., Cuthbert, M.O., Miralles, D.G., MacLeod, D., Quichimbo, E.A., Michaelides, K., 2021. Hourly potential evapotranspiration at 0.1° resolution for the global land surface from 1981-present. Sci Data 8, 224. https://doi.org/10.1038/s41597-021-01003-9

Slessarev, E.W., Lin, Y., Bingham, N.L., Johnson, J.E., Dai, Y., Schimel, J.P., Chadwick, O.A., 2016. Water balance creates a threshold in soil pH at the global scale. Nature 540, 567–569. https://doi.org/10.1038/nature20139

Soares, M., Rousk, J., 2019. Microbial growth and carbon use efficiency in soil: Links to fungal-bacterial dominance, SOC-quality and stoichiometry. Soil Biology and Biochemistry 131, 195–205. https://doi.org/10.1016/j.soilbio.2019.01.010

Tester, M., Morris, C., 1987. The penetration of light through soil. Plant, Cell & Environment 10, 281–286. https://doi.org/10.1111/j.1365-3040.1987.tb01607.x

Tisserant, E., Malbreil, M., Kuo, A., Kohler, A., Symeonidi, A., Balestrini, R., Charron, P., Duensing, N., Frei dit Frey, N., Gianinazzi-Pearson, V., Gilbert, L.B., Handa, Y., Herr, J.R., Hijri, M., Koul, R., Kawaguchi, M., Krajinski, F., Lammers, P.J., Masclaux, F.G., Murat, C., Morin, E., Ndikumana, S., Pagni, M., Petitpierre, D., Requena, N., Rosikiewicz, P., Riley, R., Saito, K., San Clemente, H., Shapiro, H., van Tuinen, D., Bécard, G., Bonfante, P., Paszkowski, U., Shachar-Hill, Y.Y., Tuskan, G.A., Young, J.P.W., Sanders, I.R., Henrissat, B., Rensing, S.A., Grigoriev, I.V., Corradi, N., Roux, C., Martin, F., 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proceedings of the National Academy of Sciences 110, 20117–20122. https://doi.org/10.1073/pnas.1313452110

Tsilimigras, M.C.B., Fodor, A.A., 2016. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Annals of Epidemiology, The Microbiome and Epidemiology 26, 330–335. https://doi.org/10.1016/j.annepidem.2016.03.002

van Swaay, C.A.M., Nowicki, P., Settele, J., van Strien, A.J., 2008. Butterfly monitoring in Europe: methods, applications and perspectives. Biodivers Conserv 17, 3455–3469. https://doi.org/10.1007/s10531-008-9491-4

Watson, D.M., 2017. Sampling effort determination in bird surveys: do current norms meet best-practice recommendations? Wildl. Res. 44, 183–193. https://doi.org/10.1071/WR16226

Watt, M., Hugenholtz, P., White, R., Vinall, K., 2006. Numbers and locations of native bacteria on field-grown wheat roots quantified by fluorescence in situ hybridization (FISH). Environmental Microbiology 8, 871–884. https://doi.org/10.1111/j.1462-2920.2005.00973.x

Watts, S.C., Ritchie, S.C., Inouye, M., Holt, K.E., 2019. FastSpar: rapid and scalable correlation estimation for compositional data. Bioinformatics 35, 1064–1066. https://doi.org/10.1093/bioinformatics/bty734

Weil, R., Brady, N., 2017. The Nature and Properties of Soils. 15th edition.

Weiss, S., Van Treuren, W., Lozupone, C., Faust, K., Friedman, J., Deng, Y., Xia, L.C., Xu, Z.Z., Ursell, L., Alm, E.J., Birmingham, A., Cram, J.A., Fuhrman, J.A., Raes, J., Sun, F., Zhou, J., Knight, R., 2016. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. ISME J 10, 1669–1681. https://doi.org/10.1038/ismej.2015.235

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E.R., Knight, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5, 27. https://doi.org/10.1186/s40168-017-0237-y

West, N.E., 1990. Structure and Function of Microphytic Soil Crusts in Wildland Ecosystems of Arid to Semi-arid Regions, in: Begon, M., Fitter, A.H., Macfadyen, A. (Eds.), Advances in Ecological Research. Academic Press, pp. 179–223. https://doi.org/10.1016/S0065-2504(08)60055-0

Wieder, W.R., Bonan, G.B., Allison, S.D., 2013. Global soil carbon projections are improved by modelling microbial processes. Nature Clim Change 3, 909–912. https://doi.org/10.1038/nclimate1951

Yang, P., van Elsas, J.D., 2018. Mechanisms and ecological implications of the movement of bacteria in soil. Applied Soil Ecology 129, 112–120. https://doi.org/10.1016/j.apsoil.2018.04.014

Young, I.M., Crawford, J.W., Nunan, N., Otten, W., Spiers, A., 2008. Chapter 4 Microbial Distribution in Soils: Physics and Scaling, in: Advances in Agronomy. Academic Press, pp. 81–121. https://doi.org/10.1016/S0065-2113(08)00604-4

Zhang, T., Shi, X.-C., Ding, R., Xu, K., Tremblay, P.-L., 2020. The hidden chemolithoautotrophic metabolism of Geobacter sulfurreducens uncovered by adaptation to formate. ISME J 14, 2078–2089. https://doi.org/10.1038/s41396-020-0673-8

Zhu, G., Jetten, M.S.M., Kuschk, P., Ettwig, K.F., Yin, C., 2010. Potential roles of anaerobic ammonium and methane oxidation in the nitrogen cycle of wetland ecosystems. Appl Microbiol Biotechnol 86, 1043–1055. https://doi.org/10.1007/s00253-010-2451-4

CHAPTER 2:

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

Alteio, L. V., Séneca, J., Canarini, A., Angel, R., Jansa, J., Guseva, K., et al. (2021). A critical perspective on interpreting amplicon sequencing data in soil ecological research. *Soil Biology and Biochemistry* 160, 108357. doi: 10.1016/j.soilbio.2021.108357.

Auguet, J.-C., Barberan, A., and Casamayor, E. O. (2009). Global ecological patterns in uncultured Archaea. *ISME J* 4, 182–190. doi: 10.1038/ismej.2009.109.

Bahram, M., Hildebrand, F., Forslund, S. K., Anderson, J. L., Soudzilovskaia, N. A., Bodegom, P. M., et al. (2018). Structure and function of the global topsoil microbiome. *Nature* 560, 233–237. doi: 10.1038/s41586-018-0386-6.

Baiser, B., Olden, J. D., Record, S., Lockwood, J. L., and McKinney, M. L. (2012). Pattern and process of biotic homogenization in the New Pangaea. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20121651. doi: 10.1098/rspb.2012.1651.

Barlow, J. T., Bogatyrev, S. R., and Ismagilov, R. F. (2020). A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. *Nature Communications* 11, 2590. doi: 10.1038/s41467-020-16224-6.

Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knight, R., and Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME J* 5, 908–917. doi: 10.1038/ismej.2010.171.

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2’s q2-feature-classifier plugin. *Microbiome* 6, 90. doi: 10.1186/s40168-018-0470-z.

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

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

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6, 1621–1624. doi: 10.1038/ismej.2012.8.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS* 108, 4516–4522. doi: 10.1073/pnas.1000080107.

Epp Schmidt, D., Dlott, G., Cavigelli, M., Yarwood, S., and Maul, J. E. (2022). Soil microbiomes in three farming systems more affected by depth than farming system. *Applied Soil Ecology* 173, 104396. doi: 10.1016/j.apsoil.2022.104396.

Epp Schmidt, D. J., Pouyat, R., Szlavecz, K., Setälä, H., Kotze, D. J., Yesilonis, I., et al. (2017). Urbanization erodes ectomycorrhizal fungal diversity and may cause microbial communities to converge. *Nature Ecology & Evolution* 1, 0123. doi: 10.1038/s41559-017-0123.

Fernandes, A. D., Macklaim, J. M., Linn, T. G., Reid, G., and Gloor, G. B. (2013). ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLOS ONE* 8, e67019. doi: 10.1371/journal.pone.0067019.

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

Fierer, N., and Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *PNAS* 103, 626–631. doi: 10.1073/pnas.0507535103.

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005.

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* 0. doi: 10.3389/fmicb.2017.02224.

Gloor, G. B., and Reid, G. (2016). Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. *Can. J. Microbiol.* 62, 692–703. doi: 10.1139/cjm-2015-0821.

Gloor, G. B., Wu, J. R., Pawlowsky-Glahn, V., and Egozcue, J. J. (2016). It's all relative: analyzing microbiome data as compositions. *Ann Epidemiol* 26, 322–329. doi: 10.1016/j.annepidem.2016.03.003.

Hoshino, T., Nakao, R., Doi, H., and Minamoto, T. (2021). Simultaneous absolute quantification and sequencing of fish environmental DNA in a mesocosm by quantitative sequencing technique. *Sci Rep* 11, 4372. doi: 10.1038/s41598-021-83318-6.

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., and Korpela, K. (2020). Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLOS ONE* 15, e0227285. doi: 10.1371/journal.pone.0227285.

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. (2015). Sparse and Compositionally Robust Inference of Microbial Ecological Networks. *PLOS Computational Biology* 11, e1004226. doi: 10.1371/journal.pcbi.1004226.

Lauber, C. L., Hamady, M., Knight, R., and Fierer, N. (2009). Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale. *Appl. Environ. Microbiol.* 75, 5111–5120. doi: 10.1128/AEM.00335-09.

Lin, H., and Peddada, S. D. (2020). Analysis of compositions of microbiomes with bias correction. *Nat Commun* 11, 3514. doi: 10.1038/s41467-020-17041-7.

López-García, A., Pineda-Quiroga, C., Atxaerandio, R., Pérez, A., Hernández, I., García-Rodríguez, A., et al. (2018). Comparison of Mothur and QIIME for the Analysis of Rumen Microbiota Composition Based on 16S rRNA Amplicon Sequences. *Front. Microbiol.* 9. doi: 10.3389/fmicb.2018.03010.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15, 550. doi: 10.1186/s13059-014-0550-8.

Martino, C., Morton, J. T., Marotz, C. A., Thompson, L. R., Tripathi, A., Knight, R., et al. (2019). A Novel Sparse Compositional Technique Reveals Microbial Perturbations. *mSystems* 4, e00016-19. doi: 10.1128/mSystems.00016-19.

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

McMurdie, P. J., and Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput Biol* 10, e1003531. doi: 10.1371/journal.pcbi.1003531.

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2016). vegan: Community Ecology Package version 2.4-0. Available at: http://CRAN.R-project.org/package=vegan.

Ovaskainen, O., and Abrego, N. (2020). *Joint Species Distribution Modelling: With Applications in R (*. Cambridge University Press.

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., and Wilding, T. A. (2014). Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources* 14, 1129–1140. doi: 10.1111/1755-0998.12261.

Pawlowski, J., Lejzerowicz, F., Apotheloz-Perret-Gentil, L., Visco, J., and Esling, P. (2016). Protist metabarcoding and environmental biomonitoring: Time for change. *European Journal of Protistology* 55, 12–25. doi: 10.1016/j.ejop.2016.02.003.

Prasse, C. E., Baldwin, A. H., and Yarwood, S. A. (2015). Site History and Edaphic Features Override the Influence of Plant Species on Microbial Communities in Restored Tidal Freshwater Wetlands. *Applied and Environmental Microbiology* 81, 3482–3491. doi: 10.1128/AEM.00038-15.

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., and Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *Gigascience* 8, giz107. doi: 10.1093/gigascience/giz107.

Schloss, P. D. (2020). Reintroducing mothur: 10 Years Later. *Appl Environ Microbiol* 86. doi: 10.1128/AEM.02343-19.

Shade, A., and Stopnisek, N. (2019). Abundance-occupancy distributions to prioritize plant core microbiome membership. *Current Opinion in Microbiology* 49, 50–58. doi: 10.1016/j.mib.2019.09.008.

Smets, W., Leff, J. W., Bradford, M. A., McCulley, R. L., Lebeer, S., and Fierer, N. (2016). A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biology and Biochemistry* 96, 145–151. doi: 10.1016/j.soilbio.2016.02.003.

Smith, J. R., Letten, A. D., Ke, P.-J., Anderson, C. B., Hendershot, J. N., Dhami, M. K., et al. (2018). A global test of ecoregions. *Nat Ecol Evol* 2, 1889–1896. doi: 10.1038/s41559-018-0709-x.

Stämmler, F., Gläsner, J., Hiergeist, A., Holler, E., Weber, D., Oefner, P. J., et al. (2016). Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome* 4, 28. doi: 10.1186/s40168-016-0175-0.

Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., et al. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution* 11, 442–447. doi: 10.1111/2041-210X.13345.

Tourlousse, D. M., Yoshiike, S., Ohashi, A., Matsukura, S., Noda, N., and Sekiguchi, Y. (2017). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Research* 45, e23. doi: 10.1093/nar/gkw984.

Tsilimigras, M. C. B., and Fodor, A. A. (2016). Compositional data analysis of the microbiome: fundamentals, tools, and challenges. *Annals of Epidemiology* 26, 330–335. doi: 10.1016/j.annepidem.2016.03.002.

Vandeputte, D., Kathagen, G., D'hoe, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., et al. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 551, 507–511. doi: 10.1038/nature24460.

Watts, S. C., Ritchie, S. C., Inouye, M., and Holt, K. E. (2019). FastSpar: rapid and scalable correlation estimation for compositional data. *Bioinformatics* 35, 1064–1066. doi: 10.1093/bioinformatics/bty734.

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27. doi: 10.1186/s40168-017-0237-y.

CHAPTER 3:

Alvarez, L., Bricio, C., José Gómez, M., Berenguer, J., 2011. Lateral Transfer of the Denitrification Pathway Genes among Thermus thermophilus Strains. Appl Environ Microbiol 77, 1352–1358. https://doi.org/10.1128/AEM.02048-10

Alves, R.J.E., Minh, B.Q., Urich, T., von Haeseler, A., Schleper, C., 2018. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. Nat Commun 9, 1517. https://doi.org/10.1038/s41467-018-03861-1

Angel, R., Matthies, D., Conrad, R., 2011. Activation of Methanogenesis in Arid Biological Soil Crusts Despite the Presence of Oxygen. PLOS ONE 6, e20453. https://doi.org/10.1371/journal.pone.0020453

Apothéloz-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., Pawlowski, J., 2017. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Molecular Ecology Resources 17, 1231–1242. https://doi.org/10.1111/1755-0998.12668

Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31, 2882–2884. https://doi.org/10.1093/bioinformatics/btv287

Bahram, M., Espenberg, M., Pärn, J., Lehtovirta-Morley, L., Anslan, S., Kasak, K., Kõljalg, U., Liira, J., Maddison, M., Moora, M., Niinemets, Ü., Öpik, M., Pärtel, M., Soosaar, K., Zobel, M., Hildebrand, F., Tedersoo, L., Mander, Ü., 2022. Structure and function of the soil microbiome underlying N2O emissions from global wetlands. Nat Commun 13, 1430. https://doi.org/10.1038/s41467-022-29161-3

Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. Nature 515, 505–511. https://doi.org/10.1038/nature13855

Barlow, J.T., Bogatyrev, S.R., Ismagilov, R.F., 2020. A quantitative sequencing framework for absolute abundance measurements of mucosal and lumenal microbial communities. Nature Communications 11, 2590. https://doi.org/10.1038/s41467-020-16224-6

Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen, E., Knight, R., Huttley, G.A., Gregory Caporaso, J., 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-

feature-classifier plugin. Microbiome 6, 90. https://doi.org/10.1186/s40168-018-0470-z

Bolhuis, H., Severin, I., Confurius-Guns, V., Wollenzien, U.I.A., Stal, L.J., 2010. Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium Microcoleus chthonoplastes. ISME J 4, 121–130. https://doi.org/10.1038/ismej.2009.99

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7, 335–336. https://doi.org/10.1038/nmeth.f.303

Cavicchioli, R., Ripple, W.J., Timmis, K.N., Azam, F., Bakken, L.R., Baylis, M., Behrenfeld, M.J., Boetius, A., Boyd, P.W., Classen, A.T., Crowther, T.W., Danovaro, R., Foreman, C.M., Huisman, J., Hutchins, D.A., Jansson, J.K., Karl, D.M., Koskella, B., Mark Welch, D.B., Martiny, J.B.H., Moran, M.A., Orphan, V.J., Reay, D.S., Remais, J.V., Rich, V.I., Singh, B.K., Stein, L.Y., Stewart, F.J., Sullivan, M.B., van Oppen, M.J.H., Weaver, S.C., Webb, E.A., Webster, N.S., 2019. Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol 17, 569–586. https://doi.org/10.1038/s41579-019-0222-5

Chase, A.B., Weihe, C., Martiny, J.B.H., 2021. Adaptive differentiation and rapid evolution of a soil bacterium along a climate gradient. PNAS 118. https://doi.org/10.1073/pnas.2101254118

Christman, G.D., Cottrell, M.T., Popp, B.N., Gier, E., Kirchman, D.L., 2011. Abundance, Diversity, and Activity of Ammonia-Oxidizing Prokaryotes in the Coastal Arctic Ocean in Summer and Winter. Applied and Environmental Microbiology 77, 2026–2034. https://doi.org/10.1128/AEM.01907-10

Coelho, L.P., Alves, R., del Río, Á.R., Myers, P.N., Cantalapiedra, C.P., Giner-Lamia, J., Schmidt, T.S., Mende, D.R., Orakov, A., Letunic, I., Hildebrand, F., Van Rossum, T., Forslund, S.K., Khedkar, S., Maistrenko, O.M., Pan, S., Jia, L., Ferretti, P., Sunagawa, S., Zhao, X.-M., Nielsen, H.B., Huerta-Cepas, J., Bork, P., 2022. Towards the biogeography of prokaryotic genes. Nature 601, 252–256. https://doi.org/10.1038/s41586-021-04233-4

Cordier, T., Forster, D., Dufresne, Y., Martins, C.I.M., Stoeck, T., Pawlowski, J., 2018. Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding applied to biomonitoring. Molecular Ecology Resources 18, 1381–1391. https://doi.org/10.1111/1755-0998.12926

Cordier, T., Lanzén, A., Apothéloz-Perret-Gentil, L., Stoeck, T., Pawlowski, J., 2019. Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring. Trends in Microbiology 27, 387–397. https://doi.org/10.1016/j.tim.2018.10.012

Crits-Christoph, A., Olm, M.R., Diamond, S., Bouma-Gregson, K., Banfield, J.F., 2020. Soil bacterial populations are shaped by recombination and gene-specific selection across a grassland meadow. ISME J 14, 1834–1846. https://doi.org/10.1038/s41396-020-0655-x

Delorme, S., Philippot, L., Edel-Hermann, V., Deulvot, C., Mougel, C., Lemanceau, P., 2003. Comparative Genetic Diversity of the narG, nosZ, and 16S rRNA Genes in Fluorescent Pseudomonads. Applied and Environmental Microbiology 69, 1004–1012. https://doi.org/10.1128/AEM.69.2.1004-1012.2003

Djemiel, C., Maron, P.-A., Terrat, S., Dequiedt, S., Cottin, A., Ranjard, L., 2022. Inferring microbiota functions from taxonomic genes: a review. GigaScience 11, giab090. https://doi.org/10.1093/gigascience/giab090

Douglas, G.M., Maffei, V.J., Zaneveld, J.R., Yurgel, S.N., Brown, J.R., Taylor, C.M., Huttenhower, C., Langille, M.G.I., 2020. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol 38, 685–688. https://doi.org/10.1038/s41587-020-0548-6

Epp Schmidt, D., Dlott, G., Cavigelli, M., Yarwood, S., Maul, J.E., 2022. Soil microbiomes in three farming systems more affected by depth than farming system. Applied Soil Ecology 173, 104396. https://doi.org/10.1016/j.apsoil.2022.104396

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

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. PNAS 103, 626–631. https://doi.org/10.1073/pnas.0507535103

Fierer, N., Wood, S.A., Bueno de Mesquita, C.P., 2021. How microbes can, and cannot, be used to assess soil health. Soil Biology and Biochemistry 153, 108111. https://doi.org/10.1016/j.soilbio.2020.108111

Gaby, J.C., Buckley, D.H., 2014. A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. Database (Oxford) 2014. https://doi.org/10.1093/database/bau001

Galand, P.E., Lovejoy, C., Hamilton, A.K., Ingram, R.G., Pedneault, E., Carmack, E.C., 2009. Archaeal diversity and a gene for ammonia oxidation are coupled to oceanic circulation. Environmental Microbiology 11, 971–980. https://doi.org/10.1111/j.1462-2920.2008.01822.x

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 0. https://doi.org/10.3389/fmicb.2017.02224

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., Egozcue, J.J., 2016. It’s all relative: analyzing microbiome data as compositions. Ann Epidemiol 26, 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003

Graf, D.R.H., Jones, C.M., Hallin, S., 2014. Intergenomic Comparisons Highlight Modularity of the Denitrification Pathway and Underpin the Importance of Community Structure for N2O Emissions. PLOS ONE 9, e114118. https://doi.org/10.1371/journal.pone.0114118

Henry, S., Baudoin, E., López-Gutiérrez, J.C., Martin-Laurent, F., Brauman, A., Philippot, L., 2004. Quantification of denitrifying bacteria in soils by nirK gene targeted real-time PCR. Journal of Microbiological Methods 59, 327–335. https://doi.org/10.1016/j.mimet.2004.07.002

Henry, S., Bru, D., Stres, B., Hallet, S., Philippot, L., 2006. Quantitative detection of the nosZ gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, narG, nirK, and nosZ genes in soils. Appl Environ Microbiol 72, 5181–5189. https://doi.org/10.1128/AEM.00231-06

Heylen, K., Gevers, D., Vanparys, B., Wittebolle, L., Geets, J., Boon, N., De Vos, P., 2006. The incidence of nirS and nirK and their genetic heterogeneity in cultivated denitrifiers. Environmental Microbiology 8, 2012–2021. https://doi.org/10.1111/j.1462-2920.2006.01081.x

Horn, M.A., Drake, H.L., Schramm, A., 2006. Nitrous Oxide Reductase Genes (nosZ) of Denitrifying Microbial Populations in Soil and the Earthworm

Gut Are Phylogenetically Similar. Appl. Environ. Microbiol. 72, 1019–1026. https://doi.org/10.1128/AEM.72.2.1019-1026.2006

Im, J., Lee, S.-W., Bodrossy, L., Barcelona, M.J., Semrau, J.D., 2011. Field application of nitrogen and phenylacetylene to mitigate greenhouse gas emissions from landfill cover soils: effects on microbial community structure. Appl Microbiol Biotechnol 89, 189–200. https://doi.org/10.1007/s00253-010-2811-0

Iwai, S., Weinmaier, T., Schmidt, B.L., Albertson, D.G., Poloso, N.J., Dabbagh, K., DeSantis, T.Z., 2016. Piphillin: Improved Prediction of Metagenomic Content by Direct Inference from Human Microbiomes. PLOS ONE 11, e0166104. https://doi.org/10.1371/journal.pone.0166104

Jang, J., Sakai, Y., Senoo, K., Ishii, S., 2019. Potentially Mobile Denitrification Genes Identified in Azospirillum sp. Strain TSH58. Applied and Environmental Microbiology 85, e02474-18. https://doi.org/10.1128/AEM.02474-18

Jian, C., Luukkonen, P., Yki-Jarvinen, H., Salonen, A., Korpela, K., 2018. Quantitative PCR provides a simple and accessible method for quantitative microbiome profiling. bioRxiv 478685. https://doi.org/10.1101/478685

Jones, C.M., Graf, D.R., Bru, D., Philippot, L., Hallin, S., 2013. The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink. ISME J 7, 417–426. https://doi.org/10.1038/ismej.2012.125

Jones, C.M., Stres, B., Rosenquist, M., Hallin, S., 2008. Phylogenetic Analysis of Nitrite, Nitric Oxide, and Nitrous Oxide Respiratory Enzymes Reveal a Complex Evolutionary History for Denitrification. Molecular Biology and Evolution 25, 1955–1966. https://doi.org/10.1093/molbev/msn146

Jun, S.-R., Robeson, M.S., Hauser, L.J., Schadt, C.W., Gorin, A.A., 2015. PanFP: pangenome-based functional profiles for microbial communities. BMC Research Notes 8, 479. https://doi.org/10.1186/s13104-015-1462-8

Kandeler, E., Deiglmayr, K., Tscherko, D., Bru, D., Philippot, L., 2006. Abundance of narG, nirS, nirK, and nosZ Genes of Denitrifying Bacteria during Primary Successions of a Glacier Foreland. Applied and Environmental Microbiology 72, 5957–5962. https://doi.org/10.1128/AEM.00439-06

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016a. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44, D457-462. https://doi.org/10.1093/nar/gkv1070

Kanehisa, M., Sato, Y., Morishima, K., 2016b. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol 428, 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

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

Louca, S., Jacques, S.M.S., Pires, A.P.F., Leal, J.S., Srivastava, D.S., Parfrey, L.W., Farjalla, V.F., Doebeli, M., 2016a. High taxonomic variability despite stable functional structure across microbial communities. Nat Ecol Evol 1, 1–12. https://doi.org/10.1038/s41559-016-0015

Louca, S., Parfrey, L.W., Doebeli, M., 2016b. Decoupling function and taxonomy in the global ocean microbiome. Science 353, 1272–1277. https://doi.org/10.1126/science.aaf4507

Mantri, S.S., Negri, T., Sales-Ortells, H., Angelov, A., Peter, S., Neidhardt, H., Oelmann, Y., Ziemert, N., 2021. Metagenomic Sequencing of Multiple Soil Horizons and Sites in Close Vicinity Revealed Novel Secondary Metabolite Diversity. mSystems. https://doi.org/10.1128/mSystems.01018-21

Meinhardt, K.A., Bertagnolli, A., Pannu, M.W., Strand, S.E., Brown, S.L., Stahl, D.A., 2015. Evaluation of revised polymerase chain reaction primers for more inclusive quantification of ammonia-oxidizing archaea and bacteria. Environmental Microbiology Reports 7, 354–363. https://doi.org/10.1111/1758-2229.12259

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M., DeLong, E.F., 2007. Quantitative distribution of presumptive archaeal and bacterial nitrifiers in Monterey Bay and the North Pacific Subtropical Gyre. Environmental Microbiology 9, 1162–1175. https://doi.org/10.1111/j.1462-2920.2007.01239.x

Mise, K., Masuda, Y., Senoo, K., Itoh, H., 2021. Undervalued Pseudo-nifH Sequences in Public Databases Distort Metagenomic Insights into

Biological Nitrogen Fixers. mSphere 6, e00785-21. https://doi.org/10.1128/msphere.00785-21

Morett, E., Moreno, S., Espin, G., 1988. Transcription analysis of the three nifH genes of Rhizobium phaseoli with gene fusions. Mol Gen Genet 213, 499–504. https://doi.org/10.1007/BF00339622

Narayan, N.R., Weinmaier, T., Laserna-Mendieta, E.J., Claesson, M.J., Shanahan, F., Dabbagh, K., Iwai, S., DeSantis, T.Z., 2020. Piphillin predicts metagenomic composition and dynamics from DADA2-corrected 16S rDNA sequences. BMC Genomics 21, 56. https://doi.org/10.1186/s12864-019-6427-1

Norton, J.M., Alzerreca, J.J., Suwa, Y., Klotz, M.G., 2002. Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. Arch Microbiol 177, 139–149. https://doi.org/10.1007/s00203-001-0369-z

Palmer, K., Drake, H.L., Horn, M.A., 2009. Genome-Derived Criteria for Assigning Environmental narG and nosZ Sequences to Operational Taxonomic Units of Nitrate Reducers. Applied and Environmental Microbiology 75, 5170–5174. https://doi.org/10.1128/AEM.00254-09

Pereyra, L.P., Hiibel, S.R., Prieto Riquelme, M.V., Reardon, K.F., Pruden, A., 2010. Detection and quantification of functional genes of cellulose-degrading, fermentative, and sulfate-reducing bacteria and methanogenic archaea. Appl Environ Microbiol 76, 2192–2202. https://doi.org/10.1128/AEM.01285-09

Pester, M., Rattei, T., Flechl, S., Gröngröft, A., Richter, A., Overmann, J., Reinhold-Hurek, B., Loy, A., Wagner, M., 2012. amoA-based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of amoA genes from soils of four different geographic regions. Environmental Microbiology 14, 525–539. https://doi.org/10.1111/j.1462-2920.2011.02666.x

Petrova, L.P., Varshalomidze, O.E., Shelud’ko, A.V., Katsy, E.I., 2010. Localization of denitrification genes in plasmid DNA of bacteria Azospirillum brasilense. Russ J Genet 46, 801–807. https://doi.org/10.1134/S1022795410070045

Philippot, L., Andersson, S.G.E., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B., Hallin, S., 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Micro 8, 523–529. https://doi.org/10.1038/nrmicro2367

Pii, Y., Crimi, M., Cremonese, G., Spena, A., Pandolfini, T., 2007. Auxin and nitric oxide control indeterminate nodule formation. BMC Plant Biology 7, 21. https://doi.org/10.1186/1471-2229-7-21

Porter, T.M., Hajibabaei, M., 2018. Scaling up: A guide to high-throughput genomic approaches for biodiversity analysis. Molecular Ecology 27, 313–338. https://doi.org/10.1111/mec.14478

Pothier, J.F., Prigent-Combaret, C., Haurat, J., Moënne-Loccoz, Y., Wisniewski-Dyé, F., 2008. Duplication of Plasmid-Borne Nitrite Reductase Gene nirK in the Wheat-Associated Plant Growth–Promoting Rhizobacterium Azospirillum brasilense Sp245. MPMI 21, 831–842. https://doi.org/10.1094/MPMI-21-6-0831

Quinn, T.P., Erb, I., Gloor, G., Notredame, C., Richardson, M.F., Crowley, T.M., 2019. A field guide for the compositional analysis of any-omics data. Gigascience 8, giz107. https://doi.org/10.1093/gigascience/giz107

Rösch, C., Mergel, A., Bothe, H., 2002. Biodiversity of denitrifying and dinitrogen-fixing bacteria in an acid forest soil. Appl Environ Microbiol 68, 3818–3829. https://doi.org/10.1128/AEM.68.8.3818-3829.2002

Rotthauwe, J.H., Witzel, K.P., Liesack, W., 1997. The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. Appl. Environ. Microbiol. 63, 4704–4712.

Sansupa, C., Wahdan, S.F.M., Hossen, S., Disayathanoowat, T., Wubet, T., Purahong, W., 2021. Can we use functional annotation of prokaryotic taxa (FAPROTAX) to assign the ecological functions of soil bacteria? Applied Sciences 11, 688. https://doi.org/10.3390/app11020688

Signorelli, S., Sainz, M., Tabares-da Rosa, S., Monza, J., 2020. The Role of Nitric Oxide in Nitrogen Fixation by Legumes. Frontiers in Plant Science 11.

Suman, J., Rakshit, A., Ogireddy, S.D., Singh, S., Gupta, C., Chandrakala, J., 2022. Microbiome as a Key Player in Sustainable Agriculture and Human Health. Frontiers in Soil Science 2.

Thiel, T., Pratte, B.S., 2014. Regulation of Three Nitrogenase Gene Clusters in the Cyanobacterium Anabaena variabilis ATCC 29413. Life 4, 944–967. https://doi.org/10.3390/life4040944

Throbäck, I.N., Enwall, K., Jarvis, A., Hallin, S., 2004. Reassessing PCR primers targeting nirS, nirK and nosZ genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiol Ecol 49, 401–417. https://doi.org/10.1016/j.femsec.2004.04.011

Tourna, M., Freitag, T.E., Nicol, G.W., Prosser, J.I., 2008. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. Environmental Microbiology 10, 1357–1364. https://doi.org/10.1111/j.1462-2920.2007.01563.x

Wang, H., Bagnoud, A., Ponce-Toledo, R.I., Kerou, M., Weil, M., Schleper, C., Urich, T., 2016. Linking 16S rRNA Gene Classification to amoA Gene Taxonomy Reveals Environmental Distribution of Ammonia-Oxidizing Archaeal Clades in Peatland Soils. mSystems 6, e00546-21. https://doi.org/10.1128/mSystems.00546-21

Wei, W., Isobe, K., Nishizawa, T., Zhu, L., Shiratori, Y., Ohte, N., Koba, K., Otsuka, S., Senoo, K., 2015. Higher diversity and abundance of denitrifying microorganisms in environments than considered previously. ISME J 9, 1954–1965. https://doi.org/10.1038/ismej.2015.9

Wemheuer, F., Taylor, J.A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T., Wemheuer, B., 2020. Tax4Fun2: prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene sequences. Environmental Microbiome 15, 11. https://doi.org/10.1186/s40793-020-00358-7

Wemheuer, F., Taylor, J.A., Daniel, R., Johnston, E., Meinicke, P., Thomas, T., Wemheuer, B., 2018. Tax4Fun2: a R-based tool for the rapid prediction of habitat-specific functional profiles and functional redundancy based on 16S rRNA gene marker gene sequences (preprint). Bioinformatics. https://doi.org/10.1101/490037

CHAPTER 4:

Bach, E.M., Ramirez, K.S., Fraser, T.D., Wall, D.H., 2020. Soil Biodiversity Integrates Solutions for a Sustainable Future. Sustainability 12, 2662. https://doi.org/10.3390/su12072662

Bickel, S., Or, D., 2020. Soil bacterial diversity mediated by microscale aqueous-phase processes across biomes. Nat Commun 11, 116. https://doi.org/10.1038/s41467-019-13966-w

Brune, A., Ludwig, W., Schink, B., 2002. Propionivibrio limicola sp. nov., a fermentative bacterium specialized in the degradation of hydroaromatic compounds, reclassification of Propionibacter pelophilus as

Propionivibrio pelophilus comb. nov. and amended description of the genus Propionivibrio. Int J Syst Evol Microbiol 52, 441–444. https://doi.org/10.1099/00207713-52-2-441

Cáceres, M.D., Legendre, P., 2009. Associations between species and groups of sites: indices and statistical inference. Ecology 90, 3566–3574. https://doi.org/10.1890/08-1823.1

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

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

Cavigelli, M., Teasdale, J.R., Conklin, A.E., 2008. Long-Term Agronomic Performance of Organic and Conventional Field Crops in the Mid-Atlantic Region. Agron. J. 100. https://doi.org/10.2134/agronj2006.0373

Chen, H., Dai, Z., Veach, A.M., Zheng, J., Xu, J., Schadt, C.W., 2020. Global meta-analyses show that conservation tillage practices promote soil fungal and bacterial biomass. Agriculture, Ecosystems & Environment 293, 106841. https://doi.org/10.1016/j.agee.2020.106841

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., Tiedje, J.M., 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37, D141–D145. https://doi.org/10.1093/nar/gkn879

Csárdi, G., Nepusz, T., 2006. The igraph software package for complex network research. undefined.

Daws, S.C., Cline, L.A., Rotenberry, J., Sadowsky, M.J., Staley, C., Dalzell, B., Kennedy, P.G., 2020. Do shared traits create the same fates? Examining the link between morphological type and the biogeography of fungal and bacterial communities. Fungal Ecology, Trait-based approaches to Fungal Ecology 46, 100948. https://doi.org/10.1016/j.funeco.2020.100948

de Graaff, M.-A., Hornslein, N., Throop, H.L., Kardol, P., van Diepen, L.T.A., 2019. Chapter One - Effects of agricultural intensification on soil biodiversity and implications for ecosystem functioning: A meta-analysis, in: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 1–44. https://doi.org/10.1016/bs.agron.2019.01.001

Degani, O., Khatib, S., Becher, P., Gordani, A., Harris, R., 2021. Trichoderma asperellum Secreted 6-Pentyl-α-Pyrone to Control Magnaporthiopsis maydis, the Maize Late Wilt Disease Agent. Biology 10, 897. https://doi.org/10.3390/biology10090897

Dong, W., Liu, E., Yan, C., Tian, J., Zhang, H., Zhang, Y., 2017. Impact of no tillage vs. conventional tillage on the soil bacterial community structure in a winter wheat cropping succession in northern China. European Journal of Soil Biology 80, 35–42. https://doi.org/10.1016/j.ejsobi.2017.03.001

Duruigbo, C.I., Obiefuna, J.C., Onweremadu, E.U., 2007. Effect of Poultry Manure Rates on Soil Acidity in an Ultisol. International Journal of Soil Science. https://doi.org/10.3923/ijss.2007.154.158

Egidi, E., Delgado-Baquerizo, M., Plett, J.M., Wang, J., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., 2019. A few Ascomycota taxa dominate soil fungal communities worldwide. Nat Commun 10, 2369. https://doi.org/10.1038/s41467-019-10373-z

Epp Schmidt, D.J., Pouyat, R., Szlavecz, K., Setälä, H., Kotze, D.J., Yesilonis, I., Cilliers, S., Hornung, E., Dombos, M., Yarwood, S.A., 2017. Urbanization erodes ectomycorrhizal fungal diversity and may cause microbial communities to converge. Nature Ecology & Evolution 1, 0123. https://doi.org/10.1038/s41559-017-0123

Fierer, N., Jackson, R.B., 2006. The diversity and biogeography of soil bacterial communities. PNAS 103, 626–631. https://doi.org/10.1073/pnas.0507535103

Fierer, N., Schimel, J.P., Holden, P.A., 2003. Variations in microbial community composition through two soil depth profiles. Soil Biology and Biochemistry 35, 167–176. https://doi.org/10.1016/S0038-0717(02)00251-1

Fierer, N., Strickland, M.S., Liptzin, D., Bradford, M.A., Cleveland, C.C., 2009. Global patterns in belowground communities. Ecology Letters 12, 1238–1249. https://doi.org/10.1111/j.1461-0248.2009.01360.x

Fierer, N., Wood, S.A., Bueno de Mesquita, C.P., 2021. How microbes can, and cannot, be used to assess soil health. Soil Biology and

Biochemistry 153, 108111. https://doi.org/10.1016/j.soilbio.2020.108111

Finney, D.M., Buyer, J.S., Kaye, J.P., 2017. Living cover crops have immediate impacts on soil microbial community structure and function. Journal of Soil and Water Conservation 72, 361–373. https://doi.org/10.2489/jswc.72.4.361

Franke-Snyder, M., Douds, D.D., Galvez, L., Phillips, J.G., Wagoner, P., Drinkwater, L., Morton, J.B., 2001. Diversity of communities of arbuscular mycorrhizal (AM) fungi present in conventional versus low-input agricultural sites in eastern Pennsylvania, USA. Applied Soil Ecology 16, 35–48. https://doi.org/10.1016/S0929-1393(00)00100-1

Gee, G., Bauder, J., 1986. 1986, Particle-size analysis. p. 383-411. In A. Klute(ed.) Methods of soil analysis, Part 1. Agron. Monogr. 9. ASA and SSSA, Madison, WI.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome Datasets Are Compositional: And This Is Not Optional. Front. Microbiol. 0. https://doi.org/10.3389/fmicb.2017.02224

Gloor, G.B., Wu, J.R., Pawlowsky-Glahn, V., Egozcue, J.J., 2016. It’s all relative: analyzing microbiome data as compositions. Ann Epidemiol 26, 322–329. https://doi.org/10.1016/j.annepidem.2016.03.003

Griffin, D.M., 1985. A Comparison of the Roles of Bacteria and Fungi, in: Leadbetter, E.R., Poindexter, J.S. (Eds.), Bacteria in Nature: Volume 1: Bacterial Activities in Perspective, Bacteria in Nature. Springer US, Boston, MA, pp. 221–255. https://doi.org/10.1007/978-1-4615-6511-6_8

Hao, J., Chai, Y.N., Lopes, L.D., Ordóñez, R.A., Wright, E.E., Archontoulis, S., Schachtman, D.P., 2021. The Effects of Soil Depth on the Structure of Microbial Communities in Agricultural Soils in Iowa (United States). Applied and Environmental Microbiology 87, e02673-20. https://doi.org/10.1128/AEM.02673-20

Helgason, B.L., Walley, F.L., Germida, J.J., 2010. No-till soil management increases microbial biomass and alters community profiles in soil aggregates. Applied Soil Ecology 46, 390–397. https://doi.org/10.1016/j.apsoil.2010.10.002

Hug, L.A., Castelle, C.J., Wrighton, K.C., Thomas, B.C., Sharon, I., Frischkorn, K.R., Williams, K.H., Tringe, S.G., Banfield, J.F., 2013. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome 1, 22. https://doi.org/10.1186/2049-2618-1-22

Iebba, V., Santangelo, F., Totino, V., Nicoletti, M., Gagliardi, A., Biase, R.V.D., Cucchiara, S., Nencioni, L., Conte, M.P., Schippa, S., 2013. Higher Prevalence and Abundance of Bdellovibrio bacteriovorus in the Human Gut of Healthy Subjects. PLOS ONE 8, e61608. https://doi.org/10.1371/journal.pone.0061608

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., Korpela, K., 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLOS ONE 15, e0227285. https://doi.org/10.1371/journal.pone.0227285

Jian, C., Luukkonen, P., Yki-Jarvinen, H., Salonen, A., Korpela, K., 2018. Quantitative PCR provides a simple and accessible method for quantitative microbiome profiling. bioRxiv 478685. https://doi.org/10.1101/478685

Kielak, A.M., Barreto, C.C., Kowalchuk, G.A., van Veen, J.A., Kuramae, E.E., 2016. The Ecology of Acidobacteria: Moving beyond Genes and Genomes. Frontiers in Microbiology 7, 744. https://doi.org/10.3389/fmicb.2016.00744

King, A.E., Hofmockel, K.S., 2017. Diversified cropping systems support greater microbial cycling and retention of carbon and nitrogen. Agriculture, Ecosystems & Environment 240, 66–76. https://doi.org/10.1016/j.agee.2017.01.040

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: Concept & review. Soil Biology and Biochemistry 83, 184–199. https://doi.org/10.1016/j.soilbio.2015.01.025

Leeuwen, J.P. van, Djukic, I., Bloem, J., Lehtinen, T., Hemerik, L., Ruiter, P.C. de, Lair, G.J., 2017. Effects of land use on soil microbial biomass, activity and community structure at different soil depths in the Danube floodplain. European Journal of Soil Biology 79, 14–20. https://doi.org/10.1016/j.ejsobi.2017.02.001

Li, F., Zhang, S., Wang, Y., Li, Y., Li, P., Chen, L., Jie, X., Hu, D., Feng, B., Yue, K., Han, Y., 2020. Rare fungus, Mortierella capitata, promotes crop growth by stimulating primary metabolisms related genes and reshaping rhizosphere bacterial community. Soil Biology and Biochemistry 151, 108017. https://doi.org/10.1016/j.soilbio.2020.108017

Liao, H.-L., Bonito, G., Rojas, J.A., Hameed, K., Wu, S., Schadt, C.W., Labbé, J., Tuskan, G.A., Martin, F., Grigoriev, I.V., Vilgalys, R., 2019. Fungal Endophytes of Populus trichocarpa Alter Host Phenotype, Gene Expression, and Rhizobiome Composition. MPMI 32, 853–864. https://doi.org/10.1094/MPMI-05-18-0133-R

Luo, J., Zhang, N., 2013. Magnaporthiopsis, a new genus in Magnaporthaceae (Ascomycota). Mycologia 105, 1019–1029. https://doi.org/10.3852/12-359

Malik, A.A., Martiny, J.B.H., Brodie, E.L., Martiny, A.C., Treseder, K.K., Allison, S.D., 2020. Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. ISME J 14, 1–9. https://doi.org/10.1038/s41396-019-0510-0

Mbuthia, L.W., Acosta-Martínez, V., DeBruyn, J., Schaeffer, S., Tyler, D., Odoi, E., Mpheshea, M., Walker, F., Eash, N., 2015. Long term tillage, cover crop, and fertilization effects on microbial community structure, activity: Implications for soil quality. Soil Biology and Biochemistry 89, 24–34. https://doi.org/10.1016/j.soilbio.2015.06.016

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput Biol 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Mestelan, S., Smeck, N., Sprunger, C., Dyck, A., Dick, W., 2021. Four decades of continuously applied tillage or no-tillage on soil properties and soil morphology. Agrosystems, Geosciences & Environment 4, e20195. https://doi.org/10.1002/agg2.20195

Nilsson, R.H., Larsson, K.-H., Taylor, A.F.S., Bengtsson-Palme, J., Jeppesen, T.S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F.O., Tedersoo, L., Saar, I., Kõljalg, U., Abarenkov, K., 2019. The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. Nucleic Acids Research 47, D259–D264. https://doi.org/10.1093/nar/gky1022

Ogle, S.M., Alsaker, C., Baldock, J., Bernoux, M., Breidt, F.J., McConkey, B., Regina, K., Vazquez-Amabile, G.G., 2019. Climate and Soil Characteristics Determine Where No-Till Management Can Store

Carbon in Soils and Mitigate Greenhouse Gas Emissions. Sci Rep 9, 11665. https://doi.org/10.1038/s41598-019-47861-7

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O'Hara, R.B., Simpson, G., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Wagner, H., 2016. vegan: Community Ecology Package version 2.4-0.

Oliverio, A.M., Bissett, A., McGuire, K., Saltonstall, K., Turner, B.L., Fierer, N., 2020. The Role of Phosphorus Limitation in Shaping Soil Bacterial Communities and Their Metabolic Capabilities. mBio 11, e01718-20. https://doi.org/10.1128/mBio.01718-20

Peng, X., Bruns, M.A., 2019a. Development of a nitrogen-fixing cyanobacterial consortium for surface stabilization of agricultural soils. Journal of Applied Phycology 31, 1047–1056. https://doi.org/10.1007/s10811-018-1597-9

Peng, X., Bruns, M.A., 2019b. Cyanobacterial Soil Surface Consortia Mediate N Cycle Processes in Agroecosystems. Frontiers in Environmental Science 6, 156. https://doi.org/10.3389/fenvs.2018.00156

Põlme, S., Abarenkov, K., Henrik Nilsson, R., Lindahl, B.D., Clemmensen, K.E., Kauserud, H., Nguyen, N., Kjøller, R., Bates, S.T., Baldrian, P., Frøslev, T.G., Adojaan, K., Vizzini, A., Suija, A., Pfister, D., Baral, H.-O., Järv, H., Madrid, H., Nordén, J., Liu, J.-K., Pawlowska, J., Põldmaa, K., Pärtel, K., Runnel, K., Hansen, K., Larsson, K.-H., Hyde, K.D., Sandoval-Denis, M., Smith, M.E., Toome-Heller, M., Wijayawardene, N.N., Menolli, N., Reynolds, N.K., Drenkhan, R., Maharachchikumbura, S.S.N., Gibertoni, T.B., Læssøe, T., Davis, W., Tokarev, Y., Corrales, A., Soares, A.M., Agan, A., Machado, A.R., Argüelles-Moyao, A., Detheridge, A., de Meiras-Ottoni, A., Verbeken, A., Dutta, A.K., Cui, B.-K., Pradeep, C.K., Marín, C., Stanton, D., Gohar, D., Wanasinghe, D.N., Otsing, E., Aslani, F., Griffith, G.W., Lumbsch, T.H., Grossart, H.-P., Masigol, H., Timling, I., Hiiesalu, I., Oja, J., Kupagme, J.Y., Geml, J., Alvarez-Manjarrez, J., Ilves, K., Loit, K., Adamson, K., Nara, K., Küngas, K., Rojas-Jimenez, K., Bitenieks, K., Irinyi, L., Nagy, L.L., Soonvald, L., Zhou, L.-W., Wagner, L., Aime, M.C., Öpik, M., Mujica, M.I., Metsoja, M., Ryberg, M., Vasar, M., Murata, M., Nelsen, M.P., Cleary, M., Samarakoon, M.C., Doilom, M., Bahram, M., Hagh-Doust, N., Dulya, O., Johnston, P., Kohout, P., Chen, Q., Tian, Q., Nandi, R., Amiri, R., Perera, R.H., dos Santos Chikowski, R., Mendes-Alvarenga, R.L., Garibay-Orijel, R., Gielen, R., Phookamsak, R., Jayawardena, R.S., Rahimlou, S., Karunarathna, S.C., Tibpromma, S., Brown, S.P., Sepp, S.-K., Mundra, S., Luo, Z.-H., Bose, T., Vahter, T., Netherway, T., Yang, T., May, T., Varga, T., Li, W., Coimbra, V.R.M., de Oliveira, V.R.T., de Lima, V.X., Mikryukov, V.S., Lu, Y., Matsuda, Y., Miyamoto, Y., Kõljalg, U., Tedersoo, L., 2020. FungalTraits: a user-friendly traits database of fungi and fungus-like stramenopiles. Fungal Diversity 105, 1–16. https://doi.org/10.1007/s13225-020-00466-2

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41, D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team, 2015. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Schmidt, R., Gravuer, K., Bossange, A.V., Mitchell, J., Scow, K., 2018. Long-term use of cover crops and no-till shift soil microbial community life strategies in agricultural soil. PLOS ONE 13, e0192953. https://doi.org/10.1371/journal.pone.0192953

Smith, D.P., Peay, K.G., 2014. Sequence depth, not PCR replication, improves ecological inference from next generation DNA sequencing. PLoS ONE 9, e90234. https://doi.org/10.1371/journal.pone.0090234

Soil Survey Staff, 2022. Web Soil Survey.

Soil Survey Staff, 2014. Keys to Soil Taxonomy. USDA-NRCS 12.

Sood, M., Kapoor, D., Kumar, V., Sheteiwy, M.S., Ramakrishnan, M., Landi, M., Araniti, F., Sharma, A., 2020. Trichoderma: The “Secrets” of a Multitalented Biocontrol Agent. Plants 9, 762. https://doi.org/10.3390/plants9060762

Spargo, J.T., Cavigelli, M.A., Mirsky, S.B., Maul, J.E., Meisinger, J.J., 2011. Mineralizable soil nitrogen and labile soil organic matter in diverse long-term cropping systems. Nutrient Cycling in Agroecosystems 90, 253–266. https://doi.org/10.1007/s10705-011-9426-4

Speirs, L.B.M., Rice, D.T.F., Petrovski, S., Seviour, R.J., 2019. The Phylogeny, Biodiversity, and Ecology of the Chloroflexi in Activated Sludge. Frontiers in Microbiology 10, 2015. https://doi.org/10.3389/fmicb.2019.02015

Steenwerth, K., Drenovsky, R., Lambert, J.J., Kluepfel, D., Scow, K., Smart, D., 2008. Soil morphology, depth and grapevine root frequency influence microbial communities in a Pinot noir vineyard. Soil Biology and Biochemistry 40. https://doi.org/10.1016/j.soilbio.2007.04.031

Stone, M.M., DeForest, J.L., Plante, A.F., 2014. Changes in extracellular enzyme activity and microbial community structure with soil depth at the Luquillo Critical Zone Observatory. Soil Biology and Biochemistry Complete, 237–247. https://doi.org/10.1016/j.soilbio.2014.04.017

Stott, D., 2019. Recommended Soil Health Indicators and Associated Laboratory Procedures (Technical Note No. 450– 03), Soil Health. USDA-NRCS.

Teasdale, J.R., Coffman, C.B., Mangum, R.W., 2007. Potential Long-Term Benefits of No-Tillage and Organic Cropping Systems for Grain Production and Soil Improvement. Agronomy Journal 99, 1297–1305. https://doi.org/10.2134/agronj2006.0362

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V., Vasco-Palacios, A.M., Thu, P.Q., Suija, A., Smith, M.E., Sharp, C., Saluveer, E., Saitta, A., Rosas, M., Riit, T., Ratkowsky, D., Pritsch, K., Põldmaa, K., Piepenbring, M., Phosri, C., Peterson, M., Parts, K., Pärtel, K., Otsing, E., Nouhra, E., Njouonkou, A.L., Nilsson, R.H., Morgado, L.N., Mayor, J., May, T.W., Majuakim, L., Lodge, D.J., Lee, S.S., Larsson, K.-H., Kohout, P., Hosaka, K., Hiiesalu, I., Henkel, T.W., Harend, H., Guo, L., Greslebin, A., Grelet, G., Geml, J., Gates, G., Dunstan, W., Dunk, C., Drenkhan, R., Dearnaley, J., Kesel, A.D., Dang, T., Chen, X., Buegger, F., Brearley, F.Q., Bonito, G., Anslan, S., Abell, S., Abarenkov, K., 2014. Global diversity and geography of soil fungi. Science. https://doi.org/10.1126/science.1256688

Treseder, K.K., Lennon, J.T., 2015. Fungal Traits That Drive Ecosystem Dynamics on Land. Microbiology and Molecular Biology Reviews 79, 243–262. https://doi.org/10.1128/MMBR.00001-15

Uehling, J.K., Entler, M.R., Meredith, H.R., Millet, L.J., Timm, C.M., Aufrecht, J.A., Bonito, G.M., Engle, N.L., Labbé, J.L., Doktycz, M.J., Retterer, S.T., Spatafora, J.W., Stajich, J.E., Tschaplinski, T.J., Vilgalys, R.J., 2019. Microfluidics and Metabolomics Reveal Symbiotic Bacterial–Fungal Interactions Between Mortierella elongata and Burkholderia Include Metabolite Exchange. Frontiers in Microbiology 10, 2163. https://doi.org/10.3389/fmicb.2019.02163

White, K.E., Cavigelli, M.A., Conklin, A.E., Rasmann, C., 2019. Economic Performance of Long-term Organic and Conventional Crop Rotations in the Mid-Atlantic. Agronomy Journal 111, 1358–1370. https://doi.org/10.2134/agronj2018.09.0604

Wolf, A.B., Vos, M., Boer, W. de, Kowalchuk, G.A., 2013. Impact of Matric Potential and Pore Size Distribution on Growth Dynamics of Filamentous and Non-Filamentous Soil Bacteria. PLOS ONE 8, e83661. https://doi.org/10.1371/journal.pone.0083661

Wu, T., Milner, H., Díaz-Pérez, J.C., Ji, P., 2015. Effects of soil management practices on soil microbial communities and development of southern blight in vegetable production. Applied Soil Ecology 91, 58–67. https://doi.org/10.1016/j.apsoil.2015.02.011

Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R., 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat Rev Micro 12, 635–645. https://doi.org/10.1038/nrmicro3330

Zeng, J., Liu, X., Song, L., Lin, X., Zhang, H., Shen, C., Chu, H., 2016. Nitrogen fertilization directly affects soil bacterial diversity and indirectly affects bacterial community composition. Soil Biology and Biochemistry 92, 41–49. https://doi.org/10.1016/j.soilbio.2015.09.018

Zuber, S.M., Villamil, M.B., 2016. Meta-analysis approach to assess effect of tillage on microbial biomass and enzyme activities. Soil Biology and Biochemistry 97, 176–187. https://doi.org/10.1016/j.soilbio.2016.03.011

CHAPTER 5:

Balboa, G.R., Ciampitti, I.A., 2020. Estimating biological nitrogen fixation in field-grown soybeans: impact of B value. Plant Soil 446, 195–210. https://doi.org/10.1007/s11104-019-04317-1

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1–48. https://doi.org/10.18637/jss.v067.i01

Bolhuis, H., Severin, I., Confurius-Guns, V., Wollenzien, U.I.A., Stal, L.J., 2010. Horizontal transfer of the nitrogen fixation gene cluster in the cyanobacterium Microcoleus chthonoplastes. ISME J 4, 121–130. https://doi.org/10.1038/ismej.2009.99

Bremner, J.M., Keeney, D.R., 1965. Steam distillation methods for determination of ammonium, nitrate and nitrite. Analytica Chimica Acta. https://doi.org/10.1016/S0003-2670(00)88973-4

Broadbent, F.E., 1965. Part 1: physical and mineralogical methods; organic matter, in: Methods of Soil Analysis, Monograph. American Society of Agronomy, pp. 1397–1400.

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

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Cavigelli, M., Teasdale, J.R., Conklin, A.E., 2008. Long-term agronomic performance of organic and conventional field crops in the mid-atlantic region. Agron. J. 100. https://doi.org/10.2134/agronj2006.0373

Coupland, D., Caseley, J.C., 1979. Presence of 14c activity in root exudates and guttation fluid from Agropyron repens treated with 14c-labelled glyphosate. New Phytologist 83, 17–22. https://doi.org/10.1111/j.1469-8137.1979.tb00721.x

Crits-Christoph, A., Olm, M.R., Diamond, S., Bouma-Gregson, K., Banfield, J.F., 2020. Soil bacterial populations are shaped by recombination and gene-specific selection across a grassland meadow. ISME J 14, 1834–1846. https://doi.org/10.1038/s41396-020-0655-x

Delves, A.C., Mathews, A., Day, D.A., Carter, A.S., Carroll, B.J., Gresshoff, P.M., 1986. Regulation of the soybean-rhizobium nodule symbiosis by shoot and root factors 1. Plant Physiology 82, 588–590. https://doi.org/10.1104/pp.82.2.588

Epp Schmidt, D.J., Kotze, D.J., Hornung, E., Setälä, H., Yesilonis, I., Szlavecz, K., Dombos, M., Pouyat, R., Cilliers, S., Tóth, Z., Yarwood, S., 2019. Metagenomics reveals bacterial and archaeal adaptation to urban land-use: N catabolism, methanogenesis, and nutrient acquisition. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02330

Fan, L., Feng, Y., Weaver, D.B., Delaney, D.P., Wehtje, G.R., Wang, G., 2017. Glyphosate effects on symbiotic nitrogen fixation in glyphosate-resistant soybean. Applied Soil Ecology 121, 11–19. https://doi.org/10.1016/j.apsoil.2017.09.015

Forlani, G., Bertazzini, M., Barillaro, D., Rippka, R., 2015. Divergent properties and phylogeny of cyanobacterial 5-enol-pyruvyl-shikimate-3-phosphate synthases: evidence for horizontal gene transfer in the Nostocales. New Phytologist 205, 160–171. https://doi.org/10.1111/nph.13022

Fujimatsu, T., Endo, K., Yazaki, K., Sugiyama, A., 2020. Secretion dynamics of soyasaponins in soybean roots and effects to modify the bacterial composition. Plant Direct 4, e00259. https://doi.org/10.1002/pld3.259

Gaby, J.C., Buckley, D.H., 2014. A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. Database (Oxford) 2014. https://doi.org/10.1093/database/bau001

Granato, T.C., Banwart, W.L., Porter, P.M., Hassett, J.J., 1983. Effect of variety and stage of growth on potential allelochemic compounds in soybean roots. J Chem Ecol 9, 1281–1292. https://doi.org/10.1007/BF00982229

Grube, M., Kalnenieks, U., Muter, O., 2019. Metabolic response of bacteria to elevated concentrations of glyphosate-based herbicide. Ecotoxicology and Environmental Safety 173, 373–380. https://doi.org/10.1016/j.ecoenv.2019.02.045

Hao, Y., Chen, H., Xu, W., Gao, J., Yang, Y., Zhang, Y., Tao, L., 2019. Roundup® confers cytotoxicity through DNA damage and mitochondria-associated apoptosis induction. Environmental Pollution 252, 917–923. https://doi.org/10.1016/j.envpol.2019.05.128

Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., Korpela, K., 2020. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. PLOS ONE 15, e0227285. https://doi.org/10.1371/journal.pone.0227285

Kepler, R., Epp Schmidt, D., Yarwood, S., Cavigelli, M., Reddy, K., Duke, S., Bradley, C., Williams, M., Buyer, J., Maul, J., 2019. Soil microbial communities in diverse agroecosystems exposed to the herbicide glyphosate. Applied and Environmental Microbiology 86. https://doi.org/10.1128/AEM.01744-19

Kremer, R., Means, N., Kim, S., 2005. Glyphosate affects soybean root exudation and rhizosphere micro-organisms. International Journal of Environmental Analytical Chemistry 85, 1165–1174. https://doi.org/10.1080/03067310500273146

Kuzyakov, Y., Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: concept & review. Soil Biology and Biochemistry 83, 184–199. https://doi.org/10.1016/j.soilbio.2015.01.025

Kuzyakov, Y., Razavi, B.S., 2019. Rhizosphere size and shape: Temporal dynamics and spatial stationarity. Soil Biology and Biochemistry 135, 343–360. https://doi.org/10.1016/j.soilbio.2019.05.011

Laitinen, P., Rämö, S., Siimes, K., 2007. Glyphosate translocation from plants to soil – does this constitute a significant proportion of residues in soil? Plant Soil 300, 51–60. https://doi.org/10.1007/s11104-007-9387-1

Lepsch, H.C., Brown, P.H., Peterson, C.A., Gaudin, A.C.M., Khalsa, S.D.S., 2019. Impact of organic matter amendments on soil and tree water status in a California orchard. Agricultural Water Management 222, 204–212. https://doi.org/10.1016/j.agwat.2019.06.002

Louca, S., Jacques, S.M.S., Pires, A.P.F., Leal, J.S., Srivastava, D.S., Parfrey, L.W., Farjalla, V.F., Doebeli, M., 2016a. High taxonomic variability despite stable functional structure across microbial communities. Nat Ecol Evol 1, 1–12. https://doi.org/10.1038/s41559-016-0015

Louca, S., Parfrey, L.W., Doebeli, M., 2016b. Decoupling function and taxonomy in the global ocean microbiome. Science 353, 1272–1277. https://doi.org/10.1126/science.aaf4507

Loveland, P., Webb, J., 2003. Is there a critical level of organic matter in the agricultural soils of temperate regions: a review. Soil and Tillage Research 70, 1–18. https://doi.org/10.1016/S0167-1987(02)00139-3

Lupwayi, N.Z., Blackshaw, R.E., 2012. Soil microbiology in glyphosate-resistant corn cropping systems. Agronomy Journal 104, 1041–1048. https://doi.org/10.2134/agronj2012.0054

Moebius, B.N., van Es, H.M., Schindelbeck, R.R., Idowu, O.J., Clune, D.J., Thies, J.E., 2007. Evaluation of laboratory-measured soil properties as indicators of soil physical quality. Soil Science 172, 895–912. https://doi.org/10.1097/ss.0b013e318154b520

Moebius-Clune, B.N., Moebius-Clune, D.J., Gugino, B.K., Idowu, O.J., Schindelbeck, R.R., Ristow, A.J., van Es, H.M., Thies, J.E., Shayler, H.A., McBride, M.B., Kurtz, K.S.M., Wolfe, D.W., Abawi, G.S., 2016. Comprehensive assessment of soil health: the Cornell framework manual.

Nguyen, D., Rose, M., Rose, T., Morris, S., Zwieten, L., 2016. Impact of glyphosate on soil microbial biomass and respiration: A meta-analysis. Soil Biology and Biochemistry 92, 50–57. https://doi.org/10.1016/j.soilbio.2015.09.014

Philippot, L., Andersson, S.G.E., Battin, T.J., Prosser, J.I., Schimel, J.P., Whitman, W.B., Hallin, S., 2010. The ecological coherence of high bacterial taxonomic ranks. Nat Rev Micro 8, 523–529. https://doi.org/10.1038/nrmicro2367

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41, D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rainio, M.J., Ruuskanen, S., Helander, M., Saikkonen, K., Saloniemi, I., Puigbò, P., 2021. Adaptation of bacteria to glyphosate: a microevolutionary perspective of the enzyme 5-enolpyruvylshikimate-3-phosphate synthase. Environmental Microbiology Reports 13, 309–316. https://doi.org/10.1111/1758-2229.12931

Reinhold, B., Bü, W., Burbano, C., Sabale, M., Hurek, T., 2015. Roots Shaping Their Microbiome: Global Hotspots for Microbial Activity. Annual review of phytopathology 53, 403–24. https://doi.org/10.1146/annurev-phyto-082712-102342

Rocha, E.P.C., 2018. Neutral theory, microbial practice: challenges in bacterial population genetics. Molecular Biology and Evolution 35, 1338–1347. https://doi.org/10.1093/molbev/msy078

Sansupa, C., Wahdan, S.F.M., Hossen, S., Disayathanoowat, T., Wubet, T., Purahong, W., 2021. Can we use functional annotation of prokaryotic taxa (FAPROTAX) to assign the ecological functions of soil bacteria? Applied Sciences 11, 688. https://doi.org/10.3390/app11020688

Schlatter, D.C., Yin, C., Hulbert, S., Burke, I., Paulitz, T., 2017. Impacts of repeated glyphosate use on wheat-associated bacteria are small and depend on glyphosate use history. Appl. Environ. Microbiol. 83, e01354-17. https://doi.org/10.1128/AEM.01354-17

Schönbrunn, E., Eschenburg, S., Shuttleworth, W.A., Schloss, J.V., Amrhein, N., Evans, J.N.S., Kabsch, W., 2001. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. Proc Natl Acad Sci U S A 98, 1376–1380.

Steinrücken, H.C., Amrhein, N., 1980. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimic acid-3-phosphate synthase. Biochemical and Biophysical Research Communications 94, 1207–1212. https://doi.org/10.1016/0006-291X(80)90547-1

Sugiyama, A., Yamazaki, Y., Yamashita, K., Takahashi, S., Nakayama, T., Yazaki, K., 2016. Developmental and nutritional regulation of isoflavone secretion from soybean roots. Bioscience, Biotechnology, and Biochemistry 80, 89–94. https://doi.org/10.1080/09168451.2015.1062714

Tzin, V., Galili, G., 2010. New insights into the shikimate and aromatic amino acids biosynthesis pathways in plants. Molecular Plant 3, 956–972. https://doi.org/10.1093/mp/ssq048

Vetterlein, D., Carminati, A., Kögel-Knabner, I., Bienert, G., Smalla, K., Oburger, E., Schnepf, A., Banitz, T., Tarkka, M., Schlüter, S., 2020. Rhizosphere spatiotemporal organization–a key to rhizosphere functions. Frontiers in Agronomy 2, 8. https://doi.org/10.3389/fagro.2020.00008

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Research 42, D643–D648. https://doi.org/10.1093/nar/gkt1209

Zablotowicz, R.M., Reddy, K.N., 2007. Nitrogenase activity, nitrogen content, and yield responses to glyphosate in glyphosate-resistant soybean. Crop Protection, Weed Science in Time of Transition 26, 370–376. https://doi.org/10.1016/j.cropro.2005.05.013

Zobiole, L.H.S., Kremer, R.J., de Oliveira Jr., R.S., Constantin, J., 2012. Glyphosate effects on photosynthesis, nutrient accumulation, and nodulation in glyphosate-resistant soybean. Journal of Plant Nutrition and Soil Science 175, 319–330. https://doi.org/10.1002/jpln.201000434

Zobiole, L.H.S., Kremer, R.J., Oliveira, R.S., Constantin, J., 2011. Glyphosate affects chlorophyll, nodulation and nutrient accumulation of “second generation” glyphosate-resistant soybean (Glycine max L.). Pesticide Biochemistry and Physiology 99, 53–60. https://doi.org/10.1016/j.pestbp.2010.10.005

Zobiole, L.H.S., Kremer, R.J., Oliveira, R.S., Constantin, J., 2010. Glyphosate affects micro-organisms in rhizospheres of glyphosate-resistant soybeans. Journal of Applied Microbiology 110, 118–127. https://doi.org/10.1111/j.1365-2672.2010.04864.x